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

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:10579360

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Embryonic stem cell-specific signatures in cancer: insights into genomic regulatory networks and implications for medicine

Jonghwan Kim and Stuart H Orkin*

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 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 reinstate 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 re-programming underscore the similarity between cancer cells and iPSCs. 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 iPSCs. The key regulators in pluripotent stem cells, their functions, and the experimental methods applied to investigate them are summarized in Table 1.

<table>
<thead>
<tr>
<th>Core transcription factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occt4 (Pou5f1)</td>
</tr>
<tr>
<td>Sox2</td>
</tr>
<tr>
<td>Nanog</td>
</tr>
<tr>
<td>Tcf3 (Tcf7l1)</td>
</tr>
<tr>
<td>Klf4</td>
</tr>
<tr>
<td>Stat3</td>
</tr>
<tr>
<td>Dax1 (Nr0b1)</td>
</tr>
<tr>
<td>Sall4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Polycomb-related factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>EzH2</td>
</tr>
<tr>
<td>Jarid2</td>
</tr>
<tr>
<td>Mrtf2</td>
</tr>
<tr>
<td>Suz12</td>
</tr>
<tr>
<td>Eed</td>
</tr>
<tr>
<td>Rnf2</td>
</tr>
<tr>
<td>Phc1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Myc-related factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myc</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>Zfx</td>
</tr>
<tr>
<td>Trap</td>
</tr>
<tr>
<td>Tip60 (Kat5)</td>
</tr>
<tr>
<td>Ep400</td>
</tr>
<tr>
<td>Dmap1</td>
</tr>
<tr>
<td>E2F1</td>
</tr>
<tr>
<td>E2F4</td>
</tr>
<tr>
<td>Cnot3</td>
</tr>
</tbody>
</table>

| Trim28 (Tif1b) | Transcription co-activator | ChIP, RNAi | [80] |

LIF, leukemia inhibitory factor.

By 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 (PRC1) 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

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

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 ESC-like 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
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; iPSc, induced pluripotent stem; MS, mass spectrometry; PcG, Polycomb-group; PRC, PcG-repressive protein complex; RNAi, RNA interference.

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

Acknowledgements
We thank Andrew J. Woo for critical reading of the manuscript. SHO is an investigator at the Howard Hughes Medical Institute. JK is supported by the National Institutes of Health (NIH)/National Institute of General Medical Sciences (NIGMS) grant K99/R00GM088384.

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.


