[Cell Cycle 8:1, 43-48; 1 January 2009]; ©2009 Landes Bioscience

Review

Stem cells do play with dice

A statistical physics view of transcription

Sol Efroni,[1] Shai Melcer,[2] Malka Nissim-Rafinia[2] and Eran Meshorer[2,*]

[1]The Faculty of Life Sciences; Bar-Ilan University; Ramat-Gan Israel; [2]Department of Genetics; The institute of Life Sciences; The Hebrew University of Jerusalem; Jerusalem Israel

Key words: embryonic stem cells, stem cells, transcription, chromatin, microarrays, genome-wide, systems biology

Embryonic stem cells display wide-spread pervasive transcriptional output. Here, we propose that multiple simultaneous transcriptional states underlay pluripotency.

Introduction

The first half of the 20th century in physics was marked with a live debate about the fate and origin of everything.[1] The deterministic view regarded fate and finite states as well defined, while the other, probabilistic view, considered chance as an important determinate. The debate, and the era in physics, became well known with Einstein's famous quote "God does not play dice with the universe",[2] where he tried to dismiss the "quantum" claim, in which a physical state is a probabilistic one. As years went by, the quantum, statistical, view took over as the major (some would say only) paradigm in physics.

The stem cell transcriptome portrays a metaphorically similar setting. The collective efforts to unequivocally characterize the stem cell molecular signature have so far produced a wide array of different profiles, with disappointing overlap. Thus, the deterministic view, regarding stem cells as well defined entities, seems to be experimentally unsuccessful. A non deterministic view, on the other hand, perceiving cells as a collection of different states, may be able to settle disconcerted experimental findings. According to this view, the stem cell transcriptome would be expected to comprise of multiple such layers of transcription, resulting in promiscuous, pervasive transcriptional output across the genome.[3,4]

Over recent years, this molecular characteristic has become of special importance in cancer research, with the discovery of a unique population in tumors, termed "tumor initiating cells" (or 'cancer stem cells'), with an enhanced ability to initiate tumors, far superior to any other subpopulation.[5] The capacity to characterize tumor initiating cells may provide the clinician with an ability to target disease causing cells without the toxicity associated with affecting other cell populations. Since embryonic stem cells (ESCs), similar to 'cancer stem cells' possess an inherent ability to form teratomas when transplanted into severe combined immunodeficient (SCID) mice, there are tight relationships between an ability to characterize stem cells and the ability to characterize tumor initiating cells and a successful signature of the phenotypes may be carried quickly into the clinic.

However, embryonic stem cells, unlike cancer stem cells or adult, tissue-specific, stem cells define an 'ex vivo' cultured state. The in vivo parallel—the inner cell mass (ICM) of the developing embryo at the blastocyst stage from which ESCs are derived—is extremely transient during development, and these cells lose their pluripotency during implantation and proceed to define specific germ layers during gastrulation. It therefore remains to be determined whether the molecular signature of ESCs is equivalent to that of the ICM. Regardless, the growing interest in ESCs in both the lab and the clinic and their enormous potential for regenerative medicine necessitates understanding their molecular layer.

In Search of a Unifying Signature—"Stemness"

Whole transcriptome assays, combined with the accessibility of ESC lines, gave the means to profile gene expression differences between stem cells and their differentiated progeny. Over the past few years, such studies led to a series of papers that gave the first clues of transcriptional "stemness", a molecular signature that would allow the community a focus on the basic building blocks of ESCs, e.g., pluripotency and self renewal. First in the current surge of papers were studies published in 2002, which demonstrated two convincing molecular signatures. The first one included a 216 gene signature, where the authors intersected lists of expressed genes from embryonic, hematopoietic and neural (murine) stem cells.[6] The signature was composed of the minimal set of probes with expression in all three stem cell groups. The authors further showed non-trivial chromosomal mapping of subsets of the signature genes. In the second study,[7] the authors found shared genes between the human and murine gene expression profiles, and determined this intersection as a subset of foremost significance to the stem cell state. Another study of the same year[8] identified candidate transcripts as important determinants of pluripotency and lineage specification.

These papers were closely followed by publications that referred to the fact that such stemness defining studies came up with distinctively different signature contents, and an alert discussion about the stem

*Correspondence to: Eran Meshorer; The Hebrew University of Jerusalem; Department of Genetics; Edmond J. Safra; Givat Ram; Jerusalem 91904 Israel; Tel.: +972.2.6585161; Email: meshorer@cc.huji.ac.il/ Sol Efroni; The Mina and Everard Goodman Faculty of Life Sciences; Bar-Ilan University; Ramat-Gan 52900 Israel; Email: solefroni@gmail.com

Submitted: 08/18/08; Revised: 09/23/08; Accepted: 10/16/08

Previously published online as a *Cell Cycle* E-publication:
http://www.landesbioscience.com/journals/cc/article/7216

cell state, the differences between different lines, and the reliability of the different technologies arose.[9-13] The discussion was later followed by studies that aimed to fine tune the experimental assays with the purpose of generating a clear phenotypic contrast. It was argued that the transcriptional similarities and dissimilarities between comparable stem cell lines may serve as a defining property.[14] In this work, the authors found significant dissimilarities between different ESC lines. Such differences have recently been independently corroborated, demonstrating how 17 human ESC lines carry diverse propensities to differentiate into specific lineages, sometimes with an up to 100-fold difference.[15] In another report, 6 hESC lines have been studied, defining their own stem cell molecular repertoire.[16] The obtained signature had, similar to previous attempts, only little overlap with the "stemness" genes defined in the preceding studies.

By utilizing a different technology, Expressed Sequence Tags (ESTs), it was shown that out of the 148,453 ESTs found in ESCs, more than 32,000 were unidentified, half of which had no close matches in UniGene.[17] The authors of this study further showed how the strong statistical scrutiny they applied has been validated using meta-comparisons with other studies. Yet again, the signature obtained had only minor overlap with earlier studies (9 genes overlap with,[6] 32 genes overlap with[7]). Using massively parallel signature sequencing (MPSS), the transcriptomes of mouse and human ESCs have been compared with the transcriptomes of their differentiating embryoid bodies (EBs),[18] with a special emphasis on some of the major signaling pathways—LIF, Wnt, FGF and the TGFβ superfamily. Comparing previous expression microarray experiments,[6,7,11] the authors show significant differences from their own, bringing us further away from a universal stem cell signature. By combining ESTs and MPSS signatures of ESCs and EBs together, a small subset of signature genes was pinned down and validated.[19] Similar studies published during the same period of time[20-23] highlighted a collection of potential "stemness" signature genes.

Two recent studies profiled the gene expression signature between embryonic germ cells and ESCs.[24,25] Interestingly, these two populations have striking similarities in gene expression profiling. Yet, the authors were able to mine the data and obtain a signature that enabled separation of the two phenotypes. The intersection between the interesting group of multipotent adult progenitor cells and ESCs was less obvious,[26] but by doing so, the authors could identify a list of hundreds of genes that change their expression between phenotypes and determine the differences between the transcriptome behaviors.

In another recent study, a meta analyses of previously published gene expression data from stem cells was performed.[27] The authors took a different approach and characterized genes through their alternative promoters and CpG richness, searching for correlations between promoter features and stem cell features. This approach positioned proliferating adult stem cells and ESCs in the same group, and quiescent stem cells and pooled tissues in another. Such data mining, enhancing feature selection according to previously ignored biological traits, opens a path to novel definitions of a so far resistant stem-cell signature.

Although a small number of genes, well identified even without transcriptome studies (such as Oct4, Sox2 and Nanog) does emerge from the combined efforts, the studies in general fail to deliver a unified molecular signature. One possibility for this heterogeneity is the dissociated origin of the samples themselves. This was suggested based on studies of non-human primates.[28] Using cells derived from 41 rhesus embryos, the authors show homogenous expression profiles, suggesting that factors such as origin, derivation conditions, handling and culture might be one of the reasons for the inherent 'noise' in the ESC system.

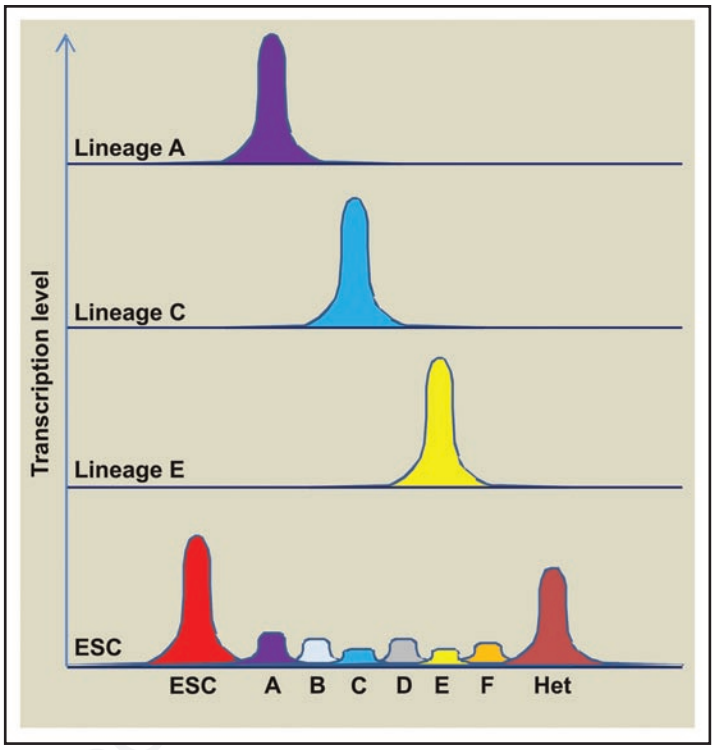

Figure 1. Pervasive transcription in ESCs. Shown are schematic diagrams of gene sets expressed in different cell lineages. Each lineage expresses a unique and characteristic gene set (lineage A, C, E, in different colors) while ESCs express their own characteristic gene set (red) plus low level transcription of the different lineages, as well as normally silenced heterochromatic regions (Het) such as satellite repeats.

We propose that 'noisy' transcription is an underlying inherent feature that distinguishes ESCs from differentiated cells (Fig. 1). Using a whole genome tiling transcription array, we measured the entire transcriptome (exonic, intronic, intergenic) expression levels in differentiating ESCs over multiple time points during 7 days and followed the transcriptional differences as a global phenomena, and not at the single locus level. We were able to demonstrate that it is the global phenomenon that marks the progressive transition from ESC into neural progenitors[29] and that the permissive transcriptional state in ESCs across the entire genome is being silenced during differentiation. This occurs in most chromosomes and at all exonic, intronic and intergenic regions, although the latter, as expected from 'promiscuous' type of transcription, show the most pronounced effect. Moreover, when expression levels of every transcript at zero time (undifferentiated ESC) was compared with consecutive time points, the overwhelming majority of transcription probes displayed reduction in their expression levels. These findings combine to demonstrate a global transcriptional behavior associated with differentiation from pluripotency to commitment. This global transcription activity may be viewed as a promising profile to the stem cell "state".

Evidence for Enhanced Global Transcription in ESC

The collection of studies over the past few years, in search of a stemness signature, left behind them a trail of evidence in support of the global transcription claim. We will show how findings embedded within the reported results serve to demonstrate how the transcriptional machinery of ESC is more "active" than in the differentiated state. Most revealing in this respect are the non biased EST and MPSS studies, as well as deep sequencing technologies, where, unlike gene expression microarrays, the entire transcriptome is collected at different resolutions.

Elevated levels of gene expression have been reported in many studies. For example, in the earlier studies, 60% of the genes were shown to be expressed in mouse ESCs[6] and over 30% in human ESCs, larger proportion than any other cell type examined.[14] By EST profiling in ESCs, the authors stumbled upon unusually high novelty rates in transcripts extracted from ESCs, and they observed that this is a general phenomenon when looking at ESCs.[8] MPSS analysis also led the authors to observe that "most genes (including tissue-specific genes) are maintained in an open state with low/undetectable levels of transcription".[19] MPSS allows the simple count of transcripts (counted by tpm—transcripts per million), and in this study the authors show that ESCs had a count of 21,874 tpm, compared with 15,336 tpm in EBs, again, demonstrating a higher transcriptional output, complexity and abundance in the undifferentiated state. The authors go on to claim that "this suggests that repression of gene expression is an important aspect of development". They further show that 2,376 genes are being downregulated during differentiation, while 1,996 are being upregulated. When the authors discuss the reasons for the major differences between studies, they report that most genes in ESCs are expressed at very low levels, with 70% of them at less than 50 tpm, which would render them practically undetectable in hybridization microarrays, for example. This is compatible with another microarray study[30] and an additional MPSS study mentioned above,[18] which shows a total of 13,824 unique signatures in mouse ESCs, compared with 9,845 in EBs. The latter study also shows 20,027 (or 23,500 in another line) unique transcripts in human ESCs contrasted with 17,278 unique transcripts in human EBs.

A recent study, mostly a tour de force of the novel methodology RNE-seq[31] mentions that when looking at the transcriptome complexity of ESCs versus EBs, the ESCs gave 53,056 transcripts from 31,872 ESC loci while EBs gave 50,881 mRNAs from 29,606 EB loci. The same study also shows that 31–37% of all transcription comes from intergenic regions and that "much of the non-exonic expression we observed in gene regions was pervasive and low", again, showing a global, low level transcription, although to a lesser extent. Taken together, these studies show that whenever transcriptional complexity has been measured, ESCs demonstrated a larger number of transcripts, originating from a larger set of unique sequences.

Non-Genic Transcription

As indicated above, undifferentiated ESCs display elevated levels of transcription throughout their genome. Non-coding regions pose a particularly interesting aspect of pervasive transcriptional output. Expression of transposable elements during an early window of mouse embryonic development was reported over 20 years ago by Francois Jacob and colleagues.[32] Using RNA FISH against ETn transposons the authors found no expression at early cleavage embryos and high levels at all pluripotent cells of the ICM at the blastocyst stage. At later embryonic stages of 6.5 and 8.5 days, expression was restricted to specific embryonic regions and gradually disappeared. A similar pattern was observed in embryonal carcinoma (EC) cells, with high ETn expression at the undifferentiated state and a marked reduction (of more than 95%) after differentiation.[32] In ESCs, expression patterns of transposable elements were reported several years later.[33] Using differential-display PCR, the authors identified EC-1 as an ESC-enriched transposable element. EC-1, they found, is highly expressed in EC and ESCs and is reduced by ~50% after 24 hours of retinoic acid treatment, and in a variety of differentiated cell types it is virtually absent. These observations were recently extended to human ESCs, where expression of the long interspersed nuclear element 1 (LINE-1) was observed.[34] Using a retrotransposition reporter assay, the authors could show that hESCs not only express but also support LINE-1 transpositions events, suggesting that retrotransposition may occur during early stages of development.

Non-genic expression in ESCs extends beyond transposons. Repetitive elements, including the major and minor satellite repeats, which are normally heterochromatic and silenced, are expressed at high levels in undifferentiated ESCs.[29] This is also true for some long intergenic regions found scattered throughout the genome, which are no longer expressed following differentiation.[29] In addition, several primary miRNAs (pri-miRNAs) are highly expressed in human and mouse ESCs, mouse EC cells, and human primary tumors, but, interestingly, their corresponding mature species are not detectable.[35] This situation is reminiscent of low-level transcription of tissue-specific genes in ESCs, of which protein products are undetectable.[29] A protein repression mechanism likely plays a role in undifferentiated ESCs, suppressing protein synthesis,[36] as well as degrading nascent transcripts,[37] although the latter is likely a distinct mechanism operating inside the nucleus. Along the same lines, the processing of several pri-miRNAs is blocked in embryonic tissues, with activation of processing occurring only as development proceeds.[35] This suggests, as is the case for protein translation, that there may be a posttranscriptional block in miRNA (miR) biogenesis, the mechanism of which remains unknown.

Among the growing number of studies showing miR regulation in ESCs, Let-7 serves as an intriguing example for ESC-specific post-transcriptional silencing.[38,39] The level of pri-let-7 is similar in ESCs and EBs, whereas the mature let-7 is undetectable in ESCs but is strongly elevated in EBs. This shift is mediated by the pluripotency factor Lin-28, which selectively blocks let-7 processing and therefore regulates the maturation of pri-let-7 in ESCs.[39] Additional noteworthy examples demonstrate that miRNAs control cell lineage determination in ESCs, likely by fine-tuning the transcriptome of differentiating cells during the commitment to a newly adopted fate. For example, miR-124a and miR-9, both of which are expressed at the primary transcript level but have undetected mature products in ESCs, regulate ESC neuronal differentiation,[40] while the similarly behaved miR-1 and miR-133 affect mesoderm commitment in mouse ESC by promoting muscle lineage decisions and repressing ectodermal and endodermal differentiation.[41] miR-21 is also present at low levels in ESCs and at higher level in EBs, but its ESC-specific suppression is mediated by REST (RE1 silencing transcription

factor).[42] This miR was shown to reduce self-renewal of ESCs upon ectopic overexpression, accompanied by a decrease in the expression levels of pluripotency factors Oct4, Nanog, Sox2 and c-Myc. These selected examples show that miRs play essential roles in ESCs and that while many of them are expressed at the primary transcript level, they are processed to a mature state only following differentiation.

But do miRs also follow the general rule of permissive transcription in ESCs? A recent report suggests they do.[43] Using extensive bioinformatic analyses of a variety of different cell types, the authors identified three classes of miRs that are enriched in undifferentiated ESCs when compared with adult tissue. Interestingly, out of the 229 miRs identified in this analysis, 50% (114 miRs) were highly enriched in undifferentiated ESCs, while only 36% (83 miRs) were enriched in the adult tissues. Whether this represents a true functional cluster of ESC-enriched miRs remains to be seen, but the overall enrichment of miRs in the undifferentiated state is in line with the overall increased transcriptional activity in ESCs.

Together, these data demonstrate extensive non-genic transcription in ESCs, but at the same time, the presence of silencing or regulating mechanisms that keep things in balance.

Is Pervasive Transcription Functionally Important?

The enhanced transcriptional activity associated with the undifferentiated state raises thoughts regarding its significance. Does global, permissive transcription play a role in stem cell identity and maintenance? Or, with regard to the correlation between heterochromatin formation and differentiation,[4,44] is it a mere byproduct of the overall decondensed chromatin structure in undifferentiated ESCs?

Advanced transcriptome research reveals a growing variety of non-coding transcripts, some of which have a relatively characterized function, related either to sequence (siRNA, miRNA, *Xist*) or structure (rRNA, tRNA, snRNA, snoRNA). Still, many transcripts have yet to be implicated with any clear activity (other than being transcribed) and are appropriately termed Transcripts of Unknown Function, or TUFs.[45] The location of many TUFs at unannotated regions upstream of known coding genes raised the possibility that they might participate in a yet undefined form of transcription-mediated chromatin remodeling.

In a comprehensive study of nuclear and cytoplasmic human RNA transcripts, classes of RNAs transcribed from gene boundaries were identified and characterized according to length and cellular location.[46] Thus, short RNAs (sRNA, under 200 nt) and most poly-adenylated long RNAs (lRNA, over 200 nt) were found both in the nucleus and the cytoplasm while other, non-poly-adenylated lRNAs were found exclusively in the cytoplasm. Furthermore, gene expression strongly correlates with the presence of certain RNAs. Most notable are the promoter-associating-sRNAs (PASRs), the presence of which tightly and positively correlate with expressed genes. Conversely, most silenced genes are devoid of any detectable PASRs. Also, some genes are only partially transcribed to the extent of the first exon and intron. This behavior particularly defines transcriptional status in hESCs,[47] possibly marking expression states by transcript length while linking global transcription to chromatin structure.

One intriguing option for the role of long intergenic transcription comes from studies of early development in Drosophila, where 30% of overall embryonic transcription stems from genomic non-coding regions.[48] The authors found alternative Transcription Start Sites (TSSs) with new 5' 'microexons' (most of which are non-coding) that precede extremely long intronic sequences ('megaintrons'), more than 10-fold longer than the average annotated intron length (averaging at ~20 kb with examples of over 100 kb). Considering the compactness and transcriptional efficiency of the fly genome, this combination seemed to serve no purpose other than transcription itself, or rather to communicate to remote regions that active transcription is taking place elsewhere.

Such large scale transcription-without-translation during fly early development, not unlike in mammalian ESCs, tempts to consider that throughout eukaryotic evolution different mechanisms have been formed, utilizing transcription and subsequently maintaining chromatin in an open state. Alternative TSSs and 'microexons' coupled with 'megaintrons' evidently serve this purpose.

However attractive this possibility is, different mechanisms might have been formed to maintain chromatin at an open configuration in order to allow access for transcription. In this view, the default state of chromatin, or its minimal energy configuration, is a compact, heterochromatin-like state, and chromatin-bound proteins, such as chromatin remodeling factors, are required to maintain chromatin in its decondensed, open form. It is possible that they would also recruit transcriptional repressors to prevent massive transcription from all open loci in ESCs. Although it is difficult to unequivocally discern between these two possibilities (Fig. 2), based on the collection of data, we find the former more attractive, considering the nature of the expressed transcripts. Permissive transcription resulting from chromatin conformation held open by proteins would likely result in a collection of transcripts of all possible sizes. On the other hand, transcription-mediated open chromatin, we argue, could be explained by low-level tissue-specific gene expression as well as defined families of non-genic, intergenic transcripts as observed. Thus, expression, albeit promiscuous, of specific transcription units of relatively defined length, maintains the open chromatin state. Along these lines, it can be argued that early development, during which multiple choices are made regarding cell identity and tissue formation requires high flexibility in protein expression profiles and the ability to quickly alter transcription patterns. This would require high accessibility of DNA sequences AND high availability of transcription machinery and so maintaining open chromatin by way of promiscuous transcription initiation satisfies both requirements with one mechanism, evidently efficient.

Conclusions

Our inability to define a uniform stem cell molecular signature led us to suggest that transcriptional 'uncertainty' is an inherent principle of the stem cell state. A stochastic, multi state behavior of stem cells is a mechanism that allows the directed silencing of some of those states—the process of differentiation. A stem cell population exhibits a wide range of transcriptional outputs of a single gene.[49] When a small fraction of that initial population is re-cultured and allowed to propagate, it resumes the same transcriptional profile of the original population,[49] suggesting that multiple transcriptional states is indeed an inherent feature of stem cells. Thus, pluripotency dictates multiple simultaneous conditions, resulting in pervasive, global transcription. We further suggest that this permissive transcription program is required to maintain an open chromatin

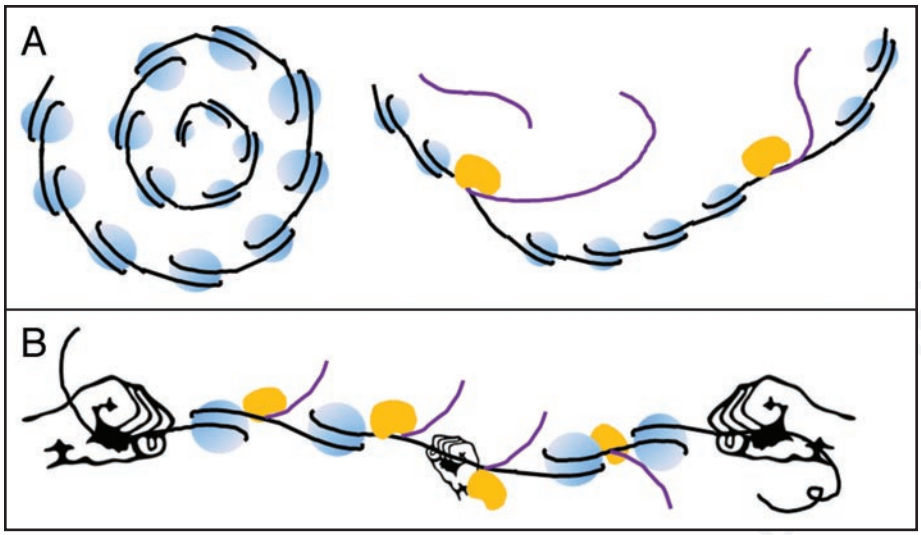

Figure 2. Models depicting the relationship between chromatin structure and global transcription in ESCs. (A) Transcription is required to maintain open chromatin in ESCs. According to this model, the 'default' structure of chromatin is a compact one (left) and transcription keeps chromatin in a decondensed form (right). (B) Open chromatin allows promiscuous transcription in ESCs. According to this model, chromatin proteins such as remodelers (depicted as hand palms) maintain chromatin in an open configuration and allows promiscuous transcription. Yellow, RNA polymerase II; purple, RNA molecules.

conformation in the pluripotent state and that differentiation and loss of pluripotency is accompanied by reduction of genome-wide transcriptional activity. It remains to be seen how global transcription is regulated and maintained and what is the role of the wide-spread expression of both low-level tissue-specific transcripts and normally silenced genomic regions such as heterochromatic satellite repetitive elements. The recent advances in sequencing and RNAi technologies should partially assist in getting some of these answers.

Acknowledgements

EM is supported by the Israel Science Foundation (grant 215/07), the European Union (grant IRG-206872), and an Alon Fellowship.

References

1. Kragh H. Quantum generations: a history of physics in the twentieth century. Princeton, N.J.: Princeton University Press 1999.
2. Einstein A, Born M, Born H. The Born-Einstein letters; correspondence between Albert Einstein and Max and Hedwig Born from 1916 to 1955. New York: Walker 1971.
3. Meshorer E, Misteli T. Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol 2006; 7:540-6.
4. Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev Cell 2006; 10:105-16.
5. Clarke MF, Dick JE, Dirks PB, Eaves CJ, Jamieson CH, Jones DL, Visvader J, Weissman IL, Wahl GM. Cancer stem cells—perspectives on current status and future directions: AACR Workshop on cancer stem cells. Cancer Res 2006; 66:9339-44.
6. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. "Stemness": transcriptional profiling of embryonic and adult stem cells. Science 2002; 298:597-600.
7. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. A stem cell molecular signature. Science 2002; 298:601-4.
8. Tanaka TS, Kunath T, Kimber WL, Jaradat SA, Stagg CA, Usuda M, et al. Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. Genome Res 2002; 12:1921-8.
9. Vogel G. Stem cells. 'Stemness' genes still elusive. Science 2003; 302:371.
10. 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.
11. Fortunel NO, Otu HH, Ng HH, Chen J, Mu X, Chevassut T, et al. Comment on "'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". Science 2003; 302:393.
12. Zipori D. The nature of stem cells: state rather than entity. Nat Rev Genet 2004; 5:873-8.
13. Eckfeldt CE, Mendenhall EM, Verfaillie CM. The molecular repertoire of the 'almighty' stem cell. Nat Rev Mol Cell Biol 2005; 6:726-37.
14. Abeyta MJ, Clark AT, Rodriguez RT, Bodnar MS, Pera RA, Firpo MT. Unique gene expression signatures of independently-derived human embryonic stem cell lines. Hum Mol Genet 2004; 13:601-8.
15. Osafune K, Caron L, Borowiak M, Martinez RJ, Fitz-Gerald CS, Sato Y, Cowan CA, Chien KR, Melton DA. Marked differences in differentiation propensity among human embryonic stem cell lines. Nat Biotechnol 2008; 26:313-5.
16. Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, et al. Gene expression in human embryonic stem cell lines: unique molecular signature. Blood 2004; 103:2956-64.
17. Brandenberger R, Wei H, Zhang S, Lei S, Murage J, Fisk GJ, et al. Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat Biotechnol 2004; 22:707-16.
18. Wei CL, Miura T, Robson P, Lim SK, Xu XQ, Lee MY, et al.Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. Stem Cells 2005; 23:166-85.
19. Miura T, Luo Y, Khrebtukova I, Brandenberger R, Zhou D, Thies RS, et al. Monitoring early differentiation events in human embryonic stem cells by massively parallel signature sequencing and expressed sequence tag scan. Stem Cells Dev 2004; 13:694-715.
20. Sato N, Sanjuan IM, Heke M, Uchida M, Naef F, Brivanlou AH. Molecular signature of human embryonic stem cells and its comparison with the mouse. Dev Biol 2003; 260:404-13.
21. Sperger JM, Chen X, Draper JS, Antosiewicz JE, Chon CH, Jones SB, et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Natl Acad Sci USA 2003; 100:13350-5.
22. Ginis I, Luo Y, Miura T, Thies S, Brandenberger R, Gerecht-Nir S, et al. Differences between human and mouse embryonic stem cells. Dev Biol 2004; 269:360-80.
23. Richards M, Tan SP, Tan JH, Chan WK, Bongso A. The transcriptome profile of human embryonic stem cells as defined by SAGE. Stem Cells 2004; 22:51-64.
24. Sharova LV, Sharov AA, Piao Y, Shaik N, Sullivan T, Stewart CL, et al. Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains. Dev Biol 2007; 307:446-59.
25. Grskovic M, Chaivorapol C, Gaspar-Maia A, Li H, Ramalho-Santos M. Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. PLoS Genet 2007; 3:145.
26. Ulloa-Montoya F, Kidder BL, Pauwelyn KA, Chase LG, Luttun A, Crabbe A, et al. Comparative transcriptome analysis of embryonic and adult stem cells with extended and limited differentiation capacity. Genome Biol 2007; 8:163.
27. Edwards YJ, Bryson K, Jones DT. A meta-analysis of microarray gene expression in mouse stem cells: redefining stemness. PLoS ONE 2008; 3:2712.
28. Navara CS, Mich-Basso JD, Redinger CJ, Ben-Yehudah A, Jacoby E, Kovkarova-Naumovski E, et al. Pedigreed primate embryonic stem cells express homogeneous familial gene profiles. Stem Cells 2007; 25:2695-704.
29. Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, et al. Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2008; 2:437-47.
30. Golan-Mashiach M, Dazard JE, Gerecht-Nir S, Amariglio N, Fisher T, Jacob-Hirsch J, et al. Design principle of gene expression used by human stem cells: implication for pluripotency. Faseb J 2005; 19:147-9.

31. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 2008; 5:613-9.
32. Brulet P, Condamine H, Jacob F. Spatial distribution of transcripts of the long repeated ETn sequence during early mouse embryogenesis. Proc Natl Acad Sci USA 1985; 82:2054-8.
33. Hirose T, Austin SJ, Jetten AM. Identification of a transposon-related RNA downregulated by retinoic acid in embryonal carcinoma and embryonic stem cells. Exp Cell Res 1995; 221:294-300.
34. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS, Moran JV. LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet 2007; 16:1569-77.
35. Chen L, Daley GQ. Molecular basis of pluripotency. Hum Mol Genet 2008; 17:23-7.
36. Sampath P, Pritchard DK, Pabon L, Reinecke H, Schwartz SM, Morris DR, Murry CE. A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. Cell Stem Cell 2008; 2:448-60.
37. Szutorisz H, Georgiou A, Tora L, Dillon N. The proteasome restricts permissive transcription at tissue-specific gene Loci in embryonic stem cells. Cell 2006; 127:1375-88.
38. Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, Wulczyn FG. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat Cell Biol 2008; 10:987-93.
39. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. Science 2008; 320:97-100.
40. Krichevsky AM, Sonntag KC, Isacson O, Kosik KS. Specific microRNAs modulate embryonic stem cell-derived neurogenesis. Stem Cells 2006; 24:857-64.
41. Ivey KN, Muth A, Arnold J, King FW, Yeh RF, Fish JE, et al. MicroRNA regulation of cell lineages in mouse and human embryonic stem cells. Cell Stem Cell 2008; 2:219-29.
42. Singh SK, Kagalwala MN, Parker-Thornburg J, Adams H, Majumder S. REST maintains self-renewal and pluripotency of embryonic stem cells. Nature 2008; 453:223-7.
43. Gu P, Reid JG, Gao X, Shaw CA, Creighton C, Tran PL, et al. Novel microRNA candidates and miRNA-mRNA pairs in embryonic stem (ES) cells. PLoS ONE 2008; 3:2548.
44. Aoto T, Saitoh N, Ichimura T, Niwa H, Nakao M. Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. Dev Biol 2006; 298:354-67.
45. Gingeras TR. Origin of phenotypes: genes and transcripts. Genome Res 2007; 17:682-90.
46. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. Science 2007.
47. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 2007; 130:77-88.
48. Manak JR, Dike S, Sementchenko V, Kapranov P, Biemar F, Long J, Cheng J, Bell I, Ghosh S, Piccolboni A, Gingeras TR. Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. Nat Genet 2006; 38:1151-8.
49. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 2008; 453:544-7.