Long Noncoding RNAs in Human Stemness and Differentiation

Title

long non-coding RNAs in human stemness and differentiation

Fatemeh Mirzadeh Azad[1,2,3,4], Isabelle Laurence Polignano*[1,2,3,4], Valentina Proserpio*[,1,3] and Salvatore Oliviero*[,1, 2]

[1] University of Turin, Department of Life sciences and Systems Biology, Torino, Italy

[2] IIGM Foundation – Italian Institute for Genomic Medicine, Candiolo, Torino, Italy

[3] Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Torino, Italy

[4] Equal contributions

* Correspondence:

valentina.proserpio@iigm.it (V. Proserpio)

salvatore.oliviero@unito.it (S. Oliviero)

Highlights

Embryonic differentiation at the gastrulation stage is a very tightly regulated process in which many long non-coding RNAs (lncRNAs) have been demonstrated to play a crucial role in gene expression modulation.

lncRNAs can regulate cellular functions at the transcriptional, post transcriptional, translational, and post-translational level.

lncRNAs exert a large repertoire of regulatory mechanisms both in the nucleus and in the cytoplasm.

Conserved lncRNAs are enriched in functional roles.

Abstract

Stem cells are known for their ability to self-renew to maintain their own population and to produce different progeny of cells through differentiation. Sustaining these two main features entails a tightly controlled network of regulators that work on different biological levels. long non-coding RNAs (lncRNAs) are one of the main regulatory factors of stem cell maintenance and differentiation. They exert their functions through different mechanisms and by interactions with different proteins, DNA, and RNA molecules. Here, we provide a broad view of the role of lncRNAs focusing on human regulations and culture model systems. We discuss the functions of lncRNAs in supporting the molecular identity of human cells in homeostasis at different stages ranging from stem cells to progenitors and to terminally differentiated tissues.

Keywords (two to six)

Long non-coding RNAs (lncRNAs)

Differentiation

Neuroectoderm

Mesoderm

Endoderm

Introduction: Stemness and early development

Stem cells are specified groups of cells that are known for two main characteristics: ability of self-renewal to sustain their own population and ability of producing different progeny of cells through differentiation. These two features are known as main pillars of “stemness”. Any cell that shows some degree of potency and self-renewal can fall into the stem cells realm [1, 2].

One of the main sources for obtaining stem cells is the inner cell mass of mammalian blastocyst (Figure 1) [3]. These cells are wired to go through gastrulation to generate embryonic **ectoderm**, **mesoderm**, and **endoderm** (see Glossary). Each of these layers can further differentiate into specialized tissues. For instance, ectoderm generates neural tube, epidermis, and neural crest. Mesoderm layer is the source of muscles, connective tissues, and bones, while endoderm gives rise to respiratory and digestive epithelia and organs related to the digestive system [4].

Given the right culture conditions, ICM derived cells can be kept at undifferentiated state as embryonic stem cells (ESCs) or cued for differentiation towards each of the three main germ layers. This level of plasticity is known as pluripotency, [3,5] which can also be achieved in primordial germ cells [6,7] and be induced in somatic cells.

Adult terminally differentiated cells can also be artificially converted into pluripotent stem cells (Box2). The first reported induced pluripotent stem cells (iPSCs) were generated by introducing OCT4, SOX2, c-MYC, and KLF4 (known as Yamanaka factors) into fibroblasts [8]. Although iPSCs retain some of their somatic original

epigenetic memory, they can express stemness markers, have the same morphology and telomerase activity of ESCs, can differentiate to different layers and are capable of self-renewal [9].

Another type of stem cell is represented by multipotent stem cells. During development, cells committed to each layer are restricted to produce a limited number of tissues. This level of potency can be seen in tissue specific progenitors, resident in adult organs also known as adult or somatic stem cells [10].

Stem cell research started in mouse models and expanded to higher mammals and humans; however, despite similarities in stemness repertoire, there are some critical differences between human and mouse stem cells and their path to differentiation. For example, unlike mESCs, hESCs do not rely on LIF and BMP4 signaling for maintaining pluripotency and, conversely, these factors push hESC towards differentiation. Moreover, ICM derived stem cells in humans are not at the same developmental stage as their mouse counterpart, as they resemble a primed state of pluripotency and are closer to mouse EpiSC [2,11,12].

Considering the differences between mice and human stem cells and emerging applications of hESCs and hiPSCs in personalized and regenerative medicine, drug discoveries and developmental studies, it seems important to focus on human specific regulations and culture model systems instead of relying on findings mainly generated in mice [13].

lncRNAs in Pluripotency

lncRNAs are a versatile group of non-coding RNAs transcribed by RNA Polymerase II, spliced and polyadenylated, longer than 200 nucleotides. They regulate different biological processes at different developmental stages. Starting from maintaining the pluripotency state, lncRNAs use different mechanisms to exert their regulatory roles (Figure 2). One of the best examples of this subset of lncRNAs is linc-ROR (regulator of reprogramming) which was first discovered in iPSC showed that linc-ROR maintained the pluripotency state of hESCs by sequestering miR-145 and releasing its negative regulation from NANOG, OCT4, and SOX2 mRNAs (Figure 1 and Table 1, Key Table) [14]. Suppressing linc-*ROR* expression in stem cells leads to exiting the stem state and subsequent upregulation of ectodermal markers SOX1 and Vimentin [15].

lncRNAs could be functioning at various levels in pluripotent cells. For instance, lncRNA T-UCstem1 that is transcribed from an ultra-conserved region, supports self-renewal in stem cells by sequestering miR-9 in the cytoplasm and through stabilization of Polycomb Repressive Complex 2 (PRC2) over bivalent promoters in the nucleus (Figure 1 and Table 1) [16].

Not only are lncRNAs important in maintaining self-renewal in hESC they can also regulate the retrospective conversion of primed stem cells to ground state of pluripotency. For example, lncRNA *CCDC144NL-AS1* knockdown converts the transcriptome of embryonic stem cells to naïve state (Figure 1 and Table 1). Although the detailed mechanism of this action has not been revealed yet, Wang et al. [12] showed that suppressing CCDC144NL-AS1 in stem cells, would upregulate

LIF/STAT3 downstream genes, inhibit MAPK pathway, and accumulate active β-catenin in the cells.

Linc-hoxa1 is a long non-coding RNA, regulated by OCT4 which interacts with PRC2 to impose repressive histone mark (H3K27me3) at HOX genes locus in embryonic stem cells (Figure 1 and Table 1) [17].

Not all lncRNAs exert their functions by guiding the regulatory elements to specific locus. For instance, lncRNA *PRESS1*, which is regulated by p53, is highly expressed in hESCs and supports stemness by scavenging histone deacetylase SIRT6 in the nucleus (Figure 1 and Table 1)[18]. Through this interaction, *PRESS1* inhibits SIRT6 chromatin localization and therefore sustains the accumulation of H3K56 and H3K9 acetylation at promoters of pluripotency genes. During differentiation, p53 suppresses *PRESS1* expression and relieves its scavenging effects from SIRT6 that promotes differentiation by removing active histone marks from stem genes promoter.

lncRNAs in Neural Differentiation

Neural differentiation is one of the most complex developmental processes which is tightly regulated by coding and non-coding genes. Several studies report the stage specific expression for lncRNAs that regulate neural differentiation. However, some lncRNAs are expressed throughout the process of differentiation but exert stage-specific functions. For example, *MEG3* is maternally imprinted and expressed in human stem cells (Figure 1 and Table 1). Suppressing this lncRNA in hESCs decreases the efficiency of their subsequent neural differentiation [19]. Further studies revealed that *MEG3* is also expressed in mature neurons and its epi-mutation is related to delayed gross motor neuron maturation in humans [20]. Mechanistically, *MEG3* controls the suppressive methylation on progenitor genes such as *PAX6* and

IRX3 in post mitotic neurons through stabilizing the interaction of JARID2 and EZH2 subunits of PRC2. The neurogenic effect of *MEG3* is not restricted to the central nervous system as this lncRNA is also expressed during differentiation of enteric nervous ganglions and it is necessary for their survival. In these cells however, MEG3 acts as ceRNA and supports the expression of GDNF by releasing negative regulation of miR-211-5p from GDNF mRNA [21].

Aside from guiding the differentiation process, lncRNAs also regulate the maintenance of neural stem cells and progenitors. *PNKY* is an evolutionary conserved lncRNAs expressed in the developing human brain and is abundant in Neural Stem Cells (NSCs) (Figure 1 and Table 1) [22]. PNKY physically interacts with PTBP1, a regulator of mRNA splicing downregulated during neuronal differentiation. PNKY together with PTBP1 controls the alternative splicing of core transcript sets in neural progenitors. Surprisingly, *PNKY* knockdown leads to an increase in terminally differentiated neuronal colonies. This phenotype is also seen in *PTBP1* knockdown models [22].

Another lncRNA that plays a pivotal function in maintaining neural stem cell state, is the chromatin associated lncRNA named *FMR4* (Figure 1 and Table 1). This lncRNA enriches in chromatin and regulates promoter accessibility of neural gene regulators such as *BDNF*, *HTR1D*, and *SEMA3E* by inducing H3K4me3 and reducing H3K27me3 level in their locus [23,24].

Changing the epigenetic profile is one of the main steps of fate determination (Box 1). Therefore, it is not surprising to find lncRNAs fine-tuning functions of chromatin modifiers in course of differentiation. The case in point is CAT7 lncRNA that was found to coprecipitate with the PRC1 complex in human neural lineages (Figure 1 and Table 1). *CAT7* knockdown affects the PRC1 regulated genes and leads to de repression of

genes like *MNX1* that are meant to be expressed in terminally differentiated motor neurons. In other words, *CAT7* temporally regulates *MNX1* expression at early steps of differentiation by recruiting PRC1 to *MNX1* gene locus [25]. *NBAT-1* is another lncRNA that exerts its regulatory role on neural differentiation by affecting PRC derived methylation (Figure 1 and Table 1). This lncRNA can interact with EZH2 to direct PRC2 methylation to NRSF/REST promoters. *NBAT-1* controls neural differentiation through suppressing NRSF/REST and releasing their negative regulation from neural-specific genes in progenitors committed for terminal differentiation [26].

lncRNAs versatility comes from their diverse structural domains and the ability to produce several transcript isoforms with different sets of domains. By Analyzing lncRNA repertoire in different types of neurons, Gendron et al. stated the cell type specificity of lncRNAs spliced variants in neural cells [27]. This particular behavior can be seen for the chromatin enriched *lncRNA-NR2F1* which shows increasing expression during neural differentiation and interacts with enhancers/promoters of genes involved in neuron maturation and axon guidance (Figure 1 and Table 1). However, the level of occupancy is not the same for all the spliced isoforms of this lncRNA and only its longer isoforms (consisting of specific domains) could interact with pro neurogenic genes. Moreover, *lncRNA-NR2F1* suppression in neural progenitors leads to loss of neural identity and ectopic expression of epidermal related genes [28]. The phenomenon of difference in isoform functionality is not limited to pro-neurogenic lncRNAs. It can also be seen in lncRNAs that fine-tune pluripotency whose aberrant expressions cause stemness loss and differentiation. As an example only one specific set of PSORS1C3 transcripts, which are coexpressed with and can mediate *OCT4* expression, show expression alteration during Neural differentiation of NTera2 cells [29]

lncRNAs in Cutaneous differentiation

Despite originating from the same germinal layer and sharing primordial regulatory pathways with neuroectoderm, epidermal formation needs specific regulators to ensure accurate differentiation. lncRNAs fit the specificity needed for cell-fate commitment and differentiation of epidermal progenitors. Despite the mechanistic ambiguity that surrounds some lncRNAs with stage specific expression during cutaneous differentiation, namely *ANCR* [30] and *BLNCR* [31], the elucidated functions of other lncRNAs during keratinocyte differentiation reflect the true heterogeneity in this class of regulators.

TINCR is a lncRNA highly expressed in epidermal progenitors that controls mRNA stability at the post-transcriptional level (Figure 1 and Table 1) [32]. Mechanistically, TINCR guides STAU1 protein to mRNAs containing a sequence motif called TINCR-box and ensures the abundance of epidermal specific regulators such as FLG, LOR, ALOXE3, ALOX12B, ABCA12, CASP14 and ELOVL3 by mediating their interaction with STAU1. TINCR-STAU1 interaction and regulatory effects are specific to and necessary to form ultrastructures like keratohyalin granules and intact lamellar bodies in terminally differentiated epidermis [32].

Another case of post-transcriptional regulation in keratinocyte terminal differentiation is ceRNA action of lncRNA *H19* (Figure 1 and Table 1). This cytoplasmic lncRNA can scavenge miR-130b-3p and release its inhibitory effect from DSG1 mRNA which is a diver regulator for terminal differentiation of epidermal progenitor cells [33].

lncRNAs are also able to fine-tune molecular switches of cutaneous differentiation. One of these fine-tuners is lncRNA *UC.291* that increases promoter accessibility in epidermis-specific genes for BAF chromatin remodeling factor (Figure 1 and Table 1).

Upon the induction of terminal differentiation lncRNA UC.291 interacts with chromatin associated ACTL6A that antagonizes BAF recruitment to the genome. By destabilizing ACTL6A interaction, UC.291 makes the epidermis-specific promoters accessible to the BAF complex. UC.291 depletion results in loss of epidermal potentiality of progenitors and, by affecting BAF dependent chromatin remodeling, loss of this lncRNA can impair the expression of pro-differentiation genes such as *LOR, FLG* and *ACE18* [34].

lncRNAs in Endoderm

Endoderm progenitors are the first cells to enter the primitive streak to form the definitive endoderm (DE) layer below the epiblast. The pluripotency masters, *OCT4*, *SOX2,* and *NANOG*, together with the well-known Wnt/β-catenin, Activin/Nodal and BMP signaling pathways form a complex regulatory network that blocks pluripotency maintenance and drive endoderm differentiation. In particular, the decrease of SOX2 levels, which expression correlates first with stemness and then with ectoderm specification, represents a key step to proceed towards endoderm direction unlocking the expression of *EOMES* first, known marker of primitive streak formation, and *FOXA2* and *SOX17* later, as specific marker for DE. Many studies have been published to precisely decipher the role of long non-coding RNAs during this process. *DEANR1* (definitive endoderm associated lncRNA1 - *LINC00261*) was identified in 2015 as a FOXA2-associated lncRNA which regulates hESCs differentiation to endoderm and pancreatic cell lineage (Figure 1 and Table 1) [35]. The expression level of this lncRNA during the differentiation process has been shown to correlate to its neighboring gene, FOXA2. DEANR1 depletion abolishes the potency of endoderm differentiation in hESCs and, interestingly, FOXA2 over expression in these cells,

could rescue the observed phenotype. Moreover, DEANR1 physically interacts with *FOXA2* promoter and assists SMAD2/3 recruitment to the locus. This observation indicates that *DEANR1* can modulate endoderm differentiation by regulating FOXA2 expression in cis [35].

Using a similar mechanism, lncRNA *GATA6-AS1*, also mediates endoderm differentiation through SMAD2/3 recruitment (Figure 1 and Table 1). This lncRNA is divergently transcribed from *GATA6* locus, and its expression is necessary for transcriptional activation of *GATA6* by SMAD2/3 [36]. It has been suggested that this mechanism of action between *GATA6* and *GATA6-AS1* could be generally applied to different cell types in which their expression is correlated.

Considering the fact that SMAD3 signaling pathway is at the core of endoderm commitment, it is not surprising to find its regulatory effects behind endoderm specific lncRNA modulators. Another evidence for this note is *DIGIT* lncRNA (Divergent to GSC, Induced by TGF-β family signaling) which is located in close vicinity to *SMAD3* enhancer and its expression is influenced by the lineage specific regulations of this locus (Figure 1 and Table 1) [37]. *DIGIT* expression level has a positive correlation with endoderm commitment and can affect its antisense protein coding gene, GSC as well. Moreover, DIGIT depletion in hESCs can compromise the efficiency of definitive endoderm formation. Although the detailed mechanism of action of this lncRNA is not yet understood, expression analysis proves that *DIGIT* manipulation can alter the differentiating cells transcriptome. Genes that are affected by DIGIT disruption include *SOX17*, *FOXA2*, *CXCR4*, *GSC* and even *DEANR1* lncRNA [37].

Generation of definitive endoderm derived cells and tissues is quite appealing for regenerative medicine; specifically, liver and pancreas progenies which are hard to obtain directly from pluripotent stem cells. In these cases, the use of mesenchymal or

adult stem cells has been suggested because of their ease of isolation, maintenance and their multipotency. MSCs are able to differentiate toward definitive endoderm and its derivative terminal tissues, however the path they undergo is not quite similar to that of hESCs and the molecular mechanisms are still under investigation.

Regarding the role of lncRNAs it has been recently demonstrated that *ANCR* (antidifferentiation non-coding RNA) plays a negative role in endoderm differentiation of human adipose-derived mesenchymal stem cells (hAMSCs) (Figure 1 and Table 1) [38]. This lncRNA is mainly cytoplasmic and controls positively ID2 mRNA stability. The ID gene family is known to be involved in pluripotency maintenance and differentiation inhibition, and in particular ID2 negatively correlates with definitive endoderm markers during differentiation. The proposed model is that ANCR directly binds to PTBP1, an RNA-binding protein which regulates the process of alternative splicing and mRNA stability. Through this interaction, ANCR enhances the accessibility of ID2 mRNA for PTBP1 leading to increasing *ID2* mRNA stability controlling its subsequent effects on differentiation paths [38].

During germ layers formation not only, biochemical factors drive the process, but also physical inputs are important. Focusing on these latter parameters, it was observed that the transcription changes in response to the different stiffness of the substrates used to culture hESCs. In particular, Chen et al. [39] found endoderm induction culturing hPSCs on soft substrates in comparison to hard ones. In addition to the upregulation of well-known DE markers, lncRNAs expression was also altered. In particular, among them, *DEANR1* was upregulated in soft substrates-induced differentiation. *LINC00458* and *LINC01356* showed a stiffness-associated expression signature in hPSCs, with respectively nuclear and cytoplasmic localization (Figure 1 and Table 1). In particular, *LINC00458* downregulation or upregulation affects

negatively or positively endoderm differentiation in response to soft substrates. The mechanism of action is based on its specific interaction with SMAD2/3 that displayed an increase in phosphorylation levels in soft substrates induced differentiation [39].

lncRNAs in Pancreas differentiation

A recent study screened all the lncRNAs displaying a dynamic regulation during pancreatic lineage differentiation [40]. Starting from a step-by-step differentiation protocol to clearly separate each stage (hESCs, definitive endoderm, pancreatic progenitors 1 and 2 and, finally, endocrine cells), the authors analyzed the expression levels of lncRNAs, their subcellular localization and translation potential highlighting how most of the cytoplasmic lncRNAs have a comparable expression level to protein-coding mRNAs and contain ORFs producing microproteins with different subcellular localizations. Within this category they identified the lncRNA *LINC00261* (also called *DEANR1*) proposing an additional role (Figure 1 and Table 1). In fact, in a previous study this lncRNA was indicated as a main player in definitive endoderm formation [35]. In this new report Gaertner et al. [40]. do not observe an effect on endoderm markers but on the regulators of beta cell differentiation MAFB and PAX4 and a variation in the number of insulin Positive cells. The mechanism proposed is based on a trans-regulatory function of *LINC00261*, supported on one hand by its mainly cytoplasmic localization, and on the other hand by the chromosomal distribution of the genes differentially regulated upon *LINC00261* depletion. Moreover, by the introduction of frameshift mutations they demonstrated that the role of *LINC00261* in pancreatic endocrine cell differentiation is unrelated to the microproteins translated from its ORFs [40].

lncRNAs in Liver differentiation

MALAT1 is involved in several physiological and pathological processes. Zhang and collaborators [41] recently proposed a novel function by highlighting *MALAT1* role during hepatic differentiation (Figure 1 and Table 1). They demonstrated that *MALAT1* acts as a scaffold between SETD2 that interacts with pSMAD2/3, and *PPM1A*, which is a specific phosphatase of p*SMAD2/3.* Thus, proposed that this lncRNA interacting with this protein complex promotes the dephosphorylation of pSMAD2/3 and the subsequent termination of TGF-β/SMADs signaling [41].

lncRNAs in Lung differentiation

Lung morphogenesis can be recapitulated by a multistep differentiation that starts from hESCs or hiPSCs followed by definitive endoderm formation, anteriorization of the foregut endoderm and distal and proximal lung epithelial cells differentiation.

Focusing on distal lung differentiation, lncRNA *RP11-380D23.2* presents two promising characteristics: it is spatially close to *PITX2* that is a main player in promoting asymmetrical development of the lungs, and it contains a binding site for PARP1 that is a chromatin-associated protein that contributes to compact the DNA (Figure 1 and Table 1) [42]. Banerjee and collaborators observed a negative correlation between the expression of PARP1 with lncRNA *RP11-380D23.2* and *PITX2*. PITX2 carries on its role in distal lung morphogenesis by interacting with the WNT signaling cascade and is positively regulated by *RP11-380D23.2*. Thus, when PARP1 represses RP11-380D23.2 expression, in turn it modulates the expression of PITX2, unpairing WNT signaling cascade [42].

lncRNAs in Mesoderm differentiation

As the word suggests, mesoderm is the middle of the three germ layers. After gastrulation mesoderm is divided into three main categories: paraxial, intermediate, and lateral mesoderm that will develop into several adult tissues and organs including smooth, skeletal and cardiac muscle, kidney, reproductive organs, connective tissue, bones and blood.

In mesoderm the master regulator is BRACHYURY (T) and its divergent lncRNA was named yylncT in 2019 by Frank and collaborators (Figure 1 and Table 1) [43]. yylncT is an example of a class of lncRNAs that are divergently transcribed from the same genomic locus of key regulators and named them yin yang (yy)lncRNAs. yylncT loss resulted in strong inhibition of T and of early mesodermal genes expression upon mesoderm induction and increase in apoptosis. Differentiation towards ectoderm or endoderm were conversely unaffected by loss of yylncT, indicating a specific fundamental role of yylncT for human pluripotent stem cells differentiation towards in mesoderm only.

Mechanistically, yylnc*T* exerts its function by directly binding to the de novo DNA methyltransferase, *DNMT3B* thus inhibiting its function and preventing DNA methylation deposition in lncyy*T/T l*ocus [43].

lncRNAs in Muscle differentiation

Skeletal muscle is a complex system mainly composed of muscle fibers, nerves, and satellite cells responsible for muscle regeneration.

The coordinate process that converts mononucleated myoblasts to polynucleated myofibers is named myogenesis. This process is regulated at the transcriptional level by several transcription and epigenetic factors (reviewed in [44]), including lncRNAs.

While many lncRNA pathways and mechanisms of action have been reported in mouse, a paucity of lncRNA roles have been assessed in human differentiation.

Within those *Linc-YY1*, despite a modest sequence conservation, displays a very conserved expression pattern between human and mouse during myogenic differentiation with an initial increase during the first steps of differentiation and a subsequent drop (Figure 1 and Table 1). *Linc-YY1* promotes muscle differentiation by binding its neighboring gene *YY1* and thus destabilizing its association with *PRC2* [45]. MyoD is the master regulator of muscle differentiation in human and mouse and the LncMyoD has been shown to regulate MyoD function in mouse, but this lncRNA does not have a human homologous. The human LncMyoD named *hLncMyoD* has been found in the same position as in the mouse (~20kb upstream of MyoD locus), but it displays a low sequence conservation (Figure 1 and Table 1) [46]. Despite this difference both human and mouse LncMyoD can restore myogenesis in LncMyoD deficient mouse myoblasts, thus demonstrating a conserved role of *hLncMyoD* and LncMyoD. These data are in agreement with those supporting that sequence conservation is not the only feature to select functionally relevant lncRNA, but positional conservation should also be taken in account when looking for conserved lncRNA [47].

lncRNAs in Cardiovascular differentiation

Cardiogenesis is a coordinated process that first converts mesoderm into cardiac mesoderm and then into heart progenitors that finally differentiate into cardiovascular lineage cells. Despite the heart being the first functioning organ in vertebrates with a very conserved development and highly conserved coding genes, non-coding RNAs

during heart development still show a very high species-specificity. Most of the current knowledge on heart embryonic development derives from studies conducted in mice, with a paucity of studies on human embryos that rarely corroborates the findings. In the last decade, with the development of more tractable and reliable in vitro model systems for in vitro cardiac development [48] from hESCs and hiPSCs some lncRNA role and regulation pathways have started to be analysed also in human.

In 2015 Pedrazzini Lab identified a novel Super Enhancer (SE) associated lncRNA with three different isoforms named *CARMEN1*, 2, and 3 (that stands for (CAR)diac (M)esoderm (E)nhancer-associated (N)oncoding RNA), that is necessary for the correct cardiac differentiation of pluripotent and human cardiac precursor cells (CPCs) isolated from fetal heart (Figure 1 and Table 1) [49]. The enhancer activity of the SE is dependent on the production of the associated lncRNA and *CARMEN* has been shown to directly interact with protein of the PRC2 complexes (SUZ12 and EZH2). *CARMEN* Knockdown in human fetal CPCs induced downregulation of cardiac differentiation markers and transcription factor master regulators as *GATA4*, *TBX5*, *MYH6* and *MYH7*, thus overall impairing the ability to produce fully differentiate cardiomyocytes. Another factor impaired by *CARMEN* loss is *EOMES*, that is a key factor mesoderm and DE differentiation suggesting a key role of *CARMEN* in early mesendoderm specification. Interestingly, *CARMEN* is conserved between human and mouse and in the promoter orthologous region of the mouse genome the Braveheart (Bvht) lncRNA is present on the opposite strand. Bvht, that has no ortholog in human, has been thoroughly investigated in mouse and its expression is essential for cardiogenesis [50].

Kurian and collaborators show that most of the transcriptomic changes happening during differentiation involve non-coding fraction of RNA and identified three new

lncRNA expressed at specific stages and with a precise function in those particular stages of human cardiovascular development named *TERMINATOR* (specific for pluripotent stem cells), *ALIEN* (in vascular progenitors only) and *PUNISHER* (in mature endothelial cells) (Figure 1 and Table 1) [51].

More recently, a human lncRNA, *HBL1* (Heart Brake LncRNA 1) has been identified by Liu and collaborators (Figure 1 and Table 1) [52]. *HBL1* is conserved amongst non-human primates but not to other vertebrate and its expression has been shown to peak in undifferentiated stem cells and drop during cardiovascular differentiation. The overexpression of HBL1 strongly impairs cardiomyocyte differentiation and, vice versa, its knockout favors differentiation. The mechanism of action of *HBL1* involves the binding in complex with AGO2 and the subsequent repression of *miR-1*, one of the most abundant and well characterized miRNA in heart development [53].

lncRNAs in blood differentiation

Blood is a very complex and articulate organ composed by many different cell types each one with specific differentiation program.

In dendritic cells (DC) Wang and collaborators have identified a new lncRNA named *lnc-DC* reporting a novel molecular mechanism for lnc-RNA functions (Figure 1 and Table 1) [54]. Perturbation experiments demonstrated that lnc-DC is crucial for a correct differentiation of DC and its loss affects the expression of several DC function-related genes, thus impairing DC antigen uptake and T cell stimulation ability. *Lnc-DC* has no in cis effect on its neighboring genes and it is mainly expressed in the cytosol where it binds STAT3. The binding with *lnc-DC* prevents STAT3 to bind to the tyrosine phosphatase SHP1 and its dephosphorylation at Y705, thus affecting the downstream

signaling cascade of STAT3. This was the first report of a regulation of signaling molecules at the post transcriptional level by a lncRNA.

For Natural Killer (NK) cells differentiation, Zhang and collaborators [55] identified several lncRNAs associated with NK maturation and function and *lnc-CD56* has been reported to positively regulate the adhesion molecule *CD56*, marker of mature NK (Figure 1 and Table 1) [55]. Similarly, HOTAIRM1 has been identified by Zhang and co-workers as a lncRNA with a crucial role during granulocyte maturation (Figure 1 and Table 1) [56].

Outstanding Questions

- LncRNAs have been shown to play regulatory roles at many different levels within a cell, regulating the cellular dynamic at the transcriptional, translational and also at the post-translational level. With so many diverse and articulate functions, their transcription is strictly tissue and stage specific. lncRNAs must be tightly regulated, as perturbation in their expression can have a multitude of adverse effects. One obvious question that still remains open is, but who regulates their transcription? Who is upstream of the lncRNAs?
- Reverse genetics implies that for uncovering the function of unknown genes one could rely on sequence and syntenic conservation. Based on this concept, using animal models for genetic manipulation and investigating the function of novel genes is justified. However, the non-coding genome is not evolutionarily conserved and many lncRNAs lack sequence homology between different species. This raises the question of whether applying the same pipeline of functional analysis that was set based on features of protein coding genes to lncRNAs characterization is correct? Can we trust in generalizing the findings

in animal models to human when it comes to non-coding genome? How much can we rely on homology-based prediction software and sequence homology search? As some of the lncRNAs function is due to their secondary and quaternary structures, can we use RNA structure analysis as a function prediction method? Can we improve our prediction algorithm to find the human hortologue of many lncRNA that have been extensively studied in mice only?

- Development and differentiation are biological events that are tightly regulated at transcription, translation, and post-translational levels. Years of research in the pre-genome area led to recognizing many protein coding genes as master regulators of stemness maintenance and fate determination. However, the new wave of non-coding genome investigation showed that not only many non-coding genes are functioning downstream of master transcription factors, but also governing or fine tuning their driver functions as well. Hence, the concept of “master regulator” should be reconsidered in this field because raising the question of “what comes first” is only valid if there is a linear cascade of regulators and not a circuit of events. Even so, the true order of importance for non-coding genes in multidimensional picture of regulators is still under question.
- lncRNAs functions are usually defined by their interactions with other biomolecules. Considering the fact that most of the experiments are designed in a simplified system, are we considering the existing uncertainty hovering over lncRNAs interactions/functions?

Figures

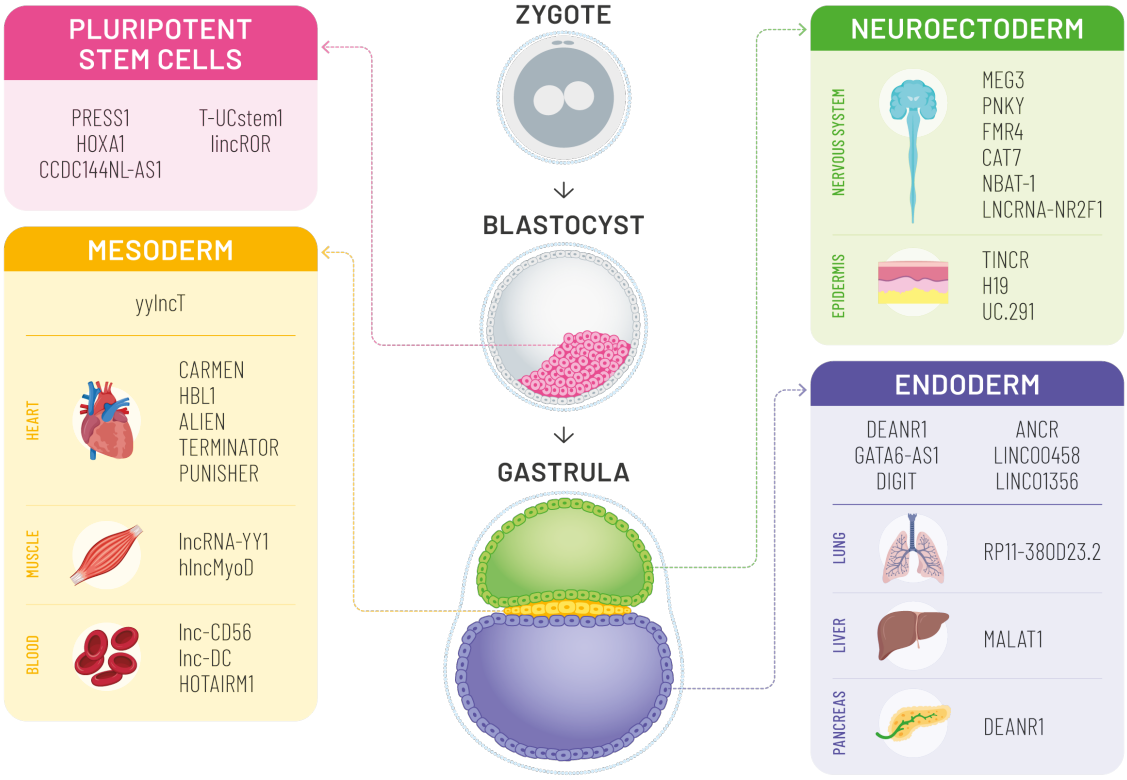

Figure 1 Schematic representation of organ differentiation from the three germ layers with functional lncRNAs involved in their regulation.

The zygote stage originates from the oocyte upon fertilization by the sperm and through several steps of symmetrical cleavage divisions it reaches the morula stage (32 cells - not shown). The blastocyst stage begins with the formation of the blastocoel cavity delimited by the trophoblast. The trophoblast derives from the outer cells of the morula, while the inner cells give rise to the inner cell mass. During the implantation of the embryo into the maternal endometrium, the inner cell mass differentiates to the bilaminar embryonic disc formed by two layers: the epiblast, which originates the embryo itself, and the hypoblast or primitive endoderm, which will give rise to the extraembryonic mesoderm and yolk sac. Gastrulation begins with the transformation of the bilaminar disc into the trilaminar embryonic one, which consists of the three

germ layers: ectoderm, mesoderm and definitive endoderm. lncRNAs have been demonstrated to play important roles in both pluripotency maintenance and differentiation. A schematic overview of the lncRNAs having a known role in human development is provided, subdivided into pluripotency (pink), neuroectoderm (green), endoderm (purple) and mesoderm (yellow) related lncRNAs, including the derivative terminal tissues, respectively nervous system and epidermis for neuroectoderm; lung, liver and pancreas for endoderm; and heart, muscle and blood for mesoderm.

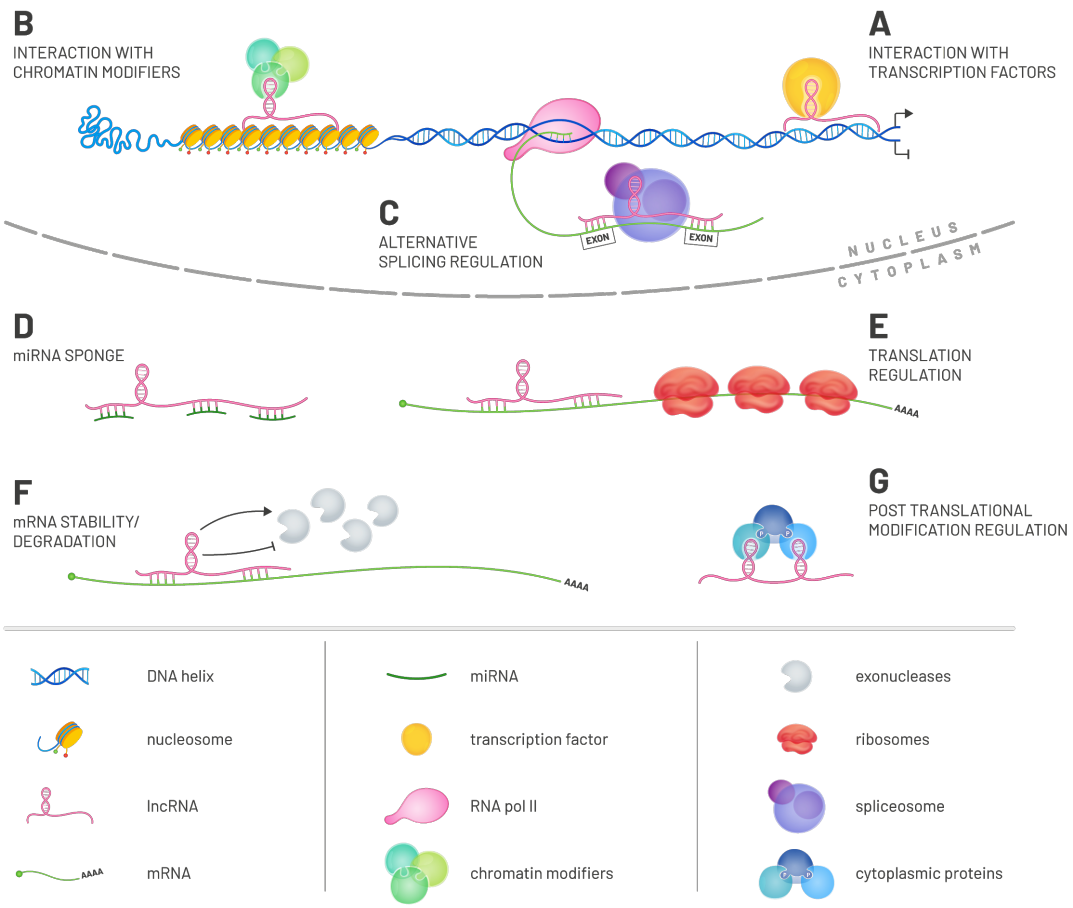

Figure 2. lncRNAs are involved in many cellular processes and exert their role through different mechanisms of action. In the nucleus lncRNAs can interact with transcription factors (A) and chromatin modifiers (B) to influence gene activation and transcription as well as chromatin state, respectively. Moreover, they can participate in alternative splicing regulation interacting with the splicing factors (C). In the cytoplasm lncRNAs influence miRNAs half-life acting as sponge (D), thus influencing the repression of miRNAs target mRNAs and creating a complex cross-talk between different RNA subtypes. They can also regulate mRNAs translation (E) and influence mRNA stability and degradation. Finally, they can also play a role in regulating post-translational modifications (G) such as phosphorylation acting as scaffold.

Germ layer or tissue	lncRNA	Localization: Nucleus (N) or cytoplasm (C)	Function	Refs

ESCs	linc-ROR	N/C	ceRNA activity, releasing negative regulation of miR-145 from OCT4, SOX2 and Nanog	[14]
ESCs	T-UCstem1	N/C	ceRNA activity, releasing negative regulation of miR-9; stabilizing PRC2 activity	[16]
ESCs	CCDC144NL-AS1	N	regulating LIF/STAT3 and MAPK pathway	[12]
ESCs	linc-hoxa1	N	interaction with PRC2	[17]
ESCs	*PRESS1*	N	interaction with SIRT6	[18]
central nervous system	MEG3	N	Interaction with EZH2, and regulation of methylation on neural markers	[19] [20]
central nervous system	PNKY	N	Interaction with PTBP1 and regulation of alternative splicing	[22]
central nervous system	FMR4	N	Regulation of histone code modification on differentiation genes	[23] [24]
central nervous system	CAT7	N	Interaction with PRC1	[25]
central nervous system	NBAT-1	N	Interaction with PRC2	[26]

central nervous system	lncRNA-NR2F1	N	Interaction with chromatin and promoter sequences	[28]
enteric nervous system	MEG3	C	ceRNA activity on miR-211-5p that targets BDNF	[21]
epidermis	TINCR	C	Interaction with Stau1 mRNA and regulation of mRNA stability	[32]
epidermis	H19	C	ceRNA activity on miR-130b-3p and DSG1	[33]
epidermis	UC.291	N	Regulation of BAF chromatin remodelling	[34]
endoderm	DEANR1 (*LINC00261)*	N	In cis activation of GATA6 throuh SMAD2/3	[35]
endoderm	GATA6-AS1	N	In cis activation of GATA6 throuh SMAD2/3	[36]
endoderm	DIGIT	N	In trans activation of GSC	[37]
endoderm	ANCR	C	Stabilization of ID2 mRNA	[38]
endoderm	LINC00458	N	Interaction with SMAD2/3	[39]
endoderm	LINC01356	C	/	[39]
pancreas	DEANR1 (*LINC00261)*	C	in trans activity	[40]
liver	MALAT1	N	lncRNA-proteins complex leads to the dephosphorylation of pSmad2/3	[41]
lung	RP11-380D23.2	N	interaction with PARP1 and regulates PITX2	[42]

mesoderm	yylncT	N	Binds DNMT3B and affects DNA methylation at the active T locus	[43]
muscle	Linc-YY1	N	Interaction with YY1 and decrease of PRC2 bindings	[45]
muscle	hlncMyoD	C	Mouse LncMyoD binds to IMPs and regulates mRNA translation	[46]
heart	CARMEN	N	Interaction with SUZ12 EZH2	[49]
heart	*TERMINATOR*	N	Essential for pluripotent stem cells survival	[51]
heart	*ALIEN*	C	Early cardiovascular development	[51]
heart	*PUNISHER*	C	Role definitive vascular development	[51]
heart	*HBL1*	C	miRNA sponge	[52]
blood	lnc-DC	C	Positive regulator of DC differentiation and maturation	[54]
blood	*lnc-CD56*	nr	Positive regulator of CD56, a NK marker	[55]
blood	HOTAIRM1	C	Mature granulocyte maturation inhibition	[56]

Box 1 Epigenetic Modifications

The two main features of all stem cells, the ability to self-renew and the potential to differentiate into multiple lineages have been demonstrated to be not just dependent, but tightly regulated by coordinate changes of the epigenetic status of cells.

The original idea proposed by Waddington, the famous “epigenetic landscape” has been extensively characterized in the context of stem cells differentiation and new key players are always joining the picture, giving rise to a well-defined epigenetic-code. In this code all the players are reversible and interdependent, they act in combination in a coordinated manner to regulate gene expression.

If at the very beginning the epigenetic word was referring mainly to DNA methylation and histone acetylation and methylation, nowadays epigenetic mechanisms include few chemical modifications of DNA, a plethora of histone chemical modifications, complex chromatin remodeling phenomena as well as lncRNA regulation of gene expression.

For many years the only DNA change thoroughly analyzed have been DNA methylation at the level of the 5th position of the pyrimidine ring of cytosine (5mC), but, more recently, other types of modifications at the same position, such as 5-hydroxymethyl (5hmC) [58], 5-formyl (5fC), 5-carboxyl (5caC) [59], and also N6-adenine (6mA) [60] have joined the epigenetic landscape.

The same happened to chromatin modifications where histone acetylation, methylation, ubiquitination and phosphorylation at precise residues of the histone core have now been joined by citrullination, O-GlcNAcylation, crotonylation, and isomerization [61].

Within the RNA categories that do not translate for proteins, the non-coding RNA, one of the most widely studied, is represented by the lncRNAs. LncRNAs are, by definition,

longer than 200nt. They are very abundant and are known to play as key regulators of several cellular processes such as transcriptional and post-transcriptional regulation, chromatin remodeling, and intracellular trafficking.

During embryonic development global changes in the epigenetic profile, mainly DNA methylation changes, have been deeply studied. While this modification is maintained in adult tissues, two waves of global demethylation and re-methylation occurs during germline production and embryo development.

The first wave occurs in the germline where a global demethylation occurs to establish gender-specific methylation patterns in germinal cells. A second wave takes place shortly after fertilization when human embryos display a quick drop in the overall methylation status. In particular, while the maternal genome experiences a passive DNA demethylation throughout several rounds of DNA replication the paternal counterpart is subjected to active demethylation process by Tet3 dioxygenase.

Subsequently, while the embryonic lineages are specified de novo methylation will occur at specific loci for each tissue.

Box 2 Culture systems for modeling early development and differentiation of hESCs

Modeling pluripotency and early developmental events in human has been the topic of many controversies through the years. Considering the ethical boundaries of culturing human embryos [62], the first breakthrough was adopting hESC derived from inner cell mass to grow on mouse embryonic fibroblasts (MEFs) as feeder layers. Although feeders support pluripotency of stem cells, The risk of transmission of animal-derived pathogens for clinical applications and the poor reproducibility

triggered the development of xeno-free feeder-free culture conditions [63,64]. Next big step was developing a culture condition that would recapitulate the early step of human embryo genesis. It started from tailoring the medium component (growth factors, minerals, lipids and vitamins) on monolayer stem cells to push them towards lineage differentiation. This method is known as 2D differentiation methods that rely on manipulating main signaling pathways using GFs, TFs and even small molecules. Although 2D culture systems are successful in maintaining stemness and are suitable for differentiation studies, they cannot recapitulate the spatial organization and cell migration occurring during the gastrulation stage. The first attempt to reproduce the 3D complexity was represented by embryoid bodies (EBs). In vitro culture of hESCs in suspension allow the cells to spontaneously aggregate with simultaneous induction of differentiation into the three germ layers. Moreover, by stimulating EBs with specific factors it was possible to force their commitment to further developmental stages or to enrich them for one of the three germinal layers [65, 66]. In the wake of EBs, **gastruloids** (see Glossary) arose as 2D or 3D organized model systems, that could mimic the spatial order of cells distribution as expected during Gastrulation. in 3D gastruloids the gradient of differentiation can be seen as they happen in animal models gastrulation [67]. In parallel Warmflash et al. proposed 2D-gastruloids showing that geometric spatial confinement is sufficient to cause self-organization patterning in hESCs [68]. They demonstrated that hESCs grown in a disk-shape region and in presence of BMP4, differentiate and self-organize forming concentric rings of germ layers. More recently this model was employed to build up a primitive “fate map” of the human primitive streak trying to summarize localization, cell migration and response to different differentiation-driving factors [69]. Moving through differentiation paths, organoids are structures that resemble the organogenesis in human embryos.

Meaning that the cells are past germ layer commitment and passing through progenitor state, they can form different specialized cells and structures that are found in functioning organs.[70, 71]

Glossary

lncRNA: long non coding RNA are classified as RNA without translational potential longer than 200 nucleotides. They can be polyadenylated or not and they are thought to have regulatory functions.

Ectoderm: the outermost of the three germ layers. During development ectoderm will both give rise to the epidermis of the skin, hair and exocrine glands and also to neuroectoderm. The neuroectoderm represents the developmental step of the nervous system formation and will form the spinal cord and the brain.

Mesoderm: the middle of the three germ layers. With the evolution of bigger vertebrates, the role of mesoderm becomes more articulate with functions of support, movement, circulation, and reproduction. Muscles, blood cells and blood vessels, bones and reproductive organs all derive from mesoderm.

Endoderm: the innermost germ layer in very early vertebrate embryos. It gives origin to all the organs responsible for the secretory/absorptive and respiratory functions as lungs, liver, pancreas and parts of the endocrine system.

Organoids: three dimensional structures obtained by culturing pluripotent (both embryonic or induced) or adult stem cells from different organs *in vitro*. Organoids final structures resemble the one of the organ of origin of the stem cells. They mimic the natural environment with different cell subtypes in contact with each other. They are the most physiological system to study complicated phenomena of an organ in a petri dish.

Gastruloids: three dimensional self-aggregated embryo-like formations that originate from embryonic stem cells cultured in vitro under appropriate conditions. They mimic the post implantation developmental step of embryos showing the anteroposterior and dorsoventral polarization peculiar to embryos.

Acknowledgements

Work in the Oliviero laboratory is supported by AIRC IG†2017†IdÆ20240, PRIN 2018, and IIGM istitutional funds. Valentina Proserpio is supported by Fondazione Umberto Veronesi -FUV-. We are extremely grateful to Fabio Servolo for his professional contribution for the figures.

References

1 Romito, A. and Cobellis, G. (2016) Pluripotent Stem Cells: Current Understanding and Future Directions. *Stem Cells Int.* 2016, 9451492

2 Wu, J. *et al.* An overview of mammalian pluripotency. , *Development*, 143. (2016) , 1644–1648

3 Thomson, J.A. (1998) Embryonic Stem Cell Lines Derived from Human Blastocysts. , *Science*, 282. 1145–1147

4 Rossant, J. and Tam, P.P.L. (2017) New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. *Cell Stem Cell* 20, 18–28

5 Liu, G. *et al.* (2020) Advances in Pluripotent Stem Cells: History, Mechanisms, Technologies, and Applications. *Stem Cell Rev Rep* 16, 3–32

6 Kossack, N. *et al.* (2009) Isolation and characterization of pluripotent human spermatogonial stem cell-derived cells. *Stem Cells* 27, 138–149

7 Ko, K. *et al.* (2009) Induction of pluripotency in adult unipotent germline stem cells. *Cell Stem Cell* 5, 87–96

8 Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676

9 Kim, K. *et al.* (2010) Epigenetic memory in induced pluripotent stem cells. *Nature* 467, 285–290

10 Eckfeldt, C.E. *et al.* (2005) The molecular repertoire of the “almighty” stem cell. *Nat. Rev. Mol. Cell Biol.* 6, 726–737

11 Schnerch, A. *et al.* (2010) Distinguishing between mouse and human pluripotent stem cell regulation: the best laid plans of mice and men. *Stem Cells* 28, 419–430

12 Wang, Y. *et al.* (2019) Long noncoding RNA CCDC144NL-AS1 knockdown induces naïve-like state conversion of human pluripotent stem cells. *Stem Cell Res. Ther.* 10, 220

13 Malan, D. *et al.* (2016) Human iPS cell model of type 3 long QT syndrome recapitulates drug-based phenotype correction. *Basic Res. Cardiol.* 111, 14

14 Loewer, S. *et al.* (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* 42, 1113–1117

15 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, 69–80

16 Fiorenzano, A. *et al.* (2018) An Ultraconserved Element Containing lncRNA Preserves Transcriptional Dynamics and Maintains ESC Self-Renewal. *Stem Cell Reports* 10, 1102–1114

17 Liu, G.-Y. *et al.* (2016)The long noncoding RNAGm15055repressesHoxagene expression by recruiting PRC2 to the gene cluster. , *Nucleic Acids Research*, 44, 2613–2627

18 Jain, A.K. *et al.* (2016) LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. *Mol. Cell* 64, 967–981
19 Mo, C.-F. *et al.* (2015) Loss of non-coding RNA expression from the DLK1-DIO3 imprinted locus correlates with reduced neural differentiation potential in human embryonic stem cell lines. *Stem Cell Res. Ther.* 6, 1
20 Yen, Y.-P. *et al.* (2018) Dlk1-Dio3 locus-derived lncRNAs perpetuate postmitotic motor neuron cell fate and subtype identity. *eLife*, 7.
21 Xia, Z. *et al.* (2019) LncRNA-MEG3 protects against ganglion cell dysplasia in congenital intestinal atresia through directly regulating miR-211-5p/GDNF axis. *Biomed. Pharmacother.* 111, 436–442
22 Ramos, A.D. *et al.* (2015) The Long Noncoding RNA Pnky Regulates Neuronal Differentiation of Embryonic and Postnatal Neural Stem Cells. , *Cell Stem Cell*, 16. , 439–447
23 Peschansky, V.J. *et al.* (2016) The long non-coding RNA FMR4 promotes proliferation of human neural precursor cells and epigenetic regulation of gene expression in trans. , *Molecular and Cellular Neuroscience*, 74, 49–57
24 Peschansky, V.J. *et al.* (2015) Changes in expression of the long non-coding RNA FMR4 associate with altered gene expression during differentiation of human neural precursor cells. , *Frontiers in Genetics*, 6.
25 Ray, M.K. *et al.* (2016) CAT7 and cat7l Long Non-coding RNAs Tune Polycomb Repressive Complex 1 Function during Human and Zebrafish Development. *J. Biol. Chem.* 291, 19558–19572
26 Pandey, G.K. *et al.* (2014) The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell* 26, 722–737
27 Gendron, J. *et al.* (2019) Long non-coding RNA repertoire and open chromatin regions constitute midbrain dopaminergic neuron - specific molecular signatures. *Sci. Rep.* 9, 1409
28 Ang, C.E. *et al.* (2019)The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. , *eLife*, 8.
29 Mirzadeh Azad, F. *et al.* (2019) lncRNA PSORS1C3 is regulated by glucocorticoids and fine-tunes OCT4 expression in non-pluripotent cells. *Sci. Rep.* 9, 8370
30 Kretz, M. *et al.* (2012) Suppression of progenitor differentiation requires the long noncoding RNA ANCR. , *Genes & Development*, 26. , 338–343
31 Tanis, S.E.J. *et al.* (2019) BLNCR is a long non-coding RNA adjacent to integrin beta-1 that is rapidly lost during epidermal progenitor cell differentiation. *Sci. Rep.* 9, 31
32 Kretz, M. *et al.* (2013)Control of somatic tissue differentiation by the long non-coding RNA TINCR. , *Nature*, 493, 231–235
33 Li, C.-X. *et al.* (2017) H19 lncRNA regulates keratinocyte differentiation by targeting miR-130b-3p. *Cell Death Dis.* 8, e3174
34 Panatta, E. *et al.* (2020) Long non-coding RNA uc.291 controls epithelial differentiation by interfering with the ACTL6A/BAF complex. *EMBO reports*
35 Jiang, W. *et al.* (2015) The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. *Cell Rep.* 11, 137–148
36 Yang, J. *et al.* (2020) GATA6-AS1 Regulates GATA6 Expression to Modulate Human Endoderm Differentiation. *Stem Cell Reports* DOI: 10.1016/j.stemcr.2020.07.014

37 Daneshvar, K. *et al.* (2016) DIGIT Is a Conserved Long Noncoding RNA that Regulates GSC Expression to Control Definitive Endoderm Differentiation of Embryonic Stem Cells. *Cell Rep.* 17, 353–365
38 Li, J. *et al.* (2019) Long noncoding RNA ANCR inhibits the differentiation of mesenchymal stem cells toward definitive endoderm by facilitating the association of PTBP1 with ID2. *Cell Death Dis.* 10, 492
39 Chen, Y.-F. *et al.* (2020) Control of matrix stiffness promotes endodermal lineage specification by regulating SMAD2/3 via lncRNA LINC00458. *Sci Adv* 6, eaay0264
40 Gaertner, B. *et al.* (2020) A human ESC-based screen identifies a role for the translated lncRNA LINC00261 in pancreatic endocrine differentiation. *Elife* 9,
41 Zhang, J. *et al.* (2020) The long-noncoding RNA MALAT1 regulates TGF-β/Smad signaling through formation of a lncRNA-protein complex with Smads, SETD2 and PPM1A in hepatic cells. *PLoS One* 15, e0228160
42 Banerjee, P. *et al.* (2018) Long Noncoding RNA RP11-380D23.2 Drives Distal-Proximal Patterning of the Lung by Regulating PITX2 Expression. *Stem Cells* 36, 218–229
43 Frank, S. *et al.* (2019) yylncT Defines a Class of Divergently Transcribed lncRNAs and Safeguards the T-mediated Mesodermal Commitment of Human PSCs. *Cell Stem Cell* 24, 318–327.e8
44 Sartorelli, V. and Caretti, G. (2005) Mechanisms underlying the transcriptional regulation of skeletal myogenesis. *Curr. Opin. Genet. Dev.* 15, 528–535
45 Zhou, L. *et al.* (2015) Linc-YY1 promotes myogenic differentiation and muscle regeneration through an interaction with the transcription factor YY1. *Nat. Commun.* 6, 10026
46 Gong, C. *et al.* (2015) A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev. Cell* 34, 181–191
47 Ulitsky, I. *et al.* (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537–1550
48 Musunuru, K. *et al.* (2010) Stem cell models of cardiac development and disease. *Annu. Rev. Cell Dev. Biol.* 26, 667–687
49 Ounzain, S. *et al.* (2015) CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *J. Mol. Cell. Cardiol.* 89, 98–112
50 Klattenhoff, C.A. *et al.* (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152, 570–583
51 Kurian, L. *et al.* (2015) Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development. *Circulation* 131, 1278–1290
52 Liu, J. *et al. (2017)* HBL1 Is a Human Long Noncoding RNA that Modulates Cardiomyocyte Development from Pluripotent Stem Cells by Counteracting MIR1. , *Developmental Cell*, 43, 372
53 Zhao, Y. *et al.* (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129, 303–317
54 Wang, P. *et al.* (2014) The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* 344, 310–313
55 Zhang, R. *et al.* (2016) A long noncoding RNA positively regulates CD56 in human natural killer cells. *Oncotarget* 7, 72546–72558
56 Zhang, X. *et al. (2009)* A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood*, 113, 2526–

2534
57 Yen, Y.-P. *et al.* LncRNA Meg3 Choreographs the Epigenetic Landscape of Postmitotic Motor Neuron Cell Fate and Subtype Identity. *SSRN Electronic Journal*.
58 Otani, J. *et al.* (2013) Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PLoS One* 8, e82961
59 Neri, F. *et al.* (2015) Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep.* 10, 674–683
60 Wu, T.P. *et al.* DNA methylation on N6-adenine in mammalian embryonic stem cells. , *Nature*, 532. (2016), 329–333
61 Zhao, Y. and Garcia, B.A. (2015) Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* 7, a025064
62 Hyun, I. *et al.* Embryology policy: Revisit the 14-day rule. , *Nature*, 533. 12-May-(2016), 169–171
63 Yu, J. *et al.* (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917–1920
64 Xu, C. *et al.* (2001) Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* 19, 971–974
65 Itskovitz-Eldor, J. *et al.* (2000) Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Mol. Med.* 6, 88–95
66 Kouskoff, V. *et al.* (2005) Sequential development of hematopoietic and cardiac mesoderm during embryonic stem cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13170–13175
67 Turner, D.A. *et al.* (2016) Organoids and the genetically encoded self-assembly of embryonic stem cells. *Bioessays* 38, 181–191
68 Warmflash, A. *et al.* (2014) A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* 11, 847–854
69 Martyn, I. *et al.* (2019) Mapping cell migrations and fates in a gastruloid model to the human primitive streak. *Development* 146,
70 Hattori, N. (2014) Cerebral organoids model human brain development and microcephaly. , *Movement disorders: official journal of the Movement Disorder Society*, 29 , 185
71 Kim, J. *et al.* (2020) Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.* 21, 571–584