Genome **Medicine**

REVIEW

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

Jia-Chi Yeo[1,2] and Huck-Hui Ng[1,2,3,4,5,]*

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

*Correspondence: nghh@gis.a-star.edu.sg
[1]Gene Regulation Laboratory, Genome Institute of Singapore, 60 Biopolis Street, Genome, Singapore 138672
Full list of author information is available at the end of the article

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 micro-arrays 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].
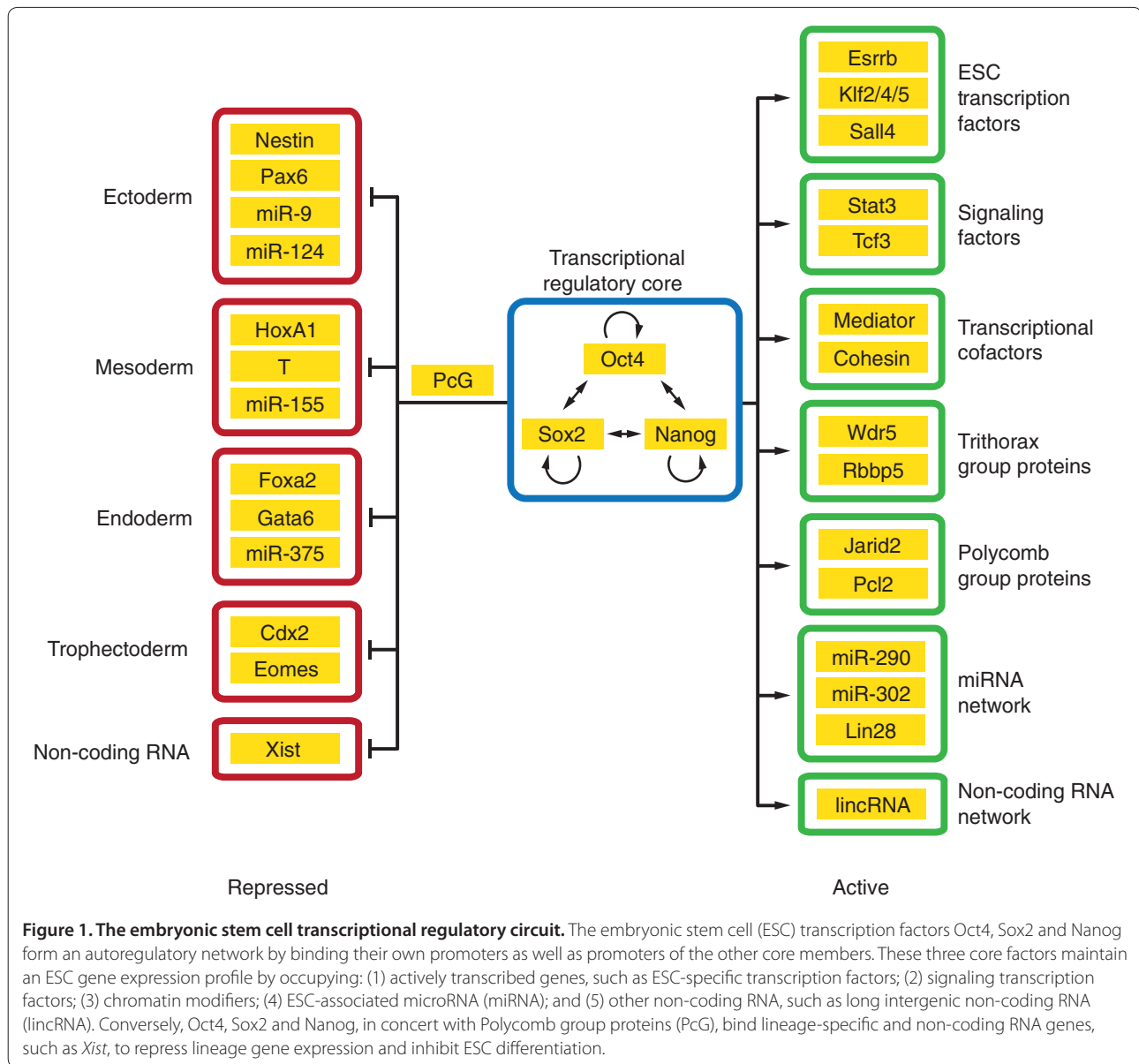
Sequence-based transcriptomic analysis on the other hand involves direct sequencing of the cDNA. Initially, Sanger sequencing techniques were used to sequence gene transcripts, but these methods were considered expensive and low throughput [7]. However, with the development of next-generation sequencing (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 immuno-precipitation 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]
	RNA-seq	[93]
miRNA expression analysis	Microarray	[11]
	RNA-seq	[18]
lncRNA expression analysis	Microarray	[12,13]
Identification of alternative splicing isoforms	Microarray	[22,94]
	RNA-seq	[95]
Mapping of protein-DNA binding	ChIP-chip	[9,10,24]
	ChIP-PET	[23]
	ChIP-seq	[15]
DNA methylation profiling	BS-seq	[68]
	MethylC-seq	[68]
	DIP-seq	[16]
Mapping of long-range chromatin interactions	ChIA-PET	[17]
	3C	[29]
Identification of RNA-protein interactions	RIP-seq	[19]
	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.

Figure 1. The embryonic stem cell transcriptional regulatory circuit. The embryonic stem cell (ESC) transcription factors Oct4, Sox2 and Nanog form an autoregulatory network by binding their own promoters as well as promoters of the other core members. These three core factors maintain an ESC gene expression profile by occupying: (1) actively transcribed genes, such as ESC-specific transcription factors; (2) signaling transcription factors; (3) chromatin modifiers; (4) ESC-associated microRNA (miRNA); and (5) other non-coding RNA, such as long intergenic non-coding RNA (lincRNA). Conversely, Oct4, Sox2 and Nanog, in concert with Polycomb group proteins (PcG), bind lineage-specific and non-coding RNA genes, such as *Xist*, to repress lineage gene expression and inhibit ESC differentiation.

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,

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]
lncRNA expression	Not determined	Differences in lincRNA expression reported. *lincRNA-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; lincRNA, long-intergenic non-coding RNA; lncRNA, long non-coding RNA; mESC, mouse embryonic stem cell; miRNA, microRNA.

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 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-TEFb 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 prognosis tumors; as *let-7* silences the expression of oncogenes *c-Myc*, *K-Ras*, *Hmga2* 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 cohesion 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.

complex to control neural crest formation [79]. Therefore, these data hint that Chd7 may partner different proteins to cooperatively regulate developmental genes. Although the mechanism behind gene regulation by Chd7 and its interacting partners is not well understood, the use of ESCs may serve as a useful system to further probe Chd7 function during development and disease.
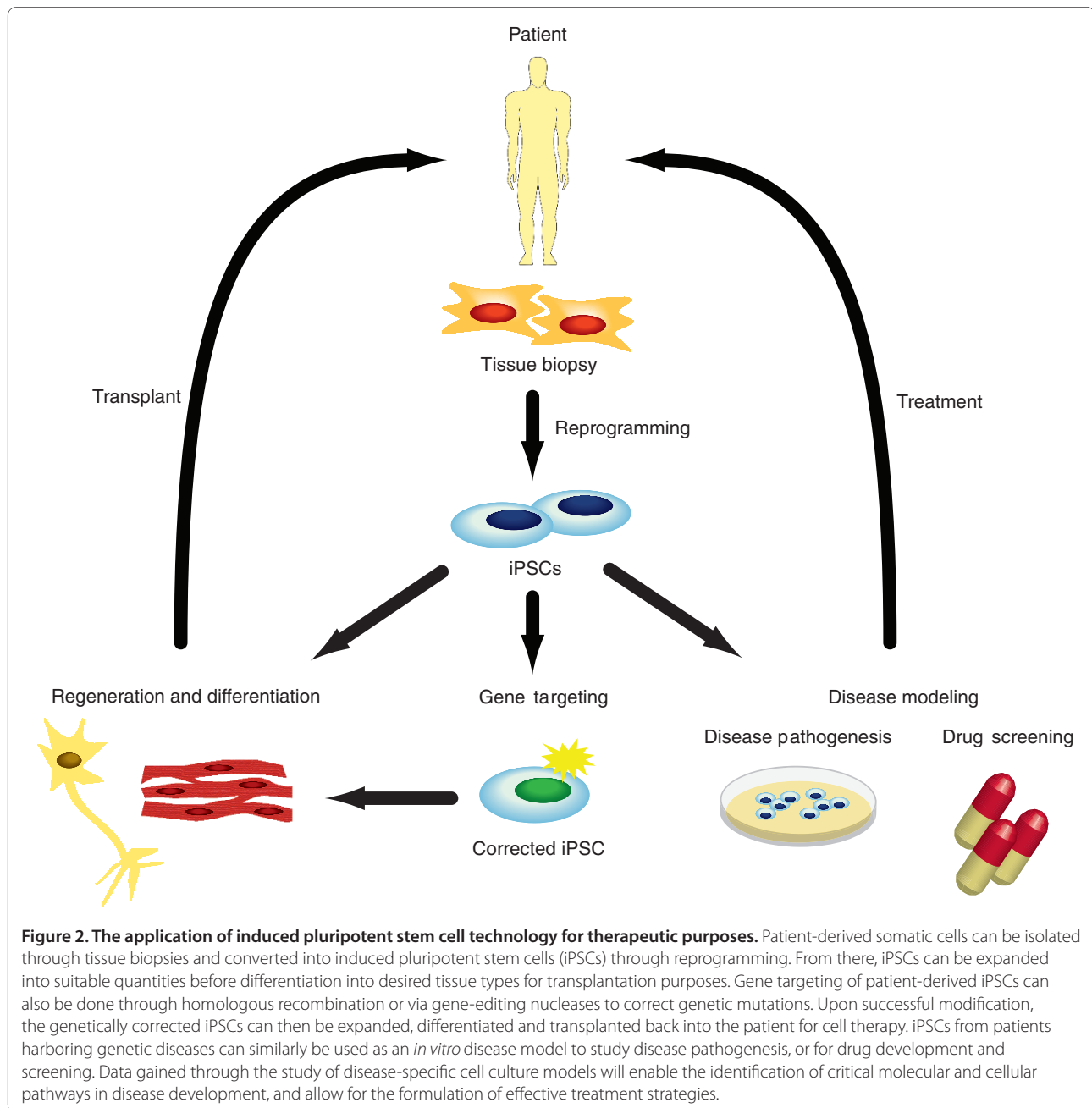
Clinical and therapeutic implications

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 Ca^{2+} 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

Figure 2. The application of induced pluripotent stem cell technology for therapeutic purposes. Patient-derived somatic cells can be isolated through tissue biopsies and converted into induced pluripotent stem cells (iPSCs) through reprogramming. From there, iPSCs can be expanded into suitable quantities before differentiation into desired tissue types for transplantation purposes. Gene targeting of patient-derived iPSCs can also be done through homologous recombination or via gene-editing nucleases to correct genetic mutations. Upon successful modification, the genetically corrected iPSCs can then be expanded, differentiated and transplanted back into the patient for cell therapy. iPSCs from patients harboring genetic diseases can similarly be used as an *in vitro* disease model to study disease pathogenesis, or for drug development and screening. Data gained through the study of disease-specific cell culture models will enable the identification of critical molecular and cellular pathways in disease development, and allow for the formulation of effective treatment strategies.

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.

Abbreviations
CHARGE, Coloboma of the eye, Heart defects, Atresia of the choanae, Retardation of growth and/or development, Genital and/or urinary abnormalities, and Ear abnormalities and deafness; ChIP, chromatin immunoprecipitation; ChIP-chip, chromatin immunoprecipitation on chip; ChIP-seq, chromatin immunoprecipitation and sequencing; DIP-seq, DNA immunoprecipitation and sequencing; ESC, embryonic stem cell; hESC, human embryonic stem cell; hiPSC, human induced pluripotent stem cell; H3K4me3, histone H3 lysine 4 trimethylation; iPSC, induced pluripotent stem cell; lincRNA, long intergenic non-coding RNA; lncRNA, long non-coding RNA; mESC, mouse embryonic stem cell; miRNA, microRNA; NGS, next-generation sequencing; ncRNA, non-coding RNA; oligo, oligonucleotide; OSKM, Oct4, Sox2, Klf4 and c-Myc; PRC2, Polycomb repressive complex 2; RIP-seq, RNA-binding protein immunoprecipitation and sequencing; RNA-seq, RNA sequencing; siRNA, short interfering RNA; SNP, single nucleotide polymorphism; TALEN, transcription activator-like effector nuclease; UTR, untranslated region; ZFN, zinc-finger nuclease.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Gene Regulation Laboratory, Genome Institute of Singapore, 60 Biopolis Street, Genome, Singapore 138672. [2]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551. [3]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597. [4]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117597. [5]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 28 Medical Drive, Singapore 117456.

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. Barrilleaux B, 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, Hochedlinger K: **Harnessing the potential of induced pluripotent stem cells for regenerative medicine.** *Nat Cell Biol* 2011, **13**:497-505.
6. Zhu H, Lensch 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 revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.
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, Fujiwara 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, Blelloch 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, Soldà G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS: **Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation.** *Genome Res* 2008, **18**:1433-1445.
13. Loewer S, Cabili MN, Guttman 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. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, 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 YL, 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 HH: **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, Cloos PA, Rappsilber J, Helin K: **TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.** *Nature* 2011, **473**:343-348.
17. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, Wong E, Sheng J, Zhang Y, Poh T, Chan CS, Kunarso G, Shahab A, Bourque G, Cacheux-Rataboul V, Sung WK, Ruan Y, Wei CL: **CTCF-mediated functional chromatin interactome 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**:610-621.
19. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE,

Borowsky M, Lee JT: **Genome-wide identification of polycomb-associated RNAs by RIP-seq.** *Mol Cell* 2010, **40:**939-953.

20. Chen M, Manley JL: **Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.** *Nat Rev Mol Cell Biol* 2009, **10:**741-754.

21. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD: **Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays.** *Science* 2003, **302:**2141-2144.

22. Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zambon AC, Vranizan K, Spindler MJ, Pico AR, Cline MS, Clark TA, Williams A, Blume JE, Samal E, Mercola M, Merrill BJ, Conklin BR: **Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation.** *Proc Natl Acad Sci U S A* 2010, **107:**10514-10519.

23. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH: **The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.** *Nat Genet* 2006, **38:**431-440.

24. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122:**947-956.

25. Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA: **Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells.** *Genes Dev* 2008, **22:**746-755.

26. Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J: **An expanded Oct4 interaction network: implications for stem cell biology, development, and disease.** *Cell Stem Cell* 2010, **6:**382-395.

27. van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA: **An Oct4-centered protein interaction network in embryonic stem cells.** *Cell Stem Cell* 2010, **6:**369-381.

28. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: **c-Myc regulates transcriptional pause release.** *Cell* 2010, **141:**432-445.

29. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA: **Mediator and cohesin connect gene expression and chromatin architecture.** *Nature* 2010, **467:**430-435.

30. Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, Jorgensen HF, Pereira CF, Leleu M, Piccolo FM, Spivakov M, Brookes E, Pombo A, Fisher C, Skarnes WC, Snoek T, Bezstarosti K, Demmers J, Klose RJ, Casanova M, Tavares L, Brockdorff N, Merkenschlager M, Fisher AG: **Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators.** *Nat Cell Biol* 2010, **12:**618-624.

31. Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D: **Jarid2 and PRC2, partners in regulating gene expression.** *Genes Dev* 2010, **24:**368-380.

32. Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K: **JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells.** *Nature* 2010, **464:**306-310.

33. Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J: **Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells.** *Cell* 2009, **139:**1290-1302.

34. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y, Orkin SH: **Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells.** *Cell* 2009, **139:**1303-1314.

35. Walker E, Chang WY, Hunkapiller J, Cagney G, Garcha K, Torchia J, Krogan NJ, Reiter JF, Stanford WL: **Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation.** *Cell Stem Cell* 2010, **6:**153-166.

36. Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, Wang J, Rendl M, Bernstein E, Schaniel C, Lemischka IR: **Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network.** *Cell* 2011, **145:**183-197.

37. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M: **Open chromatin in pluripotency and reprogramming.** *Nat Rev Mol Cell Biol* 2011, **12:**36-47.

38. Pauli A, Rinn JL, Schier AF: **Non-coding RNAs as regulators of embryogenesis.** *Nat Rev Genet* 2011, **12:**136-149.

39. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, Calabrese JM, Dennis LM, Volkert TL, Gupta S, Love J, Hannett N, Sharp PA, Bartel DP, Jaenisch R, Young RA: **Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.** *Cell* 2008, **134:**521-533.

40. Xu N, Papagiannakopoulos T, Pan G, Thomson JA, Kosik KS: **MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.** *Cell* 2009, **137:**647-658.

41. Navarro P, Chambers I, Karwacki-Neisius V, Chureau C, Morey C, Rougeulle C, Avner P: **Molecular coupling of Xist regulation and pluripotency.** *Science* 2008, **321:**1693-1695.

42. Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT: **The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting.** *Nature* 2009, **460:**128-132.

43. Navarro P, Oldfield A, Legoupi J, Festuccia N, Dubois A, Attia M, Schoorlemmer J, Rougeulle C, Chambers I, Avner P: **Molecular coupling of Tsix regulation and pluripotency.** *Nature* 2010, **468:**457-460.

44. Barakat TS, Gunhanlar N, Pardo CG, Achame EM, Ghazvini M, Boers R, Kenter A, Rentmeester E, Grootegoed JA, Gribnau J: **RNF12 activates Xist and is essential for X chromosome inactivation.** *PLoS Genet* 2011, **7:**e1002001.

45. Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A: **Klf4 reverts developmentally programmed restriction of ground state pluripotency.** *Development* 2009, **136:**1063-1069.

46. Lengner CJ, Gimelbrant AA, Erwin JA, Cheng AW, Guenther MG, Welstead GG, Alagappan R, Frampton GM, Xu P, Muffat J, Santagata S, Powers D, Barrett CB, Young RA, Lee JT, Jaenisch R, Mitalipova M: **Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations.** *Cell* 2010, **141:**872-883.

47. Zhao XY, Li W, Lv Z, Liu L, Tong M, Hai T, Hao J, Guo CL, Ma QW, Wang L, Zeng F, Zhou Q: **iPS cells produce viable mice through tetraploid complementation.** *Nature* 2009, **461:**86-90.

48. Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S: **Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.** *Nat Biotechnol* 2008, **26:**101-106.

49. Heng JC, Feng B, Han J, Jiang J, Kraus P, Ng JH, Orlov YL, Huss M, Yang L, Lufkin T, Lim B, Ng HH: **The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells.** *Cell Stem Cell* 2010, **6:**167-174.

50. Feng B, Jiang J, Kraus P, Ng JH, Heng JC, Chan YS, Yaw LP, Zhang W, Loh YH, Han J, Vega VB, Cacheux-Rataboul V, Lim B, Lufkin T, Ng HH: **Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb.** *Nat Cell Biol* 2009, **11:**197-203.

51. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell* 2007, **131:**861-872.

52. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA: **Induced pluripotent stem cell lines derived from human somatic cells.** *Science* 2007, **318:**1917-1920.

53. Maekawa M, Yamaguchi K, Nakamura T, Shibukawa R, Kodanaka I, Ichisaka T, Kawamura Y, Mochizuki H, Goshima N, Yamanaka S: **Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1.** *Nature* 2011, **474:**225-229.

54. Gonzalez F, Boue S, Izpisua Belmonte JC: **Methods for making induced pluripotent stem cells: reprogramming a la carte.** *Nat Rev Genet* 2011, **12:**231-242.

55. Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA, Morrisey EE: **Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency.** *Cell Stem Cell* 2011, **8:**376-388.

56. Miyoshi N, Ishii H, Nagano H, Haraguchi N, Dewi DL, Kano Y, Nishikawa S, Tanemura M, Mimori K, Tanaka F, Saito T, Nishimura J, Takemasa I, Mizushima T, Ikeda M, Yamamoto H, Sekimoto M, Doki Y, Mori M: **Reprogramming of mouse and human cells to pluripotency using mature microRNAs.** *Cell Stem Cell* 2011, **8:**633-638.

57. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ: **Reprogramming of human somatic cells to pluripotency with defined factors.** *Nature* 2008, **451:**141-146.

58. Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiuwu O, Richter L, Zhang J, Khvorostov I, Ott V, Grunstein M, Lavon N, Benvenisty N, Croce CM, Clark AT, Baxter T, Pyle AD, Teitell MA, Pelegrini M, Plath K, Lowry WE: **Induced pluripotent stem cells and embryonic stem**

cells are distinguished by gene expression signatures. *Cell Stem Cell* 2009, **5:**111-123.

59. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-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, Natesan S, Kono T, Shioda T, Hochedlinger 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, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger 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, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, Lynch C, Harness JV, Lee S, Barrero MJ, Ku S, Martynova M, Semechkin R, Galat V, Gottesfeld J, Izpisua Belmonte JC, Murry C, Keirstead HS, Park HS, Schmidt U, Laslett AL, Muller FJ, Nievergelt CM, Shamir R, Loring JF: **Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture.** *Cell Stem Cell* 2011, **8:**106-118.

64. Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Närvä E, Ng S, Sourour M, Hämäläinen R, Olsson C, Lundin K, Mikkola M, Trokovic R, Peitz M, Brüstle O, Bazett-Jones DP, Alitalo K, Lahesmaa R, Nagy A, Otonkoski T: **Copy number variation and selection during reprogramming to pluripotency.** *Nature* 2011, **471:**58-62.

65. Gore A, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J, Canto I, Giorgetti A, Israel MA, Kiskinis E, Lee JH, Loh YH, Manos PD, Montserrat N, Panopoulos AD, Ruiz S, Wilbert ML, Yu J, Kirkness EF, Izpisua Belmonte JC, Rossi DJ, Thomson JA, Eggan K, Daley GQ, Goldstein LS, Zhang K: **Somatic coding mutations in human induced pluripotent stem cells.** *Nature* 2011, **471:**63-67.

66. Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, Plath K, Lowry WE, Benvenisty N: **Identification and classification of chromosomal aberrations in human induced pluripotent stem cells.** *Cell Stem Cell* 2010, **7:**521-531.

67. Maitra A, Arking DE, Shivapurkar N, Ikeda M, Stastny V, Kassauei K, Sui G, Cutler DJ, Liu Y, Brimble SN, Noaksson K, Hyllner J, Schulz TC, Zeng X, Freed WJ, Crook J, Abraham S, Colman A, Sartipy P, Matsui S, Carpenter M, Gazdar AF, Rao M, Chakravarti A: **Genomic alterations in cultured human embryonic stem cells.** *Nat Genet* 2005, **37:**1099-1103.

68. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR: **Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.** *Nature* 2011, **471:**68-73.

69. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI, Yabuuchi A, Takeuchi A, Cunniff KC, Hongguang H, McKinney-Freeman S, Naveiras O, Yoon TJ, Irizarry RA, Jung N, Seita J, Hanna J, Murakami P, Jaenisch R, Weissleder R, Orkin SH, Weissman IL, Feinberg AP, Daley GQ: **Epigenetic memory in induced pluripotent stem cells.** *Nature* 2010, **467:**285-290.

70. Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ, Yu J, Hebrok M, Hochedlinger K, Costello JF, Song JS, Ramalho-Santos M: **Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells.** *Nat Cell Biol* 2011, **13:**541-549.

71. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggan K, Meissner A: **Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines.** *Cell* 2011, **144:**439-452.

72. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA: **An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.** *Nat Genet* 2008, **40:**499-507.

73. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, *et al.*: **The landscape of somatic copy-number alteration across human cancers.** *Nature* 2010, **463:**899-905.

74. Viswanathan SR, Daley GQ: **Lin28: a microRNA regulator with a macro role.** *Cell* 2010, **140:**445-449.

75. Chia NY, Chan YS, Feng B, Lu X, Orlov YL, Moreau D, Kumar P, Yang L, Jiang J, Lau MS, Huss M, Soh BS, Kraus P, Li P, Lufkin 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.

76. Puc J, Rosenfeld MG: **SOX2 and CHD7 cooperatively regulate human disease genes.** *Nat Genet* 2011, **43:**505-506.

77. Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T, Tesar P, Wei CL, Scacheri PC: **CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression.** *PLoS Genet* 2010, **6:**e1001023.

78. Engelen E, Akinci U, Bryne JC, Hou J, Gontan C, Moen M, Szumska D, Kockx C, van Ijcken W, Dekkers DH, Demmers J, Rijkers EJ, Bhattacharya S, Philipsen S, Pevny LH, Grosveld FG, Rottier RJ, Lenhard B, Poot RA: **Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes.** *Nat Genet* 2011, **43:**607-611.

79. Bajpai R, Chen DA, Rada-Iglesias A, Zhang J, Xiong Y, Helms J, Chang CP, Zhao Y, Swigut T, Wysocka J: **CHD7 cooperates with PBAF to control multipotent neural crest formation.** *Nature* 2010, **463:**958-962.

80. Liu GH, Barkho BZ, Ruiz S, Diep D, Qu J, Yang SL, Panopoulos AD, Suzuki K, Kurian L, Walsh C, Thompson J, Boue S, Fung HL, Sancho-Martinez I, Zhang K, Yates J 3rd, Izpisua Belmonte JC: **Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome.** *Nature* 2011, **472:**221-225.

81. Yazawa M, Hsueh B, Jia X, Pasca AM, Bernstein JA, Hallmayer J, Dolmetsch RE: **Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome.** *Nature* 2011, **471:**230-234.

82. Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, Li Y, Mu Y, Chen G, Yu D, McCarthy S, Sebat J, Gage FH: **Modelling schizophrenia using human induced pluripotent stem cells.** *Nature* 2011, **473:**221-225.

83. Qiang L, Fujita R, Yamashita T, Angulo S, Rhinn H, Rhee D, Doege C, Chau L, Aubry L, Vanti WB, Moreno H, Abeliovich A: **Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons.** *Cell* 2011, **146:**359-371.

84. Zwaka TP, Thomson JA: **Homologous recombination in human embryonic stem cells.** *Nat Biotechnol* 2003, **21:**319-321.

85. Buecker C, Chen HH, Polo JM, Daheron L, Bu L, Barakat TS, Okwieka P, Porter A, Gribnau J, Hochedlinger K, Geijsen N: **A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells.** *Cell Stem Cell* 2010, **6:**535-546.

86. Soldner F, Laganière 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, Urnov FD, Rebar EJ, Gregory PD, Zhang HS, Jaenisch R: **Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations.** *Cell* 2011, **146:**318-331.

87. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R: **Genetic engineering of human pluripotent cells using TALE nucleases.** *Nat Biotechnol* 2011, **29:**731-734.

88. Zhao T, Zhang ZN, Rong Z, Xu Y: **Immunogenicity of induced pluripotent stem cells.** *Nature* 2011, **474:**212-215.

89. Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ: **Natural innate and adaptive immunity to cancer.** *Annu Rev Immunol* 2011, **29:**235-271.

90. Plath K, Lowry WE: **Progress in understanding reprogramming to the induced pluripotent state.** *Nat Rev Genet* 2011, **12:**253-265.

91. Smith ZD, Nachman I, Regev A, Meissner A: **Dynamic single-cell imaging of direct reprogramming reveals an early specifying event.** *Nat Biotechnol* 2010, **28:**521-526.

92. Skarnes WC, Rosen B, West AP, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A: **A conditional knockout resource for the genome-wide study of mouse gene function.** *Nature* 2011, **474:**337-342.

93. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, Linnarsson S: **Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq.** *Genome Res* 2011, **21:**1160-1167.

94. Yeo GW, Xu X, Liang TY, Muotri AR, Carson CT, Coufal NG, Gage FH: **Alternative splicing events identified in human embryonic stem cells and neural progenitors.** *PLoS Comput Biol* 2007, **3:**1951-1967.

95. Wu JQ, Habegger L, Noisa P, Szekely A, Qiu C, Hutchison S, Raha D, Egholm M, Lin H, Weissman S, Cui W, Gerstein M, Snyder M: **Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by**

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 JC: **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 IH, Kim KS, Daley GQ, Kornblum HI, Shraiman BI, Kosik KS: **MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states.** *Cell Stem Cell* 2010, **7**:671-681.

101. Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K: **Role of the murine reprogramming factors in the induction of pluripotency.** *Cell* 2009, **136**:364-377.

102. Mäkinen N, Mehine M, Tolvanen J, Kaasinen E, Li Y, Lehtonen HJ, 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.