

# Chapter 8

## Long Noncoding RNAs in Pluripotency of Stem Cells and Cell Fate Specification

Debosree Pal and M.R.S. Rao

**Abstract** Since the annotation of the mouse genome (FANTOM project) [Kawai J et al (2001) Functional annotation of a full-length mouse cDNA collection. *Nature* 409(6821):685–690] or the human genome [An integrated encyclopedia of DNA elements in the human genome. (2012) *Nature* 489(7414):57–74; Harrow J et al (2012) GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res* 22(9):1760–1774], the roles of long noncoding RNAs in coordinating specific signaling pathways have been established in a wide variety of model systems. They have emerged as crucial and key regulators of stem cell maintenance and/or their differentiation into different lineages. In this chapter we have discussed the recently discovered lncRNAs that have been shown to be necessary for the maintenance of pluripotency of both mouse and human ES cells. We have also highlighted the different lncRNAs which are involved in directed differentiation of stem cells into any of the three germ layers. In recent years stem cell therapies including bone marrow transplantation are becoming an integral part of modern medicinal practices. However, there are still several challenges in making stem cell therapy more reproducible so that the success rate reaches a high percentage in the clinic. It is hoped that understanding the molecular mechanisms pertaining to the role of these newly discovered lncRNAs in the differentiation process of stem cells to specific lineages should pave the way to make stem cell therapy and regenerative medicine as a normal clinical practice in the near future.

**Keywords** lncRNA • Stem cells • Induced pluripotent stem cells • Pluripotency • Differentiation • Enhancer RNAs • miRNA sponge

---

D. Pal • M.R.S. Rao (✉)

Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore 560064, India

e-mail: [mrsrao@jncasr.ac.in](mailto:mrsrao@jncasr.ac.in)

## 8.1 Introduction

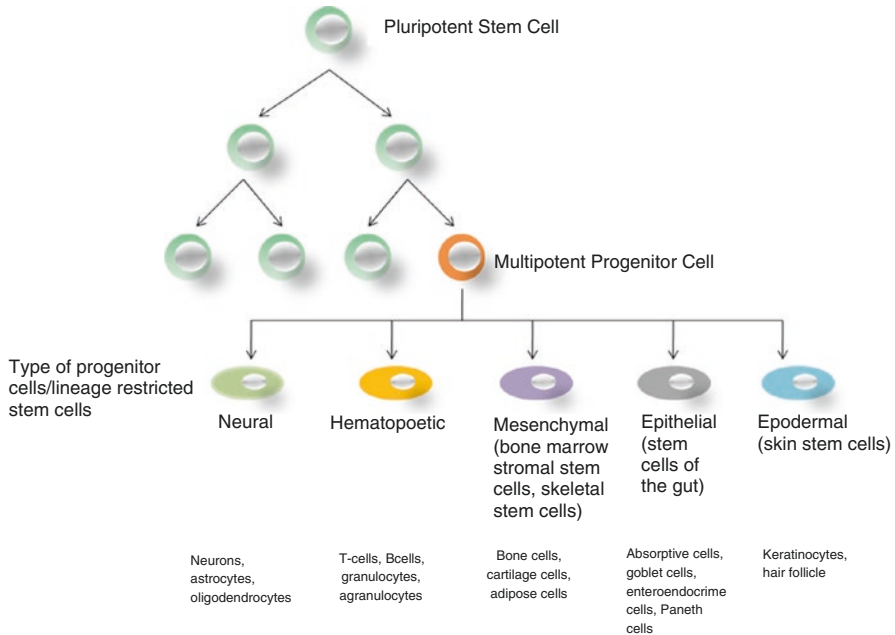
The functional relevance of long noncoding RNAs, previously thought of as by-products of transcription, is no longer a debatable topic. Even as the repertoire of lncRNAs is constantly on the rise, we ought to note that with increasing complexity of the living organisms, the percentage of the noncoding genome has also considerably increased [1]. One may attribute this feature to a concomitant increase in the genome size and hence an explosion in the proportion of “junk sequences.” But, increasing amount of evidence suggests that these noncoding transcripts play indispensable roles in the context of regulating developmental cues and signals, and their functional contribution becomes only more diverse when one moves up the evolutionary ladder. LncRNAs have been shown to participate in a wide variety of developmental processes like in regulating lineage commitment, specifying cellular identities and fates, in organogenesis, in imprinting of alleles during early development, and also in specification of the body pattern. A few of the first lncRNAs that were discovered through traditional gene mapping approaches are *Xist* [2] and *H19* [3], and interestingly enough they both play roles in regulating specific developmental processes, reiterating the aforesaid point that the evolution of the noncoding transcriptome in higher organisms has a functional significance and is not just an offshoot of the genomic size.

In the later part of the twentieth century, scientists were coalescing their efforts toward understanding how the genetic makeup of an individual regulates or predicts the development of various hereditary or familial diseases. While the field of genetics was resonating with breakthrough discoveries all over the world, cell biologists were not far behind in making discoveries that would ultimately form the basic model systems of study for the infinite complexities akin to the higher eukaryotes and mammals. In 1981, a report published by Martin Evans along with Matthew Kaufman [4] and another report published independently by Gail R. Martin [5] described the isolation of embryonic stem cells from the inner cell mass of blastocyst stage embryos and their subsequent maintenance under conditions of cell culture. These embryonic stem cells would, in the future, form the platform for carrying out research to understand the intricate signaling pathways and mechanisms governing mammalian development. They would further become the foundation for stem cell technology and stem cell therapy wherein damaged or defective tissues or organs would become replaceable due to the inherent properties of these cells (as will be discussed later). As a matter of fact, the groundwork for this technology was laid in the year 1995 by James Thomson and his colleagues at the Wisconsin Regional Primate Center (WPRC), University of Wisconsin-Madison, when they successfully isolated embryonic stem cells from the inner cell mass of rhesus monkeys making it the first report for the culture of nonhuman primate embryonic stem cells [6]. This led to the next achievement in 1998, whereby after an approval from bioethicists at the university, Thomson et al. derived human embryonic stem cells from leftover in vitro

fertilized human embryos [7] that won him the *Science's* “1999 Scientific Breakthrough of the Year” award. At the same time, the group led by John Gearhart obtained embryonic or primordial germ cells from the gonadal ridge of 5–9-week fetal tissue of electively aborted fetuses [8]. But ethical concerns over the use of human embryos for research purposes have paved the way for the generation of induced pluripotent stem cells (iPSCs), a groundbreaking discovery made independently by Thomson in his own lab [9] and Shinya Yamanaka [10] at the Kyoto University. Prof. Thomson reprogrammed adult human somatic cells into induced pluripotent stem cells by using a cocktail of four genes that were sufficient to impart “stemness” to the somatic cells. Research on the same lines carried by Yamanaka led to the identification of what is popularly known as the Yamanaka factors, namely, OCT3/4, SOX2, c-MYC, and KLF4, that could reprogram adult or embryonic fibroblasts into pluripotent stem cells. This discovery earned him the Nobel Prize for Physiology or Medicine in 2012. The implications of this discovery were immense because now theoretically, the cells from say, the skin of a person could be isolated and the clock turned backward to generate iPSCs which could be further differentiated to any cell type of the body and be used for the treatment of diseases like Parkinson's, spinal cord injury, Duchenne's muscular dystrophy, and so on, removing risks of transplants attacking their hosts.

In lieu of the importance of stem cell research, it becomes paramount to delve deeper into the mechanisms and key pathways that regulate the pluripotent nature of stem cells or guide them toward differentiation into various lineages. The term “pluripotency” has been derived from the Latin term *plurimus* meaning very many and *potens* meaning having power referring to the capability of stem cells to form various types of cells pertaining to any of the three germ layers of the body, namely, ectoderm, mesoderm, and endoderm. They also possess the power to divide and self-renew through continuous cell divisions, theoretically indefinitely (Fig. 8.1). Embryonic stem cells are those which are present in the embryo within the inner cell mass of the blastocysts, whereas adult stem cells reside in mature organs like the brain, skin, muscle, and bone marrow which act to regenerate parts of the tissues lost during processes of wear and tear or injury.

Soon after the establishment of stem cell cultures, widespread studies began on elucidating the molecular features of these cells. What factors maintain the “stemness” of these cells? What factors guide them into differentiation of either one or the other lineage? How can a bunch of similar cells give rise to an entire organism? While most of these questions have been addressed thoroughly by scientists around the world, nature never seems to exhaust us by posing new surprises and challenges. The discovery of noncoding RNAs revolutionized the understanding of the central dogma of biology and opened up a whole new avenue for exploration. Widespread studies that followed this discovery unraveled the ways in which these noncoding RNAs regulate crucial cellular pathways that govern the functioning of the individual cell and that ultimately manifests into functioning of the entire organism.



**Fig. 8.1** A pluripotent stem cell upon asymmetric division gives rise to a multipotent progenitor cell which can be of various categories as illustrated. Each category of multipotent progenitor cell or, lineage-restricted stem cell, can again undergo limited rounds of cell division or differentiate into cells of the corresponding lineage

## 8.2 Long Noncoding RNAs in Pluripotent Embryonic Stem Cells

### 8.2.1 Long Noncoding RNAs in Mouse Embryonic Stem Cells

As has been discussed in the previous chapters, lncRNAs play a significant role in modulating gene expression in several of the model systems. In this context, studies were initiated at a genome-wide level to unravel the cohort of long noncoding RNAs involved in the regulation of stem cell pluripotency. In biology, in order to understand the functional relevance of a molecule, a common approach is to selectively deplete it from the cell and observe the downstream effects with the help of techniques like microarray or RNA sequencing that shed light about the perturbations in expression of transcripts at the genome level. Guttman et al. [11] adopted such a methodology to address the function of a select class of lncRNAs known as long intergenic noncoding RNAs (lincRNAs) which as their name suggests are expressed from regions of the genomic segment present between two protein-coding genes. In this report, 226 lincRNAs were knocked down or depleted from embryonic stem cells by using short hairpin RNAs, and microarray was performed to analyze the effect. An interesting outcome of this study was that most of the lincRNAs act in

trans, at locations that are genomically farther away from their own site of transcription, adding a new dimension to the already known *cis* mechanism of action of lncRNAs. The more relevant outcome was, however, the discovery of 26 lincRNAs, knockdowns of which showed reduction in luciferase reporter activity, the expression of the luciferase gene being driven by the Nanog promoter. This observation established the fact that these lincRNAs contribute to the maintenance of pluripotency. Further experiments showed that ES cells depleted of these lincRNAs lead to loss of ES cell morphology characteristic to their pluripotent state along with a reduction in the expression of the core pluripotency factors. The fact that lincRNAs directly maintain the pluripotency of stem cells was subsequently corroborated by a more detailed analyses wherein knockdown of these lincRNAs resulted in the differentiation of stem cells toward one or the other lineage, recapitulating the phenomena that occurs when OCT4 or NANOG themselves are depleted from stem cells. It is interesting to note that at the molecular level, the lincRNAs are themselves directly regulated by the occupancy of one or more of the core pluripotency transcription factors at their promoters, establishing the importance of lncRNAs in coordinating mechanisms to maintain the pluripotent state of stem cells or repress their differentiation into various lineages.

While such holistic approaches as above have turned out to be crucial in discerning the function of the multitude of lncRNAs involved in ES cell circuitry, more direct studies with specific examples of lncRNAs have proved their indispensability for the proper functioning of ES cells. A study by Mohammed et al. [12] initiated at the genome level, to identify lncRNAs that are closely associated on the genomic loci serving as binding sites for OCT4 and NANOG, focused on two specific lncRNAs that play roles in fine-tuning the ES cell pluripotency/differentiation states. Directed knockdown of lncRNA AK028326, in essential a 3' fragment of the annotated 9 kb long lncRNA GOMAFU/MIAT, results in downregulation of Oct4 and other pluripotency markers and upregulation of markers of the trophectodermal and mesodermal lineages. Similar results were observed with lncRNA AK141205 although in this case, it was only OCT4 whose expression was concomitantly downregulated but not of Nanog. In accordance with these observations, AK028326 depletion in ES cells also resulted in a loss of ES cell colony morphology, suggesting a loss of pluripotent state, hence proving the necessity of this lncRNA in maintaining stem cell character. But an intriguing fact lay in the overexpression studies, wherein ectopic expression of these lncRNAs resulted in ES cells differentiating toward the neuroectodermal or mesodermal/ectodermal lineages, respectively. This suggests the diversity and complexity of functions of lncRNAs in stem cell biology. Basal levels of these lncRNAs might be important in maintaining the pluripotency of stem cells, whereas their overexpression may alter separate pathways altogether and guide the cells toward differentiation. *Linc86023*, named as Tc11 upstream neuron-associated lincRNA (TUNA or MEGAMIND), was similarly identified by Lin et al. [13] as a crucial molecule necessary for maintaining the pluripotent state of mouse embryonic stem cells. Being conserved remarkably across vertebrates, its loss of function resulted in altered cell morphology, reduced expression of pluripotency factors, and decreased cell proliferation, all of which are signatures of

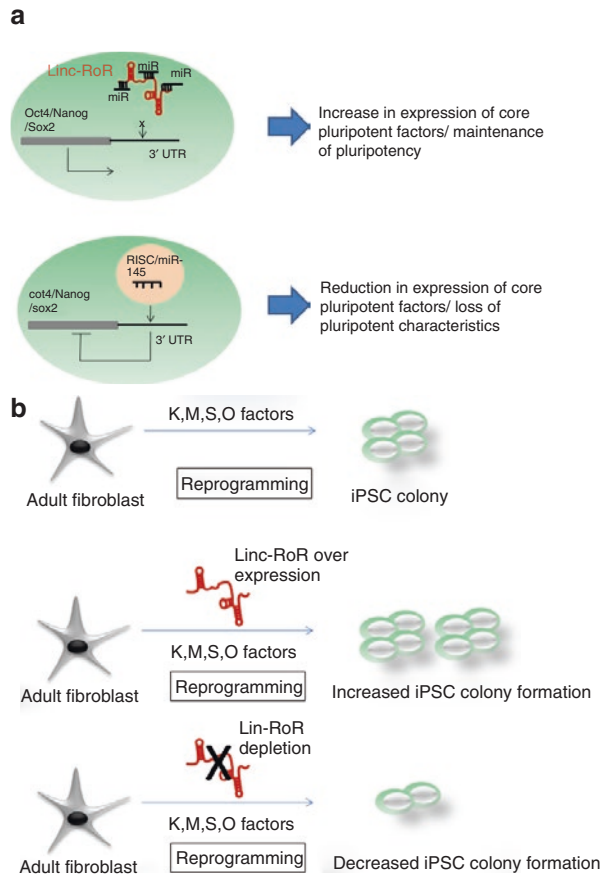
differentiation of otherwise self-renewing stem cells. TUNA was shown to form a multi-protein complex with RNA-binding proteins PTBP1, hnRNP-K, and NCL which occupy promoters of *Nanog*, *Sox2*, and *Fgf4* to maintain the pluripotent nature of stem cells. Again, in this case too, it was observed that TUNA is essential for the formation of neural precursors from stem cells in monolayer-adherent cultures, and its knockdown abolished the capacity of the stem cells to progress toward the neural lineage, emphasizing the pleiotropic nature of regulation of stem cell pathways by lncRNAs.

In another study, Chakraborty et al. [14] employed esiRNAs to downregulate around 594 previously annotated lncRNAs in mouse embryonic stem cells. The same esiRNA sequences, transcribed either in the sense or the antisense direction, were used to understand the cellular localization of the lncRNAs by FISH (fluorescent in situ hybridization). ES cells expressing GFP under the Oct4 promoter were transfected with the esiRNAs against the lncRNAs and scored for loss of GFP expression. Loss of GFP expression in the presence of esiRNA against a particular lncRNA would imply the probable involvement of that lncRNA in the maintenance of pluripotency. By this method, three lncRNAs were short-listed and were named pluripotency associated noncoding transcripts 1–3 or PANCT 1–3. Among them, PANCT 1 was characterized specifically because it showed the strongest effect on the expression of GFP. It was observed that PANCT 1 levels decreased steadily when ES cells were subjected to differentiation, and this was further confirmed by PANCT 1 knockdown studies wherein the cells showed reduction in pluripotency markers, reduction in DNA synthesis (exit from the dividing pluripotent state), and upregulation of various lineage-specific markers, suggesting a role for PANCT 1 in ES cell pluripotency regulation.

### ***8.2.2 Long Noncoding RNAs in Human Embryonic Stem Cells***

Studies on similar lines were performed in human embryonic stem cells by Ng et al. [15] who identified three lncRNAs, lncRNA\_ES1 (AK056826), lncRNA\_ES2 (EF565083), and lncRNA\_ES3 (BC026300) which had Oct4 or Nanog binding sites near their transcription start sites. OCT4 or NANOG RNAi experiments showed reduction in the expression of lncRNA ES1 and lncRNA ES2 and ES3, respectively. Downregulation of any of these three lncRNAs also resulted in loss of OCT4 expression, decrease in expression of a panel of pluripotency markers, and upregulation of genes involved in the formation of neuroectodermal, endodermal, and mesodermal markers. In accordance with studies performed before, it was observed that the lncRNAs mentioned above interact directly with either the core pluripotency factors or components of chromatin remodelers like SUZ12 (of the PRC2 complex) to determine active or silenced states of genes required for the maintenance of pluripotency or lineage differentiation. Linc-RoR (to be discussed in the next section) is yet another lincRNA that is necessary for the maintenance of the undifferentiated state of human embryonic stem cells [16]. Linc-RoR presents forth a unique example of

**Fig. 8.2** Mechanisms of action of linc-RoR. **(a)** In human embryonic stem cells, linc-RoR acts as a competing endogenous miRNA sponge and tethers away miR-145 from its targets which include the core pluripotency transcription factors, Oct4, Nanog, and Sox2; **(b)** During reprogramming, linc-RoR acts to regulate p53 DNA damage and cell apoptotic pathways to aid the formation of induced pluripotent stem cells (iPSCs)



the diverse mechanisms of action of lncRNAs. It possesses binding sites for several of the microRNAs that target and reduce the expression of the core pluripotency factors. By binding to and sequestering these miRNAs, linc-RoR acts as a “sponge” and prevents these miRNAs from degrading their target mRNAs that is required for the proper self-renewal of the human stem cells (Fig. 8.2a). Interestingly, linc-RoR transcription is itself regulated by the core transcription factors OCT4, NANOG, and SOX2, conforming to the well-known biological phenomenon of autocrine regulatory loop.

### 8.3 LncRNAs in Induced Pluripotent Stem Cells

iPSCs (induced pluripotent stem cells) are being explored as a promising candidate for stem cell-based therapies, albeit scientists are still trying to understand the pathways and regulatory mechanisms governing the framework and functioning of these



cells. In 2011, 5 years after the groundbreaking discovery of iPSCs, Loewer et al. [17] generated iPSCs from adult fibroblasts and analyzed gene expression changes on a microarray platform probing ~900 lincRNAs encoded in the human genome. About 207 lincRNAs were found to be either induced or repressed upon iPSC formation. One possible explanation for this observation is that reprogramming leads to changes in conformation of the chromatin genome wide, and opening up or compaction of protein-coding chromatin domains might directly affect the expression of the neighboring lincRNAs. However this possibility was ruled out because for each of the lincRNAs under consideration, there was no significant correlation between the neighboring protein-coding gene status. LincRNA-SFMBT2, lincRNA-VLDLR, and lincRNA-ST8SIA3 were found to be physically occupied at their promoters by Oct4, Sox2, and Nanog, indicating the functional intertwining of these lincRNAs and the core pluripotency factors in the formation of iPSCs. Furthermore it was observed that ES cells subjected to depletion of these lincRNAs by short hairpins showed a reduction in the formation of iPSC colonies in the case of lincST8SIA3, demonstrating the functional requirement of this lincRNA in iPSC formation. RACE (rapid amplification of cDNA ends) analysis recovered a transcript 2.6 kb long comprising four exons and no protein-coding activity. Overexpression of this lincRNA in fibroblasts followed by their reprogramming into iPSCs showed a two-fold increase in the formation of iPSC colonies (Fig. 8.2b). When a microarray analysis was performed upon knockdown of lincST8SIA3, it was found that genes of the p53 DNA damage response, and cell apoptotic pathways were upregulated, consistent with the phenotype observed when the lincRNA is depleted from the cells. p53 knockdown under the lincRNA knockdown conditions partially rescued the phenotype. This was one of the first reports to establish the role of a lincRNA in the formation and maintenance of iPSCs, opening up a whole new avenue of stem cell therapy and research. The lincRNA was aptly named linc-RoR or regulator of reprogramming (Table 8.1).

#### **8.4 LincRNAs in Lineage-Restricted Stem Cells and Differentiation**

While pluripotent stem cells can give rise to any of the cells specific to the three germ layers, multipotent cells are more specialized or committed in their differentiation capacity and can generate cells of a particular lineage, for example, only the neural lineage or the hematopoietic lineage. Since they possess the ability to self-renew and form a specific set of cell types, they are classified under stem cells. Multipotent stem cells exist both in the embryonic and the adult stages. In the embryonic stages, they act to generate nascent mature cells of the corresponding type, whereas adult stem cells are mainly responsible for the regeneration and repair of damaged adult tissues. In the following section, we discuss how multipotent stem cell networks are regulated by lincRNAs.



**Table 8.1** List of lncRNAs involved in maintenance of stem cell properties in mouse/human/induced stem cells

Name	Identification	Phenotype	Methods	Mechanism of action	Reference
LincRNAs containing "K4-K36" domain	ChIP followed by parallel sequencing	Loss of ES cell morphology and pluripotent characteristics	Lentiviral shRNA-mediated knockdown	LincRNAs are bound at their promoters by one or more of the pluripotency factors; they also associate with chromatin remodelers	[11, 18]
LincRNA AK028326 (3' fragment of GOMAFU/MIAT) LincRNA AK141205	Analysis of ChIP-PET results for Oct4 and Nanog	Loss of ES cell pluripotency; upregulation of markers of one or more of the three germ layers	siRNA-mediated silencing	LincRNAs are direct targets of Oct4 and Nanog, respectively; AK028326 participates in an autofeedback loop with Oct4 to regulate pluripotency	[12]
LincRNA TUNA/Megamind	Genome-scale shRNA library targeting lincRNAs	Reduced expression of pluripotent factors; decreased cell proliferation	shRNA-mediated knockdown	TUNA forms a RNP complex with PTBPL1, hnRNP-K, and NCL to occupy promoters of Nanog, Sox2, and Fgf4 in order to regulate pluripotency	[13]
PANCT 1	esiRNA library targeting 594 lincRNAs (RNAi screening)	Reduction in pluripotency markers; reduction on DNA synthesis; upregulation of lineage-specific markers	esiRNA-mediated silencing	Still to be addressed	[14]

(continued)

Table 8.1 (continued)

Name	Identification	Phenotype	Methods	Mechanism of action	Reference
LncRNA_ES1 (AK056826), LncRNA_ES2 (EF565083), and LncRNA_ES3 (BC026300)	Custom-designed microarray platform for lncRNAs	Loss of Oct4 expression; reduction in pluripotency markers; upregulation of lineage-specific markers	siRNA-mediated silencing	LncRNA ES1: expression decreased upon Oct4 RNAi LncRNA ES2 and ES3: similar effect upon Nanog RNAi; ES1 and ES2 associate with Suz12 (PRC2 component)	[15]
Linc-RoR	Previously identified	Increase in expression of core pluripotency factors or their reduction in expression	Overexpression of Linc-RoR or siRNA-mediated silencing	Functions as a competing endogenous miRNA sponge for miR-145 and prevents miR-145-mediated degradation for Oct4, Nanog, and Sox2	[16]
Linc-RoR (Linc ST8SIA3)	Microarray platform screening for ~900 lncRNAs	Increase in the formation of iPSC colonies upon overexpression and a concomitant decrease upon downregulation	Viral-mediated overexpression or shRNA-mediated knock down	Contributes to iPSC formation by regulating p53-mediated apoptotic pathways	[17]

### 8.4.1 Long Noncoding RNAs in Neural Stem Cells and Differentiation

One of the most evolutionarily susceptible and complex organs, the brain, consists of neurons that impart the sensory and motor functions and glia that act more as a support system for the cells of the brain itself. In the mammalian embryo, the fore-brain harbors the stem cells or the radial glia cells that divide and specialize to form both neurons and glia, i.e., astrocytes and oligodendrocytes. In the neonatal and subsequently in the adult stages, the quiescent neural stem cells are present in specific areas known as neurogenic niches which include the ventricular and subventricular zones and the subgranular zone of the dentate gyrus in the hippocampus [19]. In one of the genome-wide studies by Ng et al. [15], 35 lncRNAs were found which were highly expressed in mature neurons when compared to human embryonic stem cells or neural progenitors, among which knockdown of RMST (rhabdomyosarcoma 2-associated transcript), lncRNA\_N1, lncRNA\_N2, and lncRNA\_N3 led to lack of neuron generation in vitro. Overexpression studies showed the generation of an increased percentage of neurons, underlining the importance of lncRNA RMST in neuronal differentiation of human embryonic stem cells. RNA pulldown experiments revealed that RMST physically interacts with SOX2. Subsequently an overlap of the microarray datasets for siRMST and siSOX2 cells showed that they both co-regulate a specific subset of genes which are important for neurogenesis [20]. In fact, in cells where RMST was depleted by siRNA, it was observed that SOX2 binding to the target genes was ablated, underlining the importance of this lncRNA in acting as a co-regulator of SOX2-mediated neurogenesis.

Pax6 upstream antisense RNA (PAUPAR) is a lncRNA [21] situated 8.5 kb upstream of the Pax6 gene which codes for Pax6, a crucial transcription factor involved in neural progenitor cell proliferation, subtype specification, and spatial patterning in the brain. Downregulation of PAUPAR in neuroblastoma cells revealed that this lncRNA acts to maintain self-renewal of neural progenitor cells since its depletion led to increased neurite growth and increased appearance of neuronal differentiation markers in the cells. At the genic level, PAUPAR was found to be a large-scale regulator of gene expression in neural progenitor cells, affecting the expression of around 942 genes most of which belonged to synaptic regulation and cell cycle control. Interestingly, it was observed that Pax6 and PAUPAR not only co-occupy a common and distinct set of genes but also co-regulate several of them. Depletion of PAUPAR, however, does not affect the Pax6 occupancy at those genes, indicating that PAUPAR might act to recruit transcriptional coactivators at these sites of the genome and regulate their expression.

Much of the studies reported in the literature have focused on the functional significance of noncoding transcripts emanating from regions neighboring to protein-coding genes important for a specific developmental regime. lncRNA DALI [22], situated downstream from *Pou3f3* locus, exhibits concomitant expression pattern in the embryonic brain and in retinoic acid-treated ES cells with respect to *Pou3f3*, a protein known to have a role in the development of the nervous system.

In neuroblastoma cells, depletion of DALI leads to reduction in neurite growth, indicating DALI is required for proper differentiation of these cells. Genome-wide studies showed that DALI regulates genes like *E2f2*, *Fam5b*, *Sparc*, and *Dkk1* which are known to be pro-differentiation factors and negatively regulates genes that prevent the formation of neurites. An intriguing feature of this lncRNA is that it acts in cis on the neighboring *Pou3f3* gene where it physically contacts the gene at several locations as shown by 3C (chromosome conformation capture) technique. Simultaneously, it also acts in trans on genes involved in neuronal differentiation, cell cycle, neuronal projection formation, and intracellular signaling as shown by CHART-Seq (capture hybridization analysis of RNA targets). Furthermore, it also interacts with DNMT1, a DNA methyltransferase, and regulates DNA methylation at specific gene loci. DALI knockdown was shown to increase methylation at the CpG islands of *Dlgap5*, *Hmgb2*, and *Nos1* promoters, revealing an intricate network of neuronal gene regulation by lncRNA DALI.

A more recent study characterized PINKY (PNKY) lncRNA [23], a nuclear restricted neural-specific noncoding transcript, that maintains the neural stem cells of the ventricular zone in embryonic brains or ventricular-subventricular zones in adult brains. PNKY is expressed in neural stem cells but upon differentiation gets restricted specifically to the GFAP<sup>+</sup> astrocyte lineage. Knockdown of PNKY in monolayer cultures resulted in the generation of increased numbers of Tuj1<sup>+</sup> neuronal cells. When the shRNA construct of PNKY was electroporated into the embryonic brain and compared against the control brain, it was observed that the proportion of Sox2<sup>+</sup> stem cells were reduced but that of TBR2<sup>+</sup> transit-amplifying cells (an intermediate stage between stem cells and neurons) was not affected albeit there was an increase in Satb2<sup>+</sup> young neurons, indicating that PNKY maintains neural stem cells in the embryonic brain. Further exploration into its mechanism revealed that PNKY interacted with PTBP1, a repressor of neuronal differentiation. PTBP1 is known to regulate alternative splicing. Independently knocked down cells of PNKY and PTBP1 when subjected to RNA sequencing revealed that they regulate a common set of differentially perturbed genes and a common set of splice variants, suggesting a close coordination between these two molecules to maintain the neural stem cells in the brain.

#### **8.4.2 Long Noncoding RNAs in Hematopoietic Stem Cells and Differentiation**

The hematopoietic system of our body comprises of blood cells and the cells of the immune system both of which are critical for maintaining the body homeostasis. While red blood cells are the central pivots of oxygen transportation in the body and platelets of blood coagulation, white blood cells act to protect the body from the millions of pathogens it gets exposed to everyday, thereby forming the pillars of the immune system. Till and McCulloch, back in the early 1960s, [24] probed into

the components of blood that leads to its regeneration which led to the discovery of hematopoietic stem cells (HSCs). Like any other multipotent stem cells, they too can self-renew and give rise to all cell types of the blood. A mouse that has received an irradiation dose to kill its own blood-producing cells can survive if injected with these stem cells. However, HSCs can be either long-term stem cells that can constantly self-renew and support the blood system of an irradiated mouse (irradiation-depleted blood-producing cells) over several divisions or short-term progenitor or precursor cells that are restricted by the number of divisions that they can undergo. Since there are many types of blood cells, the differentiation of the HSCs has been characterized in the following manner: each stem cell can give rise to a myeloid progenitor cell and a lymphoid progenitor cell. Myeloid progenitor cells form the red blood cells, platelets, and the white blood cells which can again be divided into granulocytes (eosinophils, neutrophils, basophils) or agranulocytes (lymphocytes/macrophages). On the other hand, lymphoid progenitor cells give rise to T-lymphocytes, B-lymphocytes, and natural killer cells. HSCs have found widespread applications in the clinic. They are used for the treatment of leukemia and lymphoma wherein the patient's own blood cells are destroyed by radiation and replaced with a bone marrow transplant from a matched donor. Bone marrow transplants are also used for the treatment of genetic disorders of the blood like anemia and thalassemia.

One of the first ever lncRNAs reported to be involved in the maintenance of the hematopoiesis, specifically erythropoiesis, is lincRNA-EPS. Hu et al. [25] isolated cells from embryonic liver, a site for active erythropoiesis with cells of the erythroid lineage forming >90% of the liver and performed RNA-Seq analysis to identify the repertoire of lncRNAs which might be involved in the erythroid lineage. They concentrated their efforts on three types of cells, burst-forming erythroids, colony-forming erythroids, and Ter 119<sup>+</sup> cells that represent the three key stages of erythropoietic development and found that greater than 400 lncRNAs are perturbed during erythropoiesis. Out of these, 163 putative lncRNAs are upregulated and 42 are downregulated. They focused on those that show an increase in expression between colony-forming erythroids (progenitors) and Ter 119<sup>+</sup>-differentiated erythroblasts with an aim to understand the regulation of erythroid differentiation by lncRNAs. A probe into the functional aspects of lincRNA-EPS revealed that its depletion in erythroid progenitors led to increased apoptosis and reduction in proliferation of the progenitors in the presence of erythropoietin (erythropoietin promotes proliferation and subsequent differentiation of progenitors). This resulted in the reduced conversion of progenitors into terminally differentiated cells. On the other hand, under erythropoietin-starved conditions, progenitors that overexpressed lincRNA-EPS did not undergo apoptosis implying that lincRNA-EPS conferred anti-apoptotic phenotype to these progenitor cells. Microarray analyses in lincRNA-EPS overexpressing progenitors revealed the repression of a proapoptotic gene *Pycard*, which under normal circumstances activates caspase in apoptosis. Thus, lincRNA-EPS acts as an anti-apoptotic regulator during erythroid differentiation and development.

In a parallel study, Paralkar et al. [26] were interested in identifying the cohort of lncRNAs that are expressed in megakaryocyte-erythroid precursors from the bone marrow, megakaryocytes from cultured fetal liver progenitors, and fetal liver erythroblasts in mouse as well as in human cord blood erythroblasts. This comparative analysis identified approximately 1100 lncRNAs expressed during murine erythromegakaryopoiesis, out of which about 85% are present both in fetal and adult erythroblasts, suggesting the involvement of these lncRNAs in erythropoiesis. Interestingly, ~75% of the identified lncRNAs are expressed from promoter regions of genes, whereas ~25% are expressed from enhancer regions as evident from CHIP-Seq studies with transcription activation histone modification mark (H3K4me3) or enhancer modification mark (H3K4me1). Further CHIP-Seq studies with key erythropoietic transcription factors GATA1 and TAL1 in erythroblasts and GATA1, GATA2, TAL1, and FLI1 in megakaryocytes showed occupancy of most of the lncRNA loci with these transcription factors. Knockdown studies with shRNA constructs against several of these lncRNAs inhibited enucleation and maturation of erythroblasts into reticulocytes when the erythroblasts were subjected to differentiation in erythropoietin-containing medium. Lnc051, annotated previously as LINCRED1 along with ERYTHRA and SCARLETLTR, were a few of the candidate lncRNAs with potential roles in erythroid terminal maturation.

Eosinophils are another cell type that arise from the common myeloid progenitor and have a role to play in parasitic immunity and allergic diseases. CD34<sup>+</sup> human hematopoietic stem cells supplemented with IL-5, an eosinophil-specific cytokine for 24 h, were subjected to gene expression profiling by microarray upon which a novel transcript encoded within an intron on the opposite strand of the inositol triphosphate receptor type 1 (*Itp1*) gene was discovered [27]. It was named as EGO for eosinophil granule ontogeny lncRNA. The EGO transcript has two splice variant transcripts, EGO-A and EGO-B, and both of them are highly overexpressed upon stimulation of umbilical cord blood cells or bone marrow cells (CD34<sup>+</sup>) with IL-5 and only slightly induced in the presence of other cytokines like epoetin- $\alpha$ , SCF, GM-CSF, etc. RNA silencing experiments were performed in erythroleukemic cells to understand the functional significance of EGO lncRNA. Interestingly, it was found that levels of the eosinophil proteins MBP (major basic protein) and EDN (eosinophil-derived neurotoxin) were concomitantly reduced. CD34<sup>+</sup> umbilical cord blood cells expressing shRNA against EGO show incomplete development and die within 5 days of growth in IL-5 medium with respect to the control cells. Also, MBP and EDN levels were reduced considerably, suggesting that EGO lncRNA is necessary for the expression of these eosinophil proteins and hence normal eosinophilosis although the exact mechanism of action remains to be elucidated.

In another study, transcriptome profiling by microarray was performed on human peripheral blood neutrophils and on NB4 and HL-60 cells treated with all-trans-retinoic acid (ATRA) (cells directed toward granulocytic differentiation). This led to the identification of transcriptionally active regions between *HoxA1* and *HoxA2* genes [28]. The transcript was identified as a 483 nt RNA-spliced product from a primary transcript consisting of two exons and was subsequently named as HOTAIRM1 (HOX antisense intergenic myeloid 1). The expression of HOTAIRM1

was significantly induced when NB4 cells were treated with retinoic acid, but this phenomenon was not observed in the ATRA-resistant NB4r2 cell line. In fact, the expression of HOTAIRM1 was highly specific to the myeloid lineage as was evident by its specific upregulation in ATRA-treated NB4 or ATRA-treated K562 cells as compared to its baseline expression levels in the promyelocytic stages of NB4 cells. It was also found to exhibit low expression in hematopoietic stem or progenitor cells and was seen to be almost lacking expression in other organs like the brain, heart, pancreas, or skeletal muscle. In cells treated with shRNA against HOTAIRM1, induction of expression of *HoxA1*, *HoxA4*, and to some extent *HoxA5* was significantly attenuated in comparison to control cells, both the cell types being subjected to granulocytic differentiation by ATRA. Induction of beta2 integrin molecules, CD11B and CD18 (hallmarks of granulocyte maturation), was also abrogated, implying important roles for HOTAIRM1 in myelopoiesis. Studies by Wei et al. [29] provided insights into the mechanistic aspects whereby they observed that the transcription factor PU.1 binds to and regulates the levels of HOTAIRM1. PU.1 itself is an important transcription factor involved during myeloid differentiation, reaching highest levels in mature granulocytes and monocytes. Indeed in acute promyelocytic leukemic cells, dysregulation of HOTAIRM1 is due to the binding of PML-RAR $\alpha$  to PU.1 and subsequent prevention of PU.1-mediated transactivation of various myeloid differentiation genes.

An extensive study carried out by Hu et al. [30] was aimed at cataloging the long intergenic ncRNAs involved in T-cell maturation and differentiation. They obtained 42 subsets of T-cells which included CD4-CD8 double negative (DN), double positive (DP), single positive (SP) thymic T-cells, T-regulatory (Treg) cells from the lymph nodes of mice, and T<sub>H</sub>1, T<sub>H</sub>2, T<sub>H</sub>17 (T-helper cells), and induced Treg (iTreg) cells from in vitro cultures derived from naïve CD4<sup>+</sup> T-cells. Across all of the T-cell types, they identified 1542 genomic regions that were expressing lincRNAs individually or in clusters (more than one lincRNA expressed from the same locus). Quite intriguingly, when the data was classified based on the expression status of lincRNAs or protein-coding genes in specific subsets like only DN cells, DP<sup>+</sup>SP<sup>+</sup>Treg cells, and naïve CD4<sup>+</sup> T<sub>H</sub> cells, it was observed that 48–57% of the expressed lincRNAs were lineage specific as compared to 6–8% of mRNAs, and only 13–16% of lincRNAs were shared between subsets of T-cells in contrast to 70–80% of protein-coding transcripts. When followed over a time scale of differentiation, many of the lincRNAs were downregulated at 4 h of T-cell differentiation from naïve CD4<sup>+</sup> T-cells only to again regain the expression at 48–72 h implying their role in T-cell activation. Many of them, like LincR-Chd2-5'-74 K, remained mostly silenced after differentiation, while many others, like LincR-Sla-5'-AS, were induced at 4 h of differentiation with a gradual subsidence of expression at later stages. CHIP-Seq and knockdown studies of two important transcription factors STAT4 and STAT6 revealed that STAT4 preferentially binds to and potentially regulates lincRNAs specific to T<sub>H</sub>1 cells and STAT6 for T<sub>H</sub>2 cells. Linc-Ccr2-5'-AS was further studied whereby it was found that depletion of this lincRNA resulted in reduction of expression of *CCr 1*, 2, 3, and 5 genes (chemokine receptors), all of which are located neighboring to the lincRNA genomic locus. Moreover, in vivo depletion of this



lincRNA led to decreased migration of  $T_H2$  cells to the lung, a process which is dependent on chemokine signaling. This study along with a study conducted by Ranzani et al. [31] gives a comprehensive insight into the lincRNAs with potential regulatory functions during lymphocyte differentiation, maturation, activation, and functioning. On similar lines, Casero et al. [32] studied the lincRNA profile of ten cell types of the lymphoid lineage: (1)  $CD34^+ CD38^- Lin^-$  cells enriched in hematopoietic stem cells and obtained from the bone marrow; (2) three lymphoid progenitor populations such as common lymphoid progenitors, lymphoid-primed multipotent progenitors, and B-cell-committed progenitors from the bone marrow as well; (3)  $CD34^+$  but  $CD4 CD8$  double negative populations (Thy1, Thy2, Thy3) from the thymus; and (4) T-cell-committed populations from the thymus again. A set of 9444 lincRNA genes were identified among which 3348 are known. Yet again, most of these lincRNAs showed a highly stage-specific manner of expression, being restricted to one or the other lineage in comparison to their protein-coding counterparts. They were also positively correlated in expression with several of the protein-coding genes located either in trans or in cis to them, reinforcing the role of lincRNAs in the maintenance and/or differentiation of progenitors in the bone marrow and the thymus.

#### ***8.4.3 Long Noncoding RNAs in Muscle Stem Cells and Differentiation***

Skeletal muscle, a striated muscle tissue comprising about ~40% of the body weight, is composed of multinucleated contractile muscle cells known as myofibers which in turn are generated by the fusion of progenitor cells or myoblasts [33]. Myofibers remain constant in number in the neonatal stages, but postnatally they grow in size by the fusion of a group of stem cells known as satellite cells. Satellite cells are the stem cell population of the adult muscle tissue, being quiescent under normal physiological conditions but quickly reenter active cell division in case of muscle injury to regenerate damaged or wounded tissue. Although the regenerative capacity of muscle tissue was observed as early as the nineteenth century, it was only in 1961 that two independent studies by Alexander Mauro and Bernard Katz actually proved their presence by electron microscopy in the sublamellar region of myofibers [34]. At the molecular level, quiescent satellite cells express Pax7, and only upon activation of mitosis, they start expressing myogenic transcription factors like MYOD, MYOGENIN, MYF5, and DESMIN [34]. About 24 kb upstream of the gene-encoding transcription factor MYOD1, two regulatory regions are present for the gene itself, referred to as CE (core enhancer) and DRR (distal regulatory region). Through a series of RNA-Seq experiments, it was observed that these enhancer regions, characterized by the presence of histone modifications H3K4me1 and H3K27ac along with p300/CBP/RNAP II occupancy, are actually transcriptionally active, giving rise to enhancer RNAs or eRNAs [35].

In an approach to dissect out the role of these eRNAs, a screening was done for ten siRNAs designed against various regulatory regions upstream of *MyoD*, and interestingly enough it was observed that the levels of *MyoD* diminished drastically only in the case of siRNA targeting the CE region. It was further observed that <sup>CE</sup>eRNA acts in cis to regulate the transcription of *MyoD1* by enhancing the occupancy of RNAPol II at *MyoD1* proximal regions. On a similar note (yet with a twist in the tale), it was discovered that <sup>DRR</sup>eRNA acts in trans to enhance the expression of *MyoG* and *Myh*, thereby acting to promote myogenic differentiation. The role of eRNAs, a class of lncRNAs, was established in this study, and their mechanisms of function which mainly includes modification of chromatin organization by either causing nucleosome repositioning or by effecting recruitment of various chromatin modifiers were elucidated. Parallel studies by Mueller et al. [36] on the *MyoD* upstream locus led to further characterization of a lncRNA MUNC (*MyoD* upstream noncoding) which initiates transcription in the <sup>DRR</sup>eRNA locus. Downregulation and overexpression of MUNC in undifferentiated muscle cells in culture caused a respective decrease or increase in the levels of key myogenic transcription factors like MYOGENIN, MYH3, and MYOD itself to some extent. In vivo, when siRNA against MUNC was injected into the tibia anterior (TA) muscles of mice followed by muscle injury with cardiotoxin, it was observed that over a period of 2 weeks of muscle regeneration, the levels of MYOGENIN, MYH3, and MYOD were significantly lower in the siMUNC tissues. This was accompanied with a decrease in myofiber diameter and increase in inflammatory infiltrates in the regenerated tissue, reestablishing the importance of lncRNAs in myogenesis.

Analysis of the transcriptional start sites and promoter elements of the muscle-specific miRNA loci, pre-miRNA-133, and pre-miRNA-206 revealed the presence of lincRNA linc-MD1 [37], which indeed was the first identified muscle-specific lincRNA. Linc-MD1 is specifically activated when myoblasts, satellite cells, or MYOD-trans-differentiated fibroblasts (muscle cells derived from myoblasts) were subjected to differentiation. This lncRNA was found to be expressed in newly regenerating muscle fibers. Mechanistically, it acts as a competing endogenous RNA or ceRNA whereby it acts as a sponge or decoy to sequester miRNAs such as miR-133 and miR-135 which otherwise bind to their targets MEF2C and MAML1, both of which are important transcription factors required for myogenesis. In an independent study conducted by Legnini et al. [38], it was shown that another myogenically important RNA-binding protein, HuR, is involved in the cross talk between Linc-MD1 and miR-133. RNA interference experiments for HuR revealed a consistent decrease in the cytoplasmic accumulation of linc-MD1 and increase in the pools of miR-133a/miR-133b. A series of experiments thereafter confirmed that it is the binding of HuR to linc-MD1 that increases its presence in the cytoplasm, aiding its miRNA sponging activity at the expense of miR-133 biogenesis (miR-133 being a result of processing of linc-MD1 by Drosha). In a positive feed-forward loop, linc-MD1 and HuR regulate the differentiation of muscle progenitors and hence myogenesis.

One of the first lncRNAs to be discovered with respect to muscle differentiation was SRA (steroid receptor RNA activator). MYOD co-immunoprecipitates with p68/p72 DEAD box RNA helicases, and both of them were shown to interact with SRA in skeletal muscle cells through immunoprecipitation experiments followed by PCR to score for the associated RNA [39]. Luciferase reporter assay experiments were performed wherein the muscle-specific creatinine kinase enhancer was fused upstream of the luciferase gene and transfected into fibroblast cells along with p68, p72, or SRA expression vectors, individually or in combination. No effect was observed on the luciferase gene expression in any of the above cases. However, expression of MYOD either alone or in conjunction with either of the protein (p68/p72) or RNA (SRA) interactors enhanced the luciferase reporter activity. The highest enhancement was observed when all the three (p68/p72, SRA, and MYOD) were co-expressed, thereby establishing that p68/p72 and SRA act as transcriptional coactivators of MYOD. In fact RNA silencing experiments further proved that these three coactivators of MYOD are essential for the differentiation of muscle cells into myotubes. In another interesting study, it was shown that the SRA transcript is actually alternatively spliced to give rise to a protein counterpart SRAP [40]. In undifferentiated myoblasts versus differentiated myotubes, the ratio between the noncoding SRA and the coding SRAP is largely in favor of the noncoding counterpart. In primary human satellite cells subjected toward differentiation, a similar observation was made, SRA levels being observed to be higher than SRAP. Through a series of luciferase and chromatin immunoprecipitation experiments, SRAP was found to physically bind to SRA and prevent it from acting as the coactivator of MyoD, thus unraveling a network of proteins and RNA, fine-tuning the regulation of myogenic differentiation.

A large imprinted locus known as the Dlk1-Gtl2 (delta-like 1 homolog-gene trap locus 2) contains many protein-coding, noncoding, and paternally/maternally imprinted genes, GTL2 being one of the noncoding RNAs [41]. It is also known as MEG3 in humans. A knockout mouse was generated, the knockout locus encompassing the promoter region and exons 1–5 of the *Gtl2* gene. It was observed that while the mice carrying the deletion at the paternal locus survived and were healthy, the mice carrying the same at the maternal locus did not survive. Intriguingly enough, while the *Glt2* knockout embryos showed no abnormalities in organs like the brain, heart, liver, kidney, lung, or spleen, their skeletal muscles showed severe defects of formation. The myofibers of the paraspinal muscles were not only small and rounded with peripherally placed nuclei; they were also lower in number. It was one of the first evidences of a lncRNA being necessary in vivo for the proper development of muscles.

Genome-wide binding studies for a transcription factor Yin yang 1 (YY1), a repressor of muscle differentiation genes in proliferating myoblasts, showed that it actually binds to many intergenic loci in the genome along with previously known or unknown protein-coding loci [42]. The potential linc RNA loci were 63 in number and were named as YAM (YY1-associated muscle lincRNA). One such loci, *Yam-1*, located on chromosome 17, was found to be positively regulated by YY1 in

proliferating myoblasts. It was observed that YAM-1 was present in abundance in proliferating myoblasts or in the limb muscles of young mice displaying active myogenesis, whereas it was downregulated during myogenic differentiation of myoblasts in vitro or in vivo in older mice with reduced perinatal myogenesis. These observations were further confirmed by RNA silencing experiments. A probe into the mechanisms revealed that YAM-1 positively regulates the expression of its downstream effector miR-715 which in turn negatively regulates Wnt7b. Wnt7b is known to promote muscle differentiation. YAM-1 knockdown led to the upregulation of Wnt-7b, putting forth a mechanism whereby the anti-myogenic differentiation capacity of YAM-1 might be mediated through miR-715-mediated repression of Wnt7b. A study of the other YAMs showed that while YAM-2 and YAM-4 are pro-myogenic factors during the early stages of muscle differentiation, YAM-3 is again anti-myogenic, providing ample evidence of the tight regulation of muscle differentiation by lncRNAs.

Klattenhoff et al. [43] analyzed RNA-Seq data for the expression of lncRNAs in mouse embryonic stem cells as well as in differentiated tissues and focused on one such lncRNA AK143260. They observed that this lncRNA exhibited higher expression in the heart and hence termed it as *Braveheart* (*Bvht*). BVHT was depleted from mouse ESCs by shRNA, and the cells were subjected to in vitro cardiomyocyte differentiation by the embryoid body method. Cardiomyocytes are the muscle cells of the heart. It was observed that in the control cells, ~25% of the embryoid bodies displayed spontaneous rhythmic beating as compared to only ~5% of the knockdown cells. Global gene expression analyses by RNA-Seq in BVHT-depleted cells revealed that a multitude of transcription factors coding genes like *Mesp1*, *Hand1*, *Hand2*, *Nkx2.5*, and *Tbx20* were not activated when the cells were differentiated into the cardiac lineage, establishing the importance of BVHT in cardiac lineage specification. An ES cell line harboring a doxycycline-inducible MESP1 overexpression plasmid, when subjected to cardiac differentiation along with MESP1 induction, was able to rescue the BVHT depletion phenotype. This proved that BVHT acts upstream of MESP1 during cardiac differentiation of ES cells. Studies by Xue et al. [44] were aimed at unraveling the secondary structure of BVHT. It was shown that BVHT possesses a AGIL motif in its 5' domain. With the help of CRISPR/Cas9 system, they generated a 11 nt deletion in this motif (*bvht<sup>dagil</sup>*). Interestingly, *bvht<sup>dagil</sup>* ES cells showed significantly reduced beating during the cardiac differentiation as compared to the wild-type cells. As observed earlier with BVHT knockdown cells, *bvht<sup>dagil</sup>* cells showed a lack of activation of major cardiac transcription factors like *Nkx2.5*, *Hand2*, *Gata4*, and *Gata6*. A protein microarray was employed to understand the interaction partners of *bvht<sup>dagil</sup>* wherein CNBP or ZNF9, a zinc finger transcription factor, was found to be an interesting interacting candidate for *bvht<sup>dagil</sup>* lncRNA. These studies suggested that the lncRNA protein interaction networks are crucial components of cell fate decisions and lineage commitment.

A brief representation of the various lncRNAs involved in the maintenance and/or differentiation of stem cells for the neural, hematopoietic, and muscle lineage has been depicted in Fig. 8.3.

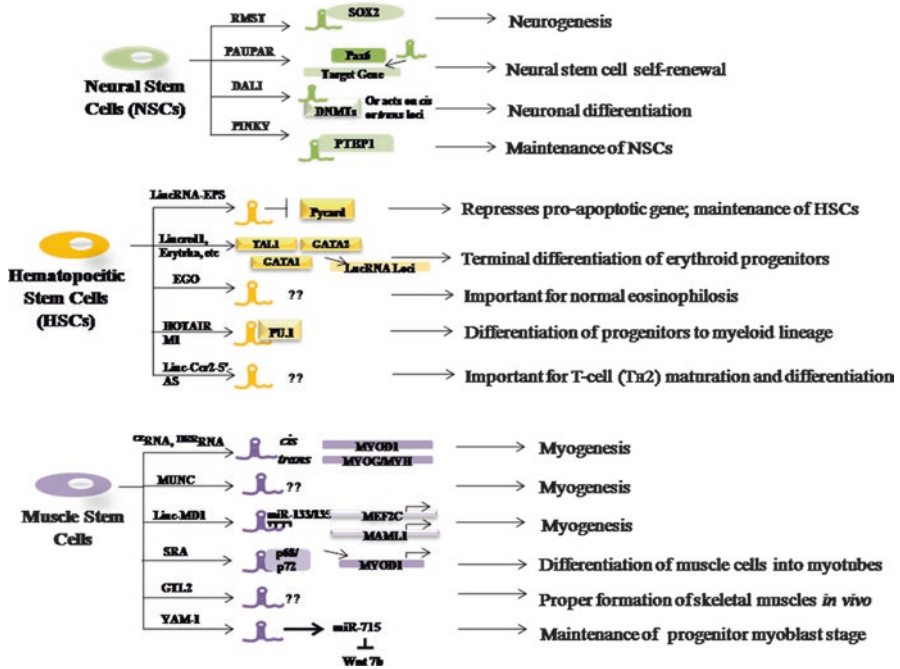


Fig. 8.3 Representative examples of lncRNAs that either maintain the stem cell state of somatic stem cells or promote their differentiation/terminal maturation. The mechanisms can either be through interaction with protein partners, regulating gene loci in cis or trans, or acting as competing endogenous RNAs

### 8.4.4 Long Noncoding RNAs in Epidermal Stem Cells and Differentiation

The skin is one of the most sturdy and versatile organs of the body in that it not only acts as a protective barrier, providing protection to the body against microbes and dehydration, but also constantly participates in maintaining homeostasis through withstanding temperature changes and providing tactile sense to the body. The stem cell niche of the skin is involved in constantly regenerating the epidermal hair and also in regenerating epidermal tissue after an injury or a wound. In the embryo, post-gastrulation, it is the neuroectoderm that gives rise to the epidermis that essentially starts as a single layer of uncommitted progenitor cells but finally forms a stratified structure, hair follicles, and the sebaceous glands or the apocrine (sweat) glands. In adults, the skin epithelium is made up of blocks, each block being made up of a pilosebaceous unit consisting of hair follicle (HF) and sebaceous gland along with the surrounding interfollicular epidermis (IFE). The HF contains multi-potent stem cells that regenerate the hair as well as supply cells for replenishing damaged ones post injury for both the hair follicle and the epidermis. The IFE contains progenitor cells too that maintain tissue integrity and self-renewal under

normal circumstances. Various types of signaling pathways including Wnt/ $\beta$ -catenin, BMP, Notch, and Shh have been implicated in the self-renewal and/or differentiation of the epidermal stem cells [45].

To understand the role of lncRNAs in keratinocyte differentiation from epidermal stem cells, Kretz et al. [46] performed high-throughput sequencing of human primary keratinocytes at various days of calcium-induced differentiation and uncovered 295 annotated and 835 unannotated putative lncRNAs. Keratinocytes are the major cell type of the epidermis. At 3 and 6 days of differentiation, the lncRNA reads obtained were compared with that of 0 day (progenitor population), and it was observed that there were significant perturbations at each of the stages of differentiation studied. To have a broader picture of previously unknown lncRNAs that may have a role to play in suppressing differentiation of various types of progenitors, RNA was obtained from keratinocytes, adipocytes, and osteoblasts in the progenitor and differentiated states and hybridized to tiling arrays. One interesting hit came in the form of the lncRNA NR\_024031, termed hitherto as ANCR (antidifferentiation noncoding RNA) which was repressed in each of the model systems studied. ANCR, located in human chromosome 4, consists of three exons, miRNA4449-encoding sequence and a snoRNA-generating sequence in the introns 1 and 2, respectively. It codes for a 855-bp-long transcript that was found to be significantly downregulated at days 3 and 6 of keratinocyte differentiation. Interestingly, the ANCR lncRNA is expressed in multiple human tissues and is concomitantly repressed in many differentiated cell types, indicating its functional relevance in the transition from progenitor to differentiated states. RNAi against ANCR in progenitor keratinocytes induced the expression of many differentiation-related genes like filaggrin, loricrin, keratin 1, small proline-rich proteins 3 and 4, involucrin, S100 calcium-binding proteins A8 and A9, and ABCA12. Microarray analyses under such conditions revealed the perturbation of 388 genes including genes responsible for epidermal differentiation, keratinization, and cornification. Furthermore ANCR was depleted in regenerated, organotypic epidermal tissue, a system recapitulating most aspects of the human epidermis. Interestingly similar results were observed, with even the epidermal basal layer expressing differentiation genes which otherwise is not known to express such genes. Thus ANCR seems to be necessary to keep differentiation-related genes from expressing in the progenitor cell niche of the epidermis and hence in maintaining the identity of keratinocyte progenitors.

This group also identified TINCR (terminal differentiation-induced ncRNA) on chromosome 19 of the human genome encoding a 3.7 kb transcript, highly expressed, by greater than 150-fold, during epidermal differentiation [47]. It was shown to be enriched in the differentiated layers of human epidermal tissue, indicating its role in the differentiation of keratinocytes. When TINCR was downregulated by RNAi in organotypic culture system, expression of key differentiation genes was perturbed in expression although the epidermis stratified normally. Transcript profiling revealed 394 genes to be affected in expression, including those involved in the formation of the epidermal barrier. Specifically, caspase-14 required for proteolysis during the formation of the barrier was reduced drastically, and protein-rich keratohyalin granules and lipid-rich lamellar bodies were ill-formed in the epider-



mis. To elucidate the mechanism of action of TINCR, an interactome analysis was done using a protein microarray consisting of approximately 9400 recombinant proteins. STAU1 protein showed the highest affinity of binding with TINCR. Although STAU1 has not been previously implicated in epidermal differentiation, it was found that STAU1 depletion recapitulated effects of TINCR depletion, and there was a significant overlap of regulated genes between siSTAU1 and siTINCR cells with a predominance of genes involved in keratinocyte differentiation. Together, TINCR and STAU1 were shown to bind to and functionally stabilize mRNAs encoding key structural and regulatory proteins necessary for keratinocyte differentiation.

#### **8.4.5 Long Noncoding RNAs in Spermatogonial Stem Cells and Differentiation**

Spermatogenesis is a physiological process which defines the formation of the spermatozoa through a series of differentiations undergone by progenitor cells referred to as spermatogonial stem cells (SSCs). In the embryonic stages, primordial germ cells (PGCs) represent a population of cells that arise in the epiblast at 7–7.5 dpc of development and migrate to the gonadal ridges at around 12.5 dpc. Once they reach the gonadal ridge, the erstwhile proliferating PGCs enter into a mitotic arrest and reenter the cell cycle only after birth. They populate the basement membrane of seminiferous tubules generating a niche comprising the Sertoli cells, Leydig cells, and surrounding interstitial cells. They undergo constant self-renewal to generate millions of spermatozoa daily. Three types of spermatogonia were initially identified based on the nuclear architecture [48]: type A consisting of a more decompacted chromatin structure, type B spermatogonia consisting of a more heterochromatic chromatin, and an intermediate type between the both. Type A spermatogonia are the undifferentiated cells further classified into three types:  $A_{\text{single}}$  ( $A_s$ ),  $A_{\text{paired}}$  ( $A_{\text{pr}}$ ), and  $A_{\text{aligned}}$  ( $A_{\text{al}}$ ) depending on the arrangement on the basement membrane of the seminiferous tubule. A single division of  $A_s$  leads to the formation of either (1) a  $A_{\text{pr}}$  that generates two  $A_s$  post-cytokinesis or (2) the two resulting cells remain connected by a cytoplasmic bridge that generates a chain of four  $A_{\text{al}}$  in the next round of division. The four  $A_{\text{al}}$  spermatogonia undergo mitotic divisions to generate 32  $A_{\text{al}}$  spermatogonia, and 4–16 such chains are finally committed to differentiation. The  $A_{\text{al}}$  spermatogonia give rise to the type B spermatogonia which generate primary spermatocytes that undergo meiosis. Two rounds of meiosis give rise to secondary spermatocytes and haploid spermatids. The haploid spermatids then undergo morphological changes through 16 steps (in mouse) finally forming the mature spermatozoa.

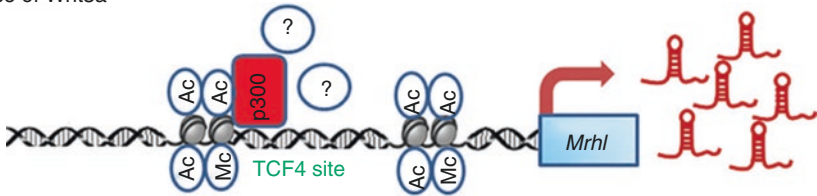
One of the first identified lncRNAs in our laboratory which was shown to have a functional role in spermatogonial physiology is MRHL (mouse recombination hotspot locus) RNA [49]. It is a 2.4 kb transcript, expressed in the adult mouse testis and processed in vitro by the Drosha machinery to a 80 nt processed transcript [50]. To gain an understanding of its function in the mammalian testis [51], the RNA was



downregulated in the mouse spermatogonial cell line (Gc1-Spg). Subsequent microarray analyses revealed a host of signaling pathways being affected, a prominent and noteworthy one being the Wnt signaling. Mass spectrometry identified p68/DDX5 helicase as one of the interacting proteins of MRHL following which it was shown that in *mrhl* RNA-depleted conditions, p68 translocates from the nucleus to the cytoplasm and aids the shuttling of Wnt signaling effector protein  $\beta$ -catenin into the nucleus resulting in subsequent activation of Wnt signaling. Thus, in mouse spermatogonial cells, *mrhl* RNA negatively regulates Wnt signaling through interaction with p68. Genome-wide occupancy studies of MRHL on the chromatin were performed through ChOP-Seq (chromatin oligoaffinity purification followed by sequencing) [52]. This study revealed that MRHL physically occupies 1400 loci among which 37 loci are regulated by this lncRNA. These loci are termed as the GRPAM loci (genes regulated by physical association of MRHL) which include genes involved in Wnt signaling, spermatogenesis, and differentiation. ChIP- and shRNA-mediated downregulation studies showed that Wnt signaling acts to downregulate MRHL RNA when spermatogonial cells are exposed to Wnt3a ligand. A detailed investigation into the mechanism of Wnt-mediated MRHL RNA downregulation revealed CTBP1 as the corepressor that increasingly occupies the promoter of *Mrhl* and establishes repressive histone modifications like H3K9me3 on the promoter leading to repression of transcription of the RNA [53]. Interestingly, it was also observed that upon Wnt treatment of spermatogonial cells, various premeiotic (*c-kit*, *Dmc1*, *Stra8*, *Lhx8*) as well as meiotic markers (*Zfp42*, *Hspa2*, *Mtl5*, and *Ccna1*) were significantly upregulated. Rescue of MRHL in trans did not abrogate these changes indicating that additional factors are necessary for the upregulation of these meiotic markers which are activated only under Wnt conditions. These studies thus proved that *mrhl* RNA acts at the chromatin level to regulate key aspects of spermatogonial differentiation initiated by Wnt signaling (Fig. 8.4).

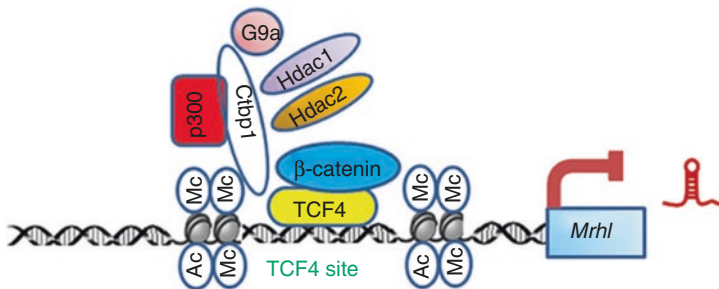
A comprehensive genome-wide study was recently carried out by Sun et al. [54] wherein they performed lncRNA microarray analysis from 6-day-old (neonatal) and 8-week-old (adult) testis. They found that out of the ~14,000 lncRNA genes represented on the microarray, ~8000 (56%) exhibited expression above background, and 37% of these (~3000 lncRNAs) showed differential expression between the two stages studied. They classified all lncRNAs perturbed into specific groups such as exonic sense or antisense, intronic sense or antisense, and bidirectional or intergenic based on their locations and directions of transcription and found interesting correlations between the expression of these lncRNAs and their neighboring protein-coding counterparts. For example, *Ccnd2*-coding gene expression occurs primarily in spermatogonia and is important for their self-renewal. Both *Ccnd2* and its associated sense lncRNA AK011429 were found to be downregulated in the adult testis tissue. Similarly, AK077193, expressed antisense to *Sycp2* (synaptonemal complex protein 2), was upregulated in the adult testis, and the expression was positively correlated with that of *Sycp2* itself, a gene required during meiosis in spermatocytes. LncRNA AK00574 was found to be specifically upregulated and highly expressed along with the protein-coding gene *Spata17* from whose intron it is transcribed in an antisense direction. *Spata17* is involved in male germ cell apoptosis

Absence of Wnt3a



Presence of Wnt3a

(Meiotic progression)



**Fig. 8.4** Model summarizing the changes occurring at the proximal promoter region of *mrhl* RNA at the TCF4 binding site upon Wnt signaling activation with respect to the binding of different proteins like  $\beta$ -catenin-, TCF4-, Ctbp1-, and Ctbp1-associated proteins (p300, G9a, Hdac1, and Hdac2). p300 binds at the TCF4 binding site even in the absence of Wnt3a, and other proteins (?) could be associated with p300 for regulation of *mrhl*RNA expression. The changes in histone modifications are also shown. In the presence of Wnt3a, *mrhl*RNA downregulation possibly leads to meiotic commitment and differentiation

in the adult testis. Although the specific functions of these lncRNAs need to be elucidated, this study has listed a cohort of lncRNAs with possible functions in male germ cell differentiation and testes development.

Similar high-throughput transcriptome analysis was performed by Li et al. [55] on primary Thy1<sup>+</sup> spermatogonial stem cell cultures in various conditions such as (1) in the presence of the growth factor GDNF, (2) 18 h post-depletion of GDNF, and (3) post 8 h reexposure to GDNF in the depleted cultures. Interestingly, normal cultures growing in the presence of GDNF showed expression of twice the number of lncRNA transcripts as compared to protein-coding mRNAs, whereas in the depleted and replenished cultures, an equal proportion of both types of transcripts was perturbed. LncRNA 033862 was found to have the most significant expression changes upon GDNF withdrawal in SSC cultures. Its expression decreased upon GDNF withdrawal for 18 h, reappeared post 8 h of GDNF reexposure, and underwent almost 97% reduction upon 30 h of GDNF removal from cultures. Tissue-specific expression analysis revealed that this RNA is highly expressed in mouse testis and brain. In the mouse testis specifically, it was expressed during the immediate postnatal stages (P1–P3) with subsequent reduction in levels at P7 and P10, indicating its role in gene regulation in the spermatogonial progenitor cells of the

testis. Indeed, in situ hybridization showed expression of this lncRNA in the spermatogonial cells located in the basement membrane of seminiferous tubules of testis. Chromatin isolation by RNA purification (ChIRP) experiments revealed that lncRNA 033862 bound physically to the *Gfra1* locus on mouse chromosome 19. lncRNA 033862 is transcribed in an antisense direction from exon 9 of *Gfra1* (GDNF family receptor). Knockdown experiments using lentiviral shRNA in SSC cultures led to increased apoptosis, significant changes in morphology with reduction in colony size and downregulation of SSC-associated self-renewal genes like *Bcl6b*, *Ccnd2*, and *Pou5f1*, and reduction in expression of *Gfra1* itself. Differentiation genes like *Stra8*, *Sycp1*, and *c-kit* were however not affected, thereby establishing that lncRNA 033862 is necessary for SSC self-renewal and maintenance. Furthermore, in vivo transplantation of the lncRNA knocked down cells into testis showed lower colonization of testis from donor cells as compared to controls. *Gfra1* encodes the co-receptor for GDNF in SSCs. The above studies proved the necessity of lncRNA 033862 in SSC maintenance and indicated that absence of GDNF signaling which led to reduction in expression of lncRNA 033862 might be the cause for transcriptional silencing of *Gfra1*, revealing an intricate role of this lncRNA in spermatogonial stem cell gene regulation.

TSX (testis-specific X-linked) is a lncRNA that is expressed from the highly characterized X-inactivation center in mammals being encoded upstream of the lncRNA locus *Xite* [56]. An expression pattern analysis revealed that while in female mice, TSX is expressed at higher levels in the brain than in the gonadal tissue; it is the reverse in males. Male gonadal tissue showed 10–100 times higher expression as compared to the brain. Isolation of male germ cells and further analyses showed that while in type A and B spermatogonia, TSX levels are comparatively lower; it is upregulated by 40-fold in the pachytene stage spermatocytes during meiosis with levels again decreasing thereafter, albeit maintaining steady-state levels in the postmeiotic stages. Generation of *Tsx* knockout mice did not affect viability of the offsprings or their Mendelian ratio although homozygous knockout female mice exhibited reduced fertility and preferred the birth of female offsprings. Closer inspection of 6-month-old testes of  $-/Y$  males showed smaller size in comparison to the wild-type ones. TUNEL experiments revealed increased apoptosis of germ cells, peaking at 14 days of development, coinciding with the first phase of pachytene stage. Further staining with SCP1 (synaptonemal complex protein 1) confirmed that it was indeed the pachytene spermatocytes that were undergoing apoptosis, thereby suggesting that lncRNA TSX might be required for germ cells to enter the meiotic phase of differentiation although its function might be redundant in the maturation of haploid spermatids during spermiogenesis.

## 8.5 Conclusions

Stem cells are an integral part of animal development. During the last two decades, we have seen an explosion in our basic understanding of stem cell biology. Stem cells are also being explored as an effective mode of human disease management

and treatment. The first stem cell therapy ever to be performed was in 1968 when clinicians successfully carried out bone marrow transplantation. Bone marrow contains multipotent stem cells that can give rise to all the types of blood cells. Since then bone marrow transplantation has formed one of the major stem cell therapies, helping millions of patients suffering from cancers like leukemia. Not very far behind was the concept of using skin stem cells to replace burnt tissue in the form of skin grafts. Limbal stem cells in the eye have also huge potential in replacing lost corneal tissue by virtue of their stem cell properties. These are some of the successful stories of stem cell therapies. There are still a number of human diseases and disorders that need to be addressed via stem cell therapies. For example, Duchenne muscular dystrophy (DMD) is a genetic disease in which skeletal muscles and often heart muscles weaken over time due to prevention of formation of dystrophin protein. As we know, muscle harbors stem cells known as satellite cells which serve as great contenders for curing such genetic diseases. On the other hand, iPSCs also possess immense potential because adult somatic cells can be reprogrammed into iPSCs which can then theoretically be directed into the generation of any type of cell such as neurons for replacement in neurodegenerative diseases like Parkinson's and Alzheimer's diseases. One of the major challenges of stem cell therapies is the generation of a pure population of cells which can be transplanted into the human body without complications of tissue rejection and immune responses. In this direction, it is very important to understand the fine details of the molecular mechanisms of differentiation processes so that we can take care of every small detail that leads to the generation of the right type of cell with the expected phenotype. In this context, the emerging lncRNAs as key regulators of lineage-specific differentiation might serve as an important tool to fine-tune the differentiation pathway. This field although very nascent provides us with potential hope in making regenerative medicine a highly successful strategy in clinical practice in the near future.

**Acknowledgments** M.R.S. Rao thanks the Department of Science and Technology for J.C. Bose and SERB Distinguished fellowships. Funding has been granted by the Department of Biotechnology (BT/01/COE/07/09).

## Glossary

**Microarray** It employs an array comprising of probes which can be DNA, cDNA, or oligonucleotides representing the sequences in a particular genome. Hybridization of query sequences to these probes can allow for the parallel analysis of gene expression for thousands of genes or for the identification of new genes.

**ChIP-Seq** Chromatin immunoprecipitation is a technique in which chromatin is isolated from cells or tissues, fragmented by sonication, chromatin associated with a particular protein is pulled down with the help of an antibody specific

against the protein of interest, and the DNA is subsequently recovered. This is followed by sequencing of the DNA to decipher genomic binding loci of the concerned protein.

**RNA-Seq** RNA Sequencing uses a population of RNA (such as polyA+) to be converted to a library of cDNAs using adapters at one or both the ends. The library is then subjected to high-throughput sequencing where each molecule is sequenced to obtain reads that are typically 30–400 bp long. The reads are then aligned to a reference genome or reference transcriptome or assembled to generate a transcriptome for the particular system used for the RNA-Seq. This accurately depicts not only the transcriptome but the expression level of each gene for that system [57].

**CHART** In capture hybridization analysis of RNA targets, the RNA is cross-linked to its genomic binding sites on the chromatin, and the genome is isolated and fragmented. The RN-bound fragments are then enriched with the help of complementary locked or O2'-methylated oligonucleotides which are immobilized by beads. The corresponding DNA or protein fractions are then eluted to analyze either loci of binding or interacting partners for the RNA of interest.

**ChIRP/ChOP** Chromatin isolation by RNA purification or chromatin oligoaffinity purification. In this case, the complementary oligonucleotides are biotinylated, and the RNA-bound chromatin fragments are enriched by magnetic streptavidin beads. The DNA associated with the RNA or the interacting proteins can then be eluted for further analysis by sequencing or mass spectrometry, respectively.

**siRNA/shRNA Mediated Knockdown** Short-interfering RNAs are double-stranded RNA molecules consisting of a 3' 2 nt overhang that activates the RNAi machinery inside the cytoplasm of cells upon delivery. After processing, one of the strands of the siRNA binds to its complementary sequence on the target mRNA leading to degradation by the RISC (RNA induced silencing complex). Short hairpin RNAs are transcribed from a plasmid in the form of a stem loop primary RNA which is processed by the Drosha machinery in the nucleus to generate siRNA.

**CRISPR/Cas9** The clustered regularly interspaced short palindromic repeats is a bacterial immune system that is used to cleave invading foreign DNA. This technique is now used for genome engineering. The CRISPR system consists of a guide RNA and a nonspecific endonuclease, Cas9. The guide RNA "guides" the Cas9 endonuclease to the target region in the genome wherein Cas9 creates double-stranded breaks. The DNA sequence is then repaired with the help of either NHEJ- or HDR-mediated repair generating indels or desired knockouts/knockins.

**Fluorescence In Situ Hybridization** FISH technique is used to label or localize regions of interest in the genome or transcriptome with the help of short sequences known as probes. These probes are most often labeled with a fluorescent tag. The probes bind to the target regions of interest by complementary hybridization, and signals can be detected by fluorescent microscopy to understand the localization/copy number of the targets.

## References

1. Huttenhofer A, Schattner P, Polacek N (2005) Non-coding RNAs: hope or hype? *Trends Genet* 21(5):289–297
2. Brown CJ et al (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71(3):527–542
3. Bartolomei MS, Zemel S, Tilghman SM (1991) Parental imprinting of the mouse H19 gene. *Nature* 351(6322):153–155
4. Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292(5819):154–156
5. Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 78(12):7634–7638
6. Thomson JA et al (1995) Isolation of a primate embryonic stem cell line. *Proc Natl Acad Sci U S A* 92(17):7844–7848
7. Thomson JA et al (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282(5391):1145–1147
8. Shablott MJ et al (1998) Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci U S A* 95(23):13726–13731
9. Yu J et al (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318(5858):1917–1920
10. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663–676
11. Guttman M et al (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477(7364):295–300
12. Sheik Mohamed J et al (2010) Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA* 16(2):324–337
13. Lin N et al (2014) An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell* 53(6):1005–1019
14. Chakraborty D et al (2012) Combined RNAi and localization for functionally dissecting long noncoding RNAs. *Nat Methods* 9(4):360–362
15. Ng SY, Johnson R, Stanton LW (2012) Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* 31(3):522–533
16. Wang Y et al (2013) Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev Cell* 25(1):69–80
17. Loewer S et al (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* 42(12):1113–1117
18. Guttman M et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235):223–227
19. Urban N, Guillemot F (2014) Neurogenesis in the embryonic and adult brain: same regulators, different roles. *Front Cell Neurosci* 8:396
20. Ng SY et al (2013) The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol Cell* 51(3):349–359
21. Vance KW et al (2014) The long non-coding RNA Paupar regulates the expression of both local and distal genes. *EMBO J* 33(4):296–311
22. Chalei V et al (2014) The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *elife* 3:e04530
23. Ramos AD et al (2015) The long noncoding RNA Pnky regulates neuronal differentiation of embryonic and postnatal neural stem cells. *Cell Stem Cell* 16(4):439–447
24. Till JE, Mc CE (1961) A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res* 14:213–222
25. Hu W et al (2011) Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes Dev* 25(24):2573–2578



26. Paralkar VR, Weiss MJ (2011) A new 'Linc' between noncoding RNAs and blood development. *Genes Dev* 25(24):2555–2558
27. Wagner LA et al (2007) EGO, a novel, noncoding RNA gene, regulates eosinophil granule protein transcript expression. *Blood* 109(12):5191–5198
28. Zhang X et al (2009) A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 113(11):2526–2534
29. Wei S et al (2016) PU.1 controls the expression of long noncoding RNA HOTAIRM1 during granulocytic differentiation. *J Hematol Oncol* 9(1):44
30. Hu G et al (2013) Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* 14(11):1190–1198
31. Ranzani V et al (2015) The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol* 16(3):318–325
32. Casero D et al (2015) Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat Immunol* 16(12):1282–1291
33. Yin H, Price F, Rudnicki MA (2013) Satellite cells and the muscle stem cell niche. *Physiol Rev* 93(1):23–67
34. Yablonka-Reuveni Z (2011) The skeletal muscle satellite cell: still young and fascinating at 50. *J Histochem Cytochem* 59(12):1041–1059
35. Mousavi K et al (2013) eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 51(5):606–617
36. Mueller AC et al (2015) MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. *Mol Cell Biol* 35(3):498–513
37. Cesana M et al (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147(2):358–369
38. Legnini I et al (2014) A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. *Mol Cell* 53(3):506–514
39. Caretti G et al (2006) The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. *Dev Cell* 11(4):547–560
40. Hube F et al (2011) Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic Acids Res* 39(2):513–525
41. Zhou Y et al (2010) Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. *Development* 137(16):2643–2652
42. Lu L et al (2013) Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. *EMBO J* 32(19):2575–2588
43. Klattenhoff CA et al (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152(3):570–583
44. Xue Z et al (2016) A G-rich motif in the lincRNA Braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol Cell* 64(1):37–50
45. Blanpain C, Fuchs E (2006) Epidermal stem cells of the skin. *Annu Rev Cell Dev Biol* 22:339–373
46. Kretz M et al (2012) Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev* 26(4):338–343
47. Kretz M et al (2013) Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493(7431):231–235
48. Luk AC et al (2014) Long noncoding RNAs in spermatogenesis: insights from recent high-throughput transcriptome studies. *Reproduction* 147(5):R131–R141
49. Nishant KT, Ravishankar H, Rao MR (2004) Characterization of a mouse recombination hot spot locus encoding a novel non-protein-coding RNA. *Mol Cell Biol* 24(12):5620–5634
50. Ganesan G, Rao SM (2008) A novel noncoding RNA processed by Drosha is restricted to nucleus in mouse. *RNA* 14(7):1399–1410
51. Arun G et al (2012) mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. *Mol Cell Biol* 32(15):3140–3152
52. Akhade VS et al (2014) Genome wide chromatin occupancy of mrhl RNA and its role in gene regulation in mouse spermatogonial cells. *RNA Biol* 11(10):1262–1279



53. Akhade VS et al (2016) Mechanism of Wnt signaling induced down regulation of mrhl long non-coding RNA in mouse spermatogonial cells. *Nucleic Acids Res* 44(1):387–401
54. Sun J, Lin Y, Wu J (2013) Long non-coding RNA expression profiling of mouse testis during postnatal development. *PLoS One* 8(10):e75750
55. Adriaens C et al (2016) p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat Med* 22(8):861–868
56. Anguera MC et al (2011) Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet* 7(9):e1002248
57. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63