From blastocyst to gastrula: gene regulatory networks of embryonic stem cells and early mouse embryogenesis

David-Emlyn Parfitt1,2,3,4 and Michael M. Shen1,2,3,4

1Department of Medicine, 2Department of Genetics and Development, 3Department of Urology, and 4Department of Systems Biology, Herbert Irving Comprehensive Cancer Center, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

To date, many regulatory genes and signalling events coordinating mammalian development from blastocyst to gastrulation stages have been identified by mutational analyses and reverse-genetic approaches, typically on a gene-by-gene basis. More recent studies have applied bioinformatic approaches to generate regulatory network models of gene interactions on a genome-wide scale. Such models have provided insights into the gene networks regulating pluripotency in embryonic and epiblast stem cells, as well as cell-lineage determination in vitro. Here, we review how regulatory networks constructed for different stem cell types relate to corresponding networks in vivo and provide insights into understanding the molecular regulation of the blastocyst–gastrula transition.

1. Introduction

In the mammalian embryo, the transition from blastocyst to gastrula is a remarkably elaborate process involving a complex series of molecular and cellular events. Many analyses of these developmental stages have typically been limited to individual genes and/or pathways and their role in specific cellular and morphogenetic events. By contrast, recent studies have focused on dissecting the molecular interactions that define the ‘regulatory logic’ of cells and tissues on a genome-wide scale. The definition and validation of such context-specific regulatory networks are likely to model experimental data more meaningfully than lists of differentially expressed genes, and place regulatory genes in a larger framework that can provide insights into their function. Below, we highlight studies that shed light on events taking place during the blastocyst–gastrula transition and have led to the elucidation of gene regulatory networks (GRNs).

2. Properties of gene regulatory networks

In general, cell types can be defined by expression of specific gene products, including transcription factors (TFs), signalling molecules and regulatory RNAs. GRN models describe predicted interactions between these gene products as context-specific patterns of gene expression and activity. Thus, GRNs can be considered as formal representations of the route from genomic information to biological processes.

GRNs are typically visualized with ‘nodes’, representing genes, and ‘edges’ between nodes, representing molecular interactions. Establishment of GRNs for a specific developmental process requires: (i) knowledge of the TFs and signalling molecules involved in the process and (ii) inference of the interactions among these components. For the first requirement, transcriptomic approaches such as microarray and RNA sequencing analyses have been a common starting point. These methods generate comprehensive gene expression profiles defining specific cell populations or developmental time points. GRNs can then be inferred and generated computationally from diverse experimental data typically provided by functional perturbation experiments, which vary depending upon the...
experimental system. GRNs for chick and *Xenopus* embryos, for example, have been constructed from expression data produced after morpholino or siRNA silencing [1,2], while network construction in the mouse has benefited from targeted knock-out lines. Such experiments reveal the impact of loss or gain of function on downstream targets, which can be used for computational inference of gene interactions, often using probabilistic graphical models and approaches based on information theory and linear regression (reviewed in [3]). A key feature of GRNs generated by such approaches is that they are scalable. Depending on the expression data provided, the resulting GRNs can provide relatively simple models of tissue-specific interactions or larger networks describing whole-genome processes. While these models are typically generated from data that have been experimentally acquired, it is important to emphasize that the utility of network identification lies in the generation of testable hypotheses about genetic relationships that direct and facilitate subsequent experimental validation.

Although this review will focus on mouse development, GRNs have provided the first truly global perspectives of development and regulatory relationships in sea urchin, *Drosophila* and *Caenorhabditis elegans*, revealing conserved gene regulatory mechanisms employed in distinct contexts [4–6]. In particular, the work of Davidson [7,8] has led to the characterization of an extensive regulatory network for sea urchin development. This GRN represents a hierarchy of regulatory steps that link components of embryonic development temporally and spatially, and has been used to define the roles of ‘high-ranking’ genes upstream of regulatory pathways, such as those regulating mesendoderm formation [9].

Indeed, GRNs for development are generally hierarchical and highly dynamic. Depending on biological context, they can involve a single genetic input leading to the activation or repression of a small number of effector genes, or progressive interactions among multiple layers of regulatory genes, ultimately leading to the activation (or repression) of ‘differentiation’ genes. Indeed, as most genes probably have limited pleiotropy [10], many operate within regulatory modules or subcircuits, examples of which include those regulated by signalling molecules such as NODAL (figure 1) [12]. Such structures result in high connectivity of particular regulatory genes, called ‘hubs’, which often constitute candidates for experimental validation of *cis*-regulatory relationships. The dynamic features of development necessitate flexibility in

**Figure 1.** GRNs for Nodal signalling in (a) mouse and (b) sea urchin. (a) A GRN that specifies mesendoderm containing a feedback loop between the NODAL precursor (proNODAL), NODAL, BMP4 and WNT3, as well as a CRIPTO-dependent feedback loop. Red indicates signals that promote mesendoderm formation, green, mesoderm formation, and blue, ectoderm formation. (Adapted from [11].) (b) A GRN for Nodal signalling in sea urchin embryogenesis, illustrating its synergistic relationship with Not in ectoderm-specific gene expression. (Adapted from [12].)
GRNs, facilitated by gene redundancy and features such as positive and negative feedback and feed-forward loops. Such feedback loops may control transcriptional noise, as well as the stability of transcriptional networks and linear dose–response relationships [13]. These features are particularly relevant in the context of signalling gradients that regulate many aspects of early embryo development.

The elucidation of GRNs in mammalian species represents a significant challenge, because the complexity of genome organization and gene regulation tends to increase with the number and diversity of cells. Despite these challenges, systems approaches should be valuable for understanding mechanisms of cell fate determination in vertebrate systems. For example, regulatory networks have been defined for neural crest cell induction [14] and mesoderm specification [15] in Xenopus, and similar networks are being constructed for mammalian development.

3. Using stem cells to model regulatory networks for mammalian development

Until recently, efforts to generate regulatory networks for mammalian development in vivo have been relatively limited, perhaps due to the small size and relative inaccessibility of the embryo. These limitations have been at least partially overcome through the analysis of stem cells in culture, which have served as paradigms for in vitro processes.

In particular, networks for the pluripotency and self-renewal capacity of embryonic stem cells (ESCs), derived from the inner cell mass (ICM) of the blastocyst, have been widely studied [16,17]. Thus, gene targeting experiments have established OCT4, NANOG and SOX2 as key TFs that regulate pluripotency in vivo and in vitro [18–20], while interactions among these TFs, their regulatory elements, and co-regulated target genes have been proposed to constitute a core transcriptional network for pluripotency [21–24]. Similarly, networks have been constructed for epiblast stem cells (EpiSCs) that are derived from the postimplantation epiblast (Epi) [25,26]. Recent analyses have also included other factors in the regulatory landscape of pluripotency. For example, ESRB, SALL4, TBX3, KLF4, KLF2 and REST have joined the ranks of TFs constituting the ‘pluripotency network’ [21,27–31]. Moreover, non-coding RNAs such as miR-134, miR-296 and miR-470 have been shown to directly regulate OCT4, Nanog and Sox2 [32], while epigenetic modifiers such as PRDM14 and WDR5 also display overlapping regulatory functions with the core pluripotency factors [33,34]. Although understanding how these molecules are functionally integrated represents a complex task, iterations of regulatory networks have been generated on transcriptional [21,24,30,35] and post-translational levels [36,37], while other studies have integrated data from multiple regulatory levels [38,39].

Several features of these networks suggest how they might operate to establish and/or maintain pluripotency. Firstly, and perhaps unsurprisingly, they are enriched for genes involved in regulation of the ICM or aspects of embryonic lineage-specific differentiation. Secondly, many genes are co-regulated and are often downregulated during ESC differentiation, suggesting their involvement in common cellular functions or pathways. Thirdly, multiple interactions among genes within these networks suggest that they affect a mutual function and that a balance between these interactions is important for maintaining pluripotency. This view is consistent with dosage-dependent effects for each of the core pluripotency factors [40–42], as well as significant intercellular differences in their expression levels in ESCs and in vivo [43–46]. Moreover, the broad range of genes present in most ESC regulatory networks implies their functional subdivision into sets of targets regulated by different regulatory genes and/or complexes. Thus, the control of target genes and signalling pathways in the context of pluripotency is more likely to be combinatorial than strictly hierarchical and represents a state of dynamic, as opposed to constant, equilibrium so that ESCs are kept in an undifferentiated state and retain the potential to undergo multi-lineage differentiation.

Classically, pluripotency has been regarded as a ‘ground state’ that is regulated by a TF network that inhibits differentiation, while the activation of one or more lineage-specifying factors can trigger differentiation [47,48]. The interpretation that the ground state is intrinsically stable was based on observations that ESC pluripotency is maintained in culture conditions that emulate the absence of ‘extrinsic instruction’ (figure 2a) [50,51]. However, this view seems to be at odds with the idea that the ground state is actively maintained. For example, the core pluripotency factors alone are tightly auto-regulated and co-occupy the promoters of thousands of targets [52].

As an alternative, a new ‘balance’ model of pluripotency [49,53] proposes that pluripotency represents a state of transcriptional competition among regulatory factors that can themselves act as lineage-specifiers in the proper context (figure 2b) [40,54,55]. Thus, ESCs are poised at the cusp of multiple opposing lineage commitment decisions, expressing many TFs required to react to differentiation cues, such as Brachyury, Hes1 and Eomes [56–58]. Given these alternative models for regulating the pluripotent state, it is important to determine which GRNs more accurately reflect how pluripotency is maintained, altered and ultimately lost during lineage-specific differentiation in stem cells in culture, as well as in the embryo.

4. Regulatory networks for pluripotency and lineage commitment in the mouse embryo

During preimplantation mouse development, successive cleavages of the zygote lead to formation of the blastocyst, containing the ICM and blastocyst cavity surrounded by the trophoderm (TE). Subsequently, the primitive endoderm (PrE) becomes apparent at the surface of the ICM in contact with the blastocyst cavity, with the remaining ICM forming the Epi. After implantation into the maternal uterus, continued cell divisions occur in concert with the morphogenetic changes associated with gastrulation. While pluripotency is maintained in cells of the Epi, it is gradually lost in descendants of the TE and PrE, and eventually elsewhere during specification of the primary germ layers.

(a) The gene regulatory network of the blastocyst

It has been long been established that Oct4, Nanog and Sox2 represent three genes that are critical to the maintenance of pluripotency in vitro and to normal ICM development in vivo [18–20,59–62]. SALL4 shares a similar role, as it binds to the conserved regulatory region in the distal
enhancer of Oct4 and activates transcription [63]; accordingly, Sall4 null embryos display decreased Oct4, Sox2 and Nanog expression [64]. Similarly, Stat3 regulates pluripotency in the ICM via the OCT4–NANOG circuitry and is dependent on leukaemia inhibitory factor (LIF) and interleukin (IL)-6 [65]. Such findings can be used to infer a simple network showing how these factors regulate the core circuitry for pluripotency (figure 3). Additional genes have been implicated in this network by their ICM-specific expression and functional characterization in ESCs. Thus, c-FOS is specifically expressed in the ICM in vivo [66] and binds to the Nanog promoter in ESCs until its expression is downregulated upon differentiation [67,68]. Similarly, Zfp322a and Egr1 are expressed specifically in the ICM [43] and are thought to regulate pluripotency by suppressing MAPK signalling [69].

However, the roles for several other genes in regulating pluripotency in vivo appear less clear. For example, Klf5 is expressed ubiquitously in the preimplantation mouse embryo [70] and is not only required for the normal transcription of Oct4, Sox2 and Nanog in the ICM, but also of Cdx2 in the TE [71,72].
Although Klf5 null embryos arrest and die at the blastocyst stage, it is unclear whether this is due to defective establishment of TE or ICM, or to a more general role for Klf5 in the cellular machinery of the early embryo. Conversely, the expression of other pluripotency factors seems to be dispensable for early development. Notably, Esrβ null embryos survive until 10.5 days post coitum (dpc) and die due to placental abnormalities [73]. Similarly, while Dpyr4, Dpyr5 and Utf1 are specifically expressed in pluripotent cells in vivo and in culture, the impact of their absence is only observed during late embryonic stages, though redundancy among these genes has not yet been investigated [74–76].

Given their common components, the GRNs modelling ESC biology and preimplantation embryo development are likely to be similar, at least to some extent. However, most likely due to lack of accessibility of experimental material, there have been few studies to date that have pursued the de novo construction of regulatory networks for pre- and postimplantation development in vivo. One example is the description of two interlinked feedback loops regulated by the NODAL precursor during mesendoderm specification in the mouse (figure 1a) [11,77]. However, the utility of techniques such as single-cell RNA sequencing has made quantitation of preimplantation transcriptomes more feasible. For example, a comparative analysis of ESC and ICM transcriptomes revealed that a number of genes are positively correlated with pluripotency in both contexts, including E-cadherin, Pim1, Fzd9 and Dazl [43]. This study also identified altered gene expression patterns during ICM outgrowth in culture, providing insights into differences between the gene networks regulating ESCs and the ICM. For example, Nodal, Lin28 and ERα are significantly upregulated during ICM outgrowth, which is interesting as these genes are positive regulators of self-renewal but are dispensable for pluripotency in ESCs, suggesting specific roles in the embryo [78–80]. Notably, genes involved in the development of TE, PrE and mesendoderm such as Cdx2, Gata6 and Tbx3 are downregulated during ICM outgrowth in vitro [43].

Further efforts to define a regulatory network for the blastocyst have analysed transcriptional profiles of pre-implantation embryos in the context of reference protein interaction networks [81], resulting in a network centred on interactions between OCT4, SOX2, NANOG, SALL4 and ILF2. The role of ILF2 during development is unclear, though interestingly it is associated with active chromatin in numerous embryonic tissues [82] and is known to interact with OCT4 and NANOG proteins in mouse ESCs [83]. Global transcriptome analysis of the embryo in the context of Oct4, Sall4 and Nanog deletion have revealed that these factors act as combinatorial regulators of developmentally important genes at preimplantation stages [84], including pluripotency factors such as Klf5 and Klf2, the epigenetic modifiers Dnmt3a and Dnmt3b, and the signalling molecules Fgf2, Fzd4 and Fzd7, as well as miRNAs with enhancers occupied by OCT4, SALL4 or NANOG in ESCs [85,86]. These data suggest that a feed-forward loop consisting of pluripotency factors, miRNAs, Rd22 and Dmnt3b regulates DNA methylation in the developing embryo and smoothes fluctuations in the expression of regulatory genes that might otherwise disrupt normal cell fate specification [84].

Interestingly, however, a number of studies indicate that miRNAs are largely dispensable for preimplantation development (reviewed by Greve et al. [87]). Indeed, the expression of several miRNA families is significantly lower in ICM cells than in ESCs [43]. While the high expression levels of the miR-290–295 cluster in both these different cell types might suggest this cluster to be an exception, blastocyst development is nonetheless normal in its absence [88]. Thus, these and other data [45] draw attention to fundamental differences between ESC and preimplantation biology. Such differences may complicate comparisons between GRNs for ESCs and the preimplantation embryo, in part perhaps because embryo GRNs operate under strict time and spatial constraints. Moreover, the de novo construction of preimplantation regulatory networks may be complicated by maternally provided transcripts and proteins, which contribute to zygotic genome activation and initial establishment of embryonic cell lineages (reviewed in [90]). Moreover, the observation that several pluripotency regulators appear to be dispensable for ICM development in vivo might reflect functionally redundant mechanisms within gene families or for genes within the same regulatory subcircuit. Indeed, such redundancy may facilitate the highly regulative nature of mammalian development, a property not generally applicable to ESCs in culture. Furthermore, reciprocal signals between the ICM and TE may also regulate their specification, and consequently networks constructed using whole blastocysts may miss features of networks specific to these lineages.

Despite these limitations, our knowledge of the transcriptional architecture of the blastocyst has nonetheless been extended by direct molecular analyses. Though correlated, the quantitative transcriptional analysis of ICM cells as they adopt properties of ESCs suggests that the regulatory network of pluripotency in the ICM operates in parallel with the developmental programme driving early differentiation events, in contrast to ESCs, where the developmental programme is silenced in favour of unrestricted self-renewal [43]. These findings raise the question of whether factors regulating pluripotent cells in the ICM and Epi also function during lineage specification. This issue has been addressed in recent studies of TE and PrE specification, as discussed below.

(b) Regulatory networks governing the first cell fate choices

Targeted deletion experiments in embryos and trophoblast stem cells have together provided relevant data for inference of a TE regulatory network that includes genes such as Cdx2, Tead4, Eomes, Id2, Elf5, Tcfyp2c, Esrβ, Sox2, Nanog and Oct4 (figure 4e) [20,73,91–95]. Reciprocal inhibition between Oct4, Nanog and Cdx2 is perhaps the most central element of this network, as CDX2 and OCT4 form a functional repressor complex in the early embryo [96], while Cdx2 is required for downregulation of Oct4 and Nanog in the TE, resulting in restriction of their expression to the ICM [92]. Thus, TE and ICM specification is contingent upon regionalized alterations in the balance of these proteins. One modulator of this balance is the activity of the Hippo signalling pathway. Specifically, cell polarity and positioning cues regulate the sub-cellular localization of LATS1/2, YAP, ANGIOMOTIN and NF2/MERLIN, which in turn regulate YAP/TEAD4-mediated expression of Cdx2, which becomes restricted to cells of the TE [94,97–100].

Other signalling events regulate maintenance of the TE fate as the embryo transitions through peri- and early
postimplantation stages. In the ICM, Oct4 and Sox2 together regulate the expression of Fibroblast growth factor 4 (Fgf4) [101], which in turn regulates cell proliferation via FGFR2-mediated ERK/MAPK signalling. Indeed, targeted disruption of ERK/MAPK signalling results in attenuated expression of Cdx2 and Eomes, and peri-implantation lethality due to impairment of trophoblast outgrowth [102]. In addition, Sox2 expression is essential for trophoblast maintenance in vivo [20] and is likely to be a downstream target of FGFR signalling, forming a regulatory circuit with Esrrb and Tcfap2c that promotes Cdx2 expression while inhibiting Oct4 [35].

Similarly, the network for PrE formation includes Oct4, Nanog, Gata6 and Fgf4 (figure 4b). Fgf4 expression becomes restricted to cells of the Epi by the time it becomes distinct from the PrE [44,46]. FGF4 secretion by Epi cells regulates the expression of Gata6, Gata4, Sox17 and Sox7 in the PrE, via FGFR2, the GRB2–MAPK pathway and interactions with GATA6 itself [103–105]. Accordingly, disruption of FGF4 signalling results in failure of PrE formation [106,107]. Indeed, differential levels of FGF4 and FGFR2 between cells of the ICM are essential for the dynamics of PrE/Epi formation [46,107–109], leading to the mutually exclusive expression of Nanog and Gata6 in the Epi and PrE, respectively [104]. In the absence of Nanog, Epi formation is specifically impaired [19] and Fgf4 expression is downregulated [110]. However, Sox17 and Gata6, which are markers of an established PrE identity, are also absent in Nanog mutant embryos [61,110], suggesting that Fgf4 expression is downstream of Nanog and is intrinsically linked to Epi identity during specification of the PrE lineage. Thus, Nanog is believed to have a non-cell-autonomous role in PrE formation through this circuitry [61,110].

In some respects, Oct4 and Sox2 could play similar roles to Nanog, as they are both required for Fgf4 expression [101]. However, unlike Sox2, Oct4 expression does not become restricted to the Epi after implantation [20,111]. Recent studies have demonstrated that Oct4 is required for the maintenance of PrE-specific gene expression, which is disrupted in the absence of zygotic OCT4 protein [59,60]. Notably, this requirement for Oct4 cannot be rescued by exogenous FGF, demonstrating that Oct4 acts downstream of FGFR signalling and plays a cell-autonomous as well as non-cell-autonomous role in PrE specification [59,60]. In addition, genome-wide expression analysis of Oct4 mutant embryos demonstrates that Gata4 and Sox17 expression are Oct4-dependent [59]. Interestingly, PrE specification appears normal in the absence of either Gata4 or Sox17 [112,113], perhaps reflecting functional redundancy among PrE genes and their parallel regulation by Oct4.

Together, these studies demonstrate two central points. First, they support the notion that the divergence of TE and PrE involves reciprocal genetic changes within GRNs shared, at least in part, with the ICM and Epi. These changes ultimately derive from dynamic and reinforcing cellular interactions, in which cell polarity and positional cues appear to trigger events separating TE from ICM, while cell division orientation, migration and apoptosis drive the divergence of PrE and Epi [114–116]. Secondly, these studies provide evidence that pluripotency factors function in networks regulating lineage-specific differentiation in the embryo. Thus, networks of genes containing at least some components common to each developmental pathway drive pluripotency and early lineage specification (figures 3 and 4). Notably, switches in the function of these common components are
mediated, at least in part, by cell-extrinsic signals that ultimately trigger lineage specification. Further evidence supporting this scenario has been provided by studies focusing on changes in the Epi during the transition towards gastrulation, as described below.

(c) Dynamics of the pluripotency network during epiblast priming and lineage specification

At the onset of gastrulation, the Epi undergoes regionalized lineage specification, beginning with the specification of mesendoderm in the primitive streak posteriorly and embryonic ectoderm in the anterior Epi. Although fate determination is ultimately concomitant with a complete loss of pluripotency, recent studies demonstrate that it is also contingent on dynamic alterations of the regulatory network for pluripotency in the Epi prior to its loss. However, analysis of these changes as they occur has been hindered by embryo inaccessibility at peri-implantation stages.

Consequently, the comparative study of ESCs and EpiSCs has proved fruitful, as these stem cell types recapitulate many features of the tissues from which they are derived. The distinct signalling pathways that regulate pluripotency and self-renewal in mouse ESCs and EpiSCs are reflected in non-identical TF binding and gene expression patterns. In particular, ESCs and EpiSCs define ‘naive’ and ‘primed’ states of pluripotency, which also exist in vivo. While expression of the core pluripotency genes Oct4, Sox2 and Nanog are associated with both naive and primed pluripotent states in the ICM and Epi [121,122], naive state markers such as Rex1, Klf4, Klf2, Esrrb, Tbx3 and Stella are downregulated after implantation in vivo [121,122]. Similarly, lineage specification markers such as Fgf5, Cer1 and T are expressed in the Epi and in EpiSCs, but are absent in the ICM and ESCs [121,122]. Thus, it seems likely that similar batteries of genes are involved in the transition between pluripotent states in vivo as in vitro. Moreover, the primed pluripotent state could be relatively stable throughout Epi development, because EpiSCs derived from pre-gastrulation embryos of different stages display similar global gene expression profiles [123].

Changes to the expression of core pluripotency genes appear to have a significant role in the transition from a naive to primed pluripotent state in vivo. Though its expression is detected postimplantation at approximately 6.5 dpc, Nanog expression is downregulated in the Epi during implantation under direct repression by Tcf7l1 [124,125]. Nanog plays a central role in the regulation of naive pluripotency, directly interacting with genes such as Esrrb, Sall4 and Tbx3 [126,127]. Reflecting this role, ESCs expressing Nanog at relatively low levels are more prone to differentiation [128], while its overexpression in EpiSCs promotes their conversion to ESCs [62], suggesting that downregulation of Nanog expression in the Epi is important to establish or maintain the primed pluripotent state. A similar scenario for Oct4 seems plausible since its expression in EpiSCs is typically lower than in ESCs or the ICM [121]. EpiSCs expressing high levels of Oct4 in culture display atypically high expression of naive markers such as Esrrb, Klf4 and Rex1, and are able to contribute to blastocyst chimaeras in vivo [121]. Therefore, the reduction in Oct4 expression that occurs upon implantation [124] could have a significant impact on the nature of pluripotency in the Epi.

Studies conducted in vivo and in vitro have suggested how other components could contribute to the regulatory network of primed Epi. For example, regulatory models constructed from transcriptional signatures of EpiSCs in the context of knock-down and overexpression experiments have highlighted the pivotal roles of Zic2/Zic3 and Otx2 in supporting the expression of Epi markers such as Fgf5, while restricting Oct4 expression levels to those permissive of the primed state [25]. Observations in the embryo support the role of Otx2 in this network [129,130]. However, efforts to validate the role of Zic2/Zic3 in vivo have been complicated by their functional overlap and the lack of a Zic2 null allele; interestingly, the most severe phenotype reported for Zic3 mutant embryos corresponds to neural plate defects, which could be linked to an earlier failure of Epi specification [131]. In addition, Tcf7l1 (Tcf3) has been characterized as an inhibitor of self-renewal and forms an auto-regulatory loop with core pluripotency factors by co-regulation of several targets [132]. In vivo and in vitro, Tcf3 expression is modulated by canonical Wnt/β-catenin signalling as well as inhibition of Gsk3 [133,134]. Thus, in the absence of Wnt signalling at peri-implantation stages, Tcf3-mediated priming of the Epi can occur through its connection with the core pluripotency circuitry [135]. Finally, the transcription factor TFE3 is thought to promote a naive pluripotent state in ESCs through direct transcriptional control of Esrrb, as well as the cis-regulatory elements of Klf2 and Tbx3 [136]. In the postimplantation embryo, Tfe3 becomes progressively restricted to the cytoplasm in the Epi, which is mediated by FOLLICULIN and its binding partners FN1P1/FNIP2, thereby precluding the transcriptional activity of TFE3 and promoting exit from the naive state [136].

Subsequent to Epi priming, a web of regionalized signalling activities initiates lineage-specific differentiation [137]. These activities include canonical Wnt/β-catenin signalling, which is essential for primitive streak formation and mesendoderm specification, while a downstream Tcfβ-mediated circuit leads to regionalized expression of several core pluripotency factors [138,139]. Specifically, although it is expressed throughout the Epi at 6.5 dpc, Sox2 becomes anteriorly restricted at 7.5 dpc [20], while re-expression of Nanog at 5.5 dpc is specific to the posterior Epi [124]. Lineage-specific gene expression is associated with specific expression levels of core pluripotency factors in differentiating ESCs, suggesting functional consequences for these dynamic changes [54,140]. For example, high levels of Sox2 are associated with neuroectoderm specification, marked by Sox1, while mesendoderm specification is associated with high levels of Otx4 and Nanog and expression of Brachyury (T). Similar associations are observed when these lineages are induced by exogenous factors in culture [54]. A network containing these genes, along with other mesendoderm and neuroectoderm factors (FOXA2 and BRN2, respectively) and the epigenetic regulators JARID2 and SUZ12 has been inferred from ESC ChIP-seq data [54,85,141] and describes how expression dynamics of core pluripotency factors may influence lineage commitment in the gastrula (figure 5).

Such a pre-gastrulation Epi GRN may also explain how the mutually exclusive expression of Sox2 and Nanog [20,124] might initiate lineage-specific gene expression during emergence of the primary germ layers. Somewhat paradoxically, Oct4 expression persists throughout the Epi until its elimination during somitogenesis, though this elimination begins in the anterior Epi [124]. However, this observation is concordant with the findings that Wnt/β-catenin-mediated primitive
streak formation is dependent on Nanog and that the association of Oct4 levels with specific lineages in differentiating ESCs is weaker than that for Nanog [140]. Indeed, in the embryo, high levels of Oct4 are required for the specification of all three primary lineages [142].

Together, these results provide evidence that lineage specification at the onset of gastrulation involves changes to the core pluripotency circuit, which in turn influences expression of downstream lineage-specifying genes. Wnt/β-catenin signalling is undoubtedly central for these alterations, given the role of this pathway during the onset of gastrulation. Indeed, Wnt-mediated expression of pluripotency factors such as Snail1 and Tcfβ is known to mediate mesendoderm specification [135,143]. Targets of other signalling pathways share similar roles in regulating different states of pluripotency. For example, in cooperation with Nodal signalling, and actively induced by Nanog, Eomes directly regulates much of the transcriptional network for definitive endoderm specification in the mouse [58]. Meanwhile, downstream of Nodal signalling, Tbx3 regulates pluripotency in ESCs and expression of Eomes and Brachyury upon differentiation towards mesendoderm [144], though definitive confirmation of this relationship in vivo has been limited to experiments in Xenopus embryos. Similarly, Otx2 stabilizes the transition from a naive to primed pluripotent state, antagonizing BMP4 signalling together with its binding partner FGF2 [145], and is thought to function in the same circuitry to regulate anterior neuroectoderm differentiation in the embryo [145,146].

5. Conclusion and future perspectives

Based on recent studies, many of the genes that control pluripotency in the earliest stages of embryonic development also regulate transitions through distinct pluripotent states, as well as specific lineage decisions throughout pre- and early postimplantation stages. Although not all such regulatory genes are active at every stage, these observations suggest that GRNs of the earliest events in mammalian development are variations on a theme, involving different combinations of the same components functioning under different constraints. Central members of the core pluripotency circuitry, including Oct4, Sox2 and Nanog, appear to be among the common components, while regionalized changes in the balance of the circuit result in activation and repression of their respective targets. Such changes are mediated in part by a gradual extinction of their expression as the accessibility of their cis-regulatory elements is decreased by nucleosome invasion [124], which alters the effect of peripheral components such as Tcfβ and Tbx3 on this circuitry, and initiates their switch to roles as lineage-specifying genes. At present, it is unclear how region-specific changes initiate this process, but alterations in cell contacts, signalling and morphogenetic events are likely to be significant.

The clarification of these and related issues should be greatly aided by the de novo construction of GRNs representing individual cell types at different stages of development. In particular, such efforts could benefit from the application of computational systems approaches that have not yet been used to study the developing embryo. For example, although many methods of co-expression analysis provide critical information about GRN architecture, they do not establish whether regulatory interactions are direct or indirect. Approaches such as Algorithm for the Reconstruction of Accurate Cellular Networks have been developed with this issue in mind, as it can eliminate indirect transcriptional interactions from inferred networks [147]. Furthermore, methods such as Modulator Inference by Network Dynamics can detect three-way interactions between gene products (such as protein kinases), downstream regulators and their targets [148].

Ideally, each predicted interaction within inferred GRNs would be validated experimentally. Ultimately, such validation would require the identification of cis-regulatory elements that underlie global gene expression dynamics. For
example, the Finding Informative Regulatory Elements algorithm can computationally identify local DNA and RNA sequence motifs that are likely to serve as regulatory elements [149]. Meanwhile,software packages such as GeNiTool [150] not only enable better visualization of regulatory networks and predicted versus observed spatial expression patterns, but also permit exploration of both cis- and trans- perturbations of GRN interactions in silico, facilitating subsequent validation experiments. Finally, gene editing tools, such as TALENs and the CRISPR/Cas system are ideally suited for the generation of mice carrying mutations in single and multiple genes, as well as reporter and conditional alleles [151–153], which will be essential for experimental validation approaches.

In the future, GRNs should facilitate the integration of upstream events and extrinsic signals with downstream sub-networks of effector genes, in order to understand context-dependent function of these networks. Although such efforts will present enormous computational challenges for modelling the transition from blastocyst to early gastrulation stages, these studies could ultimately lead to generation of whole-organism GRNs for mammalian development.

Acknowledgements. We thank Bo Li and Hui Zhao for comments on the manuscript. Funding statement. Work in the Shen laboratory is supported by grants from the NIH.

References


31. Han J et al. 2010 Tbx3 improves the germ-line competency of induced pluripotent stem cells.


