Bioinformatics Studies on Induced Pluripotent Stem Cell

Yong Wang*

Academy of Mathematics and Systems Science, National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100190, China

Abstract: The induced pluripotent stem cells (iPSCs), generated from transcription factor-induced reprogramming, hold the great promise as the next generation materials for regenerative medicine. Intensive follow-up studies have accumulated a large amount of high-throughput data in transcription, proteomics, methylation, and other levels, which makes the computational studies feasible. Here we briefly review the recent bioinformatics efforts to study iPSCs. Specifically, we will summarize several comparison studies to determine how closely human iPSCs resemble human embryonic stem cells (ESCs) from sequence, gene expression profile, chromatin structure, DNA methylation, proteomics, and function aspects. Then computational methods to assess iPSC's pluripotency in a cost-effective yet accurate way are introduced. Finally, we will indicate the further biomolecular network studies to understand the underlying mechanism for cell reprogramming and the dynamics within this biological process.

Keywords: Induced pluripotent stem cells; transcriptional similarity; gene regulatory network, cell reprogramming.

INDUCED PLURIPOTENT STEM CELL AND ITS SIGNIFICANCE

Embryonic Stem cell (ESC) is important due to the following two properties. One is the self-renewal, which is the ability to go through numerous cycles of cell division while maintaining the undifferentiated state. The other one is pluripotency, which is the capacity to differentiate into specialized cell types [1]. The ability to give rise to any mature cell type makes the medical researchers believe that ESC has the potential to dramatically change the treatment of human disease. However, ESCs, available only from 5 to 7-day-old embryos, have raised moral and ethical issues and are limited in number.

Recently, induced pluripotent stem cells (iPSCs) have shed new lights upon the above dilemma. iPSC is a type of pluripotent stem cell artificially derived from an adult somatic cell, and is similar to natural pluripotent stem cells, such as ESC, in many aspects. The success of iPSCs demonstrates that the cell can be reprogrammed to a pluripotent state by the enforced expression of defined factors (Sox2, Oct4, Klf4, and c-Myc) [2-4]. To date, somatic cells from diverse adult tissues (i.e., endoderm, mesoderm, and ectoderm origins) have been successfully reprogrammed to iPSCs by multiple strategies, including drug-inducible systems, virus-free transposon mediated, recombinant proteins, and miRNAs. Similar to the ESCs, iPSCs have the ability of self-renewal and differentiation and can be potentially used on maintaining the growth of human organs and metabolism, repairing the body's aging, and curing diseases [5, 6]. Furthermore, iPSCs do not have restrictions on the ethics and source materials. Therefore, this remarkable discovery has attracted great attention for its potential applications to drug screening and analyses of disease mechanisms, and even as next generation materials for regenerative medicine. In addition, the mechanism under cell reprogramming provides far-reaching implications for biological sciences [5, 6].

OVERVIEW OF BIOINFORMATICS STUDIES ON IPSCs

In recent five years, induced pluripotent stem cells are widely studied, for their potential therapeutic use and inherent biological interest. As a result, a large amount of high-throughput data has been accumulated in transcription, proteomics, methylation, and other levels. In Table **1**, we summarize the available gene expression data for iPSCs. And Table **2** illustrates the available epigenetics data for iPSCs.

The large scale data make the bioinformatics studies on iPSCs feasible and provide plenty resources for information mining. Here we will briefly review the recent bioinformatics efforts to study iPSCs. As illustrated in Fig. (**1**), the exiting studies can be organized according to three key scientific questions for iPSC study. It's well known that ESC has full pluripotency and we named it as pluripotent state I. Then ESC changes to somatic cells through natural differentiation and development to a differentiated state, which can be artificially reprogrammed to an induced pluripotent state by over-expressing four transcriptional factors, which we name it as pluripotent state II. Pluripotent state I gains its pluripotency by nature while pluripotent state II gets its plutipotency by artificial cell reprogramming. Naturally we will ask whether the pluripotent state I is identical with pluripotent state II. This question is central to the safe application of iPSC in medicine and bioinformatics study can provide important hints to the final answer by comparing two states' high-throughput measurements. Then the next related big question is that what's the real definition and standard for iPSCs. Since there are many ways to induce iPSCs and standardization is necessary to assess the induced pluripotent cells before safe application. Here

*Address correspondence to this author at the Academy of Mathematics and Systems Science, National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100190, China; Tel: +8610-62616659; Fax: +8610-62616659; E-mail: ywang@amss.ac.cn

1574-8936/13 $58.00+.00

Table 1. The Available Gene Expression Data for iPSCs

GEO ID	Author	Year	Refercnec	Platform	# of Samples	Organism	Data Type
GSE16654	Chin *et al.*	2009	Cell Stem Cell,5,111-123	Affymetrix	**36**	Human	Expression profiling; Non-coding RNA profiling
GSE12390	Maherali *et al.*	2008	Cell Stem Cell,3(3):340-345	Affymetrix	**21**	Human	Expression profiling
GSE13828	Ebert *et al.*	2009	Nature,457(7227):277-280.	Affymetrix	**10**	Human	Expression profiling
GSE14711	Soldner *et al.*	2009	Cell,136(5):964-77	Affymetrix	**11**	Human	Expression profiling
GSE15175	Yu *et al.*	2009	Science, 324(5928):797-801	Affymetrix	**16**	Human	Expression profiling
GSE15176	Yu *et al.*	2009	Science, 324(5928):797-801	Affymetrix	**12**	Human	Expression profiling
GSE16093	Kim *et al.*	2009	Cell Stem Cell, 4(6):472-476	Affymetrix	**5**	Human	Expression profiling
GSE9832	Park *et al.*	2007	Nature, 451(7175):141-146	Affymetrix	**16**	Human	Expression profiling
GSE23402	Guenther *et al.*	2010	Cell Stem Cell, 7(2):249-57	Affymetrix	**72**	Human	Genome binding/occupancy profiling by high throughput sequencing;Expression profiling
GSE9709	Masaki *et al.*	2007	Stem Cell Res, 1(2):105-15	Affymetrix	**13**	Human	Expression profiling
GSE22392	Chin *et al.*	2010	Cell Stem Cell,7, 263-269	Affymetrix	**10**	Human	Expression profiling
GSE25970	Bock *et al.*	2011	Cell, 144(3):439-52	Affymetrix	**43**	Human	Expression profiling

Table 2. The Available Epigenetics Data for iPSCs

Author	Year	Reference	Data	Cell type	Organism	Methods
Chin *et al.*	2009	Cell Stem Cell, 5, 111-123	Histone H3k27me3 Histone H3k4me3	**ES, iPS**	Human	ChIP-chip
Broad Institute	2010	BI Human Reference Epigenome Mapping Project	DNA methylation	**ES, iPS**	Human	ChIP-Seq
Guenther *et al.*	2010	Cell Stem Cell, 7(2):249-57	Histone H3k27me3 Histone H3k4me3	**ES, iPS**	Human	ChIP-Seq
Masaki *et al.*	2007	Stem Cell Res, 1(2):105-15	Histone H3k27me3 Histone H3k4me3	**ES, iPS**	Human	ChIP-chip
Bock *et al.*	2011	Cell,144(3):439-452	DNA methylation	**ES, iPS**	Human	ChIP-chip
Takahashi *et al.*	2007	Cell, 131, 861-872	DNA methylation	**HDF, iPS**	Human	ChIP-chip

standardization is the process of establishing a technical standard, which could be a standard definition and standard test method. Bioinformatics method has the advantage in low-cost to use the high-throughput data to standardize iPSCs. One step further, reprogramming phenomenon is interesting in biological science and bioinformatics (systems biology) study is useful to reveal the regulatory mechanism underlying cell reprogramming.

In essence, existing bioinformatics studies are driven by the rapidly accumulated high-throughput data and the three key questions in Fig. (**1**). This review is organized into three parts by summarizing the researches around the above three questions. Specifically, we will summarize several comparison studies to determine how closely human iPSCs resemble human ESCs from sequence, gene expression profile, chromatin structure, DNA methylation, proteomics, and function aspects. Then computational methods to assess iPSC's pluripotency in a cost-effective yet accurate way are introduced. Finally, we will discuss the further biomolecular network studies to understand the underlying mechanism for cell reprogramming and the dynamics within this biological process.

TO BE OR NOT TO BE: THAT IS THE QUESTION

The first fundamental unresolved issue is whether and how the generated iPSCs are molecularly and functionally similar to ESCs. For mouse, a lot of progresses have been made in recent years. Many mouse iPSCs is shown to uniformly express pluripotency markers and can activate an Oct4-GFP reporter, but most lines are incapable of tetraploid complementation, which is the gold standard of a bona fide pluripotent stem cell line [5,6]. Thus, the consistent conclusion based on these experiments is that current transcription factor-mediated reprogramming methods fail to fully recreate authentic embryonic pluripotency in the majority of differentiated mouse cells [7].

However, the question remains in human: are human iPSCs (hiPSC) molecularly and functionally exactly the same as human ESCs (hESC), or to what content have they inherited the incomplete pluripotency. It is currently not possible to use embryo complementation-based measures to assess the pluripotency of hiPSCs, so instead characterization and comparison based on high-throughput data is still the only feasible method. As far as we know, several studies have conducted sequence [8], gene

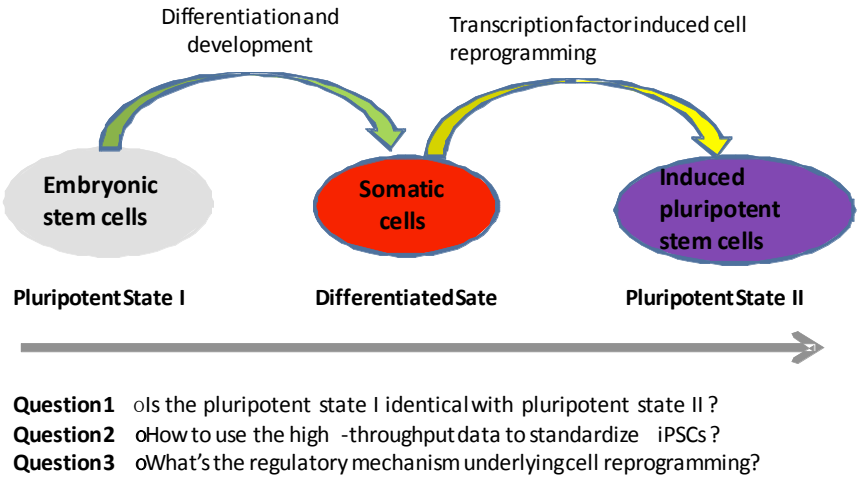

Fig. (1). Illustration of cell reprogramming and the main scientific questions for bioinformatics study on iPSCs. ESC changes to somatic cells through natural differentiation and development. And the differentiated state can be artificially reprogrammed to induced pluripotent state by over-expressing four transcriptional factors. In essence, the bioinformatics studies are driven by the rapidly accumulated high-throughput data and the three key scientific questions. This review is organized into three parts by summarizing the researches around these questions.

expression profile [9-15], chromatin structure [12], function [16], and DNA methylation [11,17] comparison of iPSC lines and ESCs to determine how closely hiPSCs resemble hESCs.

For example, it remains controversial whether the generated hiPSCs are transcriptionally similar (gene expression comparison) to embryonic stem cells. Chin *et al.* conducted a transcriptional comparison of five hiPSC lines and three hESC lines. Their results demonstrated that gene expression signature, a group of genes were consistently differentially expressed between hiPSC and hESC lines, are transcriptionally distinct from the hESC lines. And they concluded that hiPSCs should be treated as a unique subtype of pluripotent cell [10].

Recently, Guenther *et al.* have rigorously compared six hiPSC lines and six hESC lines and found only a few genes are consistently differentially expressed. Based on their results, hiPSCs have accurately reinstalled the transcriptional and epigenetic controls of ESCs and the variations observed do not serve to distinguish hiPSCs and hESCs [12]. Almost at the same time, Newman and Cooper collected data of 17 hESC lines and 67 hiPSC lines, which were produced in 7 independent experiments [13]. They then performed an unsupervised transcriptome clustering and found that hiPSCs are not grouped together with hESCs. Instead, hiPSCs and hESC lines tend to be clustered together if they are cultured in the same laboratory. Their analysis provided evidence that the observed difference is caused by the unique laboratory's culture condition and there are no consistent transcriptional differences between hiPSCs and hESCs.

As a prompt response, Chin *et al.* sticked to their original conclusions that the differences between hiPSCs and hESCs exist in gene expression [11]. Importantly, Chin *et al.* remind us that the lineage and genetic background of the starting cell type will significantly impact the properties of the resulting hiPSCs.

To solid the bioinformatic tools in the above analysis, Li *et al.* [18] recently applied the weighted P-value methods to a set of microarray data from published literature [10-15]. They aimed to integrate these data and find differentially expressed genes between hiPSC and hESC. A relatively smaller list of significant genes was found. They further combined their method with RankProd [19] and GeneMeta [20] and produced a top significantly different gene list. Pathway analysis of this list was done based on functional annotation clustering analysis using DAVID. Top functions are as follows, extracellular region, signal peptide, glycoprotein, cell migration, skeletal, face, head development, cell adhesion, extracellular matrix, endochondral bone morphogenesis, negative regulation of DNA binding, and blood vessel development.

In addition to RNA levels, epigenetic analyses have shown significant differences in the DNA methylation patterns between hiPSCs and hESCs [11,17,21]. Epigenetic reprogramming is a critical event in the generation of iPSCs [23]. In fact, the transcriptional memory of hiPSCs could be partially explained by the incomplete DNA methylation at the promotor regions of somatic genes. In addition, non-coding miRNAs have an important role in the underlying mechanisms of cell reprogramming. miRNA profiles between hESCs and hiPSCs were compared and a signature in the expression of the miR-371/372/373 cluster was found [22]. Finally, genetic integrity was also studied and it was found that the reprogramming process could induce several genomic abnormalities [10].

Meanwhile, proteomic and phosphoproteomic similarity of hESCs and hiPSCs at the protein level is also addressed. Douglas H Phanstiel *et al.* combined isobaric tagging, high-mass-accuracy mass spectrometry, and a recently developed software tool to compare two ESC lines, one iPSC line, and one fibroblast cell line. Rigorous statistical analysis revealed significant and functionally related differences between proteins and phosphorylation sites in hESCs and hiPSCs, which may reflect residual regulation characteristic of iPSCs' somatic origin [24]. Javier Munoz *et al.* present an in-depth quantitative mass spectrometry-based analysis of hESCs, two different hiPSCs and their precursor fibroblast cell lines. Their results confirmed the high similarity of hESCs and hiPSCS at the proteome level. A small group of

58 proteins, mainly related to metabolism, antigen processing, and cell adhesion, was found significantly differentially expressed between hiPSCs and hESCs [25].

Taken together, the above discussion concerns the extent of pluripotency of hiPSCs, and shows this is a topic that has substantial implications for the use of hiPSCs in the laboratory and the clinic [12-15]. Whatever, it is fundamentally urgent to find the cause for the considerable discrepancies between the results reported by these previous studies, especially from the two aspects, the lineage and genetic background of the starting cell type and lab-specific induction and culture conditions. Furthermore, it remains an open problem that whether these differences are going to be consequential. Furthermore, these results show that similarity based on high throughput profile is still the convenient and economic way when we are attempting to ascertain the pluripotency of hiPSCs. Therefore it's in pressing need for more bioinformatics studies.

COMPUTATIONALLY STANDARDIZE THE IPSCs

Another related question is how to standardize the iPSCs, i.e., to assess iPSC's pluripotency, utility, and clinical safety of iPSCs. The experimental standard for pluripotency is based on the ability to generate a complex variety of tissues in tumors. However, the generation of teratomas is technically challenging, resource-intensive, primarily qualitative and difficult to standardize. Additionally given the rapid increase in generation of iPSC in many ways, there is a pressing need for a cost-effective, animal-free alternative to the teratoma assay for assessing pluripotency in human cells. Particularly, the low cost and accessibility of microarray-based gene expression datasets makes transcription profiling an attractive alternative. The above analysis clearly shows that a well-designed expression microarray experiment can capture the fact about what happens to cells under reprogramming. This allows to use microarray data to standardize hiPSCs [11]. In addition, computational methods based on microarray data hold the promise to be able to delineate stem cell phenotypes and further predict the presence or absence of pluripotent features for unknown samples of cells.

Franz-Josef Müller *et al.* proposed a robust open-access bioinformatic method, PluriTest, to assess pluripotency in human cells based on their gene expression profiles [26]. Starting from the training gene expression data with appropriate transformation and normalization, they used nonnegative matrix factorization (NMF) for dimension reduction and to identify unexpected patterns engrained in the datasets under a machine learning framework. Then the pluripotency of an unknown, potentially pluripotent sample is assessed by comparison of a 'query gene expression profile' from the sample to the model derived from the training dataset.

To provide an informative and practically useful method for high-throughput cell-line characterization, Bock *et al.* computationally integrated several genomic assays into a scorecard [11] that measures the quality and utility of any human pluripotent cell line. Their scorecard is the combination of deviation scorecard and lineage scorecard. The deviation scorecard is based on Tukey's outlier filter, denoting all genes as putative outliers whose DNA methylation or gene expression level is significantly different. The lineage scorecard performs a parametric gene set enrichment analysis on t-scores obtained from a pairwise comparison between the iPSCs of interest and the reference of ESCs.

Similarly, Zhang B. *et al.* presented a novel supervised method for the assessment of the quality of iPSCs by estimating the gene expression profile using a 2-D ''Differentiation-index coordinate''. It consists of two ''developing lines'' that reflects the directions of ESC differentiation and the changes of cell states during differentiation. Moreover, the Distance index is defined to indicate the qualities of iPSCs, which based on the projection distance of iPSCs-ESCs and iPSCs-fibroblasts [27].

The characterization of hiPS cell lines can be carried out from different levels varying from sequence, epigenetic factors such as DNA methylation and chromatin structures, transcriptome, proteome, and even function. These sources are all important to finally standardize hiPSCs but they are important in different ways. Firstly, the current framework for gene expression data could be applied to any unbiased high-content dataset, such as global DNA methylation analysis or RNA sequencing data, provided that there is sufficient representation of a defined target phenotype in the training dataset. Secondly, integration of multi-layer data sources is expected to provide more reliable assessment. As a proof of concept study, Bock *et al.* showed that combining DNA methylation and gene expression profiling with bioinformatic comparison provides a quick and comprehensive method for excluding iPSC lines that could be problematic for an intended application. Thirdly, the current supervised framework will be further extended to unsupervised and transparent predictive model, where ESCs serve as a typical class to have pluripotency instead of gold-standard.

NETWORK BIOLOGY STUDY ON CELL REPROGRAMMING MECHANISM

Understanding the mechanism underlying cell reprogramming is one of the key steps for iPSCs before safely moving to clinical applications. Currently the underlying mechanism for cell reprogramming remains unknown and the regulatory interactions within this biological process have not been worked out. In particular from the biomolecular network viewpoint [28-35], it is not clear how the four factors initialize the reprogramming process, propagate the information in a fine tuned way, and finally lead to the dramatic phenotype changes.

Networks, especially gene regulatory networks, naturally serve as the powerful tool to understand the cell reprogramming mechanism. The reasons are in two folds. First, gene regulation is one of the dominant factors for the mechanistic picture of cell reprogramming. It's well-known that reprogramming is initialized by introducing four transcriptional factors, which will activate or depress thousand of targets and then trigger the dramastic change of gene expression landscape. If we can reconstruct the regulatory interactions during the reprogramming process, we then know how these factors regulate each other and

interact with epigenetic control factors to form a large gene regulatory network. Furthermore it helps to understand how the four factors propagate the cell reprogramming information in the gene regulatory network and lead to the dramatic phenotype changes. Secondly, a large amount of data has been accumulated in transcription level and makes the inference of large scale gene regulatory network feasible (Tables **1** and **2**). Those well-designed high throughput experiments can bring us rich information under cell reprogramming and help to reveal the causal regulatory relationships among genes.

REVERSE ENGINEERING STRATEGY

The central task for network study is to reverse engineering gene regulatory networks in an accurate and reliable manner by analyzing and integrating the available high throughput data for iPSCs. As illustrated in Fig. (**2**), integration of dynamical data and static data from different level will infer the active network underlying the multi-step reprogramming procedure, and finally lead to the understanding of mechanism.

Currently, there are several existing efforts to study gene regulatory networks for cell reprogramming. One direction is to apply the biological technologies to obtain the location (ChIP-chip/ChIP-Seq) data sets to reconstruct a part of this network [36]. In Table **3**, we listed the current available regulatory interaction data by ChIP-chip or ChIP-Seq for iPSCs. One potential limitation for application is that the assembled network from experimental data is largely biased to the well-known factors.

Recently, a novel framework called "network screening" has been applied to detect the active subnetworks for cell reprogramming by integrating ChIP-chip data or existing molecular interactions with conditional gene expression data [37]. These reconstructed networks demonstrate their power to reveal important biological insights, for example to find new important genes in reprogramming.

However, the limitations of those reconstructed network are also clear. Firstly, those networks are from existing knowledge, are small in scale, and are only part of the whole-genome network. They can only offer very limited information since cell reprogramming is such a dramatic phenotype change to lead to about 10,000 differentially expressed genes [37]. A large, even whole genome, regulatory network is necessary for a mechanistic picture. Secondly, current networks are reconstructed by using static data, i. e., the gene expression data and location data are measured after the cells get the pluripotency. As a result, those networks fail to capture the dynamic process during the induction.

As a pilot study, Duren Zhana *et al.* employed a mathematical modeling or systems biology method to reconstruct whole-genome regulatory networks [38]. Particularly, they computationally analyzed newly published time course gene expression data during reprogramming [39]. The expression data is from [39] and measure throughout reprogramming of MEF to iPSC using a Dox-inducible promoter. In this data, MEFs were treated with Dox in mES media to turn on the Oct4, Klf4, cMyc, Sox2. Total RNA was extracted at day 0 (no Dox), day 2, 5, 8, 11, 16 and 21 (with Dox) and day 30 (Dox-independent secondary iPS). Temporal analysis of this time course data revealed that reprogramming is a multi-step process that is characterized by initiation, maturation, and stabilization phases. From gene regulatory network perspective, further analysis on this dynamic data is expected to understand the process of reprogramming and in particular the master regulators and regulatory interactions that control progression to a stable pluripotent state [38].

FORWARD ENGINEERING STRATEGY

One fascinating aspect of network study is to provide a mechanistic picture of cellular reprogramming for comprehensive understanding. As illustrated in Fig. (**3**), simplified biological network serves as the starting point for theoretical studies based on differential equations, further helps to understand the mechanisms of how cellular reprogramming is achieved. With the network information, epigenetic landscape can be defined and estimated to explain cell differentiation during development and cell fate reprogramming. The landscape concept has been widely

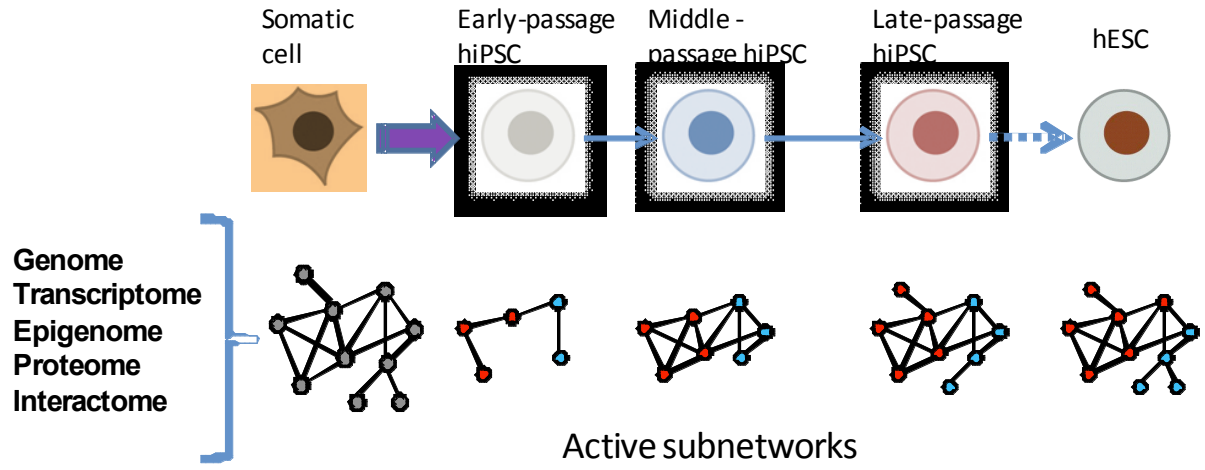

Fig. (2). Reverse engineering the dynamical networks during cell reprogramming, by integrating multi-layer omics data, helps to understand the cell reprogramming mechanism.

Table 3. The Available Regulatory Interaction Data for iPSCs

Author	Year	Reference	Methods	Cell Type	Orgnism	Number of Target Genes			
						Oct4	Sox2	Klf4	cMyc
Boyer *et al.*	2005	Cell, 122(6):947-56	ChIP-chip	ES	Human	623	1279	/	/
Loh *et al.*	2006	Nat Genet, 38(4):431-40	ChIP-chip	ES	Mouse	1083	/	/	/
Chen *et al.*	2008	Cell,133(6):1106-17	ChIP-Seq	ES	Mouse	/	/	/	/
Jiang *et al.*	2008	Nat Cell Biol, 10(3):353-60	ChIP-chip	ES	Mouse	/	/	894	/
Kim *et al.*	2008	Cell, 132(6):1049-61	ChIP-chip	ES	Mouse	783	819	1790	3542
Liu *et al.*	2008	Cell Res, 18(2):1177-1189	ChIP-chip	ES	Mouse	904	864	1505	869
Sridharan *et al.*	2009	Cell,136,364-377	ChIP-chip	ES, piPS, iPS	Mouse	847, 772, 1125	1026, 662, 755	1382, 771, 1207	2051, 1317, 2040
Huang *et al.*	2009	Cell Res, 19:1127–1138	ChIP-chip	iPS	Mouse	1388	1372	1832	2531

appreciated in protein folding/binding and more recently in genetic network analysis [40]. It has been shown as powerful model for theoretical understanding of development from a global and dynamical system viewpoint. One challenge is that the current methods of calculating network landscapes are time consuming and thus limited to small networks (often <20 genes) [40].

As a pioneering example, Chang *et al.* curated a network model (52 proteins and 124 interactions, including 85 activations and 39 repressions) to reconstruct cell reprogramming landscape. They firstly collected evidence from literature and manually constructed a genetic network involved in regulating pluripotency and hESC differentiation. Then the cell reprogramming landscape is computationally estimated, and finally reprogramming recipes are systematically searched to improve reprogramming efficiency. This work demonstrates that it is feasible to estimate the landscape in the cell-state space and monitor the trajectory of cellular reprogramming from a differentiated cell to an iPSC. In this sense, this work provides not only practical recipes for iPSC generation but also theoretical understanding of the reprogramming process. Though the network is limited in size, it shows the power of network study in cell reprogramming mechanism by introducing landscape concept.

CONCLUSION

Taken together, we demonstrated that bioinformatics efforts (computational biology or systems biology study) are helpful to clarify the differences between iPSCs and ESCs and to further reveal the cell reprogramming mechanism. We believe that the availability of this large compendium data will further drive bioinformatics studies and will finally provide a valuable baseline for the stem cell community to study gene regulation, cell reprogramming modeling, and integration with the public data resources in complex diseases, such as cancer stem cell.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

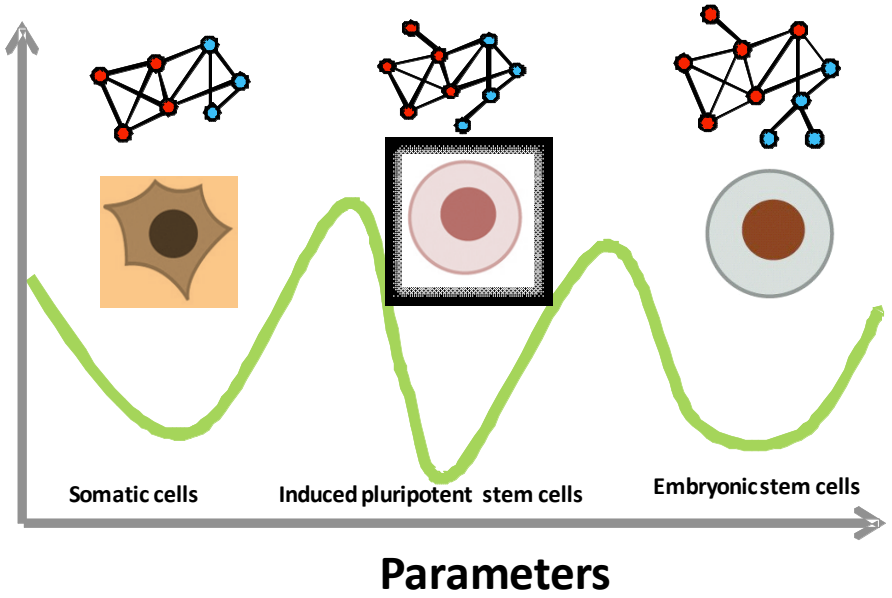

Fig. (3). Network is a powerful tool to study the dynamics from systems level by network landscape concept.

ACKNOWLEDGEMENTS

We thank Prof. Luonan Chen, Prof. Katsuhisa Horimoto, and Shigeru Saito for inspiring discussions. Also we thank the reviewers for valuable suggestions. This work is supported by National Natural Science Foundation of China (NSFC) under Grant 61171007 and 11131009.

REFERENCES

[1] Thomson JA, Itskovitz-Eldor J, Shapiro SS, *et al.* Embryonic stem cell lines derived from human blastocysts. Science 1998; 282: 1145–1147.

[2] Takahashi K, Tanabe K, Ohnuki M, *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007; 131: 861-872.

[3] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006; 126:663-676.

[4] Yu J, Vodyanik MA, Smuga-Otto K, *et al.* Induced pluripotent stem cell lines derived from human somatic cells. Science 2007; 318:1917.

[5] Amabile G, Meissner A. Induced pluripotent stem cells: current progress and potential for regenerative medicine. Trends Mol Med 2009; 15: 59-68.

[6] Li M, Chen M, Han W, Fu X. How far are induced pluripotent stem cells from the clinic? Ageing Res Rev 2010; 9: 257-264.

[7] Stadtfeld M, Apostolou E, Akutsu H, *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature 2010; 465: 175-181.

[8] Zhao X-Y, Li W, Lv Z, *et al.* iPS cells produce viable mice through tetraploid complementation. Nature 2009; 461: 86-90.

[9] Loh KM, Lim B. Recreating Pluripotency? Cell Stem Cell 2010; 7: 137-139.

[10] Laurent LC, Ulitsky I, Slavin I, *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: 106-118.

[11] Bock C, Kiskinis E, Verstappen G, *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell 2011; 144: 439-452.

[12] Chin MH, Mason MJ, Xie W, *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell 2009; 5: 111-123.

[13] Chin MH, Pellegrini M, Plath K, Lowry WE. Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. Cell Stem Cell 2010; 7: 263-269.

[14] Guenther MG, Frampton GM, Soldner F, *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell 2010; 7: 249-257.

[15] Newman AM, Cooper JB. Lab-Specific gene expression signatures in pluripotent stem cells. Cell Stem Cell 2010; 7: 258-262.

[16] Boulting GL, Kiskinis E, Croft GF, *et al.* A functionally characterized test set of human induced pluripotent stem cells. Nat Biotech 2011; 29: 279–286

[17] Lister R, Pelizzola M, Kida YS, *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 2011; 471: 68–73

[18] Li Y. Ghosh D. Assumption weighting for incorporating heterogeneity into meta-analysis of genomic data. Bioinformatics 2012; 28(6): 807-814.

[19] Choi JK, Yu U, Kim S, Yoo OJ. Combining multiple microarray studies and modeling interstudy variation. Bioinformatics 2003; 19: 84-90.

[20] Hong F, Breitling R, McEntee CW, *et al.* RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 2006; 22: 2825-2827.

[21] Polo JM, Liu S, Figueroa ME, *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol 2010; 28: 8.

[22] Wilson KD, Venkatasubrahmanyam S, Jia F, *et al.* MicroRNA profiling of human-induced pluripotent stem cells. Nature 2009; 18(5): 749-58

[23] Nishino K, Toyoda M, Yamazaki-Inoue M, *et al.* DNA Methylation Dynamics in Human Induced Pluripotent Stem Cells over Time. PLoS Genet 2011; 7(5): e1002085.

[24] Phanstiel DH, Brumbaugh J, Wenger CD, *et al.* Proteomic and phosphoproteomic comparison of human ES and iPS cells, Nat Methods 2011; 8: 821–827.

[25] Munoz Z, Low TY, Kok YJ, *et al.* The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. Mol Syst Biol 2011; 7: 550.

[26] Muller FJ, Schuldt BM, Williams R, *et al.* A bioinformatic assay for pluripotency in human cells, Nat Methods 2011; 8:315–317

[27] Zhang B, Chen B, Wu T, *et al.* Estimating the quality of reprogrammed cells using ES cell differentiation expression patterns. PLoS ONE 2011; 6(1): e15336.

[28] Jong H. de. Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 2002; 9: 67-103.

[29] Gardner TS, Faith JJ. Reverse-engineering transcription control networks. Phy Life Rev 2005; 2: 65-88.

[30] Chen L, Wang RS, Zhang XS. Biomolecular Networks: Methods and Applications in Systems Biology. John Wiley & Sons, Hoboken, New Jersey. July, 2009.

[31] Hempel H, Koseska A, Kurths J, Nikoloski Z. Inner composition alignment for inferring directed networks from short time series. Phy Rev Lett 2011; 107: 054101.

[32] Wang Y, Joshi T, Xu D, Zhang XS, Chen L. Inferring gene regulatory networks from multiple microarray datasets. Bioinformatics 2006; 22(19): 2413-2420.

[33] Wang Y, Joshi T, Xu D, Zhang XS, Chen L. Supervised inference of gene regulatory networks by linear programming. Lecture Notes Bioinform 2006; 4115: 551-561.

[34] Wang Y, Zhang XS, Chen L. A network biology study on circadian rhythm by integrating various omics data. OMICS: A. J Int Biol 2009; 13(4): 313-24.

[35] Wang Y, Zhang X-S, Xia Y. Predicting eukaryotic transcriptional cooperativity by Bayesian network integration of genome-wide data. Nucleic Acids Res 2009; 37(18): 5943-5958.

[36] Zhou Q, Chipperfield H, Melton DA, Wong WH. A gene regulatory network in mouse embryonic stem cells. Proc Nat AcadSci USA 2007; 104: 16438-16443.

[37] Saito S, Onuma Y, Ito Y, *et al.* Potential linkages between the inner and outer cellular states of human induced pluripotent stem cells, BMC Syst Biol 2011; 5(Suppl 1): S17.

[38] Zhana D, Wang Y. Saito S, Horimoto K. Inferring gene regulatory network for cell reprogramming, Proceedings of the 32nd Chinese Control Conference (CCC), 2012; 7437-7442.

[39] Samavarchi-Tehrani P, Golipour A, David L, *et al.* Functional genomics reveals a BMP-Driven Mesenchymal-to-Epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell 2010; 7(1): 64-77.

[40] Chang R, Shoemaker R, Wang W. Systematic search for recipes to generate induced pluripotent stem cells. PLoS Comput Biol 2011; 7(12): e1002300.

Received: April 1, 2012 Revised: May 10, 2012 Accepted: May 15, 2012