Establishing pluripotency in early development

Sarita S. Paranjpe, Gert Jan C. Veenstra *

Radboud University, Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 12 January 2015
Received in revised form 26 March 2015
Accepted 30 March 2015
Available online 7 April 2015

Keywords:
Pluripotency
Embryo
Chromatin
Methylation
Zygotic genome activation

ABSTRACT

The earliest steps of embryonic development involve important changes in chromatin and transcription factor networks, which are orchestrated to establish pluripotent cells that will form the embryo. DNA methylation, histone modifications, the pluripotency regulatory network of transcription factors, maternal factors and newly translated proteins all contribute to these transitions in dynamic ways. Moreover, these dynamics are linked to the onset of zygotic transcription. We will review recent progress in our understanding of the relationship between ZGA and pluripotency.

1. Introduction: routing fertilized eggs to pluripotency

At the very beginning of embryonic development two specialized and highly differentiated cells, the gametes, fuse to form the zygote which in turn produces all cell types of the organism (as well as extra-embryonic tissue in the case of mammals, see below). The parental genomes show different histone modification patterns and are subject to dramatic chromatin reorganization, DNA demethylation and remethylation after fertilization and during early development, in order to reprogram the sperm and oocyte epigenomes [1–7]. Early in development a maternal to zygotic transition (MZT) is triggered, which passes regulatory control of development from maternal to newly synthesized components [8,9]: this regulatory event can be defined as the period of time encompassing the initial degradation of maternal transcripts, zygotic genome activation (ZGA, the onset of transcription), until the first major morphological requirement for zygotic transcripts in embryonic development [9]. Following the ZGA, pluripotent cells emerge which will give rise to the three germ layers of the embryo, ectoderm, mesoderm and endoderm, however the relationship between ZGA and pluripotency is different between mammals and non-mammalian vertebrates (Fig. 1, Table 1).

In amniotic species such as mammals, the zygote starts transcribing its own genes just before the two-cell stage. Subsequently, trophoderm and primitive endoderm, cell lineages that contribute to placental development, are set up in addition to the pluripotent cells of the inner cell mass of the blastocyst that will form the organism (Fig. 1). Therefore the mammalian zygote is referred to as totipotent, being able to produce all cell types of both the embryo proper and embryonic placental tissues [10].

Although early embryonic development is strikingly different in *Xenopus* and zebrafish compared to mouse, pluripotency and subsequent germ layer commitment, patterning and convergent extension are functionally highly analogous in these species. In *Xenopus* and zebrafish early cleavage development produces a blastula embryo, which undergoes ZGA (Fig. 1). This is also referred to as the mid-blastula transition (MBT) [8,9,11]. Cell cycle lengthening and the acquisition of cell motility coincide with the onset of embryonic transcription at the MBT [11] and at this stage, at and immediately after the MBT, cells at the animal pole of the embryo are pluripotent (Fig. 1). These cells are normally fated to give rise to ectoderm (epidermal ectoderm and neural ectoderm) but can also give rise to mesoderm and endoderm derivatives when exposed to specific factors [12].

Pluripotency, as it emerges from the zygote, represents a functional cellular state, much in the same way differentiated cells have a defined set of biochemical and cellular properties. These properties emerge from a cell type-specific reading of the genomic sequence information, which is part of what is referred to as epigenetic regulation. At the molecular level this involves chemical modifications of either the DNA itself (for example methylation of cytosine residues) or the chromosomal proteins associated with genomic DNA (chromatin). The profiles of epigenetic modifications can vary between cells and developmental stages and form a molecular regulatory intermediate between genomic sequence information and biochemical and cellular properties of cells. Epigenetics therefore constitutes a developmental stage- and cell type-specific filter of genomic sequence information.

In this review we compare vertebrate studies on the pluripotent chromatin state and how it emerges during embryonic development.
We will discuss DNA methylation, histone modifications, cis-regulatory elements and transcription factors that prime the zygote for pluripotency. We will illustrate the findings from mouse, frog and fish embryo models but also discuss findings in cellular models of pluripotency such as mouse embryonic stem (ES) cells where warranted. ES cells represent stable pluripotency in vitro, whereas embryonic pluripotency is transitory. However much has been learned from these systems that also is highly relevant for pluripotency in vivo. The reader is referred to excellent reviews for other aspects of pluripotent chromatin and embryogenesis, including more detailed discussions of the MZT [8,9], chromatin interactions and complexes [13–17], and naive and primed pluripotency in cell culture [18,19].

2. Global DNA methylation dynamics in relation to pluripotency

2.1. DNA methylation dynamics in mammalian development

DNA methylation is an important epigenetic modification with a role in a variety of processes, including tissue-specific gene expression, development and cellular differentiation, carcinogenesis and aging, and specifically for mammals, genomic imprinting and X chromosome inactivation [20–22]. Methylation at the 5-position of cytosine (5mC) occurs mainly at CG dinucleotides (commonly referred to as CpG dinucleotides) in vertebrates and is a prerequisite for normal embryogenesis; the DNA methyltransferases DNMT1, DNMT3a and DNMT3b are all essential for early mouse development [23,24]. DNA methylation

Fig. 1. Embryonic development in relation to zygotic gene activation (ZGA) and pluripotency in mouse (top panel), Xenopus (middle panel) and zebrafish (bottom panel). ZGA occurs at the late 1-cell stage in mouse, but does not happen until the 12th and 10th cell cycle in Xenopus and zebrafish respectively. In mouse, before pluripotency is established two other lineages, trophoderm and primitive endoderm, need to be formed. The inner cell mass (ICM) epiblast cells of the pre-implantation blastocyst (mouse, 3.5 days, red cells) represent the ground state of pluripotency in vivo, whereas in day 6 embryos (still before gastrulation) these cells have been primed towards differentiation. In Xenopus, vegetal pole cells have been maternally specified to form endoderm and secrete signals to induce mesoderm in the marginal zone. Animal pole cells (red) correspond to the pluripotent cells of the mid-blastula embryo. Times post-fertilization are indicated for X. laevis at 22 °C. In zebrafish, the mid-blastula transition (MBT) starts two cell cycles earlier. Zebrafish has one large yolk cell which does not divide. The nuclei from blastodisc cells closest to the yolk cell form a syncytium with the yolk cell. This region has a mesendodermal fate. Zebrafish times post-fertilization are temperature-dependent and indicated at 28.5 °C.

Table 1
Overview of early development and pluripotency in mouse, Xenopus and zebrafish. Note the interspecific differences in the relationship between (1) zygotic genome activation (ZGA) and pluripotency and (2) DNA methylation and pluripotency. Abbreviations: hypoM, hypomethylated; hyperM, hypermethylated; deM, demethylated; ND, not determined.

<table>
<thead>
<tr>
<th></th>
<th>Mouse</th>
<th>Xenopus</th>
<th>Zebrafish</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZGA stage</td>
<td>2 cells</td>
<td>Blastula</td>
<td>512 cells</td>
</tr>
<tr>
<td>Pluripotent stage</td>
<td>Blastocyst (E3.5)</td>
<td>Blastula</td>
<td>512 cells</td>
</tr>
<tr>
<td>Global DNA methylation</td>
<td>hypoM at ZGA and pluripotency</td>
<td>hyperM at ZGA and pluripotency</td>
<td>hyperM at ZGA and pluripotency</td>
</tr>
<tr>
<td>Paternal DNA methylation</td>
<td>hyperM, active deM after fertilization</td>
<td>ND</td>
<td>hyperM, maintained</td>
</tr>
<tr>
<td>Maternal DNA methylation</td>
<td>Relatively hypoM, passive deM</td>
<td>ND</td>
<td>Relatively hypoM, hyperM after 16-cell stage</td>
</tr>
<tr>
<td>Histone H3 methylation</td>
<td>ZGA and later</td>
<td>ZGA and later</td>
<td>ZGA and later</td>
</tr>
<tr>
<td>Pluripotency transcription factors</td>
<td>OCT4 (POU5F1)</td>
<td>OCT4 (POU5F1)</td>
<td>OCT4 (POU5F1)</td>
</tr>
<tr>
<td>Sox2</td>
<td>Sox2</td>
<td>Sox2</td>
<td>Sox2, Sox10b</td>
</tr>
<tr>
<td>NANOG</td>
<td>Nanog</td>
<td>Nanog</td>
<td>Nanog</td>
</tr>
</tbody>
</table>

S.S. Paranjpe, G.J.C. Veenstra / Biochimica et Biophysica Acta 1849 (2015) 626–636
can be reversed passively (lack of maintenance during DNA replication) and by members of the ten–eleven translocation (TET) family of 2OG-Fe(II) dioxygenases which catalyze the hydroxylation of 5mC to generate 5-hydroxymethylcytosine (5hmC), which can be further modified to 5-formylcytosine and 5-carboxylcytosine and subsequently can be removed by base excision repair [13,25–27].

DNA methylation in somatic cells portrays a bimodal configuration, in which the majority of CpG sites are methylated; unmethylated CpGs are primarily found in clusters, known as CpG islands (CGIs), which are frequently associated with gene promoters [28,29]. Two-thirds of gene promoters in mammals are associated with CGIs. 5mC signatures are enriched at sites of DNAseI hypersensitivity, which are indicative of genomic regions bound by regulatory proteins, whereas 5mC is generally much less abundant at these locations. 5hmC is particularly enriched near transcription start sites and at active and poised enhancers.

In mouse the paternal genome derived from sperm shows 80 to 90% overall CpG methylation but is almost completely demethylated shortly after fertilization (Table 1) [30,31]. TET3, which is abundant in oocytes and zygotes, plays an important role in this active demethylation process by oxidizing 5mC, which is then passively lost during subsequent rounds of replication [32–34]. TET1 is not required for pluripotency and development, although its loss causes a decrease of 5hmC levels in ES cells and a skewed differentiation toward pluripotency and development, although its loss causes a decrease allows ES cells to colonize the placenta in embryo chimeras [38]. TET2 is important for maintaining active chromatin at the zygote stage, and zygotes, plays an important role in this active demethylation process by oxidizing 5mC, which is then passively lost during subse-

2.2. DNA methylation dynamics in non-mammalian vertebrates

Promoters in non-mammalian species tend to be much less CpG-dense and many do not meet the general sequence criteria of CGIs, but like mammalian CGI promoters they contain a cluster of unmethylated CpG dinucleotides and they also have a similar core sequence [43–45]. In zebrafish the oocyte genome is significantly hypomethylated compared to sperm, similar to mammals [46,47]. The zygote shows an average of the DNA methylation levels of sperm and oocytes immediately after fertilization, which gradually increases to the level observed in sperm by the MBT (Table 1). Genes which, against the general trend, are hypermethylated in oocytes, are demethylated during early development [47], showing that both the generally hypomethylated oocyte genome as well as specific hypermethylated oocyte genes are remodeled to a sperm-like profile during early embryogenesis. Importantly, maternal haploid embryos remodel their epigenome to a sperm-like pattern, showing that the paternal genome is not instructive for maternal genome remodeling [47]. By the time of ZGA and the emergence of pluripotent cells the zebrafish embryonic genome is relatively hypermethylated with no major changes before and after the onset of transcription. This is in striking contrast to the mouse genome which is globally hypomethylated by the time of ZGA, to be remethylated during pre-implantation development [7]. Another difference is the extent of methylene remodeling. Early reports indicated an absence of global demethylation in both fish and frogs [48,49] and the remodeling of the maternal methylome identified in recent zebrafish studies indeed does not lead to global demethylation but affects specific sequences [46,47].

In Xenopus, genomic DNA in sperm as well as embryos of early blastula and later stages are globally hypermethylated [44,49]. Depletion of maternal dnmt1 by antisense RNA during cleavage stages is associated with a decrease in the genomic 5mC content and leads to the activation of zygotic transcription approximately two cell cycles earlier than normal [50]. Hypomethylation allows the early expression of mesodermal and organizer genes such as t (brachyury), cerberus, and otx2, which are subsequently down-regulated during gastrulation of the dnmt1-depleted embryos. However, Dnm1t1-dependent gene repression was subsequently found to be independent of its catalytic activity, suggesting that Dnmt1 can also act as a potent transcriptional repressor independent of DNA methylation [51].

Of the three TET family members in mammals, only two (tet2 and tet3) are present in the Xenopus genome [52]. In contrast to mouse, in which Tet3 is abundant in oocytes and zygotes and contributes to active demethylation of sperm DNA, frog tet3 is extremely low in oocytes and cleavage stage embryos, consistent with an absence of global demethyl-

2.3. Direct and indirect reading of DNA methylation

The studies cited above indicate that the temporal paths and the extent of DNA methylation remodeling are different between mammals and fish, and that global hypomethylation is not a conserved aspect of ZGA or pluripotency among vertebrates. This may suggest that, generally, not the demethylated state, but the act of remodeling is significant for early development. DNA methylation is known to recruit methyl CpG binding proteins, which cause repression of transcription [53]. However, contrary to this general paradigm of DNA methylation-associated repression, methylated DNA can also recruit many other proteins including transcriptional activators, and methylated promoters can drive active transcription in some cases [53–57]. This diversity in methylation readout is probably highly regulated. Notable differences have been observed in DNA methylation readers in mouse ES cells, neural progenitor cells (NPCs) and brain tissue. For example, the MBD2-containing NURD complex is a reader of 5mC in NPCs and brain, but not in ES cells [56]. Moreover, in both oocytes and late stage embryos of Xenopus methylated promoter templates are strongly repressed in cis, however this is not observed in early blastula and gastrula embryos. In early embryos CpG-dense methylated promoters drive active transcription while DNA methylation is maintained. These results point to a temporal uncoupling of DNA methylation and repression of transcription during development [44].

Interestingly, in mouse ES cells, global DNA methylation status depends on the culture conditions. When grown in 2i media, the genome is globally hypomethylated in much the same way as the ICM cells of the blastocyst, whereas ES cells grown in serum plus LIF are also pluripotent but relatively hypermethylated and in this respect more similar to the
3. Histones and histone modifications

The major proteins in chromatin are the core histone proteins H2A, H2B, H3, and H4, which form nucleosome particles. The N-terminal tails undergo a diverse array of posttranslational modifications and regulate the accessibility to the underlying DNA. These modifications are deposited and removed by enzymes (‘writers’ and ‘erasers’) and are recognized by specific proteins (‘readers’) which cause a variety of functional outcomes such as transcriptional activation or repression of genes [reviewed in refs. [20,63–65]]. These chromatin states can be transmitted from mother to daughter cells. Lysine acetylation correlates with chromatin accessibility, recruitment of specific proteins and transcriptional activity, whereas lysine methylation can have different effects depending on which residue is modified. Both ES cells and early embryos have a relatively open chromatin with histones showing relatively high levels of modifications that are involved in gene activation such as histone H3 lysine 4 trimethylation (H3K4me3), acetylation of lysines 9 and 14 (H3K9ac, H3K14ac), and histone H4 acetylation (reviewed in [65,66]). H3K4me3 is found at active promoters while H3K9me3 and H3K27me3 mark repressed cell lineage-regulatory genes during pluripotency while the H3K4me3 mark could poise these genes for activation upon differentiation. In serum-grown ES cells, most bivalent promoters are transcribed at very low levels (with some exceptions, such as Klf4, Klf5 and Rex1) and are considered to be poised for activation by developmental signals. In mammals six methyltransferase proteins, SET1A, SET1B and MLL1 to MLL4, found in COMPASS-like complexes, are capable of methylating H3K4 [17]. In fibroblasts, MLL1 is required for the H3K4 trimethylation of a small subset of promoters, including a subset of Hox genes [71]. Loss of H3K4 methylation at these loci causes reduced levels of transcription initiation by RNA polymerase II. MENIN, a shared subunit of both MLL1 and MLL2 complexes, is required for most of H3K4 methylation at Hox loci, whereas this does not depend on the MLL3, MLL4 or SET1 complexes. In line with these findings, in mouse ES cells MLL2 is required for H3K4 methylation of bivalent genes that are also regulated by H3K27me3, including many of the Hox genes [72,73]. Loss of MLL2, however, has very little effect on gene activation during ES cell differentiation.

The H3K4me3 and H3K27me3 modifications are unlikely to occur on the same histone H3 tail [74]. Two copies of H3 are present in the nucleosome octamer, however, which could accommodate the two modifications. This would correspond to an asymmetrically modified nucleosome or asymmetric bivalency, rather than intra-molecular bivalency (Fig. 2). In heat maps of the genomic distribution of histone modifications of bivalent chromatin around transcription start sites, it is apparent that the H3K4me3-positive center of bivalent promoter regions contains less H3K27me3 signal compared to flanking regions [75], suggesting the presence of poly-nucleosomal bivalency (Fig. 2). In this type of bivalency adjacent nucleosomes contain the opposing histone modifications. In some cases cell populations in which bivalency is observed are heterogeneous. In this case the opposing histone modifications may be present on the same locus in different cells [76]. This represents pseudo-bivalency among cells with different patterns of epigenetic modifications (Fig. 2). Pseudo-bivalency however cannot explain the co-occurrence of H3K4 and H3K27 methylation in highly homogenous cells such as naïve pluripotent stem cells [75]. In addition, mass spectrometry-based analyses of ES cells have indicated that 15% of mononucleosomes modified with H3K4me3 also carry H3K27me3 on the second copy of histone H3 within the nucleosome [77]. Together these studies suggest that asymmetric, poly-nucleosomal and pseudo-
bivalency all contribute to the observations of co-occurring H3K4me3 and H3K27me3 in chromatin in ES cells.

During mouse early development, both H3K4me3 and H3K27me3 appear from the 2-cell stage onwards (Table 1). In the pre-implantation blastocyst, the inner cell mass (ICM) contains the pluripotent cells that will give rise to the embryo proper, whereas trophectoderm and primitive endoderm cells will contribute to extra-embryonic tissue. Trophoblast stem cells and endoderm stem cells (cultured cells derived from the embryo) show very little H3K27 methylation in contrast to ES cells, whereas these cells have similar levels of H3K4me3 [78]. Similarly, ICM and trophectoderm show different H3K27me3 levels [79]. Therefore, there is significant H3K27 methylation in contrast to ES cells, whereas these cells have similar levels of H3K4me3 [78].

In Xenopus early embryos, bivalent chromatin was observed, but not as a quantitatively dominant feature [80]. Instead, the marks accumulate by and large with different spatio-temporal kinetics during the time between the MBT and the end of gastrulation. H3K4me3 and H3K27me3 segregate spatially within the embryo (pseudo-bivalency, Figs. 2 and 3), reflecting localized gene expression in different cells. Moreover H3K4me3 precedes accumulation of H3K27me3 and classical bivalency is present at low levels but is not a dominant feature of the pluripotent chromatin state in Xenopus. Interestingly, H3K27me3 nucleates first in promoter-distal locations in blastula embryos, before accumulating in larger domains (Fig. 3), highlighting the spatio-temporal hierarchy of H3K4 and H3K27 methylation during pluripotency and germ layer induction [60,80]. Consistent with these findings, Geminin promotes Polycomb binding and H3K27 methylation and contributes to the transition from pluripotency to lineage commitment by preventing multi-lineage commitment [81]. The abundance of histone modifications has also been determined by quantitative mass spectrometry; the results showed relatively abundant H3K4me3 from early stages on, whereas H3K27me3 accumulated over time and was much less abundant at any stage compared to mouse ES cells [82].

Bivalent chromatin has been reported in early zebrafish embryos, immediately after the MBT [83]. Also in zebrafish embryos K4 accumulates first, being very abundant at the MBT, with K27 methylation accumulating significantly between the MBT and 50% epiboly (mid-gastrulation) [84]. These studies have also highlighted the accumulation of H3K4me3 before the MBT and transcriptional activation, suggesting a role for maternal factors in setting a permissive mark for transcription in preparation for the pluripotent stage of development.

---

**Fig. 2.** Different types of bivalent chromatin. Bivalent chromatin is defined by the co-occurrence of histone H3 lysine 4 and lysine 27 trimethylation (H3K4me3 and H3K27me3). In the case of intra-molecular bivalency they would co-occur on the same H3 tail; however this is unlikely to happen in vivo (see text). Asymmetric bivalency: Nucleosomes can carry the two histone modifications at two different H3 tails within the same nucleosome. Poly-nucleosomal bivalency: Adjacent nucleosomes can carry the opposing histone modifications. Pseudo-bivalency: Different cells within a population (or alleles within a cell) carry different histone modifications. This type is highly relevant in vivo where pluripotency is a transient and unstable state, leading to lineage commitment in different cells. H3K4me3 and H3K27me3 are indicated in green and red respectively.

**Fig. 3.** Histone modifications and DNA methylation in relation to pluripotency and germ layer induction in Xenopus. Overall the genome is hypermethylated, but both H4K4me3 and H3K27me3 emerge in unmethylated CpG islands with different kinetics.
3.2. Enhancer histone modification signatures

Enhancers are distinguished from promoters by robust levels of H3K4me1 while H3K4me3 is absent or very low [85]. In addition they may be decorated with H3K27 acetylation and feature the recruitment of RNAPII and the EP300/ CBP histone acetyltransferase which catalyzes H3K27 acetylation. Several hundred thousand enhancers have been identified in the human genome by profiling these histone modifications and the transcriptional coactivator EP300 [86–90]. Many of these enhancers drive cell-type-specific gene expression, for example in ES cells. Enhancers decorated with both H3K4me1 and H3K27ac are considered active enhancers, as opposed to primed enhancers only marked with H3K4me1, and the active enhancer subset is more predictive of developmental state than all enhancers [86,90]. In addition, H3K4me1-positive enhancers may be enriched for H3K27me3 rather than H3K27ac, which are referred to as poised enhancers. Profiling these histone modifications in ES cells has uncovered human enhancers that are poised in ES cells but become active during differentiation [89].

Validation of the activity of these human enhancers in zebrafish embryos showed that they are able to direct cell-type and stage-specific expression even in the absence of sequence conservation in the fish genome. During differentiation a subset of poised enhancers acquires H3K27 acetylation, RNAPII association and enhancer RNAs. These data show that developmental enhancers are epigenetically pre-marked and kept in a poised state during pluripotency. There are indications that the H3K4me1 mark is important not just for poised or primed enhancers, but also for active enhancers. For example, the Drosophila Trr complex and its mammalian ML3L and ML4L counterparts are involved in H3K4 monomethylation at enhancers [91–93]. These complexes also contain UTX (KDM6A), an H3K27 demethylase which counteracts PRC2-mediated H3K27 methylation. This is identical to the lysine cannot accommodate both acetylation and methylation simultaneously. Indeed, loss of Trr/ML4L leads to global reduction of H3K27acetylation, implicating the ML3L/4 complexes in the function of both poised and active enhancers.

Profiling of enhancers in Xenopus blastula embryos using H3K4me1 uncovered many enhancers, a subset of which recruited RNAPII, and another partially overlapping subset recruited the PRC2 catalytic subunit Ezh2 and the PRC2-associated factor Jarid2 [60], illustrating the dynamic balance of opposing activities present on enhancers during pluripotency. In zebrafish, changes in H3K27 acetylation enrichment accompany a shift from pluripotent to tissue-specific gene expression [94]. A marked increase in the number of enhancers was observed from the dome stage (blastula) to 80% epiboly (gastrula). This illustrates the lysine cannot accommodate both acetylation and methylation simultaneously. Indeed, loss of Trr/ML4L leads to global reduction of H3K27acetylation, implicating the ML3L/4 complexes in the function of both poised and active enhancers.

3.3. Histone variants and linker histones

The histone proteins are highly conserved; however there are nonallelic variants of the major histones that show a divergence ranging from almost no amino acid differences to significant changes. Localized replacement of a canonical histone with a variant can alter chromatin state. Among the core histones, H2A has the largest number of variants, including H2A.Z, MacroH2A, H2A-Bbd, H2AvD, and H2A.X [95–97]. H2A–H2B dimers can be replaced by several different H2A variants independently of replication. Histone variant incorporation is one of the mechanisms to create different states through a combination of altered chromatin structure or trans-acting factors. For instance, depletion of H2A.Z leads to reduced chromatin accessibility with a decrease in the binding of both active and repressive complexes and a reduction in the binding of the pluripotency transcription factor OCT4 to its binding sites at pluripotency genes. As a consequence the cells showed a reduction in the efficiency of self-renewal [98]. A study investigating the reprogramming of mammalian nuclei transplanted to Xenopus oocytes showed that the H3R3-dependent dehossylation of H3.3 at the oct4 locus was required for transcriptional reprogramming to a pluripotent state [99]. Another study in Xenopus investigated the role of H3.3 during embryonic development. Partial depletion of H3.3 results in abnormal development, whereas substantial depletion of both H3.1 and H3.3 showed a distinct gastrulation arrest phenotype. Using lineage marker analyses, the defect was attributed to a loss of competence to respond to mesoderm inducing cues and this defect was attributed to perturbation in chromatin organization brought about by loss of H3 and abnormal incorporation of linker histone H1A [100]. Replacement of an oocyte-specific linker histone by somatic linker histones has been shown to contribute to a loss of mesoderm competence [101]. The H1 linker histone contributes to silencing of pluripotency factors and participates in mediating changes in DNA methylation and histone marks necessary for silencing of pluripotency genes during differentiation [102]. These findings illustrate that the pluripotency gene network requires both the incorporation and loss of specific histone variants, and that remodeling chromatin state plays a role in both establishing and the exit of pluripotency.

4. Chromatin remodeling

As outlined above, chromatin state is affected by myriad proteins, including enzymes that catalyze the particular chemical modifications, proteins that recognize and bind the modifications, and the enzymes that remove this modification. In addition, chromatin state is affected by chromatin remodeling complexes that enable context-dependent access to packaged DNA. Chromatin remodelers act on nucleosomes, the basic repeating unit of packaged DNA. They use the energy of ATP hydrolysis to move, destabilize, eject, or restructure nucleosomes and thereby increase the fluidity of chromatin. There are four families of nucleosome remodelers [103,104]: (1) SWI/SNF [switching defective/sucrose nonfermenting], also called BAF [Brg/Brahma-associated factor]; (2) ISWI [imitation switch]; (3) CHD [chromodomain, helicase, and DNA binding]; and (4) INO80 [inositol requiring 80]. All these subunits of the complexes are essential for the maintenance of pluripotency and proliferation of ES cells.

The SWI/SNF (BAF) complexes of mammals contain either BRM or BRG1 as nucleosomal ATPase subunits. Whereas BRM is dispensable for mouse development, BRG1 is essential for inner cell mass and trophectoderm survival and preimplantation development [105]. esBAF is an ES cell-specific SWI/SNF complex, characterized by subunits BRG1, BAF155 and BAF60A, and the absence of BRM, BAF170, and BAF60C. esBAF is required for the maintenance of pluripotency. Like BRG1, BAF155 is important for preimplantation development, maintenance of Oct4 expression and proliferation of ES cells [106]. In ES cells, knockdown of BRG1 has been shown to downregulate pluripotency genes and upregulate differentiation genes. BRG1 not only occupied the promoters of pluripotency genes but also occupied a subset of OCT4, SOX2, NANOG and PcG protein target genes [107]. esBAF is characterized by the presence and absence of specific subunits and this complex composition has been shown to direct its specific role in the core pluripotency network [106]. In murine ES cells, esBAF has been shown to be involved in a complex regulatory circuitry. BRG1-esBAF functions positively with PcG to repress differentiation genes (all four Hox loci), but prevents PcG-mediated repression of pluripotency genes activated by ES cell transcription factors like OCT4, SOX2 and STAT3 [108].

CHD1 is necessary for the maintenance of open chromatin in ES cells [109]. Knockdown of Chd1 increases heterochromatin in ES cells.
and reduced Oct4 expression. Surprisingly this knockdown decreased the expression of only 25 genes, suggesting a functional redundancy with other chromatin remodeling complexes. A much larger group of genes was upregulated upon Chd1 knockdown, including genes involved in neurogenesis. Although ES cells with Chd1 knockdown can maintain an undifferentiated state, the cells lose the ability to differentiate into primitive endoderm and tend to differentiate into neural lineages, indicating that CHD1 is necessary for the maintenance of pluripotency [109]. More recently CHD1 has been specifically implicated in maintaining high levels of RNAPII and II in transcoding protein-coding genes and ribosomal DNA, thereby sustaining the transcriptional competence of pluripotent cells [110].

Nucleosomal remodeling deacetylase (NuRD) complexes contain Mi2 (Chd3/Chd4) nucleosomal ATPase, as well as methyl-CpG-binding proteins MBD2 or MBD3, histone deacetylase and a number of other subunits. This complex is the most abundant source of histone deacetylase in Xenopus eggs and early embryos [111]. Mouse ES cells lacking MBD3 are viable but cannot differentiate properly and show LIF-independent self-renewal [112]. A functional NuRD complex has been implicated in maintaining transcriptional heterogeneity in ES cell populations by inhibiting pluripotent gene expression and promoting exit from pluripotency [113]. Several hundred genes show increased expression in MBD3 knockdown, decreased expression in BRG1 knockdown and a wild-type like expression in double knockdowns, uncovering their opposing effects [114]. Unlike ES cells, MBD3-deficient blastocyst ICM cells grown ex vivo fail to maintain a population of Oct4-positive pluripotent cells [115]. On a similar note, MBD3 is required for reprogramming and the derivation of induced pluripotent stem (iPS) cells [116]. Depleting MBD3/NuRD complex in somatic cells impairs somatic cell reprogramming by establishing heterochromatin features and repression of key pluripotency genes [117].

Like esBAC, the INO80 nucleosomal remodeling complex is found at pluripotency genes along with Oct4, Sox2 and Nanog [118]. It mediates the recruitment of RNAPII and maintains open chromatin structure in ES cells as shown by micrococcal nuclease and DNAseI hypersensitivity assays. Knockdown of Ino80 causes a decreased expression of pluripotency factors and a loss of ES cell morphology. Expression of Ino80 is important for iPS cell colony formation. Knockdown of Ino80 in early embryos caused impaired blastocyst formation and reduced expression of pluripotency factors in the ICM [118]. The role of chromatin remodeling in Xenopus and zebrafish pluripotency has yet to be determined. In mouse, the results discussed above illustrate how mobilization of nucleosomes and/or nucleosome eviction at gene-regulatory regions is required for development and maintenance of pluripotency.

5. The pluripotency transcription factor network

The concomitant actions of cis- and trans-acting elements orchestrate the early developmental programs, influencing the cascade of events that leads to functional pluripotency. In this section we will highlight how transcription factors control the emergence of pluripotency.

5.1. The core pluripotency network in ES cells

Oct4, Sox2 and NANOG form a “core” network of pluripotency transcription factors [119,120]. They are co-recruited to regulatory elements of target genes and form auto-regulatory and feedforward loops that provide stability to pluripotency gene expression and suppress lineage-specific gene expression. These factors regulate each other and process signals to direct self-renewal and block differentiation by maintaining the expression of the pluripotency network. The binding of these transcription factors is transduced by the mediator complex to recruit general transcription factors to the promoter, which promote the recruitment of RNAPII complex. The mediator subunit MED12 interacts with Nanog and these factors work together in ES cell gene regulation [121]. In addition to the well-characterized core network of OCT4, SOX2 and NANOG, an additional set of transcription factors, including ESRRB, TFCP2L1, KLF2, and KLF4, are important contributions to the pluripotency regulatory circuitry and can reset human ES cells to ground-state pluripotency [122,123]. Three KLF transcription factors, KLF2, KLF4 and KLF5, share many transcriptional targets with NANOG and regulate Nanog and other pluripotency genes [124].

Mouse ES cells cannot only be grown in 2i or serum plus LIF conditions, conditions that correspond to naive (ground-state) and primed pluripotent states, they can also be primed by exposure to FGF2 and activin, leading to the formation of a transient epiblast-like cells (EpilCs). ES cells and EpilCs share the core transcription network of OCT4, SOX2, and NANOG. The transition between these two pluripotent states, however, is associated with widespread OCT4 relocalization accompanied by global rearrangement of enhancer chromatin landscapes [125]. Motif analyses suggested OTX2 and ZIC3/2 as mediators of primed pluripotency–specific OCT4 binding. Blocking differentiation signals and overexpressing OTX2 was sufficient to drive exit from the naive state and induce transcription of several primed pluripotency genes. OTX2 drives the early stage of ES cell differentiation in collaboration with OCT4 at enhancers [126]. ZIC3 is an activator of NANOG and contributes to maintenance of pluripotency in mouse ES cells [127,128], but in zebrafish and Xenopus embryos it functions in neural and neural crest gene expression [129,130]. Epiblast stem cells (EpiScs) are derived directly from developing mouse epiblast embryos; a subset of these cells express OCT4 in addition to SOX2. However EpiScs also express post-implantation markers and show heterogeneous expression of lineage specific genes of mesendoderm, definitive endoderm and the primitive streak [131,132].

The transcription initiation factor TBP and many of its associated factors (TAFs) of the TFIIID complex are expressed at higher levels in mouse ES cells than in somatic cells. Knockdown of these factors affected the pluripotent circuitry leading to inhibition of reprogramming of fibroblasts. Transient expression of TFIIID subunits greatly enhanced reprogramming showing that TFIIID is critical for transcription factor-mediated reprogramming and the maintenance of pluripotency [133].

5.2. Embryonic pluripotency transcription factors in non-mammalian vertebrates

The oct4 gene (pou5f1) has been duplicated early in the vertebrate lineage, giving rise to pou5f1 and pou5f3 [134]. Some vertebrates, for example marsupials and sharks, have retained both copies, whereas mammals have lost pou5f3 (Table 1). Both teleost fish and frogs have lost the other copy, pou5f1, but retained pou5f3. Additionally, there are three tandem-duplicated oct4-like pou5f3 genes in Xenopus: oct91 (pou5f3.1) which is not expressed in the oocyte but is induced at the MBT, oct25 (pou5f3.2) which is maternal but not translated until after fertilization, and oct60 (pou5f3.3) which is exclusively maternally expressed [135]. Sox2 is present in the Xenopus genome, but there is no clear orthologue of Nanog. Its activity may be taken over by ventx1 and ventx2, which are closely related and structurally and functionally very similar to Nanog. Loss of these factors leads to down-regulation of oct91 and premature differentiation, suggesting that these factors play a Nanog-like role in amphibian development [136].

In Xenopus tropicalis, the zygotic genome activates in a broad wave of new transcription extending from the 7th cleavage division (early blastula) to beyond the 12th cleavage division (MBT) [137]. The Oct25 protein levels increase dramatically in early development and Oct25 is with 400 million copies per embryo among the most abundant transcription factors at the early gastrula stage [138]. Oct25 represses Nodal and Wnt signaling and antagonizes the T-box transcription factor Vg1, thereby antagonizing mesoderm induction [139,140]. Nodal-responsive enhancers are marked with H3K4me1 and H3K27ac at the blastula stage, independent of zygotic Nodal signaling [141]. Fox1 mediates Nodal signaling together with Smad2/3, but also functions independently and co-occupies Oct25 targets, which may modulate
the transition between pluripotency and mesendoderm induction [142]. In both mouse ES cells and Xenopus embryos, neural differentiation requires Suv4-20h (Kmt5b/c) to repress Oct25, illustrating the necessity to down-regulate the pluripotency network upon exit of pluripotency [143].

Two studies in zebrafish have highlighted the role of the pluripotency network in early development and zygotic genome activation (ZGA). In zebrafish ZGA coincides with the stage of embryonic pluripotency. Using ribosome profiling it was found that zebrafish Nanog, Pou5f3 (a homolog of the mammalian pluripotency transcription factor OCT4, also referred to as Pou5f1) and SoxB1 (related to Sox2) are the most highly translated transcription factors just before zygotic genome activation sets in [144]. Pou5f3 is recruited to Sox2–POU binding sites before the MBT, and similarly Nanog binds to promoters of first wave expression genes. Combined loss of these factors resulted in developmental arrest before gastrulation and a failure to activate many embryonic genes [144,145]. These data link ZGA to the pluripotency network in zebrafish. This may be different from the situation in mouse, as genetic removal of OCT4 from mouse oocytes does not affect ZGA, totipotency–pluripotency and developmental competence [146,147].

6. Long non-coding RNA

RNA is an integral component of chromatin and plays a role in the pluripotent regulatory circuitry. Large intergenic non-coding RNAs (lincRNAs) have been identified in ES cells [148], and shRNA loss of function experiments showed that 137 of 147 lincRNAs affected the expression of an average of 175 protein-coding genes in trans [149]. Of these, 15 lincRNAs were found to be important for the expression of multiple pluripotency markers and ES cell morphology. In addition lincRNAs were found to repress gene expression related to endoderm, (neuro–)ectoderm, mesoderm and trophoderm differentiation, thereby maintaining pluripotency. Many regulatory sequences of these lincRNAs are bound by Oct4, Nanog and Sox2. Moreover, many of the lincRNAs interact with chromatin modifying enzymes, interacting with chromatin readers, writers and erasers, including the Polycomb complexes PRC1 and PRC2 [149]. PRC2 has also been found to interact with cis-acting RNA; at loci where these RNAs are made by RNAπII. Although this contributes to targeting of the complex, enzymatic activity seems to require the PRC2--associated factor Jarid2 [150,151].

One lincRNA named TUNA was shown to be highly expressed when mouse ES cells differentiated toward the neural lineages. TUNA depletion inhibited neural differentiation. TUNA was shown to occupy promoter regions of Nanog, Sox2, and Fgf4, genes that are important for pluripotency and neural lineage commitment [152].

The DNA methyltransferase Dnmt1 also interacts with many cellular transcripts, including the extra-coding CEBPA lincRNA [153]. This lincRNA adopts a stem–loop structure critical for interaction with Dnmt1 and when transcribed, acts to shield the CEBPA locus from DNA methylation. This study provides another line of evidence establishing how cells employ RNAs to modulate the deposition of repressive epigenetic marks in a genome-wide manner [153]. The WDR5 subunit of MLL complexes also interacts with RNA, including many lincRNAs important for ES cell pluripotency. WDR5 RNA binding activity is specifically required for H3K4 methylation and maintenance of pluripotency in ES cells [154]. These data highlight that non-coding RNAs, some of which are transcribed under the control of the pluripotency regulatory network, can act in cis or in trans to maintain pluripotency and repress lineage specific gene expression, and they do so, at least in part, by interactions with chromatin-modifying complexes. Developmental stage-specific lincRNAs have been identified in Xenopus and zebrafish embryos [155–159], but their role in establishing pluripotency has not been determined.

7. Perspective

By its very nature embryonic development is highly dynamic, and it is no surprise that many of the underlying molecular processes are dynamic as well. At its many different levels of organization, DNA methylation, chromatin composition, histone modifications acting at diverse genomic elements, and transcription factor networks are all acting within their own molecular logic to orchestrate development. As illustrated above, there are integral links between zygotic genome activation and the cellular state of pluripotency. In a simple scenario, exemplified by zebrafish, the core pluripotency network is instrumental for the onset of embryonic transcription while establishing the pluripotent state at the same time. In mouse, the zygote activates its genome, but first needs to produce the extra-embryonic lineages as well as the pluripotent cells of the inner cell mass of the blastocyst. Despite these differences, accumulating evidence summarized in this review shows that a great deal of functional conservation is apparent in the pluripotency network.

One of the major current challenges is to elucidate the interplay between different layers of regulation that together instruct vertebrate development. In recent years, techniques like ChIP-sequencing, RNA-sequencing, and proteomics have been applied to ES cells, their differentiated derivatives, and embryos of mouse, Xenopus, zebrafish and many others. Ongoing efforts in these areas should allow elucidating systems biology views of the regulatory networks in each of these model systems. These highly fruitful endeavors will provide further insight in the differences and similarities of pluripotency in the embryo, in culture and in different species.

Transparency document

The Transparency document associated with this article can be found in the online version.

Acknowledgements

The authors thank Dr. Hendrik Marks for critical reading. This work has been supported by grants from the US National Institutes of Health (NICHD, grant R01HD069344) and the Netherlands Organisation for Scientific Research (NWO-CW, grant 700.58.007). The authors apologize to those whose work could not be cited due to space constraints.

References


