Genome **Medicine**

REVIEW

Embryonic stem cell-specific signatures in cancer: insights into genomic regulatory networks and implications for medicine

Jonghwan Kim[1] and Stuart H Orkin[2]*

Abstract

Embryonic stem (ES) cells are of great interest as a model system for studying early developmental processes and because of their potential therapeutic applications in regenerative medicine. Obtaining a systematic understanding of the mechanisms that control the 'stemness' - self-renewal and pluripotency - of ES cells relies on high-throughput tools to define gene expression and regulatory networks at the genome level. Such recently developed systems biology approaches have revealed highly interconnected networks in which multiple regulatory factors act in combination. Interestingly, stem cells and cancer cells share some properties, notably self-renewal and a block in differentiation. Recently, several groups reported that expression signatures that are specific to ES cells are also found in many human cancers and in mouse cancer models, suggesting that these shared features might inform new approaches for cancer therapy. Here, we briefly summarize the key transcriptional regulators that contribute to the pluripotency of ES cells, the factors that account for the common gene expression patterns of ES and cancer cells, and the implications of these observations for future clinical applications.

Embryonic stem cells, cancer and genomic regulation

Embryonic stem (ES) cells are cultured cells derived from the inner cell mass of the blastocyst-stage embryo [1,2]. They exhibit two distinct properties: self-renewal, the

*Correspondence: stuart_orkin@dfci.harvard.edu
[2]Department of Pediatric Oncology, Children's Hospital and Dana Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School, Howard Hughes Medical Institute, Boston, MA 02115, USA
Full list of author information is available at the end of the article

BioMed Central

ability to maintain a proliferative state without changes in cellular characteristics; and pluripotency, the capacity to generate all of the cell types of adult organisms. Understanding how these properties are established and maintained is crucial to realizing the full potential of ES cells in basic biology and regenerative medicine.

Previously, a small cadre of transcription factors, including the homeodomain protein Oct4 (Pou5f1), SRY box-containing factor Sox2, and Nanog, were identified as key regulatory factors (or ES cell core factors) in controlling ES cell pluripotency [3-6]. Remarkably, Yamanaka and colleagues [7,8] observed that somatic cells can be reprogrammed into ES cell-like cells (induced pluripotent stem (iPS) cells) by the introduction of four transcription factors: Oct4, Sox2, Klf4, and Myc [7-11]. This observation clearly underscores the relevance of transcriptional regulatory mechanisms to pluripotency and cell fate control [12].

During the past decade, advances in high-throughput technologies, such as gene expression profiling, the global mapping of transcription factor-DNA interactions and histone modifications by microarrays or sequencing (chromatin immunoprecipitation (ChIP)-chip or ChIP-sequencing) [13], the mapping of protein-protein interactions, the identification of members of protein complexes by affinity purification followed by mass spectrometry (MS) [14], and the unbiased knockdown of genes by RNA interference (RNAi) [15], have facilitated the assembly of considerable databases of proteomic and genomic information. These new tools provide the basis for the development of a comprehensive understanding of cell states at the systems level and have been applied to dissect self-renewal and pluripotency control in ES cells, reprogramming processes, and lineage specification [16,17].

In the context of cancer biology, an important goal has been delineation of the cells that sustain cancers. Investigators have suggested that a small population of cells within a tumor may reinitiate tumor formation upon transplantation and be responsible for the maintenance of tumors and their resistance against effective

anti-cancer therapy. Such cancer stem cells, or more precisely tumor initiating cells, might arise from adult stem, or progenitor, cells or from the dedifferentiation of somatic cells [18]. It has been hypothesized that the similarities shared by stem cells and cancer cells might relate to shared patterns of gene expression regulation, which might be associated with the 'embryonic' state. Moreover, recent studies focusing on somatic cell reprogramming underscore the similarity between cancer cells and iPS cells. The acquisition of pluripotency during the reprogramming process is superficially reminiscent of the dedifferentiation proposed for some cancers [19]. In trying to account for the self-renewing properties of cancer stem cells, several investigators have defined 'ES-cell-specific expression' signatures, and these have been analyzed in diverse cancers [20-26].

In this review, we provide an overview of the current understanding of the ES-cell-specific gene expression programs that have been observed in various human cancers. We first summarize the key regulatory factors involved in controlling the self-renewal and pluripotency of ES cells, which have been thoroughly evaluated using various systems biology tools. We then discuss how these factors have contributed to our understanding of the gene expression signatures that are shared between ES cells and cancer cells. Finally, we discuss the implications of these observations for medicine.

Regulatory factors in self-renewal and pluripotency

In this section, we provide a brief overview of the key factors that regulate the self-renewal and pluripotency of ES cells, and the acquisition of pluripotency during somatic cell reprogramming. Recently, genome-scale technologies and systems-level approaches have been widely applied to investigate regulatory mechanisms in ES and iPS cells. The key regulators in pluripotent stem cells, their functions, and the experimental methods applied to investigate them are summarized in Table 1.

Core transcription factors

Initially, a few transcription factors that are critical to ES cell pluripotency, core factors Oct4, Sox2, and Nanog, were identified and functionally characterized by low-throughput methods [3-6]. Subsequently, global targets of these core factors have been identified in mouse ES cells using ChIP combined with paired-end-tag-based sequencing methods (ChIP-PET) [27] and in human ES cells using ChIP-chip [28]. The results suggested that each of the key transcription factors has numerous (>1,000) chromosomal targets, and that the factors are auto-regulated and subject to cross-regulation in an inter-connected network. A Nanog-centered map of protein-protein interactions in ES cells has also been constructed

Table 1. Genome-scale studies of self-renewal and pluripotency in ES cells

Regulators	Function	Methods	Reference(s)
Core factors			
Oct4 (Pou5f1)	ES cell core factor	ChIP, MS	[27-32,44]
Sox2	ES cell core factor	ChIP, MS	[28,32,44]
Nanog	ES cell core factor	ChIP, MS	[27-29,32,44]
Tcf3 (Tcf7l1)	Wnt signaling	ChIP	[74]
Klf4	LIF signaling	ChIP	[32,44]
Stat3	LIF signaling	ChIP	[44]
Dax1 (Nr0b1)	Negative regulation of transcription	ChIP, MS	[29,32]
Sall4	Self-renewal and pluripotency	ChIP, MS	[29,75-77]
Polycomb-related factors			
Ezh2	PRC2, repressor	ChIP, MS	[78,79]
Jarid2	Fine-tuning of PRC2	ChIP, MS	[78,79]
Mtf2	Polycomb-like protein	ChIP, MS	[78]
Suz12	PRC2, repressor	ChIP, MS	[37,38,78,79]
Eed	PRC2, repressor	ChIP	[37,38]
Rnf2	PRC1, repressor	ChIP	[37]
Phc1	PRC1, repressor	ChIP	[37]
Myc-related factors			
Myc	Proliferation	ChIP, MS	[20,44]
Max	Myc-interacting	ChIP, MS	[20]
Zfx	Self-renewal	ChIP	[44]
Trrap	Histone acetylation	RNAi	[50]
Tip60 (Kat5)	Histone acetylation	ChIP, MS, RNAi	[20,50]
Ep400	Histone acetylation	MS, RNAi	[50]
Dmap1	Histone acetylation	ChIP, MS, RNAi	[20,50]
E2F1	Regulator of cell cycle	ChIP	[44]
E2F4	Transcription activator	ChIP, MS	[20]
Cnot3	General transcription regulator	ChIP, RNAi	[80]
Trim28 (Tif1b)	Transcription co-activator	ChIP, RNAi	[80]

LIF, leukemia inhibitory factor.

using affinity purification followed by MS [29]. With the addition of the more recent Oct4-centered protein-protein interaction maps [30,31], these approaches expanded the initial ES cell core network by identifying novel interacting partners of the core factors. Using a ChIP-based method, subsequent mapping of chromosomal targets of the nine transcription factors within this expanded core network (that is, three core factors, Nanog-interacting proteins, and Yamanaka's four somatic-cell-reprogramming factors) revealed a positive correlation between transcription factor co-occupancy and target gene activity [32]. These results also provided an initial

glimpse into the unique roles of Myc in ES cells and somatic cell reprogramming. Myc has more target genes than any of the core factors, and its target genes show unique histone modification marks in their promoters.

Somatic cell reprogramming by defined factors

In the first report of somatic cell reprogramming by Yamanaka's group, mouse fibroblasts, which represent terminally differentiated cells, were reprogrammed to become pluripotent-stem-cell-like cells (iPS cells) by the introduction of four transcription factors: two core ES cell factors (Oct4 and Sox2), Klf4 and c-Myc (Myc) [7]. Successful reprogramming of human fibroblasts to iPS cells [8,10,11], together with the generation of disease-specific iPS cell lines using the cells of people with genetic disorders, provides a basis for *in vitro* culture-based studies of human disease phenotypes [33,34]. Notably, as shown by Yamanaka's initial work, the four reprogramming factors are highly expressed in ES cells. Additionally, these reprogramming factors are implicated in tumorigenesis in diverse cancer contexts [19,35]. These observations raise the hypothesis that somatic cell reprogramming, pluripotency control in ES cells, and cellular transformation might share common pathways.

Polycomb-related factors

Polycomb-group (PcG) proteins, which were first discovered in fruit flies, contribute to the repressed state of crucial developmental or lineage-specific regulators by generating a repressive histone mark. PcG proteins have essential roles in early development, as well as in ES cells [36]. Mapping of the targets of PcG-repressive protein complex (PRC)1 and PRC2 in mouse and human ES cells by ChIP-chip showed that PRC proteins occupy many common repressed target genes, including lineage-specific transcription factors [37,38]. These studies suggest that PRC proteins serve to maintain the undifferentiated state of ES cells by repressing important developmental regulators. Recent experiments involving RNA immunoprecipitation followed by sequencing (RIP-sequencing) implicate the interaction of various non-coding RNA molecules with the PRC complex in the regulation of target genes [39]. PRC proteins are also implicated in the somatic cell reprogramming process [40,41].

Myc and Myc-interacting factors

Activation of Myc, one of the most-studied oncogenes, is reported in up to 70% of human cancers [42]. Myc has numerous cellular functions and is involved in many biological pathways, including the control of self-renewal in ES cells [43]. Mapping of Myc targets in ES cells has suggested that Myc's role in maintaining the pluripotency of ES cells is distinct from that of the core factors [32,44]. Myc has many more chromatin targets than the core ES

factors, and Myc target genes are enriched in pathways that are associated with metabolism and protein synthesis. By contrast, the targets of the core factors are involved in transcription and developmental processes [32,44]. In the context of somatic cell reprogramming, Myc is a dispensable factor [45,46]; but efficient and rapid reprogramming by Myc suggests that this factor might generate a favorable environment during the reprogramming process, potentially by mediating the global alteration of chromosome structure [47-49]. Recently, Myc-interacting partner proteins and their genomic targets have been identified in ES cells [20]. These studies revealed that the Myc network is distinct from the ES cell core interaction network or the PRC network. Interestingly, an independent RNAi-based knockdown screen showed that Tip60-p400 histone acetyltransferase (HAT) complex proteins, which interact with Myc in ES cells [20], also play a crucial role in ES cell identity [50], implicating the functions of Myc-interacting proteins in the control of ES cell pluripotency and somatic cell reprogramming.

Common signatures in ES cells and cancer

Overlapping characteristics that are shared by ES cells and cancer cells have led investigators to examine the gene expression patterns that underlie these similarities [18]. We now know that one of the factors used to facilitate somatic cell reprogramming, Myc, is an established oncogene, and that the inactivation of p53 pathways, as observed in innumerable cancers, increases the efficiency of the reprogramming process [7,51-54]. These discoveries provide additional evidence that common pathways could be utilized both in the acquisition of pluripotency and in tumorigenesis. In this regard, data generated from various systems biology tools that can be used to dissect ES cell pluripotency and somatic cell reprogramming could play a crucial role in identifying the common features shared by ES cells and cancer cells. In turn, many ES-cell-specific gene sets, modules, or signatures that have been identified by systems biology studies of pluripotent stem cells have provided useful analytical tools for analyses of the gene-expression programs of human tumors and mouse tumor models. Recent analyses of ES-cell-specific signatures in human tumors are summarized in Table 2.

ES cell signatures tested in cancer

In one of the first studies aimed at revealing shared gene expression patterns, Chang and associates [22] collected large-scale data sets that had been acquired from ES cells or adult stem cells, and constructed a gene-module map. From the initial gene-module map, two modules (gene sets) that distinguish ES cells (the ESC-like module) and adult stem cells (the adult tissue stem cell module) were

Table 2. Studies of embryonic stem cell signatures in cancer

Study	Gene sets used in the study	Gene set generated by:	Tested cancers
Ben-Porath *et al.* [23]	ES cell expression profiles, Nanog, Oct4, Sox2 targets, Myc targets, and PRC targets	ES-cell-specific gene expression, and factor occupancy	Breast, glioma, and bladder cancers
Wong *et al.* [22]	ES-cell-like module, adult stem cell module	Gene module map [81]	Liver, breast, prostate, gastric, and lung cancers
Schoenhals *et al.* [26]	Nanog, Oct4, Sox2, and Klf4	ES-cell-specific gene activity in multiple cancers	Multiple cancers
Kim *et al.* [20]	Core module, PRC module, and Myc module	Factor co-occupancy within transcription sub-networks	Myeloid/lymphoid leukemia, bladder and breast cancers
Mizuno *et al.* [25]	ES cell, iPS cell, PRC2, and p53 ES cell signatures	Gene expression profiles	Breast cancer
Shats *et al.* [24]	CSR signature	Combined gene sets	Breast and lung cancers, adenocarcinoma, and medulloblastoma

defined. The activities of these two modules were tested using gene expression data sets from various human tumor samples (Table 2). Chang's group observed that the ESC-like module is activated in various human epithelial cancers. Moreover, they showed that Myc activates the ESC-like module in epithelial cells. Taking these observations together, the group proposed that the activation of an ES-cell-like transcriptional program via Myc might induce the characteristics of cancer stem cells in differentiated adult cells. Independently, Weinberg and colleagues [23] defined 13 gene sets in ES cells from previously existing large-scale data sets and placed each of these 13 data sets into one of four categories: ES-expressed, active core factor (Nanog, Oct4, and Sox2) targets, PRC targets, and Myc targets. When these data sets were tested using expression profiling data sets from human cancer patients, the activation of ES-cell-specific gene sets (such as ES-expressed) and the repression of PRC target genes were significantly enriched in poorly differentiated human tumors. A similar approach defined a consensus stemness ranking (CSR) signature from four different stem cell signatures, and also showed that the CSR signature has prognostic power in several human cancer types [24]. Notably, an active ES-cell-like expression program has been observed upon inactivation of p53 in breast and lung cancers [25]. Similar to the function of p53 in the acquisition of pluripotency during reprogramming, the inhibition of p53 or the p53 pathway increases the efficiency of somatic cell reprogramming [53]. Taken together, these studies clearly show that ES-cell-specific signatures are shared among various human cancers and animal cancer models; but the precise nature of the gene expression pathways remains unclear.

Predominant ES cell Myc module in cancer
Although ES cells and cancer cells share some properties, cancer cells do not exhibit true pluripotency like that displayed by ES cells. Furthermore, early studies failed to establish that the crucial ES-cell pluripotency genes were actually expressed in cancer cells and could account for

the apparent similarities between ES cells and cancer cells [55,56]. So how specific are the proposed ES-cell-specific gene modules? Recent findings lead to a more nuanced view of the relationship between ES cells and cancer cells. A Myc-centered regulatory network was first constructed in ES cells by combining the data sets acquired from a MS-based proteomics method as well as a ChIP-based method. When this Myc-centered regulatory network was combined with previously defined ES cell pluripotency, core and PRC networks, it was shown that the transcription regulatory program that controls ES cells can be subdivided into functionally separable regulatory units: core, PRC and Myc [20]. Such ES cell modules were defined on the basis of the target co-occupancy of the factors within the regulatory units. Subsequently, the averaged activity of the three modules (common target genes within each regulatory unit - core, PRC and Myc modules) was tested in ES cells and in various cancer types. In ES cells, the core and Myc modules are active, but the PRC module is repressed. An active Myc module is observed in many cancer types and generally predicts poor prognosis. On the other hand, the core module, which is highly active in ES cells and underlies the ES cell state, is not significantly enriched in most cancers. In contrast to the previous studies, this work suggests that the similar expression signatures of ES cells and cancer cells largely reflect the contribution of the Myc regulatory network rather than that of an ES-cell-specific core network. This conclusion is in accordance with the previous observation that Myc induces an ESC-like module in epithelial cells [22]. Note also that many genes in the previously defined ESC-like modules proposed by others [22,23] are direct target genes of Myc and are therefore likely to reinforce the common signature.

Repressive targets of PRC2 in cancer
PRC complexes (especially PRC2 proteins, including Ezh2, Eed, and Suz12) are important repressors of gene regulation that are highly expressed in ES cells. Their

downstream targets, including many lineage-specific regulators, are repressed or inactive in ES cells [37,38]. Weinberg and associates [23] observed that the target genes of PRC are also repressed in various human cancers, and that the repression of PRC target genes also predicts poorly differentiated human tumors. Interestingly, overexpression of PRC2 proteins is often observed in many different cancers; for example, Ezh2, a catalytic subunit of PRC2, has been reported to be a marker for aggressive prostate and breast tumors [57,58]. In our study of modules within ES cells, we also observed that repression of target genes by PRC is shared between ES cells and cancer cells [20]. These results strongly suggest that, in addition to the Myc network, a PRC network also generates expression signatures that are shared by ES cells and cancer cells.

ES cell core factors in cancer

Do ES cell core factors ever play a crucial role in cancer? For those cancers of germ cell origin, the expression of ES-cell-specific pluripotency factors, such as Oct4 and Nanog, is likely to be functionally relevant [59]. It has been reported that transcripts of Oct4, Nanog, and/or Sox2 may be expressed in epithelial cancers, and that their expression is correlated with tumor grade [26,60,61]. Nevertheless, the subject remains controversial because the expression of pseudogenes for Oct4 has confounded studies based on RNA expression alone [62,63]. Another key factor in ES cells, Sox2, was implicated in lung and esophageal squamous cell carcinomas; but the induction of Sox2 in a lung adenocarcinoma cell line promoted squamous traits rather than pluripotency-related characteristics. This suggests a role for Sox2 as a lineage-survival oncogene rather than as a stemness marker [60]. Our recent work has shown that the core module, which relates to ES cell core factors, is not significantly enriched in human epithelial tumors [20]. Thus, the contribution of ES-cell-specific core factors to tumor formation or maintenance is still uncertain.

Implications for cancer and medicine

The extent to which the study of pluripotent ES cells has provided insights into cancer is remarkable. In addition, the involvement of both oncogenic and tumor suppressor pathways in somatic cell reprogramming suggests that continued study of the relationship between ES cells and cancer cells is worthwhile. In this section, we discuss how ES cells might be used to accelerate the translation of basic findings into clinically relevant tests and new therapeutic approaches.

Classically, cancer cell lines have been employed as convenient biological models when investigating the characteristics of various cancers and as a platform for exploring the activity of chemotherapeutic agents. Cell lines are not usually a preferred platform for drug screening because they often represent highly selected subpopulations of cancer cells, with accumulated genetic mutations or abnormalities acquired during long-term culture. The shared signatures of ES cells and cancer cells suggest, however, that ES cells could provide an alternative system for studying pathways relevant to cancers. One strategy is depicted in Figure 1. In this scenario, genetic and/or chemical modulators that negate or alter the activities of signatures that are shared by ES cells and cancer cells may be sought in ES cells by high-throughput screening. Subsequently, selected modulators could then be re-validated in cancer cells either in culture or in various transplant protocols. A variation of this theme is the recent application of gene expression signatures to identify drugs that target specific signaling pathways (such as those for Ras, Src, and Myc) [64-66].

A particularly powerful approach is now afforded by an elegant *in silico* method based on the 'Connectivity Map' [67,68]. The Connectivity Map encompasses an expanding database of gene expression profiles from a collection of reference cell lines treated with 'perturbagens' [69]. In the original version of the Connectivity Map, cells were treated with numerous drugs, but the approach is entirely general and cells may be 'perturbed' by any chemical or genetic manipulation. In practice, the Connectivity Map database is interrogated with a gene expression signature of interest to ask whether the signature resembles the action of a perturbagen on the reference cells. As the method is performed *in silico*, it is extremely rapid.

An initial attempt to identify drugs that modulate an ES-cell-like gene expression signature has already been reported. In this instance, the Connectivity Map database was interrogated with an ES-cell signature, described as a CSR [24], to predict drugs that affect the CSR signature. Putative 'hits' were subsequently validated in human breast cancer cells. The results revealed multiple topoisomerase inhibitors, including daunorubicin, that decrease cell viability in this context [24]. We anticipate that further interrogation of the Connectivity Map database with other expression signatures could highlight agents that form the basis for novel therapeutic approaches.

Conclusions and future directions

In recent years, the utilization of emerging systems biology techniques in stem cell biology have led to considerable advances in our understanding of the regulatory networks that control the pluripotency of ES cells and the process of somatic cell reprogramming. We began with just a handful of core ES cell transcription factors, but now appreciate a more extensive list of transcription factors that are involved in the regulation of these processes. Cross-examination of large data sets generated by various tools, taken together with
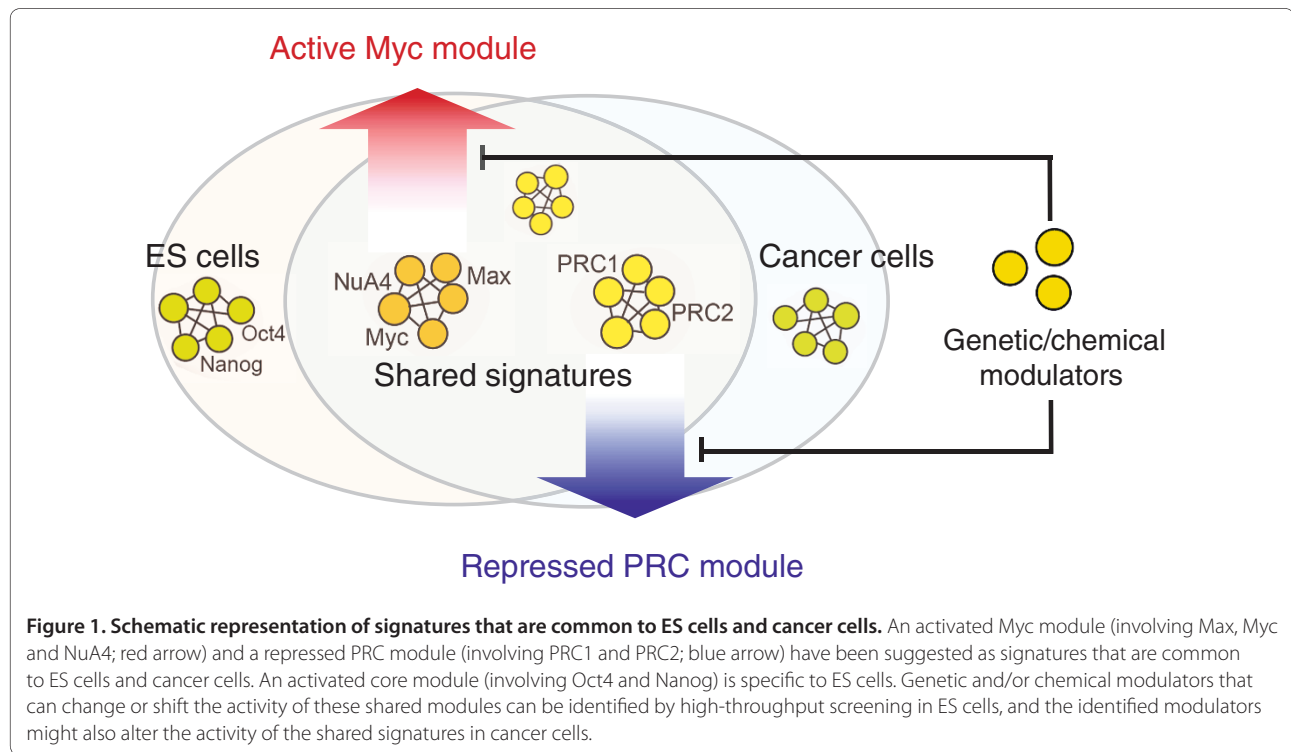
Figure 1. Schematic representation of signatures that are common to ES cells and cancer cells. An activated Myc module (involving Max, Myc and NuA4; red arrow) and a repressed PRC module (involving PRC1 and PRC2; blue arrow) have been suggested as signatures that are common to ES cells and cancer cells. An activated core module (involving Oct4 and Nanog) is specific to ES cells. Genetic and/or chemical modulators that can change or shift the activity of these shared modules can be identified by high-throughput screening in ES cells, and the identified modulators might also alter the activity of the shared signatures in cancer cells.

computational analysis, has led to an improved understanding of the gene-expression patterns that are common to ES and cancer cells. Rather than identifying core ES cell factors as contributors to shared patterns, the recent studies underscore sub-modules that refer to Myc and Polycomb transcriptional activities.

An improved understanding of the features shared by pluripotent cells and cancer cells is of potential clinical relevance. In the future, the common pathways could serve as putative targets for anti-cancer drugs, but unresolved questions remain. Recent studies describe overlapping expression signatures that are shared by ES cells and various human cancers and that also predict patient outcome, but more careful analysis needs to be performed to reveal the multiple contributions to these signatures. The heterogeneity of cancers presents a challenge to the field. Many different cell types reside within a given tumor, and tumors differ from one to another, but current methods deal poorly with cellular heterogeneity. The extent to which core ES cell pluripotency factors are involved in epithelial cancers, or in a subset of cancer stem cells, remains to be explored. If they are expressed, it is relevant to ask whether the genes or gene pathways that are controlled by ES cell core factors in cancer cells are similar to those regulated by these core factors in pluripotent stem cells.

Moreover, additional layers of regulatory mechanisms that await further characterization might be shared between ES cells and cancers. For example, microRNAs, which are crucial regulators of the pluripotent state and cell proliferation [70,71], might have patterns of regulation and downstream target genes that are common to ES and cancers cells. An improved understanding of signaling pathways that are implicated in both ES cells and cancer (or cancer stem cells) [72,73], and their connections to the regulatory networks, is also of special interest. Finally, it will be instructive to determine whether chemicals or genetic modulators could change or shift the activity of common signatures or modules shared between ES and cancer cells. The opportunities provided by these approaches could accelerate the identification and development of new cancer therapies.

Abbreviations

ChIP, chromatin immunoprecipitation; CSR, consensus stemness ranking; ES, embryonic stem; iPS, induced pluripotent stem; MS, mass spectrometry; PcG, Polycomb-group; PRC, PcG-repressive protein complex; RNAi, RNA interference.

Author details

[1]Molecular Cell and Developmental Biology, Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas 78712, USA. [2]Department of Pediatric Oncology, Children's Hospital and Dana Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School, Howard Hughes Medical Institute, Boston, MA 02115, USA.

References

1. Evans MJ, Kaufman MH: Establishment in culture of pluripotential cells from mouse embryos. *Nature* 1981, **292**:154-156.
2. Martin GR: Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 1981, **78**:7634-7638.
3. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S: The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003, **113**:631-642.
4. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A: Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003, **113**:643-655.
5. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R: Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003, **17**:126-140.
6. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A: Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998, **95**:379-391.
7. Takahashi K, Yamanaka S: Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006, **126**:663-676.
8. 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.
9. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R: *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 2007, **448**:318-324.
10. 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.
11. 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.
12. Graf T, Enver T: Forcing cells to change lineages. *Nature* 2009, **462**:587-594.
13. Hawkins RD, Hon GC, Ren B: Next-generation genomics: an integrative approach. *Nat Rev Genet* 2010, **11**:476-486.
14. Gavin AC, Maeda K, Kuhner S: Recent advances in charting protein-protein interaction: mass spectrometry-based approaches. *Curr Opin Biotechnol* 2011, **22**:42-49.
15. Dykxhoorn DM, Lieberman J: The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu Rev Med* 2005, **56**:401-423.
16. Roeder I, Radtke F: Stem cell biology meets systems biology. *Development* 2009, **136**:3525-3530.
17. Macarthur BD, Ma'ayan A, Lemischka IR: Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* 2009, **10**:672-681.
18. Reya T, Morrison SJ, Clarke MF, Weissman IL: Stem cells, cancer, and cancer stem cells. *Nature* 2001, **414**:105-111.
19. Daley GQ: Common themes of dedifferentiation in somatic cell reprogramming and cancer. *Cold Spring Harb Symp Quant Biol* 2008, **73**:171-174.
20. 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.
21. Somervaille TC, Matheny CJ, Spencer GJ, Iwasaki M, Rinn JL, Witten DM, Chang HY, Shurtleff SA, Downing JR, Cleary ML: Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell* 2009, **4**:129-140.
22. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY: Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* 2008, **2**:333-344.
23. 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.
24. Shats I, Gatza ML, Chang JT, Mori S, Wang J, Rich J, Nevins JR: Using a stem
25. cell-based signature to guide therapeutic selection in cancer. *Cancer Res* 2011, **71**:1772-1780.
25. Mizuno H, Spike BT, Wahl GM, Levine AJ: Inactivation of p53 in breast cancers correlates with stem cell transcriptional signatures. *Proc Natl Acad Sci U S A* 2010, **107**:22745-22750.
26. Schoenhals M, Kassambara A, De Vos J, Hose D, Moreaux J, Klein B: Embryonic stem cell markers expression in cancers. *Biochem Biophys Res Commun* 2009, **383**:157-162.
27. 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.
28. 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.
29. Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH: A protein interaction network for pluripotency of embryonic stem cells. *Nature* 2006, **444**:364-368.
30. 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.
31. 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.
32. 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.
33. Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW, Cowan C, Hochedlinger K, Daley GQ: Disease-specific induced pluripotent stem cells. *Cell* 2008, **134**:877-886.
34. Dimos JT, Rodolfa KT, Niakan KK, Weisenthal LM, Mitsumoto H, Chung W, Croft GF, Saphier G, Leibel R, Goland R, Wichterle H, Henderson CE, Eggan K: Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* 2008, **321**:1218-1221.
35. Werbowetski-Ogilvie TE, Bhatia M: Pluripotent human stem cell lines: what we can learn about cancer initiation. *Trends Mol Med* 2008, **14**:323-332.
36. Margueron R, Reinberg D: The Polycomb complex PRC2 and its mark in life. *Nature* 2011, **469**:343-349.
37. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R: Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 2006, **441**:349-353.
38. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA: Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 2006, **125**:301-313.
39. 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.
40. Zhang Z, Jones A, Sun CW, Li C, Chang CW, Joo HY, Dai Q, Mysliwiec MR, Wu LC, Guo Y, Yang W, Liu K, Pawlik KM, Erdjument-Bromage H, Tempst P, Lee Y, Min J, Townes TM, Wang H: PRC2 complexes with JARID2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem Cells* 2011 **29**:229-240.
41. Pereira CF, Piccolo FM, Tsubouchi T, Sauer S, Ryan NK, Bruno L, Landeira D, Santos J, Banito A, Gil J, Koseki H, Merkenschlager M, Fisher AG: ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell* 2010, **6**:547-556.
42. Nilsson JA, Cleveland JL: Myc pathways provoking cell suicide and cancer. *Oncogene* 2003, **22**:9007-9021.
43. Meyer N, Penn LZ: Reflecting on 25 years with MYC. *Nat Rev Cancer* 2008, **8**:976-990.
44. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang L, 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.
45. Wernig M, Meissner A, Cassady JP, Jaenisch R: c-Myc is dispensable for direct

reprogramming of mouse fibroblasts. *Cell Stem Cell* 2008, **2**:10-12.

46. 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.

47. Knoepfler PS: Why myc? An unexpected ingredient in the stem cell cocktail. *Cell Stem Cell* 2008, **2**:18-21.

48. Knoepfler PS, Zhang XY, Cheng PF, Gafken PR, McMahon SB, Eisenman RN: Myc influences global chromatin structure. *EMBO J* 2006, **25**:2723-2734.

49. 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.

50. Fazzio TG, Huff JT, Panning B: An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 2008, **134**:162-174.

51. Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, Blasco MA, Serrano M: The *Ink4/Arf* locus is a barrier for iPS cell reprogramming. *Nature* 2009, **460**:1136-1139.

52. Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R: Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 2009, **462**:595-601.

53. Krizhanovsky V, Lowe SW: Stem cells: the promises and perils of p53. *Nature* 2009, **460**:1085-1086.

54. Utikal J, Polo JM, Stadtfeld M, Maherali N, Kulalert W, Walsh RM, Khalil A, Rheinwald JG, Hochedlinger K: Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* 2009, **460**:1145-1148.

55. Fortunel NO, Otu HH, Ng HH, Chen J, Mu X, Chevassut T, Li X, Joseph M, Bailey C, Hatzfeld JA, Hatzfeld A, Usta F, Vega VB, Long PM, Libermann TA, Lim B: Comment on "'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* 2003, **302**:393; author reply 393.

56. Evsikov AV, Solter D: Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* 2003, **302**:393; author reply 393.

57. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RG, Otte AP, Hayes DF, Sabel MS, Livant D, Weiss SJ, Rubin MA, Chinnaiyan AM: EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 2003, **100**:11606-11611.

58. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, Rubin MA, Chinnaiyan AM: The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002, **419**:624-629.

59. Clark AT: The stem cell identity of testicular cancer. *Stem Cell Rev* 2007, **3**:49-59.

60. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, Kim SY, Wardwell L, Tamayo P, Gat-Viks I, Ramos AH, Woo MS, Weir BA, Getz G, Beroukhim R, O'Kelly M, Dutt A, Rozenblatt-Rosen O, Dziunycz P, Komisarof J, Chirieac LR, Lafargue CJ, Scheble V, Wilbertz T, Ma C, Rao S, Nakagawa H, Stairs DB, Lin L, Giordano TJ, *et al.*: SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* 2009, **41**:1238-1242.

61. Gidekel S, Pizov G, Bergman Y, Pikarsky E: Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell* 2003, **4**:361-370.

62. Zhao S, Yuan Q, Hao H, Guo Y, Liu S, Zhang Y, Wang J, Liu H, Wang F, Liu K, Ling EA, Hao A: Expression of OCT4 pseudogenes in human tumours: lessons from glioma and breast carcinoma. *J Pathol* 2011, **223**:672-682.

63. Suo G, Han J, Wang X, Zhang J, Zhao Y, Dai J: Oct4 pseudogenes are transcribed in cancers. *Biochem Biophys Res Commun* 2005, **337**:1047-1051.

64. Downward J: Cancer biology: signatures guide drug choice. *Nature* 2006, **439**:274-275.

65. Solit DB, Garraway LA, Pratilas CA, Sawai A, Getz G, Basso A, Ye Q, Lobo JM, She Y, Osman I, Golub TR, Sebolt-Leopold J, Sellers WR, Rosen N: BRAF mutation predicts sensitivity to MEK inhibition. *Nature* 2006, **439**:358-362.

66. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA Jr, Marks JR, Dressman HK, West M, Nevins JR: Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006, **439**:353-357.

67. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR: The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006, **313**:1929-1935.

68. Lamb J: The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer* 2007, **7**:54-60.

69. Hieronymus H, Lamb J, Ross KN, Peng XP, Clement C, Rodina A, Nieto M, Du J, Stegmaier K, Raj SM, Maloney KN, Clardy J, Hahn WC, Chiosis G, Golub TR: Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 2006, 10:321-330.

70. Martinez NJ, Gregory RI: MicroRNA gene regulatory pathways in the establishment and maintenance of ESC identity. *Cell Stem Cell* 2010, **7**:31-35.

71. 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.

72. Taipale J, Beachy PA: The Hedgehog and Wnt signalling pathways in cancer. *Nature* 2001, **411**:349-354.

73. Takebe N, Harris PJ, Warren RQ, Ivy SP: Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nat Rev Clin Oncol* 2011, **8**:97-106.

74. 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.

75. Yang J, Chai L, Fowles TC, Alipio Z, Xu D, Fink LM, Ward DC, Ma Y: Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. *Proc Natl Acad Sci U S A* 2008, **105**:19756-19761.

76. Rao S, Zhen S, Roumiantsev S, McDonald LT, Yuan GC, Orkin SH: Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. *Mol Cell Biol* 2010, **30**:5364-5380.

77. Lim CY, Tam WL, Zhang J, Ang HS, Jia H, Lipovich L, Ng HH, Wei CL, Sung WK, Robson P, Yang H, Lim B: Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. *Cell Stem Cell* 2008, **3**:543-554.

78. 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.

79. 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.

80. Hu G, Kim J, Xu Q, Leng Y, Orkin SH, Elledge SJ: A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* 2009, **23**:837-848.

81. Segal E, Friedman N, Koller D, Regev A: A module map showing conditional activity of expression modules in cancer. *Nat Genet* 2004, **36**:1090-1098.