

REVIEW

Standardization of human stem cell pluripotency using bioinformatics

Michael W Nestor and Scott A Noggle*

Abstract

The study of cell differentiation, embryonic development, and personalized regenerative medicine are all possible through the use of human stem cells. The propensity for these cells to differentiate into all three germ layers of the body with the potential to generate any cell type opens a number of promising avenues for studying human development and disease. One major hurdle to the development of high-throughput production of human stem cells for use in regenerative medicine has been standardization of pluripotency assays. In this review we discuss technologies currently being deployed to produce standardized, high-quality stem cells that can be scaled for high-throughput derivation and screening in regenerative medicine applications. We focus on assays for pluripotency using bioinformatics and gene expression profiling. We review a number of approaches that promise to improve unbiased prediction of utility of both human induced pluripotent stem cells and embryonic stem cells.

Introduction

Human pluripotent stem cells are promising tools to advance the study of cell differentiation and embryonic development. These cells hold promise for the development of personalized regenerative therapies. Key to these endeavors is the fundamental attributes of self-renewal and the potential to generate any human cell type, characteristics that constitute pluripotency when combined. The gold standard for human pluripotent stem cells is embryonic stem cells (ESCs), derived from preimplantation embryos in excess of clinical need. While therapies using human embryonic stem cell (hESC)-derived cells are currently in development, the ability of human adult cells to return to a pluripotent state offers the potential to personalize regenerative medicine. The landmark study by Takahashi and Yamanaka demonstrated that four transcription factors (Oct4, Klf4, Sox2, and *c-Myc*) were sufficient to convert adult cells to pluripotent cells: human induced pluripotent stem cells (iPSCs) [1,2]. Since the advent of this technology, a large number of studies have emerged demonstrating the immense power of these cells – with iPSCs having been differentiated into hematopoietic progenitors, endothelial cells, retina,

osteoclasts, islet-like cells, hepatocyte-like cells, and neurons [3].

Compared with methods for deriving ESCs, the generation of iPSCs involves management of confounds generated from resetting the adult transcriptional program. During reprogramming, the activation of multiple signaling pathways through exogenous transcription factor expression induces epigenetic changes and changes in gene expression. Prolonged expression of these factors can induce a highly variable population of reprogramming states [4]. This variability of genetic expression may combine with stochastic events involved in reprogramming to generate the inefficient and highly variable yield often observed during iPSC generation [5]. For example, while iPSC reprogramming typically results in a large number of highly proliferative cells, very few cells exhibit pluripotency [6]. Despite these inefficiencies, once derived and subjected to even minimal quality control, it is remarkable how similar these two types of pluripotent cells behave in functional assays.

How is the quality and uniformity of iPSCs and ESCs most efficiently tested? Early work established a number of empirically determined criteria, including a distinct morphology, proliferation rate, activation of pluripotent genes, expression of surface markers, silencing of reprogramming transgenes, embryoid body, and teratoma formation [7,8]. In the mouse, iPSCs and ESCs ideally form germline and tissue chimerism when injected into

* Correspondence: snoggle@nyscf.org
The New York Stem Cell Foundation, 163 Amsterdam Avenue, Box 309,
New York, NY 10023, USA

blastocysts. The most stringent assay for developmental potential is the tetraploid complementation assay, in which cells are placed in an environment where they can exclusively contribute to the entire mouse [9,10].

Because this complementation assay is not available for human cells in the context of human embryogenesis, assays for developmental potential attempt to answer the question of functionality by differentiation into mature cell types using teratoma assays. Most hESCs that have been derived and are karyotypically normal can differentiate into most cell types in these tests. Decrements in the quality of hESC lines may primarily come from problems with genome integrity. Lines with karyotypic abnormalities that confer growth advantages tend to differentiate less well in teratoma assays (reviewed in [11]). The primary measure of quality of hESCs may therefore be genomic integrity rather than stringent measures of differentiation potential.

While several groups have demonstrated fundamental similarities in biomarkers among stem cell lines (see for example [12,13]), these tests are time consuming, are difficult to perform for large numbers of cell lines, and test performance can vary from laboratory to laboratory. Concomitant with the effort to determine whether there are molecular and functional differences of consequence between iPSCs and hESCs, many sensitive bioinformatic assays have been developed that are starting to replace the embryological and teratoma assays used to characterize pluripotency. Recent work has focused on establishing better pluripotency standards for the *a priori* selection of cell lines. In this review, we consider several major bioinformatic approaches that have been used to assess the quality of pluripotent stem cells and we provide a nonexhaustive overview of the results obtained using several approaches.

Bioinformatic assays for pluripotency

In the absence of stringent embryological assays for pluripotency in human pluripotent stem cells, there has been much progress over the last few years in developing genome-wide assays and associated bioinformatic methods for their analysis. These methods originally focused on identifying global transcriptional profiles that characterize the pluripotent state relative to differentiated cells and tissues. With advancement in sequencing technologies has also come the global analysis of the epigenome. Together with analysis of various noncoding RNAs, all of these assays have been used to address the question of pluripotency identity at the molecular level.

With the development of iPSC technology, the focus has turned to characterizing differences among pluripotent stem cells. The current view is that, whether due to different derivation strategies or genetic differences, pluripotent stem cell lines can vary. For example, while

most studies find iPSCs to be quite similar to hESCs at the molecular level, the challenge has been to identify subtle differences that might have functional consequences. Efforts to characterize this variation have resulted in a number of algorithms used to assess line-to-line differences in pluripotent stem cells.

Gene expression profiling

Gene expression profiling using DNA microarrays was the first method of global molecular analysis applied to map the transcriptome of pluripotent stem cells [14-17] and has become a standard assay of pluripotency in many studies. Various classification algorithms have been used to group lines into similar transcriptional states. For example, samples of cultured pluripotent stem cells can be distinguished from multipotent stem cell populations and differentiated cell types [18].

Significant progress has been made in applying these analysis methods to discriminate more subtle differences in pluripotent stem cells. For example, initial studies comparing iPSCs and hESCs suggested that the two populations of cells are statistically different [19-21], and this difference, although significantly decreased, persists into later passages. However, more recent studies have found global similarities with small differences between iPSCs and hESCs [2,22-24]. Changes in gene expression signatures are not limited to mRNA; they have also been observed in both miRNA and long intergenic noncoding RNA [25-27]. However, it is still not clear whether this variation is due to different growth conditions, laboratory-to-laboratory variation [28], heterogeneity in iPSC quality [20], or small sample sizes [19].

Can these methods be used on their own to identify a normal pluripotent cell? Finding a unique gene expression profile that consistently varies in pluripotent cells has been difficult [22]. However, as the sample sizes of these studies are relatively small compared with, for example, gene expression in cancer studies, where sample sizes can be in the hundreds to thousands [29], the approaches used in the above studies may not be sufficiently powered to find consistent but small differences.

As the availability of well-curated samples increases, it should become possible to make more reliable biological distinctions. For example, the availability of larger datasets makes more advanced methods based on machine learning possible for classifying pluripotent stem cell lines. This approach is taken for PluriTest, an algorithm that makes use of training sets containing large numbers of undifferentiated, differentiated, normal and abnormal human stem cell lines and tissues. The large sample size allows the algorithm to construct bioinformatic models for assessing the quality of novel pluripotent stem cells based only on DNA microarray gene expression measurements [30]. To generate the model,

two principal component vectors were calculated that first separate pluripotent from differentiated states and, second, distinguish abnormal from normal expression signatures from a large training set of almost 500 samples. The samples used for training were curated for microarray data quality and contained hESCs, germ cell tumor samples, primary cell lines, and somatic tissues.

As reported, the resulting PluriTest algorithm could successfully distinguish independent samples of germ cell tumors from hESCs as well as distinguish reprogrammed from partial reprogrammed iPSCs. The algorithm was also able to distinguish parthenogenetic stem cell lines from hESCs, presumably because of differences at imprinted loci. This suggests it should be possible to distinguish abnormal samples from normal samples and to classify them as undifferentiated or differentiated. Others have reported using PluriTest to characterize iPSCs [31-33]. Additionally, the algorithm has been replicated with mouse ESCs and can predict the response to nanog overexpression, which results in shifts in the pluripotent state consistent with differentiation of the inner cell mass of the blastocyst to an epiblast-like state characteristic of the implanting embryo [31].

While the algorithm can distinguish pluripotent states, samples identified as abnormal currently need further analysis to identify the particular cause. However, Williams and colleagues note that this strategy could also be applied to other types of data describing stem cell lines, such as epigenetic status [34]. How sensitive the algorithm is for abnormalities such as copy number variations or translocations is also not clear. Because machine learning techniques are dependent on the quality and breadth of variability of the training dataset used to construct the model, including tests of genetic integrity, for example, could improve predictions of functional quality of the lines.

Epigenetic profiles

A promising route to providing standardized assays for iPSC and ESC pluripotency and differentiation is to understand the epigenetic landscape that is common to both systems and connect it to gene regulation. Epigenetic comparisons via technology such as chromatin immunoprecipitation have thus been used to develop the transcription factor binding, histone modification and DNA methylation profiles of human iPSCs and ESCs (recently reviewed in [35,36]).

Again it has been informative to look at progress in the ability to distinguish epigenetic differences between iPSCs. Initial attempts using this approach yielded inconsistent results when comparing ESCs and iPSCs. Screening for transcriptional differences in early (passage 5) and late (passage 28) iPSCs as compared with ESCs, chromatin immunoprecipitation analysis showed similar

bivalent H3K chromatin domain marks that are enriched in pluripotent cells [19]. However, in a subsequent study using six independent ESC lines and six independent iPSC lines and measuring histone H3K4me3 and H3K27me3 modifications via chromatin immunoprecipitation as a readout for transcriptionally active or repressed domains of the genome, respectively, no significant phenotypic differences in the chromatin marks were reported [37]. In contrast, another report showed that while H3K27 repressive marks were similar, a small fraction of repressive H3K9me3 marks were unique to iPSCs [38]. However, the functional consequences of these differences are still not clear.

While assaying histone modifications can identify poised transcriptional states characteristic of pluripotency, studies of genome-wide methylation can provide a complementary view of the epigenetic state as they usually anti-correlate. DNA methylation to generate single nucleotide genome-wide maps has been generated for the pluripotent state of hESCs and iPSCs [22,33,39]. Although a robust general test for pluripotency when assaying core pluripotency associated genes, global DNA methylation comparison studies have also given mixed empirical results. Using patterns of DNA methylation across ~66,000 CpG sites from iPSCs, while globally similar, differences between iPSCs and ESCs at methylation of CpG sites were observed when a hierarchical clustering analysis was performed [40]. Genes analyzed from iPSCs were less methylated than fibroblasts and ESCs, which was attributed in part to epigenetic spillover from the overexpression of transcription factors that were introduced to the iPSCs via integrated viral transgenes. Additionally, measurement of differentially methylated regions from late-passage iPSCs shows that, when compared with ESCs, iPSCs have 92% hypomethylated CpGs [23] – although this value may be skewed due to the small number of ESC samples analyzed. Additionally, differential methylation between pluripotent and somatic tissue samples has been found, mainly at imprinted loci, some of which could be explained by differences in culture conditions among lines tested [33]. Reprogramming iPSCs may also introduce aberrant and inefficient methylation [41], which may have potential functional influences during and after differentiation [33].

Inefficient DNA methylation in iPSCs combined with the stochastic nature of novel epigenetic aberrations in these cells may not show a phenotype until after differentiation when altered gene expression leads to dysfunctional cell states [33,42]. This in part may be the explanation for iPSC bias toward donor-cell-related lineages [41]. In mouse iPSCs, however, the promoter methylation pattern was correlated with donor cell origin at early passage numbers but not after subsequent

passaging [43], suggesting further completion of reprogramming over time or selection for pre-existing fully reprogrammed cells within cultures over time. This may not be the case in human pluripotent stem cell cultures because recent reports found that aberrant methylation can sometimes be gained at imprinted loci during culture [33]. Importantly, after directed differentiation into multiple tissues, such aberrant methylation patterns persist in the differentiated cells [33]. Again, it seems the functional consequences of epigenetic alterations must be further explored.

Despite these inconsistencies, current technology for monitoring epigenetics is clearly quite sensitive to small changes that might have functional consequences. Combining methylation mapping and gene expression signatures by algorithm may therefore be possible to more robustly infer the cell state. Bock and colleagues performed a number of statistical tests against previously published datasets [19,22,26,42] to show that there are small but significantly detectible differences in gene expression and DNA methylation in some but not all iPSC cell lines compared with hESC lines [22]. Their best performing classifier used a support vector machine learning algorithm trained on a combination of DNA methylation and gene expression data from ESC lines versus iPSC lines. Using 20 hESC lines and 12 iPSC lines, this method was able to correctly classify hESC lines, but was only moderately successful at classifying iPSC lines. On average, the method could predict iPSC gene signatures with 81% accuracy and 91% specificity but only moderate sensitivity (61%). While combining gene expression and methylation, this study used far fewer training samples for modeling compared with PluriTest. Whether the use of a bigger dataset for training the classifiers will improve these predictions is therefore important to determine. Additionally, like earlier studies, it is not clear whether these differences will have substantial functional consequences during or after differentiation.

This combinatorial approach has recently been shown to predict the cell state during hematopoietic stem cell differentiation [44]. Bock and colleagues intersected gene expression and DNA methylation to find a small number of loci that showed consistent negative correlations. Particular loci were indicative of known differentiation stages. Using this approach combined with a gene signature indicative of the proliferation state, they could predictively identify differentiation stages in the well-defined system of hematopoiesis in the adult mouse. This integrative approach highlights the value in combining datasets from different assays that produce complex data to gain predictive power. Whether this approach has utility in determining pluripotency status and differentiation potential in human pluripotent stem cells will be important to determine.

The scorecard approach

The selection of application-suitable cell lines that accurately differentiate into intended cell types, as currently practiced, is a labor-intensive process that requires the teratoma assay as well as low-resolution tests for pluripotency [7]. The bioinformatic approaches discussed above mainly interrogate the undifferentiated state of pluripotent stem cells. But what about the cells' ability to differentiate? Recently, an additional approach that combines gene expression and epigenetic measures with an *in vitro* differentiation assay has been proposed by Bock and colleagues [22].

This group first generated a deviation scorecard that assesses DNA methylation and gene expression profiles relative to a set of reference standard hESC lines to identify lines that deviate by outlier detection methods. The result is a list of outlier genes for each line. Genes are then highlighted that could be screened for their probable effect on performance in functional assays. To test this scorecard, genes were screened that would lead to aberrant function for motor neurons if the iPSC line was differentiated toward that fate. The hypermethylation of one such gene, GRM, a glutamate receptor expressed in motor neurons, was discovered. This quick test allowed Bock and colleagues to rule out the use of one cell line that might have been used to differentiate motor neurons.

To obtain an overall score for differentiation potential, a quantitative embryoid body differentiation assay that uses high-throughput transcript counting was used to gain a predictive measure of differentiation potential of pluripotent stem cell lines. Bock and colleagues used a nondirected embryoid body differentiation assay in which the embryoid bodies were grown for the 20 ESC lines and 12 iPSC lines and the RNA was collected and probed for expression levels of 500 marker genes. From this assay, a quantitative gene expression profile of embryoid bodies from the hESC reference lines was determined. Finally, the cell line-specific differentiation propensity was calculated for each of the germ layers using a bioinformatic algorithm that calculates differentiation propensity for multiple lineages relative to the performance of reference lines. In functional verification tests, the lineage scorecard was able to correctly classify iPSC lines based on their ability to differentiate into ISL1-positive motor neurons in directed differentiation assays.

Importantly, in a parallel but independent study by Boulting and colleagues, the differentiation propensity of these lines was compared with functional motor neuron differentiation efficiency and the cells were subjected to a number of relevant functional tests [45]. There was a statistically significant correlation of the lineage scorecard-based predictions with functional assays [45]. Important to note, however, is that Boulting and colleagues also

found that lines which performed poorly in the embryoid body assay in a forced directed differentiation protocol achieved similar functional results, suggesting that even lines which perform poorly relative to reference lines could be useful under the right conditions.

Taken together, these results suggest that integrating multiple high-content assays can predict functional outcomes in differentiating iPSCs. Additionally, the lineage scorecard approach should also be amenable to screening for a cell line's ability to differentiate into specific lineages by selecting more specific gene sets and recalibration to reference standards. As the number of lines screened increases, it should be possible to identify the most frequent gene expression and epigenetic aberrations, which should further lower the cost of these assays.

Conclusion

Observed variation in both hESCs and iPSCs may have a number of causes, including differences in *in vitro* culture as well as inherent genetic or epigenetic differences. In the process of pursuing a consistent profile of pluripotency, multiple methods have emerged that promise to correctly classify stem cell lines. In most of the current studies, only a relatively small number of hESC lines have been used as references and the genetic diversity of available hESC lines is probably much more limited than the available iPSC lines [46]. Further, several recent reports suggest that some of the differences between iPSCs and hESCs can be erased by altering culture conditions, prolonged culturing, or the stoichiometry of the reprogramming factors [19,43,47]. Even the same lines cultured in different laboratories can develop laboratory-specific signatures [22,28]. There is thus clearly still a great degree of method standardization needed to achieve accurate comparisons, and care should be taken when comparing results across studies.

While there is still significant work to be done to standardize the culture and assays for stem cells and

their differentiation, there has been much progress in the molecular and bioinformatic assays needed to monitor these steps (Table 1). The speed and scale of these assays is currently experiencing logarithmic growth, thereby reducing costs [48]. Refining these assays will greatly improve our ability to standardize the protocols used for deriving iPSCs as well as their differentiation into *bona fide* differentiated cell types needed for disease modeling and cell therapies.

Regardless of the source of variation, better methods are needed to assess pluripotency and the differentiation potential of human pluripotent stem cells. These methods will be particularly important in advancing the use of stem cells for therapeutic intervention. The inefficiency of current methods for generating a consistent core set of general-purpose iPSC lines severely limits the interpretation of data generated from iPSCs. For instance, iPSCs have recently been used to uncover 596 differentially expressed genes in schizophrenia, of which only 25% had been previously implicated in the disorder, but these data are confounded by variations in epigenetic memory that occur in iPSCs and possibly from cell culture techniques that vary from laboratory to laboratory [49]. A recent publication on a phenotype for Rett syndrome used only four fibroblast lines to report changes in neuronal function in iPSCs derived from these patients [50]. The development of cost-effective strategies for assessing quality will greatly improve our power to detect phenotypic differences in disease, particularly when quantitative traits are involved.

There are a number of therapeutic avenues for pluripotent stem cells. If the goal is to generate disease-specific cells from patients in order to study disease pathways and advance towards patient-specific interventions, then high-throughput derivation, culture, and analysis protocols must be in place to reduce experimental noise during phenotypic analysis. These protocols must allow researchers to determine what lines have the least amount of epigenetic variability and the highest propensity

Table 1 Summary of bioinformatic studies used in assessing induced pluripotent and embryonic stem cell pluripotency

Bioinformatic approach	Distinguish iPSCs from ESCs?	Functional assay?	Notes
Gene expression profiling [14-17]	Inconsistent [19-24]	No	No unique gene expression profile may be due to small sample size or heterogeneity in iPSC quality. [19,20,28]
Epigenetic profiling [35,36]	Inconsistent [19,37,38]	No	Further exploration of the functional consequences of epigenetic alterations is needed. [33]
Combinatorial profiling (methylation mapping and gene expression signatures) [22,44]	Yes [22]	No	No functional differences have been linked to the detectable differences in gene expression and DNA methylation used in these studies
Scorecard profiling (gene expression and epigenetic measures with <i>in vitro</i> differentiation) [22,45]	Yes [22,45]	Yes	Differentiation propensity was linked to motor neuron differentiation efficiency and functional relevance [22,45]

ESC, embryonic stem cell; iPSC, induced pluripotent stem cell.

for efficient and high yield differentiation. Additionally, in order to create libraries of knockout iPSCs and ESCs to study the roles of individual genes in disease, it is important to note which genes are highly variable from line to line, and to eliminate lines with too much variability in genes that may be important for function. This elimination must be done on large numbers of lines across multiple patients, within a shorter time frame and more cost-effectively than most protocols currently deliver. Alternatively, for assessing the quality and consistency of cells intended for transplantation, sensitive and robust assays must be available to monitor these products for reliability. For these purposes, algorithmic approaches such as those discussed above may be the best available tools for researchers to screen and scale multiple lines for regenerative medicine applications.

Note

This article is part of a thematic series on *Clinical applications of stem cells* edited by Mahendra Rao. Other articles in the series can be found online at <http://stemcellres.com/series/clinical>.

Abbreviations

ESC: Embryonic stem cell; hESC: Human embryonic stem cell; iPSC: Induced pluripotent stem cell; miRNA: MicroRNA.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors are grateful to members of the NYSCEF Laboratory for critical review of the manuscript. SAN's laboratory is funded by the New York Stem Cell Foundation, the Charles Evans Foundation, NYSTEM contract C024179 and C026185, and NIH sub-award 0255-5191-4609.

Published: 25 April 2013

References

1. Takahashi K, Yamanaka S: Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006, **126**:663–676.
2. 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.
3. Bilic J, Izpisua Belmonte JC: Concise review: Induced pluripotent stem cells versus embryonic stem cells: close enough or yet too far apart? *Stem Cells* 2012, **30**:33–41.
4. Young RA: Control of the embryonic stem cell state. *Cell* 2011, **144**:940–954.
5. Buganim Y, Faddah D, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R: Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 2012, **150**:1209–1222.
6. Meissner A, Wernig M, Jaenisch R: Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol* 2007, **25**:1177–1181.
7. Daley GQ, Lensch MW, Jaenisch R, Meissner A, Plath K, Yamanaka S: Broader implications of defining standards for the pluripotency of iPSCs. *Cell Stem Cell* 2009, **4**:200–201. author reply 202.
8. Brivanlou AH, Gage FH, Jaenisch R, Jessell T, Melton D, Rossant J: Stem cells. Setting standards for human embryonic stem cells. *Science* 2003, **300**:913–916.
9. Boland MJ, Hazen JL, Nazor KL, Rodriguez AR, Gifford W, Martin G, Kupriyanov S, Baldwin KK: Adult mice generated from induced pluripotent stem cells. *Nature* 2009, **461**:91–94.
10. Zhao XY, Li W, Lv Z, Liu L, Tong M, Hai T, Hao J, Guo CL, Ma QW, Wang L, Zeng F, Zhou Q: iPSC cells produce viable mice through tetraploid complementation. *Nature* 2009, **461**:86–90.
11. Gokhale PJ, Andrews PW: The development of pluripotent stem cells. *Curr Opin Genet Dev* 2012, **22**:403–408.
12. Initiative TISC: Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nat Biotechnol* 2011, **29**:1132–1144.
13. International Stem Cell Initiative, Adewumi O, Aflatoonian B, Ahrlund-Richter L, Amit M, Andrews PW, Beighton G, Bello PA, Benvenisty N, Berry LS, Bevan S, Blum B, Brooking J, Chen KG, Choo AB, Churchill GA, Corbel M, Damjanov I, Draper JS, Dvorak P, Emanuelsson K, Fleck RA, Ford A, Gertow K, Gertsenstein M, Gokhale PJ, Hamilton RS, Hampl A, Healy LE, Hovatta O, et al: Characterization of human embryonic stem cell lines by the international stem cell initiative. *Nat Biotechnol* 2007, **25**:803–816.
14. Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, Joshi BH, Ginis I, Thies RS, Amit M, Lyons I, Condie BG, Itskovitz-Eldor J, Rao MS, Puri RK: Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood* 2004, **103**:2956–2964.
15. 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**:404–413.
16. Sperger JM, Chen X, Draper JS, Antosiewicz JE, Chon CH, Jones SB, Brooks JD, Andrews PW, Brown PO, Thomson JA: Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A* 2003, **100**:13350–13355.
17. Suárez-Fariñas M, Noggle SA, Heke M, Hemmati-Brivanlou A, Magnasco MO: Comparing independent microarray studies: the case of human embryonic stem cells. *BMC Genomics* 2005, **6**:99.
18. Müller FJ, Laurent LC, Kostka D, Ulitsky I, Williams R, Lu C, Park IH, Rao MS, Shamir R, Schwartz PH, Schmidt NO, Loring JF: Regulatory networks define phenotypic classes of human stem cell lines. *Nature* 2008, **455**:401–405.
19. Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiwu O, Richter L, Zhang J, Khvorostov I, Ott V, Grunstein M, Lavon N, Benvenisty N, Croce CM, Clark AT, Baxter T, Pyle AD, Teitell MA, Pelegriani M, Plath K, Lowry WE: Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 2009, **5**:111–123.
20. Maherali N, Ahfeldt T, Rigamonti A, Utikal J, Cowan C, Hochedlinger K: A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* 2008, **3**:340–345.
21. Soldner F, Hockemeyer D, Beard C, Gao Q, Bell GW, Cook EG, Hargus G, Blak A, Cooper O, Mitalipova M, Isacson O, Jaenisch R: Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* 2009, **136**:964–977.
22. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggan K, Meissner A: Reference maps of human ES and iPSC cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 2011, **144**:439–452.
23. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR: Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 2011, **471**:68–73.
24. Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR: Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS One* 2009, **4**:e7076.
25. Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, Manos PD, Datta S, Lander ES, Schlaeger TM, Daley GQ, Rinn JL: Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* 2010, **42**:1113–1117.
26. Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, Kono T, Shioda T, Hochedlinger K: Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* 2010, **465**:175–181.
27. Lakshminathu U, Davila J, Hart RP: miRNA in pluripotent stem cells. *Regen Med* 2010, **5**:545–555.

28. Newman AM, Cooper JB: **Lab-specific gene expression signatures in pluripotent stem cells.** *Cell Stem Cell* 2010, **7**:258–262.
29. Rakyan VK, Down TA, Balding DJ, Beck S: **Epigenome-wide association studies for common human diseases.** *Nat Rev Genet* 2011, **12**:529–541.
30. Müller F-J, Schuldt BM, Williams R, Mason D, Altun G, Papapetrou EP, Danner S, Goldmann JE, Herbst A, Schmidt NO, Aldenhoff JB, Laurent LC, Loring JF: **A bioinformatic assay for pluripotency in human cells.** *Nat Methods* 2011, **8**:315.
31. Macarthur BD, Sevilla A, Lenz M, Müller FJ, Schuldt BM, Schuppert AA, Ridder SJ, Stumpf PS, Fidalgo M, Ma'ayan A, Wang J, Lemischka IR: **Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity.** *Nat Cell Biol* 2012, **14**:1139–1147.
32. Mariani J, Simonini MV, Palejev D, Tomasini L, Coppola G, Szekeley AM, Horvath TL, Vaccarino FM: **Modeling human cortical development in vitro using induced pluripotent stem cells.** *Proc Natl Acad Sci U S A* 2012, **109**:12770–12775.
33. Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Müller FJ, Wang YC, Boscolo FS, Fakunle E, Dumevska B, Lee S, Park HS, Olee T, D'Lima DD, Semechkin R, Parast MM, Galat V, Laslett AL, Schmidt U, Keirstead HS, Loring JF, Laurent LC: **Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives.** *Cell Stem Cell* 2012, **10**:620–634.
34. Williams R, Schuldt B, Müller F-J: **A guide to stem cell identification: progress and challenges in system-wide predictive testing with complex biomarkers.** *Bioessays* 2011, **33**:880–890.
35. Rada-Iglesias A, Wysocka J: **Epigenomics of human embryonic stem cells and induced pluripotent stem cells: insights into pluripotency and implications for disease.** *Genome Med* 2011, **3**:36.
36. Sindhu C, Samavarchi-Tehrani P, Meissner A: **Transcription factor-mediated epigenetic reprogramming.** *J Biol Chem* 2012, **287**:30922–30931.
37. Guenther MG, Frampton GM, Soldner F, Hockemeyer D, Mitalipova M, Jaenisch R, Young RA: **Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells.** *Cell Stem Cell* 2010, **7**:249–257.
38. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR, Ren B: **Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.** *Cell Stem Cell* 2010, **6**:479–491.
39. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462**:315–322.
40. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J, Daley GQ, Eggan K, Hochedlinger K, Thomson J, Wang W, Gao Y, Zhang K: **Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming.** *Nat Biotechnol* 2009, **27**:353–360.
41. Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ, Yu J, Hebrok M, Hochedlinger K, Costello JF, Song JS, Ramalho-Santos M: **Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells.** *Nat Cell Biol* 2011, **13**:541–549.
42. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nat Genet* 2009, **41**:1350–1353.
43. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger K: **Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells.** *Nat Biotechnol* 2010, **28**:848–855.
44. Bock C, Beerman I, Lien W-H, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A: **DNA methylation dynamics during in vivo differentiation of blood and skin stem cells.** *Mol Cell* 2012, **47**:633–647.
45. Boulting GL, Kiskinis E, Croft GF, Amoroso MW, Oakley DH, Wainger BJ, Williams DJ, Kahler DJ, Yamaki M, Davidow L, Rodolfa CT, Dimos JT, Mikkilineni S, MacDermott AB, Woolf CJ, Henderson CE, Wichterle H, Eggan K: **A functionally characterized test set of human induced pluripotent stem cells.** *Nat Biotechnol* 2011, **29**:279–286.
46. Stefanova VT, Grifo JA, Hansis C: **Derivation of novel genetically diverse human embryonic stem cell lines.** *Stem Cells Dev* 2012, **21**:1559–1570.
47. Carey BW, Markoulaki S, Hanna JH, Faddah DA, Buganim Y, Kim J, Ganz K, Steine EJ, Cassady JP, Creighton MP, Welstead GG, Gao Q, Jaenisch R: **Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells.** *Cell Stem Cell* 2011, **9**:588–598.
48. Ståhl PL, Lundeberg J: **Toward the single-hour high-quality genome.** *Annu Rev Biochem* 2012, **81**:359–378.
49. Brennand KJ, Simone A, Jou J, Gelboin-Burkhardt C, Tran N, Sangar S, Li Y, Mu Y, Chen G, Yu D, McCarthy S, Sebat J, Gage FH: **Modelling schizophrenia using human induced pluripotent stem cells.** *Nature* 2011, **473**:221–225.
50. Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y, Chen G, Gage FH, Muotri AR: **A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells.** *Cell* 2010, **143**:527–539.

doi:10.1186/scrt185

Cite this article as: Nestor and Noggle: Standardization of human stem cell pluripotency using bioinformatics. *Stem Cell Research & Therapy* 2013 4:37.