In recent years, long noncoding RNAs (lncRNAs) have emerged as an important class of regulators of gene expression. lncRNAs exhibit several distinctive features that confer unique regulatory functions, including exquisite cell- and tissue-specific expression and the capacity to transduce higher-order spatial information. Here we review evidence showing that lncRNAs exert critical functions in adult tissue stem cells, including skin, brain, and muscle, as well as in developmental patterning and pluripotency. We highlight new approaches for ascribing lncRNA functions and discuss mammalian dosage compensation as a classic example of an lncRNA network coupled to stem cell differentiation.

Introduction
Efforts to understand how tissues are patterned during development and maintained by stem cells throughout life have traditionally focused on the protein-coding genome. Over the past decade, however, our understanding of the noncoding genome and its impact on cell fate has dramatically expanded. Contrary to previous notions of genome organization and function, the identification of thousands of long and short noncoding RNAs (ncRNAs) has revealed that much of the genome is in fact transcribed. Long noncoding RNAs (lncRNAs) are operationally defined as transcripts of greater than 200 nucleotides that function by means other than coding for proteins; lncRNAs are typically transcribed by RNA polymerase II and are frequently spliced and polyadenylated (reviewed by Rinn and Chang, 2012). As a class, lncRNAs tend to be expressed at lower levels and are predominantly localized in the nucleus, in contrast to messenger RNAs, which are abundant and enriched in the cytoplasm (Derrien et al., 2012). Notwithstanding these generalizations, lncRNAs exhibit a wide range of expression levels and distinct cytotoxic localizations, reflecting a large and diverse class of regulators (reviewed by Batista and Chang, 2013). Several well-studied examples of lncRNAs suggest that they can operate through distinct modes, including as signals, scaffolds for protein-protein interactions, molecular decoys, and guides to target elements in the genome or transcriptome (Wang and Chang, 2011). The discovery of novel lncRNAs has historically outpaced their functional annotation; however, efforts to more specifically ascribe function to either previously identified or novel lncRNAs have increased in recent years. Stem cells offer an attractive system for studying lncRNA function since previous findings have suggested that lncRNA expression is more cell-type-specific than mRNA expression (Cabili et al., 2011), leading to the possibility that lncRNAs may be key regulators of cell fate.

Here we review recent developments that illuminate the roles of lncRNAs in stem cell biology. We explore efforts to characterize the functions of lncRNAs in the development and patterning of several somatic tissues, including skin, brain, and musculature. Additionally, we examine how lncRNAs contribute to the pluripotent state and can be used to assess reprogramming status.

LncRNAs in Adult Tissue Stem Cells
Skin: An Ideal Model
Studying the biology of tissues at the molecular level necessitates robust model systems. While there are few systems that are suitable for detailed molecular characterization, well-developed human models exist for the skin based on ex vivo tissue regeneration that can also be grafted in vivo (Sen et al., 2010; Truong et al., 2006). Such models provide cellular material for molecular and biochemical studies that would be otherwise inaccessible and offer a system for testing the function of lncRNAs. Surveying the pattern of gene expression during epidermal differentiation, Khavari and colleagues discovered two key lncRNAs, ANCR and TINCR, that are expressed in epidermal stem cells and their terminally differentiated progeny, respectively (Kretz et al., 2012, 2013) (Figure 1). Antidifferentiation noncoding RNA (ANCR) provides a prime example of an lncRNA that controls the differentiation state of a somatic stem cell (Kretz et al., 2012). Specifically, ANCR depletion results in ectopic differentiation of epidermal stem cells, implying that ANCR’s role is to suppress the differentiation pathway in the epidermis and maintain the stem cell compartment.

While ANCR appears to inhibit differentiation, a different lncRNA termed terminal differentiation-induced noncoding RNA (TINCR) promotes epidermal differentiation (Kretz et al., 2013). TINCR is kept at very low levels in epidermal stem cells, but it is dramatically induced upon differentiation. Mechanistic studies of TINCR revealed that TINCR is a cytoplasmic lncRNA that interacts with the RNA-binding protein (RBP) STAU1 and converts STAU1 into an mRNA stability factor (Figure 1). Together, TINCR and STAU1 bind to and functionally stabilize mRNAs that encode structural and regulatory proteins critical for terminally differentiated keratinocytes. Additionally, TINCR expression is downregulated in human squamous cell carcinoma, providing evidence that lncRNAs can functionally regulate healthy and disease tissues.

The development of two techniques made these insights possible: (1) RNA interactome analysis (RIA), which allows the retrieval and unbiased discovery of RNAs interacting with an lncRNA of interest, and (2) protein microarray hybridization, which allows rapid discovery of direct RBP partners of an
InCRNA (Kretz et al., 2013). Moreover, both ANCR and TINCR were identified from large-scale expression profiling studies, suggesting that many additional lncRNAs may be identified and characterized using this system. Indeed, the differentiation of the skin is a multistep and highly regulated process that could benefit from the diverse set of lncRNAs hiding in the genome. The development of techniques such as RIA and the implementation of protein microarrays facilitated the functional characterization of TINCR but are applicable to uncovering mechanisms of other lncRNAs. Within the skin, the regulated and sequential expression of lncRNAs is clearly essential for their function; thus, understanding what controls the spatiotemporal expression of lncRNAs, such as ANCR and TINCR, should be the focus of future studies.

**Regulation in the Brain**

Transcription and alternative splicing in the brain appear to be the most complex among all organs (Mehler and Mattick, 2007; Mercer et al., 2008). An early example of lncRNAs controlling neural cell fates involves the Evf2 lncRNA and the Dlx5/6 genomic locus (Bond et al., 2009). Evf2 is transcribed antisense to Dlx6, which encodes a transcription factor, and is located immediately downstream of the Dlx5 genomic locus. The act of transcribing Evf2 can control the levels of Dlx6 in cis, and after disengaging the polymerase, Evf2 acts in trans to modulate the methylation of the Dlx5/6 enhancer and transcription of Dlx5. Therefore, by regulating the cellular levels of the Dlx5 and Dlx6 transcription factors, Evf2 controls GABAergic interneuron activity (Berghoff et al., 2013; Bond et al., 2009). A different study characterizing another lncRNA important for neural differentiation found that an enhancer region of the gene encoding the Neurogenin 1 transcription factor was transcribed and produced an lncRNA that positively regulated Neurogenin 1 expression (Onoguchi et al., 2012). These few examples begin to build the case that lncRNAs play an important role in neural biology.

The starting point of many lncRNA studies is unbiased gene expression analysis, which can reveal novel lncRNAs and their expression pattern in a developmental context. Recent large-scale efforts have employed next generation sequencing (“-seq”) technologies, from RNA-seq to chromatin immunoprecipitation (ChIP-seq), to identify transcripts and define their genomic positions (reviewed by Rinn and Chang, 2012). In the mouse brain, Lim and colleagues isolated three separate regions, subventricular zone (SVZ), olfactory bulb (OB), and the dentate gyrus (DG), and subjected these samples to short-read RNA-seq and ChIP-seq (Ramos et al., 2013). Over 3,600 novel lncRNAs were identified, and clustering of the lncRNAs and mRNAs by their expression patterns revealed that the lncRNAs were more tissue specific than mRNAs, consistent with previous reports (Cabili et al., 2011). Application of CaptureSeq, a technique that circumvents some drawbacks of short-read sequencing (Mercer et al., 2012), to further characterize the transcriptome of adult SVC tissue doubled the number (to ~7,000) of novel lncRNAs identified. To functionally validate the cataloging effort, two lncRNAs were identified by selecting loci marked by H3K4me3, which is associated with expressed genes, in NPC-SVC cells. This search identified Six3os and Dlx1as for further testing. Notably, Six3os has been previously reported to control retinal development (Rapicavoli et al., 2011). To characterize the neural role of Six3os and Dlx1as, SVZ neural progenitor cells were challenged in a 7-day differentiation assay with short hairpin RNAs targeting the two lncRNAs or control shRNAs. Depletion of Six3os lncRNA leads to fewer Tuj1 (neuron marker)- and OLG2 (oligodendrocyte marker)-positive cells, whereas depletion of Dlx1as specifically affected the number of Tuj1-positive cells (Figure 1). While the molecular mechanisms of these lncRNAs were not explored, Six3os has been shown to physically interact with Ezh2, a component of the Polycomb repressive complex 2 (PRC2), to repress specific genes in retinal cells (Rapicavoli et al., 2011). These examples illustrate that mapping spatiotemporal patterns of lncRNAs can highlight functional transcripts. Larger-scale validation efforts will be required to fully realize the extent of lncRNA regulation in the different regions of the brain.

A complementary approach identifies potential lncRNA regulators based on their loss-of-function phenotypes in large-scale depletion studies (Guttman et al., 2011; Lin et al., 2014). Rana and colleagues targeted 1,280 mouse lncRNAs and identified 20 lncRNAs that were required for the maintenance of mouse embryonic stem cell (mESC) pluripotency. One lncRNA, named TUNA, was previously identified as megamind in zebrafish.
TUNA/Megamind depletion in zebrafish led to altered neurodevelopment and impaired locomotor response (Ulltzy et al., 2011; Lin et al., 2014). TUNA is highly conserved in human and fish, is required for the maintenance of pluripotency, and is also expressed in the brain, spinal cord, and eyes in adult tissues. Indeed, TUNA expression was increased when mESCs differentiated toward the neural lineages, and TUNA depletion inhibited neural differentiation of ESCs (Figure 1). Purifying proteins that associate with in vitro-transcribed TUNA identified hnRNP-K, Nucleolin (NCL), and PTBP1 as interaction partners. Importantly, depletion of several of these proteins phenocopied TUNA depletion (Lin et al., 2014). An important caveat to consider is that while the candidate approach characterized TUNA, Six3os, and Dlx1as IncRNAs as successful validation of genome-wide screens, such approaches leave the function of thousands of other transcripts, many of which may play important roles, unaddressed.

Many IncRNAs have been implicated in the regulation of chromatin states (Rinn and Chang, 2012), but direct evidence for their association has only recently been possible through the development chromatin isolation by RNA purification (ChIRP; and others methods discussed below) (Chu et al., 2011). ChIRP uses DNA capture probes to retrieve a specific IncRNA with its associated genomic DNA targets, and together with deep sequencing can generate a genome-wide map of IncRNA-chromatin interactions. Careful optimization of in vivo crosslinking, both of the chemical crosslinking agent and duration, and selection of proper oligonucleotide probes are important to obtain reliable measurement. This process often includes multiple but distinct DNA capture probe sets, probes targeting irrelevant RNAs as negative controls, and positive control regions to assay during pilot experiments (Chu et al., 2011). Successful implementation of ChIRP has revealed the IncRNA TUNA occupies promoter regions of Nanog, Sox2, and Fgf4 genes that are important for pluripotency and neural lineage commitment (Lin et al., 2014). Together with its protein partners and its chromatin localization, TUNA may regulate gene expression at both the transcriptional and post-transcriptional level. Thus, TUNA represents an IncRNA that is important for at least two cell states (ESC pluripotency and neural differentiation) and probably operates through multiple molecular mechanisms. This example highlights the concept that a single IncRNA can, under different cellular context and protein partners, function to control multiple molecular pathways.

IncRNAs and Muscle

LncRNAs also control development of mesodermal tissues and have similarly benefited from large-scale sequencing efforts to identify functionally important transcripts. One example of a heart-specific IncRNA named Braveheart was first functionally characterized as a key factor involved in cardiac lineage commitment because its depletion resulted in a severe reduction in the number of spontaneous beating cardiomyocytes formed during embryoid body differentiation (Klattenhoff et al., 2013) (Figure 1). Further characterization of Braveheart found that it interacts with Suz12, a subunit of PRC2, and acts in trans to regulate heart-specific differentiation genes such as MesP1. The regulation of master drivers of cardiac differentiation, such as MesP1 by Braveheart, offers new tools toward the goal of achieving highly efficient and reproducible in vitro reprogramming (Burridge et al., 2012). Producing cardiomyocytes from induced pluripotency stem cells (iPSCs) or directly from other differentiated cell types may benefit from engineering specific IncRNA expression during in vitro production.

While small interfering RNA (siRNA) knockdown of IncRNAs (used in most of the discussed work) often provides a great deal of insight into function, off-target effects and incomplete depletion must always be considered. As with protein-coding genes, knockout (KO) strategies offer potential remedies to these siRNA-related issues, but the specific strategy employed is critical (discussed below: Developmental Patterning by IncRNAs). Utilizing this concept, Herrmann and colleagues inserted a premature polyadenylation (polyA) signal into the IncRNA Fendrr’s locus to promote depletion of the full-length Fendrr RNA (Grote et al., 2013). Initial characterization of Fendrr found it expressed in the caudal end of the lateral plate mesoderm (LPM), which develops into the structures like the heart and body wall. Fendrr KO resulted in embryonic lethality at embryonic day 13.75, abdominal wall defects, and pooling of blood in the right atrium. By partnering with both activating (mixed-lineage Leukemia [MLL], WDR5) and silencing (PRC2) chromatin complexes, Fendrr was proposed to modulate the epigenetic landscape during development (Figure 1). More recently ChIRP was used to show that Fendrr physically associates with the promoters of FoxF1 and Pitx2 mRNAs, two genes-repressed by Fendrr (Grote and Hermann, 2013; Grote et al., 2013). Fendrr therefore represents a dual-function IncRNA that may control both positive and negative chromatin modifying complexes to guide development.

Long RNAs Controlling Small RNAs

The differentiation of a myoblast progenitor cell (MB) to a fully differentiated muscle cell is a highly regulated process that relies on Ying Yang 1 (YY1), a multifunctioning transcription factor (Deng et al., 2010; Lu et al., 2012). Examination of YY1’s chromatin binding pattern in MBs revealed that it bound the promoter of many ncRNA loci, and these target noncoding genes were named YY1-associated muscle IncRNAs (Yam) (Lu et al., 2013). Characterization of one of these IncRNAs, Yam1, identified it as a key regulator of myogenesis, as it was able to repress key muscle differentiation genes including myogenin, Tnni2, and α-actin, (Figure 1). Furthermore, Yam1 increased levels of microRNA-715 (miR-715), which targets Wnt7b, a protein that normally promotes muscle differentiation (Lu et al., 2013). Yam1 thus provides evidence that in muscle IncRNAs can modulate the levels of both mRNAs and other ncRNAs, such as miRNAs, providing additional network control to cells.

The regulation of miRNA networks reveals an additional mechanism through which IncRNAs exert control. Recently, multiple IncRNAs have been shown to act as competing endogenous RNAs (ceRNAs), where the IncRNAs are proposed to bind to and compete miRNAs away from cognate mRNA targets (Tay et al., 2014). Pseudogene IncRNAs are prime candidates for the ceRNA mechanism because they may share multiple miRNA binding sites, allowing more effective competition with cognate mRNAs. The ceRNA hypothesis requires that ceRNAs are expressed highly enough and have sufficient numbers of miRNA binding sites to substantially affect the pool of cellular miRNAs. Recent work exploring the dynamics of miRNA-regulated gene repression has shown that it is highly susceptible to thresholds. In certain contexts, small concentration changes of miRNA-mRNA or
miRNA-ceRNA pairs can substantially modulate the gene expression network (Bosia et al., 2013; Mukherji et al., 2011). Moreover, one example of a ceRNA, linc-MD1, has been previously shown to regulate muscle differentiation through its ability to sponge miR-133 and miR-135 away from the miRNAs MAML1 and MEF2C (Cesana et al., 2011). These two miRNAs are important transcriptional activators of the muscle differentiation program. Linc-MD1 itself contains an miR-133b, which represses muscle differentiation when processed. Recent molecular characterization of this network revealed the RBPs HuR bound to linc-MD1 and the levels of linc-MD1 positively correlated with HuR protein abundance (Legnini et al., 2014) (Figure 1). HuR controlled the fate of linc-MD1, as cellular depletion of HuR favored the processing of linc-MD1 into miR-133b, tipping the balance in favor of the miRNA over the ceRNA. HuR has known roles in myogenesis and its interaction with linc-MD1 fine-tunes the levels of miRNAs important in the muscle differentiation program. Together, these studies explore IncRNA functions in muscle tissue and help to expand the possible modes of IncRNA functions within the already complex system of miRNA-mediated gene regulation.

Developmental Patterning by IncRNAs

IncRNAs also orchestrate the patterning of cells into tissues and organs during development. HOTAIR IncRNA was one of the first characterized IncRNAs that acts at distance (in trans) to modulate Hox gene expression (Rinn et al., 2007). HOTAIR is a repressive IncRNA and serves a scaffold between two distinct chromatin modification complexes (Rinn et al., 2007; Tsai et al., 2010). Other Hox-encoded IncRNAs such as HOTTIP, Mistral, and HOTAIRm1 were shown to regulate different members of HoxA genes (Bertani et al., 2011; Wang et al., 2011; Zhang et al., 2008). For example, HOTTIP is expressed in distal anatomic structures and activates the expression of HOXA9-HOXA13 genes to promote distal limb development (Wang et al., 2011). Characterization of these IncRNAs has often occurred through overexpression or siRNA knockdown studies. While these strategies often yield relevant results, transcriptional modulation is often not complete, especially using siRNA (or even short hairpin RNA), necessitating alternative methods.

Recently there have been a number of studies utilizing gene KO to understand IncRNA biology (Grote et al., 2013; Li et al., 2013; Sauvageau et al., 2013). At least three KO strategies have been reported: (1) insertion of a polyA signal near the transcription start sites; (2) insertion of a reporter gene under the control of the endogenous promoter; and (3) complete deletion of the IncRNA locus. The latter is the most dramatic and may, in addition to removing the IncRNA exons/intron structure, remove unknown regulatory elements. Insertion of a reporter gene has the advantage of being able to monitor expression of the IncRNA throughout development; however, depending on which sequences are replaced, it may also carry similar drawbacks as the deletion strategy. Finally, insertion of a polyA signal near the transcription start sites likely has the least off-target effects; however, background expression from the IncRNA locus could still result from not removing downstream sequences, cryptic start sites, or inefficient polyA tailing and cleavage.

Elucidating IncRNA Tissue Patterning by KO Models

Recent efforts have begun to utilize full KO strategies to characterize additional IncRNAs including Hox encoded candidates. The developmental functions of mouse Hotair were investigated by full IncRNA locus deletion in the mouse (Li et al., 2013). Loss of Hotair resulted in aberrant patterning of the skeletal system during development, as was evident in abnormalities in the wrist and spine, including a switch of vertebral segment identity called homeotic transformation. Further, genome-wide characterization of the Hotair KO mouse confirmed that murine Hotair acted similarly to human HOTAIR, namely as a trans-acting IncRNA controlling histone modification at specific genomic loci (Li et al., 2013). More recently, in an effort to dramatically expand the number of IncRNA KOs, Rinn and colleagues used the reporter gene approach to generate 18 separate IncRNA knockout mice (Sauvageau et al., 2013). By replacing IncRNA exonic regions with a LacZ construct, both KO and tagging was achieved. Three of the 18 IncRNAs (Fendrr, Peril, and Mdgt) showed variable penetrance and lethality. The Mdgt and Pint KO lead to abnormally low body weight and slower growth. The detailed characterization of the IncRNA Brm1b revealed its role in cortical development; specifically, this IncRNA was important for the embryonic patterning in certain areas of projection neurons. By creating a large number of IncRNA KO mice and characterizing many of their functions in vivo, this study helped to solidify the functional importance of IncRNAs. While thousands of IncRNAs remain to be genetically tested, new and more facile genome-editing tools should speed future characterization (Mali et al., 2013).

Sauvageau et al. also generated a new Fendrr KO mouse (Sauvageau et al., 2013). Under these conditions, Fendrr was expressed much more widely than previously observed and most highly in the developing lung. Fendrr KO resulted in perinatal lethality, as Fendrr−/− embryos either failed to initiate breathing or stop breathing within 5 hr of birth, neither of which was observed in wild-type pups. While the most striking phenotype of this KO was pulmonary, heart septal defects were also apparent even though their LacZ construct did not stain the heart for expression. This discrepancy is an important example of the possible phenotypic difference achieved by differential KO strategies such as reporter construct replacement or early polyA termination (Grote et al., 2013; Sauvageau et al., 2013). Specifically, addition of the polyA sites resulted in minimal disruption of the endogenous Fendrr locus, but extremely low levels of Fendrr were still detectable (Grote and Herrmann, 2013). On the other hand, the LacZ construct replaced ~20 kb of the genome, resulting in a complete lack of Fendrr transcripts; however, this large replacement may have removed other functional elements from the genome responsible for regulating other genes. Therefore, while both approaches confirmed loss of the IncRNA transcript, additional investigation is necessary and careful consideration of the cellular outcomes from any particular targeting strategy must be included in the experimental design.

Single-Cell Analysis of IncRNA Function

Most transcript-profiling experiments of IncRNAs have employed bulk measurements, reporting results from an average of thousands or millions of cells. Recent work at the single-cell level has revealed how much heterogeneity exists even within a “clonal” population of cells (Buganim et al., 2012; Shalek et al., 2013). Thus, it follows that examination of the noncoding genome and its function at the single-cell level could also reveal
Importantly, this study highlights two key and common binding protein B (PURB) and exert transcriptional silencing of ally, lincHOXA1 was found to partner with purine-rich element-

to efficiently lower chromatin-associated transcripts. Function-

to methods differ in their capacity to reduce lincHOXA1 levels on

H-mediated cleavage of the target RNA). The two depletion

both siRNA and antisense oligonucleotides (ASO, via RNase

single-cell level, the function of lowly expressed lncRNAs (Maa-

brought to light the importance of carefully examining, at the

level, ascribed a positive correlation to the expression of lincHOXA1 and a nearby

mRNA HOXA1. Surprisingly, however, single-cell analysis re-

vealed an anticorrelation, and specifically a switch-like relationship was observed such that if a cell had above ten copies of

lincHOXA1, HOXA1 was repressed. Knockdown studies used both siRNA and antisense oligonucleotides (ASO, via RNase

H-mediated cleavage of the target RNA). The two depletion

methods differ in their capacity to reduce lincHOXA1 levels on

the chromatin versus total levels, with siRNA treatment unable to efficiently lower chromatin-associated transcripts. Function-

ally, lincHOXA1 was found to partner with purine-rich element-

binding protein B (PURB) and exert transcriptional silencing of

HOXA1. Importantly, this study highlights two key and common

that many unknown features of known lncRNAs may be re-

LncRNAs Regulation of Pluripotency

The richness of the IncRNA regulatory landscape is perhaps best exemplified in ESCs, where the noncoding transcriptome has been under intense study. The expansive number of genomic data sets, both RNA- and chromatin-based, now available in ESCs provides a rich database to characterize lncRNA function. Recent progress in understanding IncRNA control of pluripotency and dosage-compensation mechanisms have revealed intimate connections between lncRNAs and chromatin state (Figure 2). Some of the most studied lncRNA binding proteins belong to chromatin modification complexes, including PRC2 and MLL, which act to suppress and activate, respectively, transcription through methylation of histone protein.

Transition between Cell States

Characterization of the transcriptome of ESCs has revealed many lncRNAs that participate in the regulation of the pluripotent state (Guttman et al., 2011, 2009; Lin et al., 2014; Ng et al., 2012; Sheik Mohamed et al., 2010). Through a comprehensive “perturb-and-measure” strategy, Guttman et al. showed that dozens
of IncRNAs are required for the setting the gene expression patterns of mouse ESCs or the first step of differentiation toward different germ layers (Guttman et al., 2011). A subset of these IncRNAs bound one or more chromatin modification complexes, including readers, writers, or erasers of repressive histone modifications.

In contrast, the “regulator of reprogramming” IncRNA (lincRNA-RoR) was identified as an important factor for the reprogramming process as its depletion or overexpression led to a lower or higher efficiency of reprogramming fibroblasts to iPSCs, respectively (Loewer et al., 2010). However only recently was the molecular mechanism investigated (Wang et al., 2013). Pull-down experiments with lincRNA-RoR specifically isolated miR-145-5p, 181a-5p, and 99b-3p, as well as the miR-targeting protein Argonaute2 (Ago2). These miRs have been previously shown to regulate core pluripotency factors such as Pou5f1, Sox2, and NANOG, suggesting that lincRNA-RoR might act as a ceRNA. Indeed, functional assays revealed that lincRNA-RoR regulated the mature form of miR-145, characteristic of a ceRNA. Loss of lincRNA-RoR caused human ESCs (hESCs) to differentiate toward mesoderm and ectoderm, while overexpression conferred a differentiation defect. Additionally, in the context of cancer, a rapidly proliferative state similar to ESCs, lincRNA-RoR was recently shown to act in a regulatory loop suppressing the expression of the tumor suppressor p53 (Zhang et al., 2013). Together, this characterization of lincRNA-RoR further advances the idea that each IncRNA may control many pathways in different cellular contexts including tumor growth and core pluripotency gene network utilizing a ceRNA mechanism.

**Activation of the Epigenome with IncRNAs**

To date, the vast majority of IncRNAs have annotated functions in repressive complexes, with only a few examples of activating or enhancing IncRNAs (Wang et al., 2011; Zhang et al., 2009). HOTTIP, named due to its location at the distal “tip” of the HOXA gene cluster, enforces an active chromatin state by recruiting the WDR5 subunit of the MLL complex (Wang et al., 2011) (Figure 2). The HOTTIP locus comes into spatial proximity with its target genes, and all the while the expression level of HOTTIP remains near one copy per cell (Wang et al., 2011). The low copy number of HOTTIP ensures that HOTTIP acts precisely in cis on target genes defined by proximity in three-dimensional nuclear space but not broadly on other genes. More recently, biochemical characterization of the interaction between WDR5 and HOTTIP revealed a specific RNA-binding pocket of WDR5 and that RNA binding could stabilize chromatin-associated WDR5 (Yang et al., 2014). This finding suggested that in vivo, not only the localization, but also the half-life of WDR5 could be modulated by HOTTIP. Given that WDR5/MLL acts at many genomic loci, RNA immunoprecipitation-seq (RIP-seq) was used to identify over 1,400 WDR5 interacting RNAs, including many coding and noncoding RNAs. An IncRNA-binding pocket on WDR5 was discovered, and a specific mutation of the RNA-binding pocket selectively abrogated RNA binding but no other functions of the WDR5-MLL complex (Yang et al., 2014). This selective WDR5 mutant revealed that RNA binding is important for the temporal stability of the active chromatin mark H3K4me3 over time and maintenance of ESC pluripotency. These studies suggest a generalizable mechanism for functional MLL/WDR5-RNA interaction. Specifically, HOTTIP acts in cis and is expressed at far too low levels per cell to globally modulate the MLL/WDR5 chromatin localization. The RIP-seq of WDR5 in mESCs (which do not express HOTTIP) revealed that more than one thousand cellular RNAs could interact with and may modulate the chromatin modification complex. Because WDR5 targets over 10,000 genomic sites (Ang et al., 2011), whether the three-dimensional organization of the genome facilitates IncRNA coregulation of the mESC self-renewal program remains to be addressed in future studies.

**Epigenetic Repression through IncRNA-PRC2 Interactions**

Unlike activating chromatin complexes, chromatin-modifying complexes that repress transcription have been more extensively studied in the context of IncRNA interactions, resulting in a richer set of known interactions. The focus of many of these studies has been the PRC2 complex, responsible for depositing H3K27me3, which plays roles pluripotency, differentiation, XCI, and diseases such as cancer (Margueron and Reinberg, 2011). An initial survey of the RNA-interactome of Ezh2 yielded more than 9,000 target RNAs using RIP-seq in mESCs (Zhao et al., 2010). Recently, two studies have revisited this observation to further clarify the interplay between RNA and PRC2 (Davidovich et al., 2013; Kaneko et al., 2013) (Figure 2). Biochemical interaction and photoactivated RNA-crosslinking experiments suggest that Ezh2 can interact with numerous RNAs, including the 5’ end of nascent RNAs that are actively transcribed. The apparently specific interactions of PRC2 with several IncRNAs in lysate and in vivo are not recapitulated in vitro by the core PRC2 complex alone. The promiscuous RNA binding of Ezh2 may be modulated by additional proteins, such as Jarid2 and others, to facilitate higher degrees of specificity in vivo (Davidovich et al., 2013; see below). Moreover, Ezh2 may scan the genome surveying the transcriptional status of its targets. Actively transcribed regions may continually push Ezh2 away via its elongating mRNAs, while silent regions or those stably bound by IncRNAs (generated in trans) can be silenced. This proposed mechanism reinforces the status quo of gene transcription and silencing and is consistent with the known genetic role of Polycomb group proteins in chromatin state maintenance.

A similar RNA surveillance mechanism is also employed by the DNA methylase DNMT1 that interacts with many cellular transcripts, including the nonpolyadenylated extruding CEBPA (ecCEBPA) IncRNA. The ecCEBPA IncRNA adopts a characteristic stem-loop structure critical for interaction with DNMT1 and, when transcribed, acts to shield the CEBPA locus from DNA methylation (Di Ruscio et al., 2013) (Figure 2). These two examples provide evidence that cells employ RNAs to modulate the deposition of repressive epigenetic marks in a genome-wide manner. Nonetheless, recognition of the potentially broad interactions between RNA and PRC2 highlights the need for high-quality in vivo controls and validation of RNA-protein interactions. Methodological choice is critical as each assay type has its own strengths and weaknesses, which will impact the results obtained and conclusions drawn.

While PRC2 operates in a wide range of cell types, certain subunits, such as JARID2, are specifically expressed and partner with PRC2 in ESCs and certain dividing cells, including cancer cells (Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009). These initial studies established JARID2’s capacity to regulate
the stability of the PRC2 complex as well as its enzymatic activity (Figure 2). Further expanding the cellular functions of JARID2, in vitro RNA-binding assays and in vivo PAR-CLIP suggest that JARID2 directly interacts with cellular RNAs (Kaneko et al., 2014). JARID2 and Ezh2 reproducibly crosslinked to 106 and 165 IncRNAs, respectively, and 53 IncRNAs were commonly bound. The MEG3 IncRNA was bound by both subunits of PRC2; however, the RNA-binding region (RBR) of JARID2 provided the largest contribution of MEG3 binding to PRC2. Additionally, cellular levels of MEG3 contribute to PRC2’s chromatin association, as low expression of MEG3 resulted in the loss of PRC2 subunits from specific loci leading to derepression of the nearby genes. Finally, the in vitro interaction between JARID2 and Ezh2 was facilitated by HOTAIR and MEG3, and Ezh2’s chromatin association was shown to be partially dependent on JARID2’s RNA-binding domain. Thus, JARID2, an ESC-specific subunit of PRC2, appears to modulate the localization of PRC2, and thus the chromatin state, in an RNA-dependent manner. While this study offers an additional layer of regulation with respect to the Polycomb complex, little is known about the other RNA targets of JARID2, which may significantly contribute to its cellular function. Additionally, studies to rigorously interrogate the enzymatic properties of the PRC2 complex inside cells with and without its RNA partners will be very informative.

An IncRNA Network to Control Dosage Compensation
Dosage compensation of genes encoded on the X chromosome is accomplished by divergent strategies in different species; however, the use of IncRNAs is a common feature. In Drosophila, dosage compensation is achieved by precisely upregulating the X chromosome in males by 2-fold (Lucchesi et al., 2005). A desire to understand how dosage compensation operates fueled the development of ChiRP and CHART, genomic tools that map the chromatin association of IncRNAs (Chu et al., 2011; Simon et al., 2011). Initially, ChiRP and CHART were applied to the Drosophila roX2 IncRNA, which provided evidence that roX2 co-occupies genomic loci with the known dosage compensation protein factors on the X chromosome. Importantly, they proved that mapping the genomic locations of IncRNAs can generate novel hypotheses for functions of IncRNAs. While studies in Drosophila and other model systems have provided key insights into mechanisms of dosage compensation, we will focus on recent investigations conducted in mammalian cells.

Xist Spreading
In mammals, the strategy for dosage compensation is reverse from Drosophila: female cells selectively repress one entire chromosome by upregulating the repressive IncRNA Xist (Lee, 2012). Xist is transcribed from the X-inactivation center (XIC) and is responsible for physically coating and silencing the X chromosome targeted for the Barr body (the Inactive X, Xi). Another IncRNA, Tnix, is transcribed from the active X chromosome (Xa) and enforces silencing of Xist (Lee, 2012). These two IncRNAs, together with others described below, form a complex RNA-protein regulatory network that controls X chromosome dosage compensation in mammals.

Traditional techniques such as immunofluorescence (IF) and RNA fluorescence in situ hybridization (FISH) have been widely applied to study X chromosome inactivation (XCI) and have arrived at a consensus mechanism: elevated Xist expression from the future Xi leads to a cloud-like coating of Xist on Xi, and finally epigenetic silencing and chromatin compaction. While informative, IF and RNA-FISH studies had resolution limitations, and as was true for the roX2 RNA, specifically mapping the genomic locations of Xist held the promise of answering mechanistic features of its function. Recently, application of CHART and the development and application of RAP (a method with similar principles as ChiRP and CHART) to the Xist IncRNA defined its precise chromatin association (Engreitz et al., 2013; Simon et al., 2013). Together, the studies revealed that the initiation of Xist spreading occurs from the Xist locus to distinct sites across the X chromosome that are not directly adjacent to its locus. These regions are highly accessible by DNaseI footprinting and contain many genes that are actively transcribed prior to silencing. Once Xist is deposited on these early sites, it proceeds to spread and coat the rest of the chromosome to fully silence all but a few genes that escape XCI. It is proposed that the initial deposition process is mediated through higher-order chromatin architecture (Engreitz et al., 2013); however, experimental design differences between the two studies described above make it difficult to directly compare the chromatin conformation results measured. While further investigation is clearly needed to solidify and refine these results, using high-resolution genomic tools (ChiRP, CHART, or RAP) can provide critical insight into IncRNA-controlled systems previously hidden from view.

Mechanisms of Xist Regulation
Intense study of the Xist regulatory network has uncovered many novel IncRNAs in and around the XIC, often illuminating novel mechanistic concepts for how IncRNAs function. Within the IncRNA network that controls Xist, Tsix and Jpx oppose each other’s function by repressing or activating, respectively, the transcription of Xist (Lee, 2000; Tian et al., 2010). Recently, additional characterization of the Jpx pathway revealed an unexpected interplay between the IncRNA Jpx and CCCTC-binding factor (CTCF), a major DNA-binding protein involved in higher-order chromosomal folding and interactions (Sun et al., 2013). During female mESC differentiation, CTCF is lost from the Xist locus, therefore allowing allele-specific Xist upregulation. Molecular characterization of this regulatory loop revealed that CTCF directly binds Jpx and this interaction can titrate CTCF from its DNA targets. Within the conceptual framework of dosage compensation, this puts Jpx and CTCF as central players in the balance between activation and silencing of the X chromosome. Cellular levels of Jpx, as partially determined by the number of X chromosomes, would control the ability of CTCF to bind and inhibit transcription at the Xist locus only under conditions when XCI is required. Another recent study more globally characterized the RNA-binding capacity of CTCF and found a multitude of RNA targets, including Wrap53, an IncRNA that controls the induction of the tumor suppressor p53 upon DNA damage (Saldaña-Meyer et al., 2014). Interestingly, biochemical characterization of CTCF’s protein domains revealed that the RBR and RNA promoted multimerization of CTCF.

While Xist is modulated by CTCF localization and the spatio-temporal deposition of Xist has been initially defined through CHART and RAP, how Xist interacts with protein effectors of XCI remains poorly understood. The repeat A (RepA) domain of
Xist has been reported to mediate the interaction with the PRC2 complex (Zhao et al., 2008). Recent characterization of the JARID2 subunit of PRC2 also implicates it in functionally interacting with Xist (da Rocha et al., 2014). Specifically, the authors observed JARID2 and other PC2 subunits co-occupying genomic regions on the X, and a requirement for JARID2 for the deposition of H3K27me3. Further, Xist deletion experiments defined the RepB and RepF regions within the RNA as responsible for JARID2 targeting to the X. Interestingly, this function was not dependent on its previously identified RBR (Kaneko et al., 2014), suggesting that JARID2 is a multifunctioning RNA-binding protein that mediates the association of PRC2 to the X through Xist. These examples suggest that within the context of XCI, as well as during other critical cellular decisions, lncRNAs (such as Xist) can act to modulate chromosome architecture and chromatin modification patterns.

**XCI as a Marker of Reprogramming**

The ability to transform differentiated cells back into pluripotent cells holds tremendous possibilities for regenerative medicine, but many hurdles still remain before this technology is fully matured (Sánchez Alvarado and Yamanaka, 2014). Because biallelic X activation is a key epigenetic marker of pluripotency, the status of Xist and Xist-mediated gene silencing (or lack thereof) can be exploited to phenotype ESCs and iPSCs (Figure 3). Careful analysis of human iPSCs derived from female cells revealed that many carried an X0 failing to undergo X chromosome reactivation (XCR), and are epigenetically dynamic, suggesting that the derivation of hiPSCs may not result in pristinely pluripotent cells as desired (Tchieu et al., 2010). A subsequent study used X-inactivation markers to segregate populations of hiPSCs and found that female-derived iPSCs are likely to be less stable in culture than male-derived cells (Anguera et al., 2012). Indeed, erosion of dosage compensation has been observed in female hiPSCs over time in culture, significantly impacting the potential use of these cells for modeling X-linked disease (Mekhoubad et al., 2012). More recent work characterized XCR in the context of iPSC reprogramming and found PRDM14, involved in the ESC pluripotency network, controls Xist silencing (Payer et al., 2013). With the help of Tsix, PRDM14 represses Xist activators (Rnf12 and Jpx) and the Xist locus itself by recruiting PRC2, placing PRDM14 expression as a marker for XCR. Work from Heard and colleagues also explored how Xist status can directly regulate ESC differentiation, notably within the framework of the primed/metastable and ground/naive states, with the latter representing a more primordial state of pluripotency. Schultz et al. reported that an X-linked inhibitor of MAPK signaling couples the status of X chromosomes to ESC differentiation. In the ground state where both X chromosomes are active, MAPK is inhibited concomitantly with other molecular changes that block ESC differentiation (Schulz et al., 2014). Upon X chromosome inactivation in the primed state, the relief of MAPK inhibition leads to high MAPK signaling and the capacity to proceed with differentiation. Therefore the characteristic expression of Xist and X-silencing genes provides new ways to evaluate the efficiency and ultimately control of reprogramming during iPSC generation. Combining traditional pluripotency markers with new markers like X-inactivation will be critical to achieve the standardization and consistency necessary for clinical application of iPSC technologies.

**Lessons and Future Prospects**

While the myriad examples to date highlight the functions of a small fraction of known lncRNAs, they illustrate the principle that lncRNAs are intimately involved in the specification, self-renewal, differentiation, and patterning of stem cells and their differentiated progenies. It is reasonable to anticipate that similar principles will be uncovered in many additional organ systems and cell types. A frequently asked question is “Why RNA”? lncRNAs exhibit exquisite cell-type- and organ-specific expression patterns, in fact, to a greater extent than mRNAs. Evolution has probably taken advantage of this fertile soil of cell-type- and state-specific transcription to evolve regulatory functions. Thus, one area of future investigation should focus on the regulation of lncRNA expression—what exactly makes them different and endows them with such state-specific expression? A second challenge for the field is the need to predict the functions of lncRNAs from primary sequence. Finally, understanding how the structure of lncRNAs guides their function remains largely unexplored. As has been true for protein biochemistry, understanding the physical conformations lncRNAs adopt inside cells will undoubtedly uncover novel functional domains and structural elements responsible for their cellular activities.

**REFERENCES**


We apologize to colleagues whose work was not discussed due to space constraints. We thank M. Wernig, E. Heard, R.C. Spitale, A.J. Rubin, and members of the H.Y.C. lab for comments. This work was supported by NIH Medical Scientist Training Program (R.A.F.) and NIH (RO1-CA119750, RO1-ES023168) and California Institute for Regenerative Medicine (H.Y.C.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.


