Published in final edited form as:

Drug Discov Today Dis Models. 2012 ; 9(4): e161–e170. doi:10.1016/j.ddmod.2012.02.003.

Human ESC/iPSC-based “Omics” and Bioinformatics for Translational Research

Gerd Müller¹, Kirill V. Tarasov², Rebekah L. Gundry³, and Kenneth R. Boheler^{2,4}

¹Molecular Oncology, Medical School, University of Leipzig, Leipzig, Germany

²Molecular Cardiology and Stem Cell Unit, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA 21224

³The Department of Biochemistry and The Biotechnology and Bioengineering Center, Medical College of Wisconsin, Milwaukee, WI 53226

⁴Stem Cell and Regenerative Medicine, LKS Faculty of Medicine, University of Hong Kong, Hong Kong

Abstract

The establishment of human embryonic stem cell lines (hESCs) created the basis for new approaches in regenerative medicine and drug discovery. Despite the potential of hESCs for cell based therapies, ethical controversies limit their use. These obstacles could be overcome by induced pluripotent stem cells (iPSCs) that are generated by reprogramming somatic cells. Before iPSCs can be used for clinical applications, however, they must be thoroughly analyzed for aberrations in the genome, epigenome, transcriptome, and proteome. Here, we review how ‘omics’ technologies can be employed for a quantitative and definitive assessment of these cells.

Introduction

Pluripotent stem cells (PSCs) differentiate into all cell types found in the body. The best characterized and standard for PSCs are embryonic stem cells (ESC)[1], but experimentally-derived PSCs, known as induced PSCs (iPSCs), can be generated from almost any type of somatic cell through forced expression of pluripotency-promoting transcription factors or microRNAs [2–6]. The ease of generating iPSCs has fostered the idea of immunologically compatible patient-derived cells. iPSCs thus may represent a viable alternative to human ESCs (hESCs) as the primary source of pluripotent cells for regenerative medicine; however the advantages of iPSCs are counterbalanced by unresolved questions involving differences between the two cell types. Potential iPSC line defects include chromosomal abnormalities, altered gene expression and unanticipated aberrations in the epigenetic landscape and immunogenicity [7]. Taken together, these differences demonstrate that iPSCs must be carefully analyzed on molecular, cellular and functional levels before entering the clinic (Figure 1). Omics approaches, including genome- and proteome-based, offer platforms for fully characterizing and standardizing putative iPSC lines to address these issues of heterogeneity and safety.

Corresponding Author: Kenneth R. Boheler, Ph.D., Molecular Cardiology and Stem Cell Unit, Gerontology Research Center, National Institute on Aging, NIH, 5600 Nathan Shock Drive, Baltimore, MD 21224-6825, bohelerk@grc.nia.nih.gov, Phone: 410-558-8095, Fax: 410-558-8150.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Applications of ‘Omics’ to PSCs

The human genome project, which began in 1990, led to major technological advancements that included improved sequencing of the genome and a routine analysis of a cell’s DNA (SNPs, copy number variation, mutations), methylation and histone state (epigenome), RNA abundance (transcriptome) or protein content (proteome). Collectively these analyses, among others, have been termed ‘omics’ research endeavors that are unique from traditional experimental designs. This is because ‘omics’ approaches are often large-scale and data-driven, as opposed to purely hypothesis driven [8]. Data generated from ‘omics’ approaches, when combined across platforms, are useful in describing biological relationships related to experimental and cellular fluctuations. Consequently, the integration of multiple ‘omics’ approaches to understand a cell’s phenotype, permits an ‘integrated system’ that more fully describes a cell’s response to defined variables. Finally, ‘omics’ approaches require significant statistical and computational efforts to model dynamic systems that by their very nature interact on multiple levels within a cell. Extraction of valuable biological information from ‘omics’ data is challenging, but the results, when properly analyzed, show great potential in addressing some of the current problems associated with transplantation and stem cell-based therapies [9].

A quantitative and definitive assessment of human iPSC lines should be possible through the use of ‘omics’ techniques. Genome-wide evaluations will be useful for defining the state of putative iPSC lines, and robust statistical techniques should be valuable in pin-pointing possible differences/aberrations in lines relative to “gold-standard” ESC lines (Figure 2A). More specifically, genome-wide DNA sequencing should uncover any spontaneous DNA mutations that may result during reprogramming, while microarray analysis of RNA samples or RNA-Seq experiments can provide insights on variations in gene expression that may be indicative of residual epigenetic memory. Chromatin immunoprecipitation experiments (ChIP-chip or ChIP seq) and DNA methylation studies can reveal variations in chromatin structure and transcription factor binding. Proteomic studies may also be valuable in defining variations in protein levels between cells, but perhaps more importantly, this technique may be of great value in the development of immunophenotyping techniques that can be employed to isolate and characterize “authentic” iPSC lines. By studying cells at the ‘omics’ level it should be possible to obtain fingerprints of iPSCs for comparisons with standard ESC lines and to assess how the reprogramming process affects biological processes, thus solving many of the current problems surrounding possible differences among these forms of PSCs.

Genome and Epigenome

DNA mutations—Similar to problems observed in the sheep Dolly [10], cell autonomous genetic defects are present in reprogrammed iPSCs. Specifically, iPSC lines show an enrichment of mutations and chromosomal defects that may be introduced during the reprogramming process, independent of the reprogramming vectors, or by culture adaption (Figure 2B) [11–14]. As an example, Gore et al. sequenced the protein coding exons (exomes) of twenty-two human iPSC lines generated by 5 independent methods and nine parental fibroblast lines. On average, they found five protein-coding point mutations in the regions sampled or an estimated six protein coding point mutations per exome. Although the majority of the mutations were non-synonymous, nonsense or splice variants, many of the mutations occurred in genes with causative effects in cancers. At least half of the mutations were present in the parental fibroblast, but the remainder occurred spontaneously either during or after reprogramming. Howden et al. also assessed whether human iPSCs isolated from a patient with gyrate atrophy increased its mutational load with reprogramming [15]. In these cells, no abnormalities were detected by standard G-band metaphase analysis;

however, array comparative genomic hybridization and exome sequencing identified two deletions, one amplification, and nine mutations in protein coding regions when compared against the parental patient fibroblast cell line. They then performed exome sequencing on a gene-targeted iPSC clonal line that corrected the *OAT* point mutation present in this patient's DNA, and on a cassette-free iPSC clone. Somewhat surprisingly, the genomes proved remarkably stable, as no additional mutations or copy number variations were identified, excluding the targeted correction in the *OAT* locus and a single synonymous base-pair change. These findings led to the conclusion that iPSCs carry a significant mutational load from the parental line, but clonal events and prolonged culturing do not lead to a substantial increase in mutations.

Mutations may however occur at a higher frequency than previously reported. In unpublished work presented in a recent Stem Cell Research Symposium at the NIH, Paul Liu presented results from deep whole genome sequencing and high-density SNP array analysis. He reported results from episomal-vector reprogrammed hiPSC lines derived from two tissues of a single adult donor. The data revealed over 1000 single nucleotide substitutions in each iPSC cell line when compared to the parental cell sources. Although the majority of mutations were in non-coding sequences, 6 and 12 mutations, respectively were located in coding regions, and 34 and 22 variations respectively were identified in the 5' or 3' untranslated regions of genes. Another 362 and 709 differences were observed in intronic regions that might affect gene expression. The majority of mutations were not present in the exome, and SNP analysis was not sufficient to identify these mutations. Since the specific point mutations were not conserved among the two iPSC lines, sequence substitutions were not related to susceptible "hot spots" during reprogramming. Importantly, these results established the viability of iPSC line whole-genome sequencing, which has now become cost-effective and more widely accessible to researchers in this field.

DNA Methylation—Closely affiliated with the genome sequence are DNA modifiers (i.e., DNA methyltransferases (DNMTs)) that add methyl groups at the 5'-position of cytosine. During the reprogramming process, somatic cells reset their pattern of DNA methylation to an ESC-like state. More specifically, DNA of transcriptionally active genes like pluripotency and housekeeping genes is hypomethylated [16,17], while silenced genes are hypermethylated [17,18].

Various groups have now reported that ESCs exhibit unique methylation patterns and that iPSCs show modest variations in this pattern. This was perhaps best illustrated through the use of different cell types isolated and reprogrammed from the same mouse. Among these genetically identical iPSC clonal lines, the DNA methylation profiles reflected the cell type of origin, suggesting the presence of residual parental cell "epigenetic memory". Functionally, methylation pattern permitted efficient differentiation of iPSCs into the somatic cell type of origin, but these same lines showed reduced efficiencies of differentiation into other lineages. Although the causes are still unclear, it became apparent that these effects resulted from insufficient methylation (silencing) of genes normally expressed in the somatic cells from which the iPSC are derived, and insufficient demethylation (activation) of ESC specific genes [19–22]. In support of this assumption, the addition of 5-azacytidine (Aza), a DNA methylation inhibitor, to established iPSCs lines increased their differentiation potential and made them more like ESC [21]. Moreover, *pou5f1* and *nanog* gene promoters, which are highly methylated in somatic cells, remain largely methylated in partially reprogrammed cells (Figure 2C); however, exposure to Aza reactivates endogenous *pou5f1* gene expression. Prolonged cultivation of iPSCs also diminished differences in the methylation patterns between iPSCs and ESCs [19,22,23]. These latter findings suggest that reprogramming does not completely reverse the epigenetic landscape in early clonal isolates, a finding confirmed by Kim et al. [21], and that chromatin

remodeling is a gradual process that takes place over an extended period of time. Thus the state of DNA methylation at a genome level, and especially at some specific gene loci at a function of time, is likely to be indicative of the degree of reprogramming of iPSCs.

From an ‘omics’ perspective, current DNA methylation assays are often limited in their ability to characterize a large number of genomic targets. To overcome this limitation, investigators have described methods that capture genomic targets for single-molecule bisulfite sequencing at a single-nucleotide resolution. From these large-scale studies, chromosome-wide methylation patterns were similar, but cytosine methylation proved to be slightly more prevalent in the pluripotent cells than in the fibroblasts [24]. These data, in agreement with other studies, ultimately showed that iPSCs intrinsically display more methylation than embryonic stem cells. Further improvements in whole genome methylation sequencing based studies have now made it feasible to fully characterize the methylation status of iPSCs [25–27].

Histone Modifications—Somatic cells reprogrammed to iPSCs reset post-translational histone modifications to an ESC-like state, and like DNA methylation, histone modifications differ between ESCs and iPSCs. Histone deacetylase inhibitors (HDACs), such as trichostatin A (TSA) and valproic acid (VPA), used to promote active histone acetylation marks, can enhance iPSC reprogramming efficiency [28,29]. To maintain pluripotency, the promoters of genes encoding *Nanog*, *Oct4*, and *Sox2* show a high degree of histone 3 lysine 4 (H3K4) methylation resulting in a transcriptionally active state [18,30,31]. Partially reprogrammed iPSCs, however, retained some histone modification patterns of its parental cells, indicating an association of histone modification in iPSC memory status.

In genome-wide studies performed by ChIP-chip or ChIP-Seq based techniques, bivalent patterns of histone methylation have been described that distinguish PSCs from somatic cells (Figure 2C). In bivalent promoters, histone 3 is methylated at both lysines 4 and 27 (H3K27). Both modifications are prevalent at transcription start sites of numerous developmental genes whose expression is repressed in ESCs [30,31]. Although H3K4 methylation is associated with gene activation and H3K27 methylation by polycomb group protein complexes typically results in gene repression, bivalent promoters tended to be repressed, but at times “leaky” [32,33]. With differentiation, this pattern switches from a bivalent state to a monovalent state, which results either in transcriptionally active genes characterized by H3K4 methylation or to non-transcribed genes with H3K27 methylation state [34]. Thus the poised state was seen as a central requirement for undifferentiated ESCs to maintain their developmental potential. A number of other histone modifications are also known to affect gene activity, including the repressive H3K9me3, H4K20me3 marks and multiple targets of histone acetylation [33,35,36].

Recent results from Roeder and colleagues provide new insights into these regulatory mechanisms [37]. By studying components of SET1/MLL family complexes, these authors found that depletion of Dpy-30 and RbBP5 leads to a generalized reduction in H3K4 methylation in mouse ESC. In fact, H3K4 methylation levels were significantly reduced on the promoters of key stemness genes like *nanog*, *pou5f1*, *klf4*, and *sox2*, but mRNA abundance was not altered. The proliferation rate level of alkaline phosphatase in ESCs was also unaffected by knockdown of Dpy-30. In contrast, Dpy-30 knockdown resulted in profound defects when ESCs were forced to differentiate by LIF withdrawal or retinoic acid stimulation. The reduction in H3K4 methylation resulted in defects in lineage specification and impaired plasticity in transcriptional reprogramming. These data in particular illustrate why large-scale analyses may be required to characterize iPSCs prior to therapeutic usage. More specifically, modest changes in epigenetic regulation may not adversely affect PSC

self-renewal; however, some differences may have profound effects on the progeny generated upon differentiation.

Transcriptome

Microarrays and RNA-Seq—Whole-genome expression profiling is a commonly employed approach to compare and characterize different cell populations (Figure 2B). Consequently, microarray and RNA-Seq gene expression profiling have been employed to compare iPSCs with parental cells or ESCs. During the reprogramming process, many genes expressed in ESC are reactivated, including endogenous *pou5f1*, *nanog*, *lin28* and *fgf4*. A large percentage of these genes are only up-regulated during late stages of reprogramming, and in partially reprogrammed cells, a great deal of heterogeneity has been observed. Although the majority of genome-wide studies suggest that ESCs and iPSCs are nearly identical, Chin et al. reported that these cells could be distinguished by gene expression signatures [22]. Wang et al. [38] compared whole-genome microarray datasets from five studies [22,39–42] to show that differences may be related to the methods for reprogramming. Although most of the iPSC lines appeared similar to ESC lines, the transcriptomes of iPSCs generated with retroviruses differed much more from ESCs than those generated with episomal-reprogramming vectors.

Bock et al. recently created an integrated reference map of DNA methylation and gene expression patterns through a comprehensive evaluation of 20 human ESC and 12 iPSC lines [43]. The distribution of DNA methylation and mRNA abundance was calculated for individual genes or genomic regions. A “reference corridor” was developed that defined the range of methylation/expression for a given gene in pluripotent cells that could be used to establish transcriptome-based criteria to categorize pluripotent cells. When extended to other lines, this reference facilitates the identification of genes that fall outside the “reference corridor”. Although individual outliers may result from cultivation conditions, outliers could also be indicative of inappropriately functioning genes. Once identified, iPSC lines that inappropriately express these transcripts may need to be classified as failing to meet established transcriptome-based criteria for authentic pluripotent cells. Also, by combining this method with differentiation assays, this reference makes it possible to assay the quality and the usability of a given cell line for further applications.

Müller et al. subsequently made use of more than 450 genome-wide transcriptional profiles of stem cell lines, differentiated cells, and adult human tissues [44]. In these analyses, expression profiles from 223 hES and 41 hiPS cell lines were included. They developed a web-based open access tool called “PluriTest”, which allows for the identification of pluripotent cell lines based on gene expression data with a high degree of reliability. In addition, the algorithm is able to discriminate between pluripotent germ cell tumor lines and normal PSCs as well as between fully and partially reprogrammed iPSC lines. Currently, “PluriTest” is only able to process gene expression data, however, this algorithm should be applicable to methylation analyses or RNA sequencing data as well, which would further improve the reliability of these predictions.

Transcriptional profiles of ESCs are not only characterized by the expression levels of certain genes. The presence of alternatively spliced transcripts that encode protein variants essential for ESC self-renewal, pluripotency and differentiation are also critical [45–49]. Since up to 94% of human multi-exon genes are alternatively spliced [49], the number of unique proteins in a cell is much higher than the number of genes. Alternative splicing can influence protein binding affinities, enzymatic activity, localization, mRNA half-life and splice products that are quickly degraded by the nonsense-mediated mRNA decay [50,51]. Specifically, splice forms for OCT4 (Oct4b and Oct4b1), SALL4 (Sall4a and Sall4b), FOXP1 (FoxP1-ES and FoxP1) Tcf3 and Dnmt3b, which are active in ESCs, modulate

pluripotency versus cell-type specification [52–56]. The mechanisms of the regulation of ESC-specific alternative splicing remain to be elucidated; however, transcriptomic based analyses that assay the presence or absence of these splice variants are possible, principally by microarrays, but also by RNA-Seq.

microRNAs (miRs)—The discovery by Yu et al. that Lin-28 was critically involved in somatic cell reprogramming provided the first evidence that microRNAs (miRs) were essential to PSCs [57]. This is because, Lin 28 specifically blocks processing of pri-let-7, a miR that is critical to the regulation of developmental genes activated during early differentiation [58]. MiRs are single-stranded RNA molecules of 21–24 nucleotides that are fully or partially complementary to one or more mRNA molecules [59]. When associated with their targets, miRs generally repress translation or promote mRNA degradation to effectively down-regulate targeted gene expression. More recently, Judson et al. showed that the introduction of miRs specific to ESCs into somatic cells enhanced the production of mouse iPSCs [60]. More importantly, Anokye-Danso et al. recently showed that over-expression of the miR302/367 cluster, in the presence of HDAC inhibitors, reprogrammed mouse and human somatic cells to an iPSC state without the need for exogenous transcription factors like OCT4, SOX2 or NANOG. MiRs and the proper regulation of miR expression in PSCs is therefore critical to the proper function of iPSCs [6].

To date, a total of 1424 and 720 miRNA sequences have been identified in the human and mouse genomes, respectively (www.mirbase.org) [61]. Only a subset of these has been found in hESC or mESCs by cloning and sequencing from small RNA libraries [62,63]. Because a majority of miRs are expressed in somatic cells and only a minority in PSCs, these molecules can be used to evaluate the status of putative hiPSCs. Obviously, the presence of miRs in iPSCs that are typically only found in somatic cells would be contraindicated of fully reprogrammed iPSCs, and would bring their therapeutic viability into question.

Proteome and Metabolome

Morphologically and functionally, good quality iPSCs are nearly indistinguishable from hESCs, but as described above, a number of molecular indices show differences ranging from subtle to profound. The proteomic landscape of PSCs is not yet clearly defined, and until recently, the similarity of human ESCs and iPSCs at the protein level was unexplored (Figure 2D). To address this issue, we performed a broad-based comparison of more than 30 published proteomic studies of undifferentiated human and mouse PSCs [64]. These analyses resulted in a comprehensive resource of 7,471 and 7,281 proteins identified in mouse and human, respectively, of which 3,114 were found in both species. Unexpectedly, 63% of proteins were found in only one or two datasets, illustrating the variability among studies to describe and quantify this critical proteome. Also in 2011, Phanstiel et al. compared the proteomes and phosphoproteomes of two ESC lines, one iPSC line and one fibroblast cell line [65]. Statistical analyses revealed subtle, but significant and functionally related differences between proteins and phosphorylation sites in human ESCs and iPSCs. A number of these differences were thought to reflect residual regulation characteristics of an iPSCs' somatic origin. The authors also developed the Stem Cell–Omics Repository (SCOR), a resource designed to collate and display quantitative information across multiple planes of measurement. These recently established resources are expected to be very useful for this rapidly growing field of investigation.

The rate of enzymatic reactions in cells is also regulated by substrate concentration and their products, and in most organisms, no direct relationship exists between cellular metabolites (i.e., intermediates and products of metabolism) and gene function. Moreover, metabolite concentrations within cells vary as a consequence of genetic or physiological changes [66];

consequently, metabolomics, which focuses on the end-products of gene expression (metabolites) as well as other small proteins, toxins, chemicals, and organic compounds, may represent one approach that can provide functional insights regarding iPSC line states and variations relative to ESCs. To date, however, there are very few reports regarding PSCs and metabolomics. Only recently did Panopoulos et al. report that cellular bioenergetics of somatic cells convert from an oxidative state to a glycolytic state in reprogrammed cells, and that human iPSCs share a pluripotent metabolomic signature with ESCs that is distinct from parental cells [67]. They also identified several metabolites that differ between iPSCs and ESCs and novel metabolic pathways that play a critical role in regulating somatic cell reprogramming [67], thus validating the role of metabolomics in the identification of metabolic differences among PSCs. Metabolomics should, however, be considered a “cousin” to proteomics: the strategies and technologies are similar, but this “omic” technology really measures distinct types of biomolecules as well as small peptides.

Perhaps more importantly than either global proteomic or metabolomic approaches are focused analyses of PSC subproteomes. In particular, we have advocated the need for focused analyses of the surface proteome (i.e., surfaceome) of PSCs [64,68]. We expect surface proteins to be uniquely informative of a biological state for specific cell types, as evidenced by the use of surrogate markers to define hematopoietic stem cell (HSC) phenotypes. In fact, immunophenotyping, a process in which the functional potential of a cell is related to its surface marker expression pattern, has been used extensively to isolate subsets of bone marrow-derived HSCs for clinical interventions. Proof-of-principle studies for this concept were published in 2009 [69] where the authors used a targeted chemoproteomic strategy to identify 341 cell surface glycoproteins, including 53 CD-annotated proteins from mouse ESCs. The result of this antibody-independent strategy confirmed the expected decrease in LIF receptor and increase in FGF receptor 2 abundance during differentiation into the neural lineage. Such targeted strategies, when extended to reprogrammed cells, are likely to foster the rapid isolation and characterization of more homogeneous and therapeutically viable patient compatible iPSCs and will accelerate the development of disease models and clinical strategies for cell replacement therapy to treat human disease (Figure 3).

How could “omics” strategies be used routinely in the clinic?

Successful therapeutic approaches developed with ESC- or patient derived iPSC-progeny are predicted to be a future mainstay of modern medicine. Experiments in animal models have already proven that such therapeutic approaches hold a promising potential for regenerative medicine [70, 71], and recent results in macular degeneration suggest that the day is rapidly approaching for therapeutic applications in human [72].

Before these cells and their derivatives can be routinely employed clinically, careful molecular, immunological and functional assays must be performed. Conventional assays like G-banding are not sufficient since only large genetic abnormalities can be detected. In contrast, a combination of “omics” techniques represents a promising approach to assess these cells – particularly in pre-clinical stages. However, to date there are no definitive criteria on how to best define pluripotent cells and their specific cell derivatives, as well as assess the functional consequences of potential alterations in PSC genomes, epigenomes, transcriptomes, proteomes and metabolomes. Moreover, genetic defects are generally sporadic, thus complicating standard clinically accessible analyses. Even if a certain cell line has been evaluated with a combination of “omics” technologies and aberrations relative to a potential gold standard cell line have been pinpointed, it is not clear which changes or what degree of genetic changes are still acceptable for clinical use. Therefore, as an important step towards the clinical use of ESCs and iPSCs, well defined standards must be formulated to best identify cells suited for transplantation in patients and to minimize patient’s risks in

terms of tumorigenicity and immunogenicity. Thus, generally accepted guidelines regarding generation, expansion, manipulation, purification and evaluation of stem cells and stem cell derivatives must be developed (reviewed in [73]) before “omics” technologies can be routinely applied preclinically. But with that said, one “omics” technology has the potential of facilitating the daily use of PSCs and their derivatives for therapeutics. This is based on the identification of cell surface proteins as surrogate markers of a cell’s phenotype and/or function analogous to that already described for HSCs. The generation of non-genetic, immunophenotyping methods (Figure 3) for the isolation of defined cell states should permit the efficient isolation of desired cell types for clinical applications.

Conclusions

Differences between ESC and iPSC genomes, epigenomes, transcriptomes, proteomes and metabolomes are well established. While we have emphasized the differences, good quality iPSCs are almost identical to ESCs, but currently, there are no fully accepted criteria to make this determination in human cells. Omics approaches, especially when coupled with bioinformatics tools (Table 1) and functional assays are likely to play a critical role in establishing, characterizing and eventually defining which populations of iPSCs are appropriate for translational research.

Acknowledgments

The authors are supported by 4R00HL094708-03 (RLG), the Innovation Center at the Medical College of Wisconsin (RLG), the Intramural Research Program of the NIH, National Institute on Aging (KRB), and NIH Induced Pluripotent Stem Cell Center (NiPSCC) Pilot Study Award (KRB).

References

1. Wobus AM, Boheler KR. Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol Rev.* 2005; 85 (2):635–678. [PubMed: 15788707]
2. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006; 126(4):663–676. Epub 2006 Aug 2010. [PubMed: 16904174]
3. Okita K, et al. Generation of germline-competent induced pluripotent stem cells. *Nature.* 2007; 448(7151):313–317. Epub 2007 Jun 2006. [PubMed: 17554338]
4. Boheler KR. Pluripotency of human embryonic and induced pluripotent stem cells for cardiac and vascular regeneration. *Thrombosis and Haemostasis.* 2010; 104 (1):23–29. [PubMed: 20458433]
5. Lin SL, et al. Regulation of somatic cell reprogramming through inducible mir-302 expression. *Nucleic Acids Research.* 2011; 39 (3):1054–1065. [PubMed: 20870751]
6. Anokye-Danso F, et al. Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. *Cell Stem Cell.* 2011; 8 (4):376–388. [PubMed: 21474102]
7. Zhao TB, et al. Immunogenicity of induced pluripotent stem cells. *Nature.* 2011; 474 (7350):212–U251. [PubMed: 21572395]
8. Robert C. Microarray analysis of gene expression during early development: a cautionary overview. *Reproduction.* 2010; 140 (6):787–801. [PubMed: 20833752]
9. Perkins D, et al. Advances of genomic science and systems biology in renal transplantation: a review. *Seminars in Immunopathology.* 2011; 33 (2):211–218. [PubMed: 21318414]
10. Wilmut I, et al. Viable offspring derived from fetal and adult mammalian cells [see comments] [published erratum appears in *Nature* 1997 Mar 13; 386(6621):200]. *Nature.* 1997; 385:810–813. [PubMed: 9039911]
11. Mayshar Y, et al. Identification and Classification of Chromosomal Aberrations in Human Induced Pluripotent Stem Cells. *Cell Stem Cell.* 2010; 7 (4):521–531. [PubMed: 20887957]
12. Gore A, et al. Somatic coding mutations in human induced pluripotent stem cells. *Nature.* 2011; 471 (7336):63–U76. [PubMed: 21368825]

13. Hussein SM, et al. Copy number variation and selection during reprogramming to pluripotency. *Nature*. 2011; 471 (7336):58–U67. [PubMed: 21368824]
14. Laurent LC, et al. Dynamic Changes in the Copy Number of Pluripotency and Cell Proliferation Genes in Human ESCs and iPSCs during Reprogramming and Time in Culture. *Cell Stem Cell*. 2011; 8 (1):106–118. [PubMed: 21211785]
15. Howden SE, et al. Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108 (16):6537–6542. [PubMed: 21464322]
16. Weber M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*. 2007; 39 (4):457–466. [PubMed: 17334365]
17. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 2008; 454 (7205):766–U791. [PubMed: 18600261]
18. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007; 448 (7153):553–U552. [PubMed: 17603471]
19. Polo JM, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnology*. 2010; 28 (8):848–U130.
20. Ohi Y, et al. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biology*. 2011; 13 (5):541–U328.
21. Kim K, et al. Epigenetic memory in induced pluripotent stem cells. *Nature*. 2010; 467 (7313):285–U260. [PubMed: 20644535]
22. Chin MH, et al. Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. *Cell Stem Cell*. 2009; 5 (1):111–123. [PubMed: 19570518]
23. Nishino K, et al. DNA Methylation Dynamics in Human Induced Pluripotent Stem Cells over Time. *Plos Genetics*. 2011; 7(5)
24. Deng J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotechnology*. 2009; 27 (4):353–360.
25. Bock C, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnology*. 2010; 28 (10):1106–U1196.
26. Butcher LM, Beck S. AutoMeDIP-seq: A high-throughput, whole genome, DNA methylation assay. *Methods*. 2010; 52 (3):223–231. [PubMed: 20385236]
27. Li N, et al. Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods*. 2010; 52 (3):203–212. [PubMed: 20430099]
28. Huangfu DW, et al. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature Biotechnology*. 2008; 26 (11):1269–1275.
29. Huangfu DW, et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnology*. 2008; 26 (7):795–797.
30. Zhao XD, et al. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell*. 2007; 1 (3):286–298. [PubMed: 18371363]
31. Pan GJ, et al. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell*. 2007; 1 (3):299–312. [PubMed: 18371364]
32. Ringrose L, Paro R. Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. *Annual Review of Genetics*. 2004; 38:413–443.
33. Ringrose L, et al. Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of Polycomb complexes. *Molecular Cell*. 2004; 16 (4):641–653. [PubMed: 15546623]
34. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125 (2):315–326. [PubMed: 16630819]
35. Marion RM, et al. Telomeres Acquire Embryonic Stem Cell Characteristics in Induced Pluripotent Stem Cells. *Cell Stem Cell*. 2009; 4 (2):141–154. [PubMed: 19200803]
36. Mali P, et al. Butyrate Greatly Enhances Derivation of Human Induced Pluripotent Stem Cells by Promoting Epigenetic Remodeling and the Expression of Pluripotency-Associated Genes. *Stem Cells*. 2010; 28 (4):713–720. [PubMed: 20201064]

37. Jiang H, et al. Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. *Cell*. 2011; 144 (4):513–525. [PubMed: 21335234]
38. Wang Y, et al. A Transcriptional Roadmap to the Induction of Pluripotency in Somatic Cells. *Stem Cell Reviews and Reports*. 2010; 6 (2):282–296. [PubMed: 20336394]
39. Lowry WE, et al. Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105 (8): 2883–2888. [PubMed: 18287077]
40. Maherali N, et al. A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell*. 2008; 3 (3):340–345. [PubMed: 18786420]
41. Soldner F, et al. Parkinson's Disease Patient-Derived Induced Pluripotent Stem Cells Free of Viral Reprogramming Factors. *Cell*. 2009; 136 (5):964–977. [PubMed: 19269371]
42. Yu JY, et al. Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. *Science*. 2009; 324 (5928):797–801. [PubMed: 19325077]
43. 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):439–452. [PubMed: 21295703]
44. Muller FJ, et al. A bioinformatic assay for pluripotency in human cells. *Nature Methods*. 2011; 8 (4):315–U354. [PubMed: 21378979]
45. Lemischka IR, Pritsker M. Alternative splicing increases complexity of stem cell transcriptome. *Cell Cycle*. 2006; 5 (4):347–351. [PubMed: 16479168]
46. Pritsker M, et al. Diversification of stem cell molecular repertoire by alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102 (40):14290–14295. [PubMed: 16183747]
47. Salomonis N, et al. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107 (23):10514–10519. [PubMed: 20498046]
48. Yeo GW, et al. Alternative splicing events identified in human embryonic stem cells and neural progenitors. *Plos Computational Biology*. 2007; 3 (10):1951–1967. [PubMed: 17967047]
49. Wang ET, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456 (7221):470–476. [PubMed: 18978772]
50. Stamm S, et al. Function of alternative splicing. *Gene*. 2005; 344:1–20. [PubMed: 15656968]
51. Lewis BP, et al. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100 (1):189–192. [PubMed: 12502788]
52. Atlasi Y, et al. OCT4 Spliced Variants Are Differentially Expressed in Human Pluripotent and Nonpluripotent Cells. *Stem Cells*. 2008; 26 (12):3068–3074. [PubMed: 18787205]
53. Cauffman G, et al. POU5F1 isoforms show different expression patterns in human embryonic stem cells and preimplantation embryos. *Stem Cells*. 2006; 24 (12):2685–2691. [PubMed: 16916925]
54. Lee J, et al. The human OCT-4 isoforms differ in their ability to confer self-renewal. *Journal of Biological Chemistry*. 2006; 281 (44):33554–33565. [PubMed: 16951404]
55. Rao S, et al. Differential Roles of Sall4 Isoforms in Embryonic Stem Cell Pluripotency. *Molecular and Cellular Biology*. 2010; 30 (22):5364–5380. [PubMed: 20837710]
56. Gabut M, et al. An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell*. 2011; 147 (1):132–146. [PubMed: 21924763]
57. Yu J, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007; 318(5858):1917–1920. Epub 2007 Nov 1920. [PubMed: 18029452]
58. Viswanathan SR, et al. Selective blockade of MicroRNA processing by Lin28. *Science*. 2008; 320 (5872):97–100. [PubMed: 18292307]
59. Bushati N, Cohen SM. microRNA Functions. *Annu Rev Cell Dev Biol*. 2007; 21:21.
60. Judson RL, et al. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature Biotechnology*. 2009; 27 (5):459–461.
61. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*. 2011; 39:D152–D157. [PubMed: 21037258]

62. Houbaviy HB, et al. Embryonic stem cell-specific MicroRNAs. *Developmental Cell*. 2003; 5 (2): 351–358. [PubMed: 12919684]
63. Suh MR, et al. Human embryonic stem cells express a unique set of microRNAs. *Developmental Biology*. 2004; 270 (2):488–498. [PubMed: 15183728]
64. Gundry RL, et al. Pluripotent stem cell heterogeneity and the evolving role of proteomic technologies in stem cell biology. *Proteomics*. 2011; 11 (20):3947–3961. [PubMed: 21834136]
65. Phanstiel DH, et al. Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nature Methods*. 2011; 8 (10):821–U884. [PubMed: 21983960]
66. Raamsdonk LM, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol*. 2001; 19:45–50. [PubMed: 11135551]
67. Panopoulos AD, et al. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res*. 2012; 2012(1):168–177. [PubMed: 22064701]
68. Gundry RL, et al. A novel role for proteomics in the discovery of cell-surface markers on stem cells: Scratching the surface. *Proteomics Clinical Applications*. 2008; 2 (6):892–903. [PubMed: 19526049]
69. Wollscheid B, et al. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnology*. 2009; 27 (4):378–386.
70. Hanna J, et al. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science*. 2007; 318:1920–1923. [PubMed: 18063756]
71. Wernig M, et al. Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proc Natl Acad Sci U S A*. 2008; 105:5856–5861. [PubMed: 18391196]
72. Schwartz SD, et al. Embryonic stem cell trials for macular degeneration: a preliminary report. *Lancet*. 2012 Jan 24.2012 [Epub ahead of print].
73. Goldring CE, et al. Assessing the safety of stem cell therapeutics. *Cell Stem Cell*. 2011; 8:618–628. [PubMed: 21624806]

Highlights

- Pluripotent stem cell heterogeneity
- Genomics, transcriptomics, epigenomics, proteomics, metabolomics
- Omics techniques to evaluate and standardize pluripotent stem cells
- Cell surface proteins to characterize pluripotent stem cells

Figure 1.

Reprogramming of somatic cells (human fibroblasts (Fbs)) to induced pluripotent stem cells (hiPSCs). A number of reprogramming factor combinations are useful for generating iPSCs, including the 7 factors in episomal constructs used here (in blue). Typical characteristics of starting somatic cells and putative iPSCs are shown. HiPSC lines should be considered putative until a full analysis of potency is performed. This requires an analysis based on morphology, expression of pluripotency transcription (Txn) factors (OCT4, NANOG, SOX2), expression of surface markers (SSEA3, SSEA4, Tra-1-60, Tra-1-81) and teratoma assays. Alternatively “omic” based techniques, as described in the text, may be invaluable to quantitatively assess the quality of these cells.

Figure 2.

Omics approaches and PSCs. A) Human ESCs in culture and after immunostaining with cell surface antibodies to stage specific antigens. B) Omics techniques to evaluate DNA or RNA sequences or transcript abundance by microarrays reveal possible genetic mutations, changes in RNA expression or splicing. C) Central to epigenetic control is DNA chromatin and the nucleosome, which can be modified through a variety of enzymes and pathways. Epigenomic techniques are useful for monitoring changes to DNA methylation or histone methylation, both of which can affect “epigenetic memory”. In the example to the right, the methylation states of histone 3 at residues lysine 4 and 27 are shown. The simultaneous methylation of both residues is associated with a “poised” state, where transcription is generally low, but upon differentiation, gene transcription either can be rapidly activated or repressed. Additional traits that may differ between ESCs and iPSCs are DNA methylation (shown at the left), which represses gene transcription. D) Proteomic based techniques, while only modestly used to date on PSCs, hold great potential for characterizing differences among lines, as well as identifying proteins useful for isolating live cells with a well-defined degree of pluripotency.

Figure 3.

Targeted proteomic strategies for studying surface proteins, post-translational modifications, and protein isoforms are likely to contribute to the development of functionally defined stem cell populations that are applicable for therapy, disease modeling, and drug development.

Table 1

Selected Bioinformatics websites relevant to “Omics” research

URL	Key Features
General	
http://seqanswers.com/wiki/Software/list	Summary of useful software for data analysis
Pluripotency Test	
http://pluritest.org	Bioinformatic assay for pluripotency based on microarray data
Genomic/Transcriptomic	
http://genecodes.com	Sequencher software for DNA sequence assembly and analysis tools for DNA data sets
http://www.astrdbio.com	GenoMiner- NGS data analysis
http://www.avadis-ngs.com	Avadis NGS - desktop software platform for NGS (RNA-Seq, DNA-Seq, and ChIP-Seq analysis).
http://www.biobase-international.com	Genome Trax - identification of human genome variations of functional significance.
http://www.clcbio.com	CLC Genomics Workbench for analyzing and visualizing Next Generation Sequencing data
http://www.dnastar.com	Lasergene for next-gen sequence assembly and analysis.
http://www.genomatix.de	Genomatix Mining Station - mapping of NGS reads onto genomes, transcriptomes and splice junction libraries.
http://www.integromics.com	SeqSolve - for analysis of Next Generation Sequencing data
http://www.omicssoft.com	Array Studio - statistics and visualization for high dimensional data (NGS, microarray, SNP, CNV)
http://www.partek.com	Partek Genomics Suite - for analysis of microarray- and NGS data
http://www.phenosystems.com	GensearchNGS - software solution for Next Generation Sequencing
http://www.realtimegenomics.com	RTG Investigator - software for NGS sequence analysis
http://www.softgenetics.com	NextGENe - for analysis of next generation sequencing data
http://www.spiralgenetics.com	Spiral Studio - for analysis of next generation sequencing dataset
Proteomic	
http://scor.chem.wisc.edu/	Stem Cell Omics repository
http://www.ebi.ac.uk/pride/	General proteomic data repository that contains stem cell data
https://proteomecommons.org/tranche/	General proteomic data repository that contains stem cell data
http://www.peptideatlas.org/	Peptide data repository, especially useful for developing quantitative MS assays
http://gpsdb.thegpm.org/	General proteomic data repository that contains stem cell data

NGS- next generation sequence