Transcriptomic analysis of pluripotent stem cells: insights into health and disease

Jia-Chi Yeol,2 and Huck-Hui Ng1,2,3,4,5,*

Abstract
Embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) hold tremendous clinical potential because of their ability to self-renew, and to differentiate into all cell types of the body. This unique capacity of ESCs and iPSCs to form all cell lineages is termed pluripotency. While ESCs and iPSCs are pluripotent and remarkably similar in appearance, whether iPSCs truly resemble ESCs at the molecular level is still being debated. Further research is therefore needed to resolve this issue before iPSCs may be safely applied in humans for cell therapy or regenerative medicine. Nevertheless, the use of iPSCs as an in vitro human genetic disease model has been useful in studying the molecular pathology of complex genetic diseases, as well as facilitating genetic or drug screens. Here, we review recent progress in transcriptomic approaches in the study of ESCs and iPSCs, and discuss how deregulation of these pathways may be involved in the development of disease. Finally, we address the importance of these advances for developing new therapeutics, and the future challenges facing the clinical application of ESCs and iPSCs.

Keywords Embryonic stem cells, gene expression, induced pluripotent stem cells, pluripotency, regenerative medicine, therapy, transcriptional regulation, transcriptomics.

Stem cell transcriptomics and transcriptional networks
Embryonic stem cells (ESCs) have the unique ability to self-renew and differentiate into cells of all three germ layers of the body. This capacity to form all adult cell types, termed ‘pluripotency’, allows researchers to study early mammalian development in an artificial setting and offers opportunities for regenerative medicine, whereby ESCs could generate clinically relevant cell types for tissue repair. However, this same malleability of ESCs also renders it a challenge to obtain in vitro differentiation of ESCs to specific cell types at high efficacy. Therefore, harnessing the full potential of ESCs requires an in-depth understanding of the factors and mechanisms regulating ESC pluripotency and cell lineage decisions.

Early studies on ESCs led to the discovery of the core pluripotency factors Oct4, Sox2 and Nanog [1], and, increasingly, the use of genome-level screening assays has revealed new insights by uncovering additional transcription factors, transcriptional cofactors and chromatin remodeling complexes involved in the maintenance of pluripotency [1]. The study of ESC transcriptional regulation is also useful in the understanding of human diseases. ESCs, for instance, are known to share certain cellular and molecular signatures similar to those of cancer cells [2], and deregulation of ESC-associated transcriptional regulators has been implicated in many human developmental diseases.

Despite the promising potential, the use of human ESCs (hESCs) in clinical applications has been slow because of ethical, immunological and tumorigenicity concerns [3]. These ethical and immunogenicity issues were seemingly overcome by the creation of induced pluripotent stem cells (iPSCs), whereby exogenous expression of Oct4, Sox2, Klf4 and c-Myc in differentiated cells could revert them to pluripotency [4]. However, the question of whether these iPSCs truly resemble ESCs is still actively debated and remains unresolved [5]. Nevertheless, the application of iPSCs as an in vitro human genetic disease model has been successful in revealing novel molecular disease pathologies, as well as facilitating genetic or drug screenings [6].

In this review, we describe recent advances in understanding the ESC and iPSC transcriptional network, and also discuss how deregulation of ESC pathways is implicated in human diseases. Finally, we address how the knowledge gained through transcriptional studies of ESCs and iPSCs has impacted translational medicine.
**Transcriptomic approaches for studying stem cells**

The transcriptome is the universe of expressed transcripts within a cell at a particular state [7]; and understanding the ESC transcriptome is key towards appreciating the mechanism behind the genetic regulation of pluripotency and differentiation. The methods used to study gene expression patterns can be classified into two groups: (1) those using hybridization-based approaches, and (2) those using sequencing-based approaches (Table 1).

For hybridization-based methods, the commonly used ‘DNA microarray’ technique relies on hybridization between expressed transcripts and microarray printed oligonucleotide (oligo) probes from annotated gene regions [7]. In addition to allowing the identification of highly expressed genes, microarrays also enable the study of gene expression changes under various conditions. However, microarrays have their limitations, whereby prior knowledge of genomic sequences is required, and cross-hybridization of oligo probes may lead to false identification [7]. Subsequently, later versions of microarrays were modified to include exon-spanning probes for alternative-spliced isoforms, as well as ‘tiling arrays’, which comprise oligo probes spanning large genomic regions to allow for the accurate mapping of gene transcripts [7,8]. Indeed, conventional microarrays and tiling arrays have been instrumental in advancing our understanding of ESC transcriptional regulation (Table 1) through the mapping of ESC-associated transcription-factor binding sites (chromatin immunoprecipitation (ChIP)-chip) [9,10], identification of microRNA (miRNA) regulation in ESCs [11], as well as the identification of long non-coding RNA (lncRNA) [12] and long intergenic non-coding RNA (lincRNA) [13,14].

Sequence-based transcriptomic analysis on the other hand involves direct sequencing of the cDNA. Initially, Sanger sequencing techniques were used to sequence gene transcripts, but these methods were considered expensive and low throughput [7]. However, with the development of next-generation sequencing (NGS), such as the 454, Illumina and SOLiD platforms, it is now possible to perform affordable and rapid sequencing of massive genomic information [8]. Importantly, NGS when coupled with transcriptome sequencing (RNA-seq) offers high-resolution mapping and high-throughput transcriptome data, revealing new insights into transcriptional events such as alternative splicing, cancer fusion-genes and non-coding RNAs (ncRNAs). This versatility of NGS for ESC research is evident through its various applications (Table 1), such as chromatin immunoprecipitation coupled to sequencing (ChIP-seq) [15], methylated DNA immunoprecipitation coupled to sequencing (DIP-seq) [16], identification of long-range chromatin interactions [17], miRNA profiling [18], and RNA-binding protein immunoprecipitation coupled to sequencing (RIP-seq) [19].

**Table 1. Transcriptomic approaches for studying stem cells**

| Objective                                   | Method                  | Reference |
|---------------------------------------------|-------------------------|-----------|
| DNA sequencing                              | NGS                     | [65]      |
| mRNA expression analysis                    | Microarray              | [58]      |
| miRNA expression analysis                   | RNA-seq                 | [93]      |
| mRNA expression analysis                    | Microarray              | [11]      |
| miRNA expression analysis                   | RNA-seq                 | [18]      |
| LncRNA expression analysis                  | Microarray              | [12,13]   |
| Identification of alternative splicing isoforms | Microarray               | [22,94]   |
| Identification of alternative splicing isoforms | RNA-seq                 | [95]      |
| Mapping of protein-DNA binding              | ChiP-chip               | [9,10,24] |
| Identification of DNA methylation profiling | BS-seq                  | [68]      |
| DNA methylation profiling                   | MethylC-seq             | [68]      |
| Identification of protein-DNA interactions   | DIP-seq                 | [16]      |
| Mapping of long-range chromatin interactions| ChIA-PET                | [17]      |
| Identification of RNA-protein interactions  | RIP-seq                 | [19]      |
| Identification of RNA-protein interactions  | RIP and direct RNA quantification | [14] |

BS-seq, bisulfite sequencing; ChIA-PET, chromatin interaction analysis with paired-end tag sequencing; ChiP-chip, chromatin immunoprecipitation on chip; ChiP-PET, chromatin immunoprecipitation with paired-end tag sequencing; ChiP-seq, chromatin immunoprecipitation and sequencing; DIP-seq, DNA immunoprecipitation and sequencing; MethylC-seq, methylcytosine sequencing; NGS, next-generation sequencing; RIP, RNA-binding protein immunoprecipitation; RIP-seq, RNA-binding protein immunoprecipitation and sequencing; RNA-seq, RNA sequencing; 3C, chromosome conformation capture.

Transcriptomics has been instrumental in the study of alternative splicing events. It has been suggested that around 95% of all multi-exon human genes undergo alternative splicing to generate different protein variants for an assortment of cellular processes [20], and that alternative splicing contributes to higher eukaryotic complexity [21]. In mouse ESCs (mESCs) undergoing embryoid body formation, exon-spanning microarrays have identified possible alternative splicing events in genes associated with pluripotency, lineage specification and cell-cycle regulation [22]. More interestingly, it was found that alternative splicing of the Serca2b gene during ESC differentiation resulted in a shorter Serca2a isoform with missing miR-200 targeting sites in its 3’-UTR. Given that miR-200 is highly expressed in cardiac lineages, and that Serca2a protein is essential for cardiac function, the results suggest that during mESC differentiation some genes may utilize alternative splicing to bypass lineage-specific miRNA silencing [22]. With the largely uncharacterized nature of alternative splicing in ESCs, and the availability of high-throughput sequencing tools, it would be of interest to further dissect these pathways.
Transcriptional networks controlling ESCs
The core transcriptional regulatory network
In ESCs, the undifferentiated state is maintained by the core transcription factors Oct4, Sox2 and Nanog [1]. Early mapping studies revealed that Oct4, Sox2 and Nanog co-bind gene promoters of many mESC and hESC genes [23,24]. Importantly, the core transcription factors were found to maintain pluripotency by: (1) activating other pluripotency factors, while simultaneously repressing lineage-specific genes via Polycomb group proteins; and (2) activating their own gene expression, as well as that of each other. Therefore, with this autoregulatory and feed-forward system, Oct4, Sox2 and Nanog constitute the ESC core transcriptional network (Figure 1) [23,24]. Subsequent studies on additional ESC-related transcription factors using ChIP-based transcriptomics led to the discovery of transcription factors associating into an ‘Oct4’ or ‘Myc’ module [10,15].

The expanded pluripotency network
Apart from Oct4, Sox2 and Nanog, the Oct4 module also includes the downstream transcription factors of the LIF, BMP4 and Wnt signaling pathways: Stat3, Smad1 and Tcf3 [15,25]. Indeed, Stat3, Smad1 and Tcf3 co-occupy certain regulatory regions with Oct4, Sox2 and Nanog, thus establishing the pathway in which external signaling can affect ESC transcriptional regulation [15,25]. Mass spectrometry has also facilitated the study of protein-protein interaction networks of core transcription factors [26,27], revealing that Oct4 can interact with a diverse
population of proteins, including transcriptional regulators, chromatin-binding proteins and modifiers, protein-modifying factors, and chromatin assembly proteins. Importantly, knockdown of Oct4 protein levels is known to cause the loss of co-binding activity of other transcription factors [15,27], suggesting that Oct4 serves as a platform for the binding of its interacting protein partners onto their target genes.

The Myc module consists of transcription factors such as c-Myc, n-Myc, Zfx, E2f1 and Rex1, and is associated with self-renewal and cellular metabolism [10,15]. Approximately one-third of all active genes in ESCs are bound by both c-Myc and the core transcription factors [28]. However, unlike Oct4, Sox2 and Nanog, which can recruit RNA polymerase II via coactivators such as the Mediator complex [29], c-Myc rather appears to control the transcriptional pause release of RNA polymerase II, via recruitment of a cyclin-dependent kinase, p-TEFb [28]. It is therefore proposed that Oct4-Sox2-Nanog selects ESC genes for expression by recruiting RNA polymerase II, while c-Myc serves to regulate gene expression efficiency by releasing transcriptional pause [1]. This may thus account for the reason why overexpression of c-Myc is able to improve the efficiency of iPSC generation, and how c-Myc could be oncogenic. In fact, the Myc module rather than the Oct4 module in ESCs was recently found to be active in various cancers, and may serve as a useful tool in predicting cancer prognosis [9].

Besides targeting transcription factors to regulate gene expression, Oct4 is also known to affect the ESC chromatin landscape. Jarid2 [30-34] and Pcl2/Mtf2 [30,31,34-35] have been identified as components of the Polycomb Repressive Complex 2 (PRC2) in ESCs, and regulated by the core ESC transcription factors [10,15]. From these studies, Jarid2 is suggested to recruit PRC2 to its genomic targets, and can also control PRC2 histone methyltransferase activity [30-34]. The second protein Pcl2 shares a subset of PRC2 targets in ESCs [34-35] and appears to promote histone H3 lysine 27 trimethylation [35]. Knockdown of Pcl2 promotes self-renewal and impairs differentiation, suggesting a repressive function of Pcl2 by suppressing the pluripotency-associated factors Tbx3, Klf4 and Foxd3 [35]. Oct4 has also been demonstrated to physically interact with Wdr5, a core member of the mammalian Trithorax complex, and cooperate in the transcriptional activation of self-renewal genes [36]. As Wdr5 is needed for histone H3 lysine 4 trimethylation (H3K4me3), Oct4 depletion notably caused a decrease in both Wdr5 binding and H3K4me3 levels at Oct4-Wdr5 co-bound promoters. This indicates that Oct4 may be responsible for directing Wdr5 to ESC genes and maintaining H3K4me3 open chromatin [36]. As chromatin structure and transcriptional activity can be altered via addition or removal of histone modifications [37], the ability of Oct4, Sox2 and Nanog to regulate histone modifications expands our understanding of how the core transcriptional factors regulate chromatin structure to ultimately promote a pluripotent state.

### Pluripotent transcription factor regulation of non-coding RNA

ncRNAs are a diverse group of transcripts, and are classified into two groups: (a) lncRNAs for sequences more than 200 nucleotides in length; and (b) short ncRNAs for transcripts of less than 200 bases [38].

miRNAs that are about 22 nucleotides in length are considered to be short ncRNAs. In ESCs, miRNA expression is also regulated by the core transcription factors (Figure 1), whereby the promoters of miRNA genes, which are preferentially expressed in ESCs, are bound by Oct4, Sox2, Nanog and Tcf3 factors. Similarly, miRNA genes involved in lineage specification were occupied by core transcription factors in conjunction with Polycomb group proteins, to exert transcriptional silencing [39]. Examples of these silenced miRNA genes include let-7, which targets pluripotency factors Lin28 and Sall4 [11], as well as miR-145, which is expressed during hESC differentiation to suppress the pluripotency factors OCT4, SOX2 and KLF4 in hESCs [40].

The lncRNA Xist, which performs a critical role in X-chromosome inactivation, is silenced by the core ESC factors along intron 1 of the mESC Xist gene (Figure 1) [41]. Similarly, ESC transcription factors also regulate the expression of the Xist antisense gene Tsix [42,43]. However, it was found that deletion of Xist intron 1 containing the Oct4-binding sites in ESCs did not result in Xist derepression [44]. Epiblast-derived stem cells and hESCs that express Oct4 are known to possess an inactive X-chromosome [45], and interestingly, pre-X inactivation hESCs have been derived from human blastocysts cultured under hypoxic conditions [46]. Therefore, it is likely that the ESC transcriptional network indirectly regulates X-chromosome activation status via an intermediary effector.

Recently, lincRNAs have been demonstrated to both maintain pluripotency and suppress lineage specification, hence integrating into the molecular circuitry governing ESCs [14]. Pluripotency factors such as Oct4, Sox2, Nanog and c-Myc have also been found to co-localize at lincRNA promoters, indicating that lincRNA expression is under the direct regulation of the ESC transcriptional network. Interestingly, mESC lincRNAs have been found to bind multiple ubiquitous chromatin complexes and RNA-binding proteins, leading to the proposal that lincRNAs function as ‘flexible scaffolds’ to recruit different protein complexes into larger units. By extension of this concept, it is possible that the unique lincRNA signature of each cell type may serve to bind protein
complexes to create a cell-type-specific gene expression profile.

Cellular reprogramming and iPSCs

The importance of the transcriptional regulatory network in establishing ESC self-renewal and pluripotency was elegantly demonstrated by Takahashi and Yamanaka [4], whereby introduction of four transcription factors Oct4, Sox2, Klf4 and c-Myc (OSKM) could revert differentiated cells back to pluripotency as iPSCs. iPSCs were later demonstrated to satisfy the highest stringency test of pluripotency via tetraploid complementation to form viable ‘all-iPSC’ mice [47].

However, reprogramming is not restricted to the four OSKM factors only. Closely related family members of the classical reprogramming factors such as Klf2 and Klf5 can replace Klf4, Sox1 can substitute for Sox2, and c-Myc can be replaced by using N-myc and L-myc [48]. However, Oct4 cannot be replaced by its close homologs Oct1 and Oct6 [48], but can be substituted using an unrelated orphan nuclear receptor, Nr5a2, to form mouse iPSCs [49]. Similarly, another orphan nuclear receptor, Esrrb, was demonstrated to replace Klf4 during iPSC generation [50]. Human iPSCs (hiPSCs), aside from the classical OSKM factors [51], can also be generated using a different cocktail of factors comprising OCT4, SOX2, NANOG and LIN28 [52]. Recently, the maternally expressed transcription factor Glis1 replaced c-Myc to generate both mouse iPSCs and hiPSCs [53]. Glis1 is highly expressed in unfertilized eggs and zygotes but not in ESCs; thus, it remains to be determined if other maternally expressed genes could similarly reinitiate pluripotency.

While certain transcription factors may be replaced with chemicals during the reprogramming process, they all still require at least one transcription factor [54]. Recently, however, the creation of hiPSCs and mouse iPSCs via miRNA without additional protein-encoding factors was reported [55,56]. By expressing the miR-302-miR-367 clusters, iPSCs can be generated with two orders of magnitude higher efficiency compared with conventional OSKM reprogramming [55]. Similarly, iPSCs could be formed by transfecting miR-302, miR-200 and miR-369 into mouse adipose stromal cells, albeit at lower efficiency [56]. The ability of miRNAs to reprogram somatic cells is intriguing, and it would be of great interest to determine the gene targets of these reprogramming miRNAs.

Expression profiling of ESCs and iPSCs

The question of whether pluripotent iPSCs truly resemble ESCs is an actively debated and evolving field, with evidence arguing both for and against iPSC-ESC similarity. As such, further research using better controlled studies is needed to resolve this issue. Here, we summarize and present the key findings that address this topic.

Initially, it was believed that hiPSCs were similar to hESCs [52,57], but subsequent studies argued otherwise as differential gene expression [58], as well as DNA methylation patterns [59], could be distinguished between hiPSCs and hESCs (Table 2). However, these differences were proposed to be a consequence of comparing cells of different genetic origins [60], laboratory-to-laboratory variation [61], and the iPSC passage number [62]. Later, hiPSCs were described to contain genomic abnormalities, including gene copy number variation [63,64], point mutations [65] and chromosomal duplications [66] (Table 2). However, whether these genomic instabilities are inherent in hiPSCs only, or a consequence of culture-induced mutations, as previously described in hESCs, is still not certain [67]. Extended passages of iPSCs appeared to reduce such aberrant genomic abnormalities, possibly via growth outcompetition by healthy iPSCs [64], but this was contradicted by a separate study that found that parental epigenetic signatures are retained in iPSCs even after extended passaging [68]. Indeed, this ‘epigenetic memory’ phenomenon was also reported in two earlier studies, whereby donor cell epigenetic memory led to an iPSC differentiation bias towards donor-cell-related lineages [62,69]. The mechanism behind this residual donor cell memory found in iPSCs was attributed to incomplete promoter DNA methylation [70]. Surprisingly, knockdown of incompletely reprogrammed somatic genes was found to reduce hiPSC generation, suggesting that somatic memory genes may play an active role in the reprogramming process [70]. Differences in ncRNA expression were also found between iPSCs and ESCs (Table 2). For instance, the aberrantly silenced imprinted Dlk1-Dio3 gene locus in iPSCs results in the differential expression of its encoded ncRNA Gtl2 and Rian, and 26 miRNAs, and consequent failure to generate ‘all-iPSC’ mice [60]. Upregulation of lincRNAs specifically in hiPSCs was also reported [13]. Expression of lincRNA-RoR with OSKM could also enhance iPSC formation by twofold, suggesting a critical function of lincRNA in the reprogramming process [13].

As these reported variations between hESCs and hiPSCs could be attributed to small sample sizes, a recent large-scale study by Bock et al. [71] profiled the global transcription and DNA methylation patterns of 20 different hESC lines and 12 hiPSC lines. Importantly, the study revealed that hiPSCs and hESCs were largely similar, and that the observed hiPSC differences were similar to normally occurring variation among hESCs. Additionally, Bock et al. established a scoring algorithm to predict lineage and differentiation propensity of hiPSCs. As traditional methods of screening hiPSC quality rely on time-consuming and low-throughput teratoma assays,
the hiPSC genetic scorecard offers researchers a quick assessment of the epigenetic and transcriptional status of pluripotent cells. This may be especially useful for the rapid monitoring of cell-line quality during large-scale production of iPSCs [71].

**Deregulation of transcriptional networks in disease**

Blastocyst-derived ESCs possess an innate ability for indefinite self-renewal, and can be considered a primary untransformed cell line. Unlike primary cell cultures with limited in vitro lifespans, or immortalized/tumor-derived cell lines that do not mimic normal cell behavior, ESCs thus offer a good model for studying cellular pathways. ESC transcriptomics have indeed advanced our understanding into the molecular mechanisms affecting certain human diseases.

For instance, it was previously reported that cancer cells possess an ESC-like transcriptional program, suggesting that ESC-associated genes may contribute to tumor formation [72]. However, this expression signature was shown to be a result of c-Myc, rather than from the core pluripotency factors (Table 3) [9]. As c-Myc somatic copy-number duplications are the most frequent in cancer [73], the finding that c-Myc releases RNA polymerase II from transcriptional pause [28] offers new understanding into the transcriptional regulatory role of c-Myc in ESCs and cancer cells. Another pluripotency-associated factor, Lin28, which suppresses the maturation of pro-differentiation let-7 miRNA, is also highly expressed in poorly differentiated and low prognosis tumors [74]. Importantly, let-7 silences several oncogenes, such as c-Myc, K-Ras, Hmga2 and the gene encoding cyclin-D1, suggesting that Lin28 deregulation may promote oncogenesis [74].

Aside from cancer, mutations in ESC-associated transcriptional regulators can cause developmental abnormalities. The Mediator-cohesin complex, which occupies 60% of active mESC genes, is responsible for regulating gene expression by physically linking gene enhancers to promoters though chromatin loops [29]. Notably, the binding pattern of Mediator-cohesin onto gene promoters differs among cell types, indicative of cell-type-specific gene regulation [29]. In hESCs, Mediator was also revealed to be important in the maintenance of pluripotent stem cell identity during a genome-wide siRNA screen, suggesting an evolutionarily conserved role [75]. Given this important gene regulatory function of the Mediator-cohesin complex in mESCs and hESCs, mutations in these proteins are associated with disorders such as schizophrenia, and Opitz-Kaveggia and Lujan syndromes [29]. Interestingly, the Cornelia de Lange syndrome, which causes mental retardation due to gene dysregulation rather than chromosomal abnormalities, is associated with mutations in cohesin-loading factor Nipbl [29]. Therefore, it is proposed that such developmental syndromes may arise as a result of the failure to form appropriate enhancer-promoter interactions.

Mutations in core ESC transcription factor SOX2 and the ATP-chromatin remodeler CHD7 result in developmental defects such as SOX2 anophthalmia (congenital absence of eyeballs) and CHARGE syndrome, respectively [76]. Although a direct association between CHARGE syndrome and ESCs is not known, mESC studies revealed that Chd7 co-localizes with core ESC factors and p300 protein at gene enhancers to modulate expression of ESC-specific genes [77]. It is thus possible that CHARGE syndrome may arise due to CHD7 enhancer-mediated gene dysregulation. In neural stem cells, Chd7 is able to bind with Sox2 at the Jag1, Gli3 and Mycn genes, which are mutated in the developmental disorders Alagille, Pallister-Hall and Feingold syndromes [78]. Similarly, Chd7 has been described to interact with the PBAF

---

**Table 2. Transcriptomic comparisons between induced pluripotent stem cells and embryonic stem cells**

| Characteristic          | Mouse iPSCs                                      | Human iPSCs                                      |
|-------------------------|-------------------------------------------------|-------------------------------------------------|
| mRNA expression         | Distinct from mESCs at lower passages, donor cell gene expression still present [62,69]; closely resemble mESCs at late passages [62] | Distinct from hESCs at lower passages [58], with residual donor gene expression [70,96,97]; closely resemble hESCs at late passages [58,98] |
| miRNA expression        | miRNA encoded within the imprinted Dlk1-Dio3 locus is aberrantly silenced [60] | Small number of differences reported [58,99], but variation between hESCs and hiPSCs comparable to somatic and cancer cells [100] |
| IncRNA expression       | Not determined                                   | Differences in lncRNA expression reported. lncRNA-RoR enhances reprogramming by twofold [13] |
| DNA methylation status  | Distinct from mESCs at lower passages, donor cell DNA methylation pattern still present [62,69]; closely identical to mESCs at late passages [62] | Differences in DNA methylation reported [59,68,70], but not in all hiPSCs [71] |
| Genome status           | Not determined                                   | Possess gene copy number deletions and duplications [63-64], somatic coding mutations [65], and chromosomal duplications [66] |

hESC, human embryonic stem cell; hiPSC, human induced pluripotent stem cell; iPSC, induced pluripotent stem cell; lncRNA, long-intergenic non-coding RNA; miRNA, microRNA.
Physically links the Oct4/Sox2/Nanog-bound gene enhancers to ESC pluripotency by binding and inhibiting the maturation of pro-differentiation let-7 miRNA; LIN28 is also a hiPSC reprogramming factor [74].

SOX2
A core ESC transcription factor together with Oct4 and Nanog. Regulates the expression of pluripotency genes, and suppresses lineage-specific genes [23,24]; Sox2 is also an iPSC reprogramming factor [4].

CHD7
Binds with core ESC factors and p300 at gene enhancers to modulate ESC-specific gene expression [77].

Mediator
Physically links the Oct4/Sox2/Nanog-bound gene enhancers to active gene promoters via chromatin looping [29], necessary for normal gene activity.

Cohesin
Proposed to bind and stabilize the Oct4/Sox2/Nanog enhancer-promoter chromatin loops [1]; necessary for normal gene activity.

Nipbl
Binds with mediator complex to allow loading of cohesin and formation of stable chromatin loop [29].

Table 3. Dysregulation of transcriptional networks in stem cells and disease

| Gene/protein | Role in ESCs | Role in disease |
|--------------|--------------|-----------------|
| c-MYC        | Involved in the expression of self-renewal genes [101]; recruits p-TEFβ to initiate transcriptional pause release of RNA polymerase II [29] | Most common gene duplication in cancer [73]; c-Myc appears to be responsible for the gene expression signature of cancer cells [9] |
| LIN28        | Maintains ESC pluripotency by binding and inhibiting the maturation of pro-differentiation let-7 miRNA; LIN28 is also a hiPSC reprogramming factor [74] | Highly expressed in poorly differentiated and low prognostic tumors; as let-7 silences the expression of oncogenes c-Myc, K-Ras, Hmgα2 and the gene encoding cyclin-D1, Lin28 suppression of let-7 miRNA may thus promote oncogenesis [74] |
| SOX2         | A core ESC transcription factor together with Oct4 and Nanog. Regulates the expression of pluripotency genes, and suppresses lineage-specific genes [23,24]; Sox2 is also an iPSC reprogramming factor [4] | Mutation in SOX2 causes anophthalmia (congenital loss of eyeballs) in humans. Proposed to cooperate with CHD7 to regulate genes involved in Alagille, Pallister-Hall and Feingold syndromes [76] |
| CHD7         | Binds with core ESC factors and p300 at gene enhancers to modulate ESC-specific gene expression [77] | Mutations in CHD7 result in CHARGE syndrome; proposed to cooperate with SOX2 to regulate genes involved in Alagille, Pallister-Hall and Feingold syndromes [76] |
| Mediator     | Physically links the Oct4/Sox2/Nanog-bound gene enhancers to active gene promoters via chromatin looping [29], necessary for normal gene activity | Mutations in Mediator are associated with Opitz-Kaveggia, Lujan, and transposition of the great arteries syndromes; also implicated in schizophrenia, colon cancer progression [1] and uterine leiomyomas [102] |
| Cohesin      | Proposed to bind and stabilize the Oct4/Sox2/Nanog enhancer-promoter chromatin loops [1]; necessary for normal gene activity | Cohesin mutations implicated in Cornelia de Lange syndrome, whereby patients exhibit developmental defects and mental retardation due to dysregulation of gene expression [29] |
| Nipbl        | Binds with mediator complex to allow loading of cohesin and formation of stable chromatin loop [29] | Nipbl mutations implicated in Cornelia de Lange syndrome, whereby patients exhibit developmental defects and mental retardation due to dysregulation of gene expression [29] |

ESC, embryonic stem cell; hiPSC, human induced pluripotent stem cell; iPSC, induced pluripotent stem cell; miRNA, microRNA.

Clinical and therapeutic implications

The development of hiPSC technology offers the unique opportunity to derive disease-specific hiPSCs for the in vitro study of human disease pathogenesis (Figure 2). A major advantage of using disease-specific hiPSCs is that they allow the capture of the patient’s genetic background and, together with the patient’s medical history, will enable the researcher to uncover the disease genotypic-phenotypic relationship [6]. A number of patient-derived hiPSC disease models have been established, including those for Hutchinson Gilford Progeria, Timothy syndrome, schizophrenia and Alzheimer’s disease [5,80-83], and these have been useful in understanding the cellular mechanisms behind these illnesses. For example, transcriptional profiling of schizophrenia neurons derived from iPSCs have identified 596 differentially expressed genes, 75% of which were not previously implicated in schizophrenia [82]. This highlights the potential of disease-specific iPSCs in unlocking hidden pathways. Additionally, the use of disease cell lines can facilitate drug design and screening under disease conditions (Figure 2) [6]. One such example is the drug roscovitine, which was found to restore the electrical and Ca2+ signaling in Timothy syndrome cardiomyocytes [81].

The self-renewing ability of hiPSCs means that a potentially unlimited source of patient-specific cells can be generated for regenerative purposes (Figure 2). Importantly, hiPSCs, when coupled with gene targeting approaches to rectify genetic mutations, can be differentiated into the desired cell type and reintroduced to the patient (Figure 2) [5]. However, unlike mESCs, hESCs and hiPSCs cannot be passaged as single cells and have very poor homologous recombination ability [84]. Circumventing this problem may require the conversion of hiPSCs into a mESC-like state, which is more amenable to gene targeting [85]. Alternatively, recent reports of successful gene targeting in human pluripotent stem cells using zinc-finger nucleases (ZFNs) [86], and transcription activator-like effector nucleases (TALENs) [87], presents another option for genetically altering hiPSCs for cell therapy. Albeit that there are concerns of off-target effects, the advantage of using nuclease-targeting approaches is that they do not necessitate the conversion of hESCs and hiPSCs into mESC-like states prior to genomic manipulation.

While it has been assumed that iPSCs generated from an autologous host should be immune-tolerated, Zhao et al. [88] recently demonstrated that iPSCs were immunogenic...
and could elicit a T-cell immune response when transplanted into syngeneic mice. However, it should be distinguished that in the Zhao et al. study undifferentiated iPSCs were injected into mice, rather than differentiated iPSC-derived cells, which are the clinically relevant cell type for medical purposes. Furthermore, the immune system is capable of ‘cancer immunosurveillance’ to identify and destroy tumorigenic cells [89]. Hence, it may be possible that the observed iPSC immunogenicity could have arisen through cancer immunosurveillance against undifferentiated tumor-like iPSCs, and that iPSC-derived differentiated cells may not be immunogenic. It would thus be necessary to experimentally verify if iPSC-derived differentiated cells are immunogenic in syngeneic hosts.

**Conclusions and future challenges**

Understanding and exploiting the mechanisms that govern pluripotency are necessary if hESCs and hiPSCs are to be successfully translated to benefit clinical and medical applications. One approach for understanding hESCs and hiPSCs would be to study their transcriptomes,
and, through various approaches, we have learnt how the core pluripotency factors create an ESC gene expression signature by regulating other transcription factors and controlling chromatin structure and ncRNA expression.

Current methodologies to generate iPSCs are inefficient, suggesting that significant and unknown epigenetic barriers to successful reprogramming remain [90]. However, defining these barriers is difficult, as existing transcriptomic studies rely on average readings taken across a heterogeneous cell population. This therefore masks essential rate-limiting transcriptional and epigenetic remodeling steps in iPSC formation. Future studies in elucidating the iPSC generation process may thus adopt a single-cell approach [91], which will offer the resolution needed to define key reprogramming steps. Future efforts should also be focused upon improving hiPSC safety for human applications, through the use of stringent genomic and functional screening strategies on hiPSCs and their differentiated tissues [3]. Only with well-defined and non-tumorigenic iPSC-derived tissue would we then be able to assess the transplant potential of iPSCs in personalized medicine.

In addition to generating disease-specific iPSCs from patients, the use of gene-modifying nucleases to create hESCs harboring specific genetic mutations may be a forward approach towards studying human disease pathogenesis [86]. With the recent creation of approximately 9,000 conditional targeted alleles in mESCs [92], it would be of tremendous scientific and clinical value to likewise establish a hESC knockout library to study the role of individual genes in disease and development. Furthermore, while SNP and haplotype mapping may be useful in associating diseases with specific genetic loci, the use of ZFNs or TALENs to recreate these specific gene variations in hESCs may offer an experimental means of verifying the relationship of SNPs or haplotypes with diseases.

Acknowledgments
The authors thank all the members of the Ng laboratory for their comments on this manuscript.

Author details
1. Gene Regulation Laboratory, Genome Institute of Singapore, 60 Biopolis Street, Genome, Singapore 138672. “School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551. “Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597. “Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117597. “NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 28 Medical Drive, Singapore 117456.

Published: 27 October 2011

References
1. Young RA: Control of the embryonic stem cell state. Cell 2011, 144:940-954.
2. Ben-David U, Benvenisty N: The tumorigenicity of human embryonic and induced pluripotent stem cells. Nat Rev Cancer 2011, 11:268-277.
3. Barrillex M, Knoepfler PS: Inducing iPSCs to escape the dish. Cell Stem Cell 2011, 9:103-111.
4. Takahashi K, Yamanaka S: Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006, 126:663-676.
5. Wu SM, Hochedinger E: Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nat Cell Biol 2011, 13:947-505.
6. Zou H, Lentsch MW, Cahan P, Daley GQ: Investigating monogenic and complex diseases with pluripotent stem cells. Nat Rev Genet 2011, 12:266-275.
7. Wang Z, Gerstein M, Snyder M: RNA-seq: a revolutionay tool for transcriptomics. Nat Rev Genet 2009, 10:567-63.
8. Marguerat S, Bahler J: RNA-seq from technology to biology. Cell Mol Life Sci 2010, 67:569-579.
9. Kim J, Woo AJ, Chu J, Snow JW, Fujihara Y, Kim CG, Cantor AB, Orkin SH: A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell 2010, 143:313-324.
10. Kim J, Chu J, Shen X, Wang J, Orkin SH: An extended transcriptional network for pluripotency of embryonic stem cells. Cell 2008, 132:1049-1061.
11. Melton C, Judson RL, Bleiholder R: Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature 2010, 463:621-626.
12. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, Sunkim SM, Crowle ML, Grimmmond SM, Perkins AG, Mattick JS: Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 2008, 18:1433-1445.
13. Lowe SR, Cabili MN, Guttmann M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, Manos PD, Datta S, Lander ES, Schlaeger TM, Daley GQ, Rinn JL: Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet 2010, 42:1113-1117.
14. Guttmann M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amj T, Meissner A, Regev A, Rinn J, Root DE, Lander ES: lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 2011, 477:295-300.
15. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YI, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HK: Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 2008, 133:1106-1117.
16. Williams K, Christensen J, Pedersen MT, Johansen JV, Cioss PA, Rappssler J, Helin K: TET1 and hydroxymethylation in transcription and DNA methylation fidelity. Nature 2011, 473:343-348.
17. Handikou L, Xiu H, Li G, Nigan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, Wong E, Sheng J, Zhang Y, Poh T, Chan CS, Kurunio G, Shahab A, Bourque G, Cachex-Ratafoul N, Sung WK, Ruan Y, Wei CL, Onder T: CTCF-mediated functional chromatin interaction in pluripotent cells. Nat Genet 2011, 43:630-638.
18. Morin RD, O’Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA: Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 2008, 18:510-521.
19. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarm J, Song JL, Kingston RE,
cells are distinguished by gene expression signatures. Cell Stem Cell 2009, 5(1):111-123.
59. Doi A, Park IH, Wen B, Murakami P, Ayee MJ, Irazary R, Herb L, Badd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP: Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 2009, 41:1350-1353.
60. Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natasean T, Kono T, Shioda T, Hochegener K: Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature 2010, 465:175-181.
61. Newman AM, Cooper JB: Lab-specific gene expression signatures in pluripotent stem cells. Cell Stem Cell 2010, 7:258-262.
62. Polo JM, Liu S, Figueroa ME, Kulalet W, Emirini S, Tan K, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natasean T, Wagers AJ, Melnick A, Evans T, Hochegener K: Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol 2010, 28:848-855.
63. Laurent LC, Ulltakay I, Stavin I, Tran H, Schork A, Money R, Lynch J, Celec D, Smith Z, Montserrat N, Panopoulos AD, Rieu S, Wilbert MD, Yu J, Kirschke EF, Izpisua Belmonte JC, Ross DJ, Thomson JA, Eggan K, Goldstein LS, Yizha GL, Soldan F: Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. Nature 2011, 472:221-225.
64. Viswanathan SR, Daley GQ: Lin28: a microRNA regulator with a macro role. Cell 2010, 140:445-449.
65. Chiu NY, Chan YS, Feng B, Lu X, Orlow YI, Moreau D, Kumar P, Yang L, Jiang J, Lau MS, Hoss SS, Kraus F, Li P, Lu X, Luhf T, Lim B, Clarke ND, Bard F, Ng HH: A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature 2010, 468:316-320.
66. Puc J, Rosenfeld MG: SOX2 and CHD7 cooperatively regulate human disease genes. Nat Genet 2011, 43:505-506.
67. Scherf MP: Handlko A, Aktar-Zaibi D, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flick P, Crawford GE, Laframboise T, Ties P, Wei CL, Scharfen PC: CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. PLoS Genet 2010, 6:e1000103.
68. Engelen E, Aljini U, Byun JC, Jho J, Gontank AM, Moen M, Szumska D, Kock C, van Ijcken W, Dekkers DH, Demmers J, Rijikers E, Bhattacharya S, Philippsen S, Pevny LH, Grosveld F, Rottier RI, Lenhard B: Port RA: Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. Nat Genet 2011, 43:607-611.
69. Bajpai R, Chen DA, Rada-Iglesias A, Zhang J, Xiong Y, Helms J, Chang CP, Zhao Y, Swigut T, Wysoczka J: CHD7 cooperates with PBAF to control multipotent neural crest formation. Nature 2010, 463:956-962.
70. Liu GH, Barkho BZ, Rieu S, Dev D, Qi J, Yang SL, Panopoulos AD, Suzuki K, Kuranil W, Clark T, Thompson J, Bouse S, Fung H, Sancho-Martinez I, Zhang K, Yates J, Izpisua Belmonte JC: Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. Nature 2011, 472:221-225.
71. Buecker C, Chen HH, Polo JM, Daleron L, Bu L, Baratke TS, Okvietke P, Porter A, Grimbau J, Hochegener K, Geisbin N: A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell Stem Cell 2010, 6:535-546.
72. Soldner F, Lagenaritte J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, Khurana V, Golbe LI, Myers RH, Lindquist S, Zhang L, Guschin D, Fong LK, Vu BJ, Meng X, Uroso FD, Redor EJ, Gregory PD, Zhang L, Haensch J, Rappold J: Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 2011, 146:318-331.
73. Hockemeyer D, Wang H, Kian S, Lai CS, Wang D, McCarther S, Sebat J, Gage FH: Modelling schizophrenia using human induced pluripotent stem cells. Nature 2011, 473:221-225.
74. Quag L, Willett R, Yamashita J, Angulo S, Rhinn H, Rhee D, Dvodec J, Chau L, Aubrey L, Vanti WB, Moreno H, Abelovich A: Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. Cell 2011, 146:359-371.
75. Zwaka TP, Thomson RH: Homologous recombination in human embryonic stem cells. Nat Biotechnol 2003, 21:319-321.
76. Eberly S, Chen HH, Polo JM, Daleron L, Bu L, Baratke TS, Okvietke P, Porter A, Grimbau J, Hochegener K, Geisbin N: A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell Stem Cell 2010, 6:535-546.
77. Soldner F, Laganeritte J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, Khurana V, Golbe LI, Myers RH, Lindquist S, Zhang L, Guschin D, Fong LK, Vu BJ, Meng X, Uroso FD, Redor EJ, Gregory PD, Zhang L, Haensch J, Rappold J: Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 2011, 146:318-331.
78. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Sebat J, Gage FH: Modelling schizophrenia using human induced pluripotent stem cells. Nature 2011, 473:221-225.
79. Quag L, Willett R, Yamashita J, Angulo S, Rhinn H, Rhee D, Dvodec J, Chau L, Aubrey L, Vanti WB, Moreno H, Abelovich A: Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. Cell 2011, 146:359-371.
80. Zwaka TP, Thomson RH: Homologous recombination in human embryonic stem cells. Nat Biotechnol 2003, 21:319-321.
81. Huey B, Lackard J, Atal L, Lowry WE: Benvenisty N: Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 2010, 7:521-531.
82. Maitra A, Arking DE, Shivapuran N, Ikeda M, Stavitsky V, Kassaee K, Sui G, Cutler DJ, Liu Y, Bimbble SN, Noaksson K, Hyfler J, Schulz TC, Zeng X, Freed WJ, Crock J, Abrahams S, Colman A, Sarvey P, Matsus A, Carpenter M, Gazdar AF, Rao M, Chakravarti A: Genomic alterations in cultured human embryonic stem cells. Nat Genet 2005, 37:1099-1103.
83. Lister R, Pedetola M, Kida YS, Hawkins RD, Nery JR, Henn G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thompson JA, Essam RM, Ecker JR: Hotspots of aberrant epigenetic reprogramming in human induced pluripotent stem cells. Nature 2011, 471:65-73.
84. Buecker C, Chen HH, Polo JM, Daleron L, Bu L, Baratke TS, Okvietke P, Porter A, Grimbau J, Hochegener K, Geisbin N: A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. Cell Stem Cell 2010, 6:535-546.
85. Fisker D, Lagenaritte J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, Khurana V, Golbe LI, Myers RH, Lindquist S, Zhang L, Guschin D, Fong LK, Vu BJ, Meng X, Uroso FD, Redor EJ, Gregory PD, Zhang L, Haensch J, Rappold J: Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 2011, 146:318-331.
86. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Sebat J, Gage FH: Modelling schizophrenia using human induced pluripotent stem cells. Nature 2011, 473:221-225.
short, long, and paired-end sequencing. Proc Natl Acad Sci U S A 2010, 107:5254-5259.
96. Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR: Transcriptional signature and memory retention of human-induced pluripotent stem cells. PLoS One 2009, 4:e7076.
97. Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu J. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. PLoS One 2010, 5:e8975.
98. Guenther MG, Frampton GM, Soldner F, Hockemeyer D, Mitalipova M, Jaenisch R, Young RA: Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell 2010, 7:249-257.
99. Wilson KD, Venkatasubrahmanyam S, Jia F, Sun N, Butte AJ, Wu JC. MicroRNA profiling of human-induced pluripotent stem cells. Stem Cells Dev 2009, 18:749-758.
100. Neveu P, Kye MJ, Qi S, Buchholz DE, Clegg DO, Sahin M, Park H, Kim KS, Daley GQ, Komblum H, Shraiman BI, Kosik KS. MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states. Cell Stem Cell 2010, 7:671-681.
101. Sritharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Pliath K: Role of the murine reprogramming factors in the induction of pluripotency. Cell 2009, 136:364-377.
102. Mäkinen N, Mehlke M, Toivanen J, Kaasinen E, Li Y, Lehtonen H, Gentile M, Yan J, Enge M, Taipale M, Vahteristo P, Aaltonen LA: MED12, the Mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. Science 2011.

Cite this article as: Yeo J-C, Ng H-H. Transcriptional analysis of pluripotent stem cells: insights into health and disease. Genome Medicine 2011, 3:68.