

Utilizing Regulatory Networks for Pluripotency Assessment in Stem Cells

Björn Brändl¹ · Bernhard M. Schuldt¹ · Lena Böhnke¹ · Oliver Keminer³ ·
Lea A. I. Vaas³ · Rainer Fischer² · Franz-Josef Müller¹ · Ole Pless³

Published online: 5 July 2016
© Springer International Publishing AG 2016

Abstract Pluripotency is a term in cell biology describing a unique state present in distinct stem cell lines, which were either established from the inner cell mass of the mammalian embryo or derived from somatic cells that have been reprogrammed to induced pluripotent stem cells. Pluripotent stem cells are continuously self-renewing, and their differentiation capacity enables them to develop into all derivatives of the three germ layers of a gastrulating embryo (endoderm, ectoderm, mesoderm). Both human embryonic stem cells (hESC) and human-induced pluripotent stem cells (hiPSC) are virtually indistinguishable, at least based on their global RNA expression patterns. Yet, after these in vitro cell cultures have been generated, the cell lines' pluripotent properties may change considerably on the genetic and/or epigenetic level as a consequence of long-term propagation. Among other unphysiological changes, cell lines might acquire aneuploidies, loose physiological imprinting marks, or develop

differentiation biases favoring one cell lineage over the other. As a result, stem cell researchers have to continuously monitor each stem cell line's integrity, transcriptional profile, and functional properties. Regulatory transcription factors, protein-protein interactions, and signaling networks govern the pluripotent state. As a consequence, emerging small- and large-scale perturbations to these gene regulatory networks mediate the outlined unfavorable changes to the pluripotent phenotype. Here, we describe a reliable bioinformatic framework called PluriTest for confirmation and assessment of pluripotency as an animal-free, fast, and inexpensive way based on genome-wide transcriptional RNA profiles from microarrays. Additionally, we discuss future developments using RNA expression profiling for pluripotency assessment.

Keywords hiPSC · Pluripotency · PluriTest · Transcriptional profiling · Microarray · RNA-seq

This article is part of the Topical Collection on *Age-related Stem Cell Modifiers*

Franz-Josef Müller and Ole Pless contributed equally to this work.

✉ Franz-Josef Müller
franz-josef.mueller@uksh.de

Björn Brändl
BjoemBraendl@gmx.de

Bernhard M. Schuldt
bernhard@bernhardschuldt.de

Lena Böhnke
l.boehnke@hotmail.de

Oliver Keminer
Oliver.keminer@ime.fraunhofer.de

Lea A. I. Vaas
Lea.Vaas@ime.fraunhofer.de

Rainer Fischer
rainer.fischer@ime.fraunhofer.de

¹ Zentrum für Integrative Psychiatrie (ZIP) gGmbH, Niemannsweg 147, 24105 Kiel, Germany

² Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Forckenbeckstraße 6, 52074 Aachen, Germany

³ Fraunhofer Institute for Molecular Biology and Applied Ecology - ScreeningPort (IME-SP), Schnackenburgallee 114, 22525 Hamburg, Germany

Introduction: Large-Scale iPSC Cell Generation and the Need for Pluripotency Assessment

Since Shinya Yamanaka and colleagues [1, 2] and James Thomson and colleagues [3] have reprogrammed, for the first time, human somatic cells in vitro towards pluripotency with defined factors, human-induced pluripotent stem cells (hiPSC) lines have become a powerful tool for studying stem cell fates and developmental processes. Furthermore, hiPSCs have been used to generate human disease models in vitro [4] and are already—10 years after their original description—being applied in clinical tissue regeneration [5•]. Current state-of-the-art methods for reprogramming and culturing PSCs are highly refined and thus enable to derive thousands of hiPSC lines every year [6•]. Moreover, large stem cell consortia have been funded and formed, planning to generate and bank up to 25,000 iPSC lines in the next 4 years for the purpose of disease modeling and drug discovery [7]. In every single case, once a pluripotent stem cell line has been established, the researcher has to confirm bona fide pluripotent differentiation potential in every given iPSC line before any further experiments should be started. Table 1 summarizes the currently available methodologies for pluripotency assessment of stem cell preparations.

Experimental Approaches for Pluripotency Assessment

In this section, we will briefly outline the two prevailing experimental approaches to prove pluripotency in murine and human pluripotent stem cells. The gold standard for mouse pluripotent stem cells is the ability for germline transmission, which demonstrates the ability to generate all somatic cell types and germ cells in vivo [15]. Experimentally, pluripotency can be tested most rigorously using the tetraploid complementation assay as the ultimate benchmark [16]. Successful completion of this test yields mice fully derived

from a given stem cell line if it was truly pluripotent. For hiPSC, such a rigorous assay is—for obvious reasons—not available. The teratoma assay is the “next best gold standard” for testing pluripotency in human stem cell preparations [15]. Methodically, human stem cells are injected into immunodeficient mice in order to form teratomas in vivo. A teratoma is an experimental tumor entity that consists of cells of all three germ layers (mesoderm, endoderm, ectoderm), primarily from tissues such as neurons, heart muscle, and secretory epithelia. The teratoma assay has several disadvantages with respect to time, complexity, and animal usage [17]. First, the teratoma assay requires several weeks from injection of the cells to final histological results. Second, a mouse facility and skilled technicians are necessary in order to obtain reliable test outcomes. Alternatively, this experiment has to be outsourced to commercial service providers, incurring additional costs. Third, established protocols varying from different stem cell facilities may result in incomparable results among different studies. Fourth, even if the teratoma assay is carried out without technical inconsistencies, it has been reported that qualitatively, some stem cell lines may form all three germ layers but at the same time may be quantitatively biased towards only one or two germ layers [14, 18, 19]. Last but not the least, experimentation using animals is expensive, inapplicable for screening large banks of stem cell lines, and ethically problematic [20]. In summary, the assay does not add substantial information on the genetic or epigenetic and resulting transcriptional integrity and possible differentiation biases of a tested stem cell line.

Protein Biomarkers for Pluripotency Assessment

Initial hope that specific cellular subtypes could be easily defined by a small set of significantly up-regulated genes (biomarker) has proven to be elusive, as these limited gene sets are not specific enough for the rigorous characterization of pluripotency. For example, POU5F1 (often referred to by

Table 1 Assays for pluripotency assessment

Assay type	Complexity	Cost	Standardizable	Sensitive	Specific	Ref.
qRT-PCR (“marker”)	Low	Low	Yes	Yes	No	[8]
FACS	Low	Low	No	Yes	No	[8]
Immunostaining	Low	Low	Yes	Yes	No	[8]
TaqMan scorecard (EB + qRT-PCR)	High	Medium	difficult	Yes	Yes	[9]
PluriTest	Medium	Medium	Yes	Yes	Yes	[10]
Epi-Pluri score	Medium	Medium	Yes	Yes	No	[11]
Embryoid body formation	High	Medium	difficult	Yes	No	[12]
Directed differentiation	Medium-high	Medium	Yes	Yes	Yes	[12]
Teratoma formation	Very high	High	difficult	No	Yes	[13]
TeratoScore (teratoma + microarray analysis)	Very high	High(er)	difficult	No	Yes	[14]

its former name OCT4) is a well-characterized pluripotency master transcription factor that is always expressed in stem cells but, in some cases, in nullipotent cell lines, i.e. cells that fail to develop into all cell types [21–23]. POU5F1/OCT4 is for example expressed in teratocarcinoma cell lines obtained from malignant human tissues [21, 24]. These cell lines express many if not nearly all of the genes usually associated with genomically normal hESC and hiPSC [10, 21, 24]. Historically, before murine [25] and then human [26] ESC became available, pluripotency was first discovered and studied in pluripotent stem cell lines derived from teratocarcinomas [27–30]. This makes clear that a simplistic curation of a gene list with bona fide pluripotency associated genes (such as OCT4/POU5F1, NANOG, LIN28, and REX1) may be sensitive enough to detect even very few PSCs in large populations of differentiated, somatic cell types [31] but lacks specificity to discriminate, e. g., fully reprogrammed, genetically, and epigenetically intact and truly pluripotent PSC lines from those with limited differentiation potential or even gross karyotypic abnormalities [32].

Gene Regulatory Networks for Pluripotency Assessment

This insight has led us and others to develop genome-wide bioinformatic approaches utilizing “complex” biomarkers based on gene regulatory networks (GRNs) reconstructed from transcriptional microarray databases such as the stem cell matrix [10, 24]: First, a global model of the transcriptional patterns present in a cell state of interest is trained from gene expression data originating from microarray or RNA-seq experiments. For this, a network of statistically and/or experimentally supported correlations and/or links between genes is being trained and sub-networks out of the global transcriptional graph are being selected that are suited best for discriminating one cell type from all other cell types [24]. For the identification of discriminant GRNs, it is paramount to have a valid definition of the cell type to be analyzed. So far, our understanding of the overall and global organization of the transcriptional landscape is incomplete, yet we expect significant advances in this exciting research area in the near future [33]. Without a reliable definition of a cell identity or definable state, an alternative approach proposed by us is to use high-content gene expression profiles with unsupervised clustering methods to identify (stem) cell types just from the measurable transcriptomic structure [24]. This approach ensures that only robustly differentiable cellular phenotypes are used for GRN discovery. We, for example found that hPSC are—based on their transcriptomic features—distinct from any other stem cell type thus contradicting a previously postulated, quintessential “stemness” as a salient and unifying feature of all stem cell types [34, 35]. Conversely, based on the global

transcriptomic structure, we were unable to differentiate fibroblasts from mesenchymal stem cells, while discovering an unexpected heterogeneity among stem cells all considered to belong to the neural lineage [24]. Once a stem cell phenotype has been delineated or can be functionally defined (e.g., by the teratoma assay), it is possible to reconstruct gene regulatory networks useful for discriminating one cell type from another [10, 24]. Independently, several bioinformatic and theoretical studies have proven the validity of network reconstruction approaches from high-content datasets [36–40]. In the following next step after network reconstruction (and in some cases exploratory analysis), the computationally derived discriminatory networks can be exploited as complex “biomarkers” for predictive tasks, such as estimating differentiation potential in stem cell lines as it has been outlined previously by us and others [10, 24, 32, 41–43].

Epigenetic Marks for Pluripotency Assessment

The first genome-wide reference map of DNA methylation became available in 2011, summarizing differences among a well-defined set of human ESC and iPSC lines [18, 19]. Here, the authors utilized DNA methylation data (obtained by reduced-representation bisulfite sequencing (RRBS)) from undifferentiated hPSC to derive a linear model of differentiation and a set of marker genes to predict an associated lineage bias. This work has received much attention and has moved towards commercial application (TaqMan[®] hPSC Scorecard[™] Assay provided by ThermoFisher Scientific in a revised version [9]). Furthermore, pluripotency assessment using computational models built on DNA methylation states of cells have already been implemented and could potentially be combined with PluriTest technologies [11]. The complexity of histone tail modifications and the “histone code” [44] of pluripotent cells has so far not been utilized for hPSC pluripotency assessment and quality control. Upon reprogramming of somatic cells into iPSCs, rapid, genome-wide changes (in particular at euchromatic H3K4me2 marks) have been reported at more than a thousand loci including large subsets of promoters and enhancers of pluripotency-related or developmentally regulated genes [45]. These chromatin regulatory events precede transcriptional changes within the corresponding loci and could therefore be used as very early predictors for loss of pluripotency and initiation of differentiation trajectories.

The Basic PluriTest Concept

Conceptually, the PluriTest algorithm is built on a stem cell classification and prediction framework based on non-negative matrix factorization (NMF) [46]. To this end, all

features of a given microarray platform (in the current technical implementation: probes on Illumina HT12v3 or HT12v4 arrays, 23,000 genes per sample) are “condensed” to relatively few dimensions (sometimes also referred to as “metagenes”) based on the information content and correlation patterns of feature groups in order to separate pluripotent stem cells from differentiated somatic cells. Such metagenes can be also interpreted in a network-context as sub-networks of features co-regulated by a “hidden” factor, such as, e.g., a single or more transcription factors binding to the promoter regions of each feature represented by a certain metagene [47]. In 2011, we reported a bioinformatic assay for pluripotency termed PluriTest [10]. The PluriTest platform enables screening of PSC lines in a high-throughput fashion based on their global gene expression patterns. As training dataset for the development of the PluriTest gene expression model, microarray expression profiles from teratoma assay validated hESC and hiPSC lines were used and contrasted with PSC-derived *in vitro* preparations, somatic stem cell lines, and terminally differentiated cell lines extracted from tissues. PluriTest results can be obtained within a few days after a stem cell sample has been prepared in the lab, while the process of uploading to the PluriTest website (www.pluritest.org) and the server-side microarray raw data processing requires only a few minutes. The PluriTest platform offers a microarray-based approach to match messenger RNA (mRNA) samples against a well-established data set of characterized pluripotent and non-pluripotent cell lines termed as stem cell matrix 2 (SCM2). The SCM2 database consists of more than 450 genome-wide transcriptional profiles including 223 hESC, 41 iPSC lines, somatic cell lines, and tissue. This dataset contains only optimally quality controlled, normalized, transformed, and filtered data for subsequent bioinformatic applications. Up to 12 samples can be uploaded in parallel with a subsequent transformation and normalization step using a lumi Bioconductor algorithm [48] that reduces microarray chip features by a principal component analysis termed non-negative matrix factorization (NMF). An additional step includes the comparison of the training data set (SCM2) against the reduced data set from uploaded user samples. The results are calculated in several minutes and illustrated in a PluriTest report page as summary. In principle, there are two closely related parameters depicted as Pluripotency Score and Novelty Score that summarize data analysis from samples based on empirically determined thresholds. The Pluripotency Score describes the similarity of a given sample compared to the SCM2. If highly positive values are reported (green value in Pluri Raw), this indicates that the sample is much more similar to pluripotent cells from SCM2 than somatic cells or tissue from this matrix (red value in Pluri Raw). Any samples that are specified to have a similar pluripotent signature compared to pluripotent cells from SCM2 may not necessarily reflect intact iPSC or hESC cell lines, as partially reprogrammed, karyotypically abnormal or

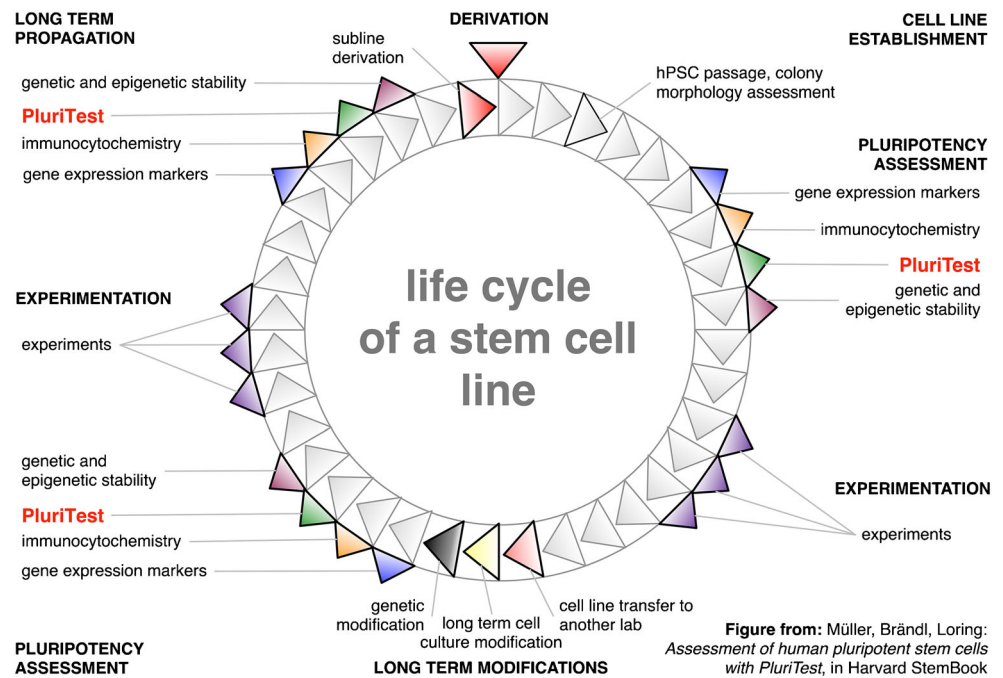
teratocarcinoma cells sometimes show similar high Pluripotency Scores due to their “pluripotent-like” expression pattern. A second parameter—the Novelty Score—was included to the PluriTest assay to match the test samples against well-established and karyotypically normal cells from the stem cell matrix to yield higher resolution of the similarity between samples and pluripotent cells from this matrix. If high Novelty Scores are observed, this indicates that test samples are more dissimilar to iPSC and hESC from the reference matrix. Alternatively, cells may contain an unknown pluripotent signature that currently cannot be explained or compared to highly characterized iPSC and hESC lines. Low Novelty Scores in contrast show a high similarity in their expression profiles thus qualifying them as pluripotent.

Future Development of PluriTest

Even if *in vivo* testing of hPSC in animals is omitted from the basic characterization of a clonal stem cell line, the process is still expensive and laborious and requires a large array of equipment, consumables, and expertise [12]. We have previously outlined the concepts of the lifecycle of a stem cell line (Fig. 1, [49]). A more detailed view at the basic characterization of a newly established pluripotent stem cell line illustrates the large toolset of complimentary and at times competing technologies for this task (Fig. 2a). We predict that within the near future, this resource intensive process can be consolidated into only a few, sequencing-based, analytical strategies (Fig. 2b). As sequencing costs have come down considerably, our envisioned stem cell characterization and quality control process will report—through predefined, “pipelined” bioinformatic analysis tools—additional sophisticated readouts to stem cell scientists. We believe that the first step towards this direction will be expanding tertiary analysis tools such as PluriTest towards accommodating of second generation sequencing data sets.

In recent years, next-generation sequencing technologies, including high-throughput RNA sequencing (RNA-seq), have revolutionized transcriptomic discoveries. The qualitative assessment of nucleotide-based signals, including genomic variation in health and disease, identification of transcript variants specific to defined cell types, identification of long non-coding RNAs (among other previously unknown RNA species), and detection of pathogens in complex body fluids, has been a tremendous achievement and vast improvement over microarray technologies. Therefore, RNA-seq is currently displacing microarrays as the preferred method for gene expression profiling [50]. As far as the quantitative assessment of biological systems is concerned, several open questions remain with both next-generation sequencing (NGS) and microarray technologies. Based on our experience, working with a diverse set of sequencing-based readouts of cell identity

Fig. 1 Life cycle of a pluripotent stem cell line. An hPSC line progresses through general stages throughout its life cycle: establishment, experimentation, long-term modification, and long-term propagation. An hPSC line can “fork” into different paths, for example, clonal lines and reporter lines have been derived from the WA09 line (e.g., WA09.2). At each of the illustrated steps, a thorough pluripotency assessment is recommended (Fig. 1 and the associated figure legend are from Müller et al. [49])



(including RNA-seq) and the data generated by the SEQC/MAQC-III consortium [51], RNA-seq (with spike in controls provided by the External RNA Control Consortium (ERCC) [52], and a broad variety of alignment and quantification strategies) does—somewhat surprisingly—not offer any predictive advantage over microarrays in the assessment of endpoints, even though all the mentioned qualitative readouts would suggest otherwise [53]. Genome-wide, unbiased datasets (from arrays or NGS technologies) capture so many data points that the resulting predictions can be extremely good—so good indeed that they perform equally well—even though RNA-seq would offer much more data depth on the qualitative level. NGS technology is superior when calling the genomic identity of a cell line using SNPs—information clearly unattainable from gene expression microarrays. In cancer, structural genomic information, e.g., on fusion transcripts provided by NGS, might allow for the rational development/application of targeted therapies. The endpoint prediction of genomic integrity of “normal” hiPSC lines, however, which stands at the core of all current PluriTest results, would not benefit from such information.

This technological refinement requires major adaptations of the original implementation of PluriTest and also provides a chance to implement novel functionalities based on the original PluriTest/SCM2 concept. Expanding this assay to RNA-seq is a logical evolution, yet the main challenge is not conceptual but technical: microarray platforms have matured to a highly reproducible and standardized technology, while “RNA-seq” can only be considered an umbrella term for a group of related technologies, comparable to the term “PCR”

encompassing genomic PCR, conventional RT-PCR, TaqMan PCR, and digital droplet PCR. A wide variety of RNA-seq platforms, protocols, and performance capabilities exists which has created the need for comprehensive reference data [51, 54]. Despite recent technological advancements with regard to sample handling and processing, e.g., the automation for DNA/RNA fragmentation and size selection or the library preparation workflow, we and others have observed an enormous technical variability among batches even when RNA-seq datasets were generated by the same core facility. This issue becomes significantly worse when datasets from different sites are combined. Overall, the field has already achieved a high concordance rate with regard to intra-platform and inter-platform variability for expression level comparisons [54]. Other challenges remain, e.g., the reliable detection of splice junctions and resulting transcript variants between RNA-seq platforms [54]. Therefore, it will be the key to further refine metrics for expression level comparisons and methods for identification of differentially expressed transcripts. Several approaches have been developed to date, but as of yet, no consensus exists on the best processing pipeline to use [55, 56]. Furthermore, due to the reduction of starting RNA material, required fascinating technological advancements have emerged including single cell transcriptomics, in particular transcriptional profiling of individual cells using nanoliter droplets [57••, 58••]. These “Drop-seq” technologies will enable researchers to take into account the heterogeneity of hPSC cultures and the relative contribution of (a) karyotypically abnormal cells which might gain a growth advantage and

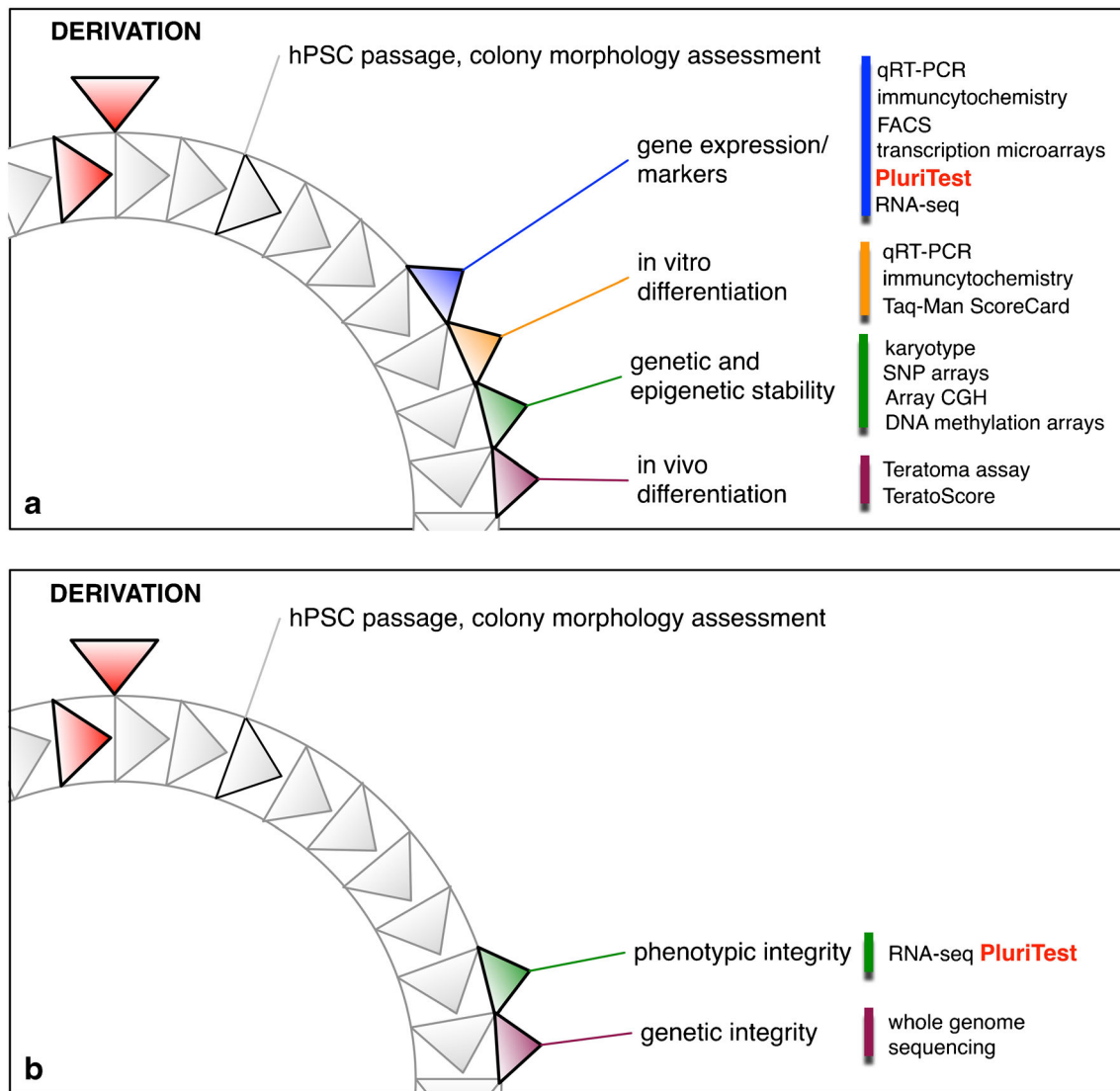


Fig. 2 Basic characterization of hPSC lines today and in the near future. **a** After a clone stem salon has been derived, it has to undergo several complementary characterization and quality control steps following

(b) cells which have entered early differentiation programs and have lost pluripotency.

We are confident that the PluriTest concept can be exploited to not only interrogate the pluripotent state of a given (single) cell but could also allow for the assessment of transcriptomic (and associated genomic) stability. Moreover, the routine application of human iPSC-derived terminally differentiated cells (e.g., cardiomyocytes, hepatocytes, neurons) in disease modeling, drug development, and clinical application for regenerative medicine is often hampered by insufficient maturity of the respective cell type. Major advances with regard to robustness, effectiveness, scalability, and reproducibility of differentiation paradigms for ectodermal (e.g., neurons [59]), mesodermal (e.g., cardiomyocytes [60, 61]), or endodermal (e.g., pancreatic β cells [62]) lineages from hiPSCs have been achieved in the past years. In addition, it

consensus in the stem cell field. We predict that as outlined in **b**, this process can be consolidated in only a few, high-throughput sequencing-based methods and subsequent tertiary analysis pipelines

will be of utmost importance to identify residual “contaminating” pluripotent cells in preparations of differentiated cells from hiPSC for clinical application in regenerative medicine [5•].

Conclusions

With tens of thousands of hiPSC lines being generated, the current need for reliable, unbiased, cost-efficient, and animal-free assays for pluripotency assessment of stem cell characteristics is very high. RNA expression signatures—as previously demonstrated using the PluriTest platform—provide a basis for pluripotency assessment of stem cells. Future applications of the concept extend to the interrogation of pluripotency and differentiation characteristics for in vitro disease modeling,

drug development, and clinical application for regenerative therapy.

Acknowledgments This work was supported by the Innovative Medicines Initiative (IMI) project “European Bank for induced Pluripotent Stem Cells” (EBiSC) and from the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research, grant 13GW0128 “PluriTest2”).

Compliance with Ethical Standards

Conflict of Interest Björn Brändl, Bernhard M. Schuldt, Lena Böhnke, Oliver Keminer, Lea A. I. Vaas, Rainer Fischer, and Ole Pless declare that they have no conflict of interest. Franz-Josef Müller is named inventor on the patent US Patent 8,442,772 B2 “Compositions and methods for defining cells”, which describes key aspects of PluriTest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Takahashi K et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 2007;131(5):861–72.
2. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126(4):663–76.
3. Yu J et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;318(5858):1917–20.
4. Grskovic M et al. Induced pluripotent stem cells—opportunities for disease modelling and drug discovery. *Nat Rev Drug Discov*. 2011;10(12):915–29.
5. Kamao H et al. Characterization of human induced pluripotent stem cell-derived retinal pigment epithelium cell sheets aiming for clinical application. *Stem Cell Reports*. 2014;2(2):205–18. **This study describes the first application of iPS cells in a clinical trial.**
6. Paull D et al. Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. *Nat Methods*. 2015;12(9):885–92. **This study describes an automated platform for iPS cell generation.**
7. McKernan R, Watt FM. What is the point of large-scale collections of human induced pluripotent stem cells? *Nat Biotechnol*. 2013;31(10):875–7.
8. International Stem Cell, I et al. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat Biotechnol*. 2007;25(7):803–16.
9. Tsankov AM et al. A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. *Nat Biotechnol*. 2015;33(11):1182–92.
10. Muller FJ et al. A bioinformatic assay for pluripotency in human cells. *Nat Methods*. 2011;8(4):315–7.
11. Lenz M et al. Epigenetic biomarker to support classification into pluripotent and non-pluripotent cells. *Sci Rep*. 2015;5:8973.
12. Marti M et al. Characterization of pluripotent stem cells. *Nat Protoc*. 2013;8(2):223–53.
13. Peterson SE et al. Teratoma generation in the testis capsule. *J Vis Exp*. 2011;57, e3177.
14. Avior Y, Biancotti JC, Benvenisty N. TeratoScore: assessing the differentiation potential of human pluripotent stem cells by quantitative expression analysis of teratomas. *Stem Cell Reports*. 2015;4(6):967–74.
15. Bradley A et al. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*. 1984;309(5965):255–6.
16. Nagy A et al. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A*. 1993;90(18):8424–8.
17. Muller FJ et al. A call to standardize teratoma assays used to define human pluripotent cell lines. *Cell Stem Cell*. 2010;6(5):412–4.
18. Bock C et al. Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell*. 2011;144(3):439–52.
19. Boulting GL et al. A functionally characterized test set of human induced pluripotent stem cells. *Nat Biotechnol*. 2011;29(3):279–86.
20. Buta C et al. Reconsidering pluripotency tests: do we still need teratoma assays? *Stem Cell Res*. 2013;11(1):552–62.
21. Josephson R et al. Qualification of embryonal carcinoma 2102Ep as a reference for human embryonic stem cell research. *Stem Cells*. 2007;25(2):437–46.
22. Scholer HR et al. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J*. 1989;8(9):2551–7.
23. Scholer HR et al. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J*. 1990;9(7):2185–95.
24. Muller FJ et al. Regulatory networks define phenotypic classes of human stem cell lines. *Nature*. 2008;455(7211):401–5.
25. Evans M. Origin of mouse embryonal carcinoma cells and the possibility of their direct isolation into tissue culture. *J Reprod Fertil*. 1981;62(2):625–31.
26. Thomson JA et al. Embryonic stem cell lines derived from human blastocysts. *Science*. 1998;282(5391):1145–7.
27. Andrews PW et al. A comparative study of eight cell lines derived from human testicular teratocarcinoma. *Int J Cancer*. 1980;26(3):269–80.
28. Martin GR, Evans MJ. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A*. 1975;72(4):1441–5.
29. Pera MF, Blasco Lafita MJ, Mills J. Cultured stem-cells from human testicular teratomas: the nature of human embryonal carcinoma, and its comparison with two types of yolk-sac carcinoma. *Int J Cancer*. 1987;40(3):334–43.
30. Stevens LC, Little CC. Spontaneous testicular teratomas in an inbred strain of mice. *Proc Natl Acad Sci U S A*. 1954;40(11):1080–7.
31. Kuroda T et al. Highly sensitive in vitro methods for detection of residual undifferentiated cells in retinal pigment epithelial cells derived from human iPS cells. *PLoS One*. 2012;7(5), e37342.
32. Williams R, Schuldt B, Muller FJ. A guide to stem cell identification: progress and challenges in system-wide predictive testing with complex biomarkers. *Bioessays*. 2011;33(11):880–90.
33. Lukk M et al. A global map of human gene expression. *Nat Biotechnol*. 2010;28(4):322–4.
34. Ivanova NB et al. A stem cell molecular signature. *Science*. 2002;298(5593):601–4.
35. Ramalho-Santos M et al. “Stemness”: transcriptional profiling of embryonic and adult stem cells. *Science*. 2002;298(5593):597–600.
36. Feizi A, Bordel S. Metabolic and protein interaction sub-networks controlling the proliferation rate of cancer cells and their impact on patient survival. *Sci Rep*. 2013;3:3041.
37. Gupta R et al. A computational framework for gene regulatory network inference that combines multiple methods and datasets. *BMC Syst Biol*. 2011;5:52.

38. Marbach D et al. Wisdom of crowds for robust gene network inference. *Nat Methods*. 2012;9(8):796–804.
39. Marbach D et al. Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci U S A*. 2010;107(14):6286–91.
40. Marbach D et al. Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks. *Genome Res*. 2012;22(7):1334–49.
41. Cahan P et al. Cell Net: network biology applied to stem cell engineering. *Cell*. 2014;158(4):903–15.
42. Cahan P et al. Defining cellular identity through network biology. *Cell Cycle*. 2014;13(21):3313–4.
43. Morris SA et al. Dissecting engineered cell types and enhancing cell fate conversion via Cell Net. *Cell*. 2014;158(4):889–902.
44. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001;293(5532):1074–80.
45. Koche RP et al. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell*. 2011;8(1):96–105.
46. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999;401(6755):788–91.
47. Brunet JP et al. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A*. 2004;101(12):4164–9.
48. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24(13):1547–8.
49. Muller FJ, Brandl B, Loring JF. Assessment of human pluripotent stem cells with PluriTest. In: *StemBook*. Cambridge (MA): Harvard Stem Cell Institute; 2008.
50. McGettigan PA. Transcriptomics in the RNA-seq era. *Curr Opin Chem Biol*. 2013;17(1):4–11.
51. Consortium SM-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol*. 2014;32(9):903–14.
52. Baker SC et al. The external RNA controls consortium: a progress report. *Nat Methods*. 2005;2(10):731–4.
53. Zhang W et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol*. 2015;16:133.
54. Li S et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol*. 2014;32(9):915–25.
55. Finotello F, Di Camillo B. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Brief Funct Genomics*. 2015;14(2):130–42.
56. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform*. 2015;16(1):59–70.
57. Klein AM et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015;161(5):1187–201. **This study provides the first description of how transcriptional profiles from single cells (here embryonic stem cells) can be generated using nanoliter droplets.**
58. Macosko EZ et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161(5):1202–14. **This study provides the first description of how transcriptional profiles from single cells (here murine retina cells) can be generated inexpensively in a significantly larger scale than before using nanoliter droplets.**
59. Chambers SM et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol*. 2009;27(3):275–80.
60. Kempf H et al. Cardiac differentiation of human pluripotent stem cells in scalable suspension culture. *Nat Protoc*. 2015;10(9):1345–61.
61. Rao J et al. Stepwise clearance of repressive roadblocks drives cardiac induction in human ESCs. *Cell Stem Cell*. 2016;18(3):341–53.
62. Pagliuca FW et al. Generation of functional human pancreatic beta cells in vitro. *Cell*. 2014;159(2):428–39.