

## Exploring the Neural Reprogramming using Bioinformatics Approaches

Thileepan Sekaran<sup>1,\*</sup>, Rajkumar P. Thummer<sup>1,\*</sup> and Frank Edenhofer<sup>1,2\*</sup>

<sup>1</sup>Stem Cell Engineering Group, Institute of Reconstructive Neurobiology, University of Bonn - Life & Brain Center and Hertie Foundation, Sigmund-Freud Str. 25, 53105 Bonn, Germany

<sup>2</sup>Stem Cell and Regenerative Medicine Group, Institute of Anatomy and Cell Biology, Julius-Maximilians-University Würzburg, Koellikerstrasse 6, D-97070 Würzburg, Germany

\*These authors have contributed equally to this work

### Abstract

Recent studies demonstrate that mammalian cells can be artificially reprogrammed by ectopic expression of transcription factors in an unforeseen straightforward manner. Patient-derived reprogrammed cells hold great potential for biomedical applications such as cell replacement therapy, drug toxicity studies and disease modeling. Somatic cells such as fibroblasts can be dedifferentiated into so-called induced pluripotent stem cells (iPSCs) that are similar to embryonic stem cells (ESCs) by overexpression of Oct4, Sox2, Klf4 and cMyc. However, clinical applications using iPSCs carry the risk of tumor formation due to incomplete differentiation. More recently, it has been demonstrated that transcription factor-driven reprogramming enables direct conversion of fibroblasts into neurons, cardiomyocytes, hepatocytes as well as neural progenitors. Various groups elaborated protocols for the direct conversion of fibroblasts into multipotent induced neural stem cells (iNSCs) using different combinations of transcription factors and media conditions. These studies have shown that iNSCs exhibit morphology, gene expression and self-renewing capacity similar to NSCs derived from primary tissue. Moreover, these iNSCs differentiated into neurons, astrocytes and oligodendrocytes indicating multipotency of these cells. Here, we compare the gene expression profile of reprogrammed cells reported in these studies to determine the similarity in expression profile between the generated iNSCs using bioinformatics approaches. We provide a general workflow that can be applied to evaluate the status of reprogrammed cell populations. Using hierarchical clustering analysis and principal component analysis (PCA), we show that iNSCs reported by Thier et al. resemble more closely to those of Han et al. On the other hand, iNSCs generated by Ring et al. are relatively similar to 4F iNSC (late) of the study by Han et al. as judged by hierarchical clustering analysis. Our study demonstrates that bioinformatics approaches are particularly valuable to robustly assess the transcriptional status of reprogrammed cells and to anticipate their cellular functionality.

**Keywords:** iNSCs; Cellular reprogramming; Transcription factors; Bioinformatics; Gene expression; Microarrays

### Introduction

Cellular reprogramming describes a phenomenon of converting a somatic differentiated cell into a different cell type either somatic; or a multipotent and pluripotent, respectively, stem cell. A multipotent cell could be further differentiated into various but limited cell types, while a pluripotent cell could be differentiated into any desired cell type (Figure 1). This process holds great promise for biomedical applications to use it as a potential treatment for human diseases and study cellular pathology in the cell culture dish [1-8]. Studies demonstrated that differentiated cells can be reprogrammed to an embryonic-like pluripotent state [9] or a multipotent state [10-12] and by that generated a lot of interest in understanding the molecular basis of cell fate determination and developmental biology. Takahashi and Yamanaka [9] achieved a breakthrough by converting somatic fibroblasts into pluripotent induced pluripotent stem cells (iPSCs) by overexpression of four transcription factors, challenging the view that cell differentiation is an irreversible process. However, for successful therapeutic applications, iPSCs need to be efficiently differentiated into the desired cell type. Disease-related applications of iPSCs harbor the risk of tumor formation due to incomplete differentiation [13-15]. Therefore, researchers have recently focused on reprogramming somatic cells into multipotent stem cells that have a more restricted developmental plasticity with less or no tumorigenic potential and are closer to the desired cell type that is required to be engineered [10-12]. Therefore, direct reprogramming of somatic cells into lineage-

restricted multipotent stem cells should obviate iPSC generation and bypass the hurdle of unwanted iPSCs in differentiated cultures.

To understand the cellular reprogramming process from a somatic state to a multipotent or pluripotent state, bioinformatics approaches have been utilized to extract meaningful information from vast amount of data [16-19]. Attempts have been made to use genomic high-throughput technologies such as gene expression microarrays to assess genome-wide mRNA expression, chromatin immunoprecipitation with microarray technology (ChIP-on-ChIP) and ChIP with massive parallel DNA sequencing (ChIP-seq) to assess protein-DNA interactions to understand the cellular conversion process and provide us with new insights towards the reprogramming process [20,21]. Bioinformatics approaches have been used to explore and understand this complex and vast amount of biological data generated from high-throughput technologies. It integrates computational and statistical techniques

**\*Corresponding author:** Frank Edenhofer, Stem Cell and Regenerative Medicine Group, Institute of Anatomy and Cell Biology, Julius-Maximilians-University Würzburg, Koellikerstrasse 697070 Würzburg, Germany Tel: +49 931 31 88113; Fax: +49 931 31 82712; E-mail: [frank.edenhofer@wuerzburg.de](mailto:frank.edenhofer@wuerzburg.de); Web: [www.reprogramming.eu](http://www.reprogramming.eu)

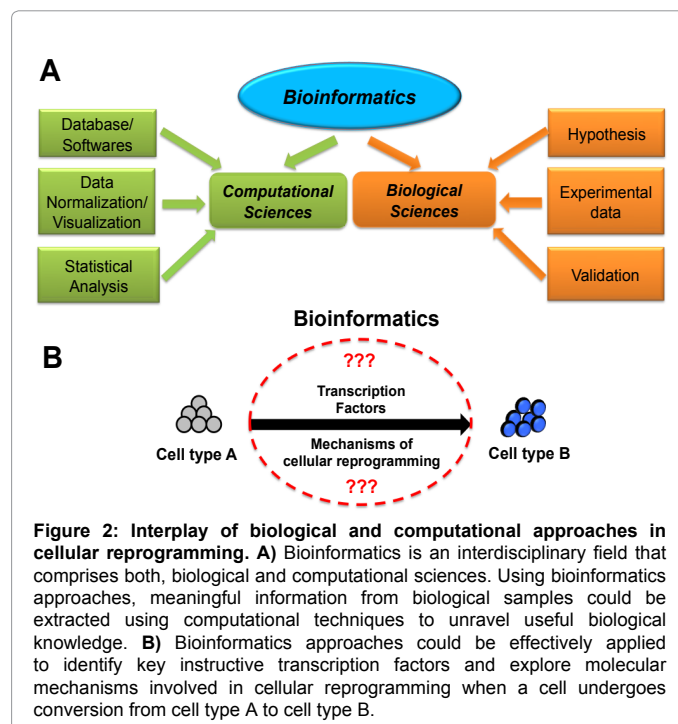
**Received** July 28, 2013; **Accepted** August 24, 2013; **Published** August 26, 2013

**Citation:** Sekaran T, Thummer RP, Edenhofer F (2013) Exploring the Neural Reprogramming using Bioinformatics Approaches. J Stem Cell Res Ther 3: 146. doi:10.4172/2157-7633.1000146

**Copyright:** © 2013 Sekaran T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

to explore and extract meaningful information from biological data (Figure 2A) [22]. By that, bioinformatics approaches could be used to identify novel transcription factors and help to understand the mechanisms of cell reprogramming during the cell conversion process (Figure 2B).

Regulation of gene expression comprises of a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products that could be either protein or RNA. To understand the function of a cell, as well as higher levels of biological organization, we need to know the components that constitute it. With the inventions of high throughput technology like microarrays, gene expression of thousands of mRNAs across different samples can be analyzed simultaneously and compared to each other [23]. In order to elucidate the properties of stem cells many researchers are comparing gene expression analysis of pluripotent cells with multipotent and/or unipotent cells [24]. Transcriptional profiling of multipotent hematopoietic stem cells (HSCs) and neural stem cells (NSCs) with pluripotent embryonic stem cells (ESCs) resulted in identification of genes that were differentially expressed among these three stem cell groups [25]. This knowledge could provide information about gene signatures of different cell types that could be used to determine the identity of artificially generated novel cell types. To unravel the wealth of information from the complex datasets, bioinformatics approaches have been deployed to explore the gene expression data. Clustering algorithms aims to group the genes based on the similar pattern of gene expression [26] while Principal Component Analysis [PCA; [27]] converts the gene expression data sets into diagonalized “eigen genes” and “eigenarrays” spaces in such a way that eigengenes are unique and orthogonal superpositions of the genes. Later the gene expression data is sorted based on eigengenes and eigenarrays reducing the features of the data to their principal components and determining the characteristic pattern of the dataset [24]. On the downside, existence



**Figure 2: Interplay of biological and computational approaches in cellular reprogramming.** A) Bioinformatics is an interdisciplinary field that comprises both, biological and computational sciences. Using bioinformatics approaches, meaningful information from biological samples could be extracted using computational techniques to unravel useful biological knowledge. B) Bioinformatics approaches could be effectively applied to identify key instructive transcription factors and explore molecular mechanisms involved in cellular reprogramming when a cell undergoes conversion from cell type A to cell type B.

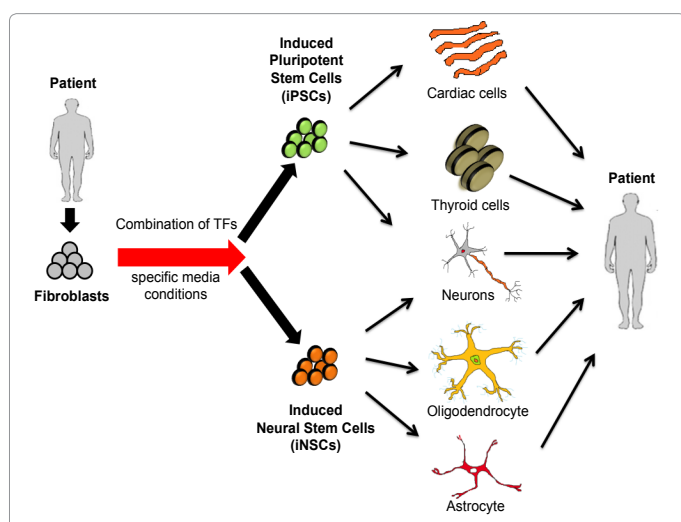
of several microarray platforms including Illumina Bead Arrays, Affymetrix chips, Agilent, Nimblegen, and other platforms renders the comparison of data sets from different platforms complicated and error-prone.

In this study, we analyzed gene expression datasets of recently reported induced neural stem cells (iNSCs) generated from fibroblasts by three different groups [10-12] using bioinformatics approaches. The datasets were generated from different platforms. Using hierarchical clustering analysis and principal component analysis (PCA) we demonstrate iNSCs generated by Thier et al. [12] resemble closely to those of Han et al. [10], whereas Ring et al. [11] iNSCs appear distinct from these cells. Thus, coupling gene expression data with bioinformatics approaches will provide us with information about the transcriptional status of the reprogrammed cells.

## Materials and Methods

Three freely available microarray datasets, namely GSE30500 [10], GSE37859 [11] and GSE36484 [12] were downloaded from Gene Expression Omnibus (GEO) [28]. All three datasets were recently reported in the context of direct conversion of fibroblasts into multipotent iNSCs using a slightly diverse combination of transcription factors (Figure 3) and specific media conditions. The details of microarray datasets downloaded from GEO are mentioned in table 1.

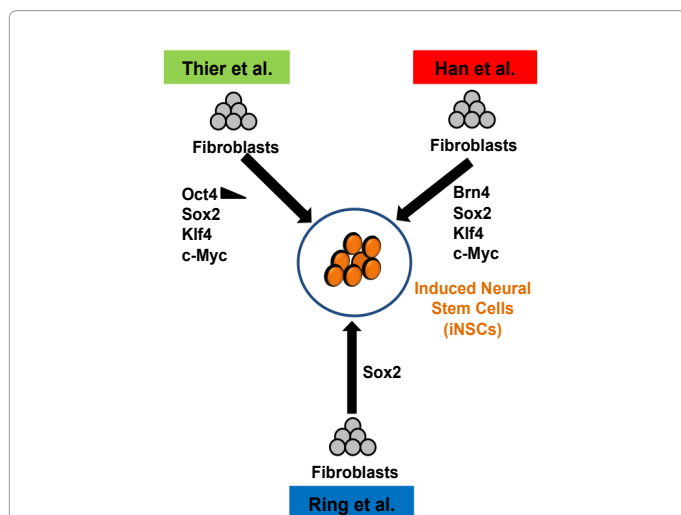
All the three microarray datasets were subjected to quality control, data transformation, integration, normalization and analysis as shown in Figure 4. Due to diversified nature of microarray platforms (Table 1) and the formats of the raw data, the datasets were first examined to determine whether they are in the same comparable range. From our Box Plot analysis we observed that the heterogeneous datasets were not in the same comparable range (Figure S1), as the expression values submitted to GEO from [12] was measured on a linear scale, whereas the other two datasets, [10,11], were measured in log2 scale.



**Figure 1: Harnessing cellular reprogramming for biomedical applications.** Schematic representation of cell reprogramming of patient-derived fibroblasts into induced pluripotent stem cells (iPSCs) and induced neural stem cells (iNSCs) using a combination of transcription factors (TFs) and specific media conditions. iPSCs can be further differentiated to all three germ layers: mesoderm (exemplified here by cardiac cells), endoderm (thyroid cells) and ectoderm (neurons), whereas multipotent iNSCs can be differentiated into neurons, oligodendrocytes and astrocytes. Pre-differentiated cells can be either transplanted into patients to treat disease or study the pathology in the cell culture dish under standardized conditions.

The expression values of the former were transformed into log2 scale using standard R functions and Bioconductor packages to bring all the datasets in the same comparable range for further analysis as performed in Heider et al. [Figure 5; [29]]. The microarray datasets were then integrated using a virtualArray, a Bioconductor package that can combine raw datasets based on the recently updated annotations from NCBI (National Centre Biotechnology Information) [29]. The redundant probes that represent the same gene were collapsed and only those gene symbols, which were common across all three datasets, were identified and a single combined “Expression Set” (comprising of 12,818 genes) was generated [29]. Quantile normalization between the arrays was performed on this expression set [30] and subjected to further analysis.

For hierarchical clustering analysis, the normalized “Expression Set” was clustered using agglomerative (using “virtualArrayHclust” function) hierarchical clustering on Euclidean distance matrix and average linkage method [29]. Many other distance measures other than

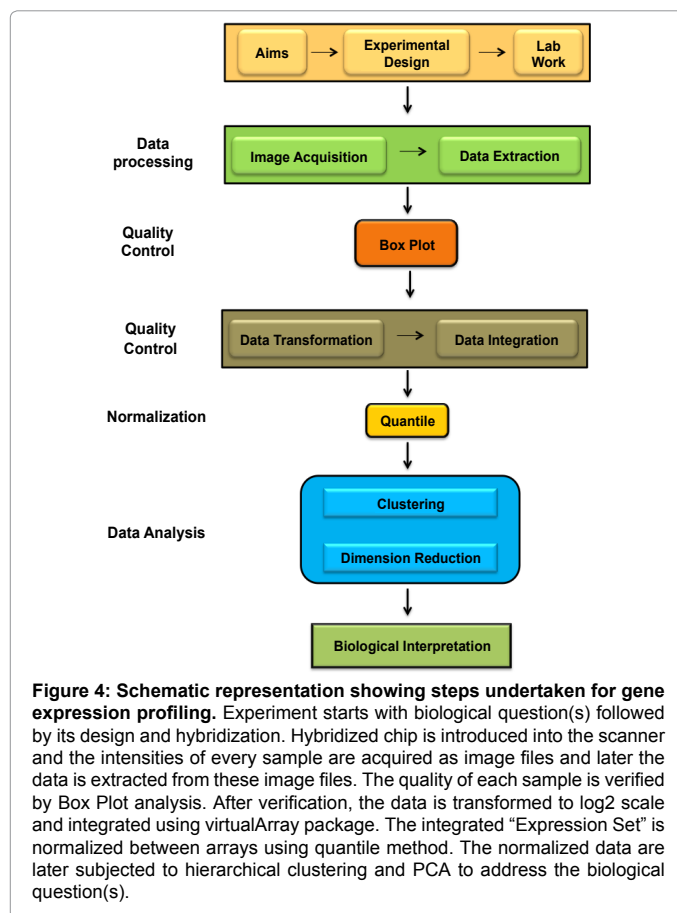


**Figure 3: Transcription factor combinations used by three different groups to generate induced neural stem cells (iNSCs).** Three different groups [10-12] recently converted fibroblasts into iNSCs using different transcription factor combinations. Han et al. used Brn4, Sox2, Klf4 and c-Myc while Ring et al. used Sox2 only for conversion into iNSCs. Thier et al. constitutively expressed Sox2, Klf4 and c-Myc, and Oct4 as a recombinant protein (▲) for the first 5 days to generate iNSCs. All three groups demonstrated in their respective studies that these iNSCs exhibit morphology, gene expression and the ability to self-renew similar to NSCs derived from primary tissue. Furthermore, these iNSCs differentiated into neurons, astrocytes and oligodendrocytes suggesting multipotency of these cells.

GEO datasets	Number of Samples	Samples Types (No. of replicates)	Platform (Chip type)
GSE36484 [12]	21	Fibroblasts (4) ESCs (3) NSCs (3) iNSCs (11)	Illumina (Mouse WG-6 v2)
GSE37859 [11]	12	Fibroblasts (3) NSCs (3) iNSCs (3)	Affymetrix (Mouse Gene 1.0 ST)
GSE30500 [10]	9	Fibroblasts (1) NSCs (2) 4 Factor iNSCs (late) (2) 4 Factor iNSCs (early) (2) 5 Factor iNSCs (2)	Illumina (Mouse Ref-8 v2)

The table represents list of gene expression datasets, sample types and its count and the platform and Chip used for generating the datasets.

**Table 1:** GEO datasets information of three groups.



**Figure 4: Schematic representation showing steps undertaken for gene expression profiling.** Experiment starts with biological question(s) followed by its design and hybridization. Hybridized chip is introduced into the scanner and the intensities of every sample are acquired as image files and later the data is extracted from these image files. The quality of each sample is verified by Box Plot analysis. After verification, the data is transformed to log2 scale and integrated using virtualArray package. The integrated “Expression Set” is normalized between arrays using quantile method. The normalized data are later subjected to hierarchical clustering and PCA to address the biological question(s).

Euclidean distance matrix are used but the Euclidean distance matrix with the commonly used average linkage method is more appropriate for normalized microarray data [31].

For Principal Component Analysis (PCA), the normalized “Expression Set” was also exported as tab-delimited text file from R and then imported into Partek Genomics Suite [32] for PCA with default parameters; correlation dispersion matrix and normalized eigen vector scaling and visualized in 3D scatterplot. The correlation method adjusts the data to be standardized to a mean of zero (mean centered) and a standard deviation of one. This adjustment is performed during the computation and does not modify the original data [32]. The eigenvectors are the direction cosines of the new axes (PCs) relative to the old (original variables), thus they define the rotations of the original axes [32].

The entire analysis was carried using Bioconductor packages, R, a statistical programming environment and Partek genomics Suite that runs in a 32-bit operating system comprising of 4GB RAM (Random Access Memory), a processor of Intel® Core (TM)2 Duo CPU T6500@2.10 GHz.

## Results and Discussion

Bioinformatics approaches are increasingly embraced by the stem cell research community because of its potential to uncover the molecular fingerprint of a stem cell state in a short period of time. These approaches are also valuable to reveal critical information about the data obtained from different laboratories and experimental conditions. In this study, we have compared the gene expression profiling data

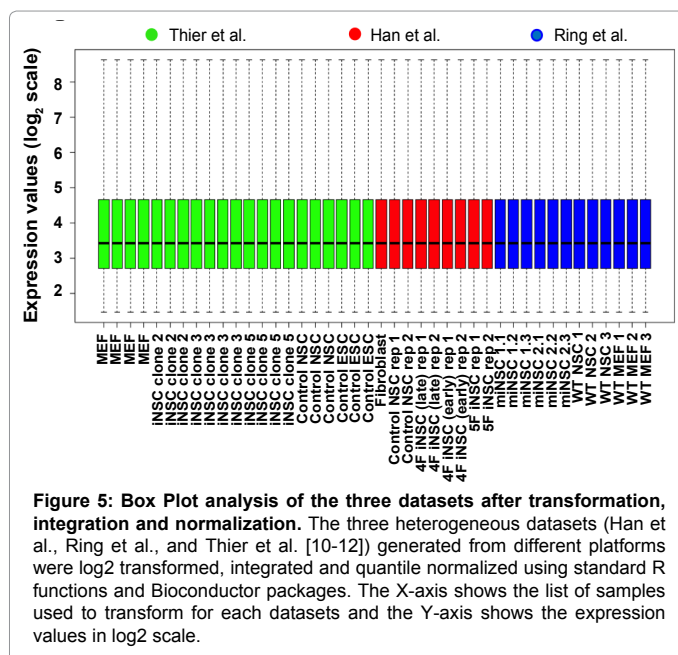
of three different studies that report direct conversion of fibroblasts into multipotent iNSCs using different combination of transcription factors and media conditions [10-12] (Figure 3). Han et al. used a cocktail of 4 factors (4F), Brn4, Sox2, Klf4 and c-Myc, and 5 factors (5F) that included the listed 4 factors plus E47/Tcf3 [10]. They analyzed the whole genome profile of 4F early- and a late-passage (in short 4F (early) and 4F (late)) as well as 5F to evaluate the reprogramming level of the entire transcriptome [10]. Sox2 only was employed by Ring et al. to convert fibroblasts into iNSCs [11]. Finally, Thier et al. [12] used Oct4 recombinant protein for the first 5 days along with constitutive expression of Sox2, Klf4 and c-Myc to accomplish a similar cell conversion process.

In this study, we have outlined cross platform comparison of three

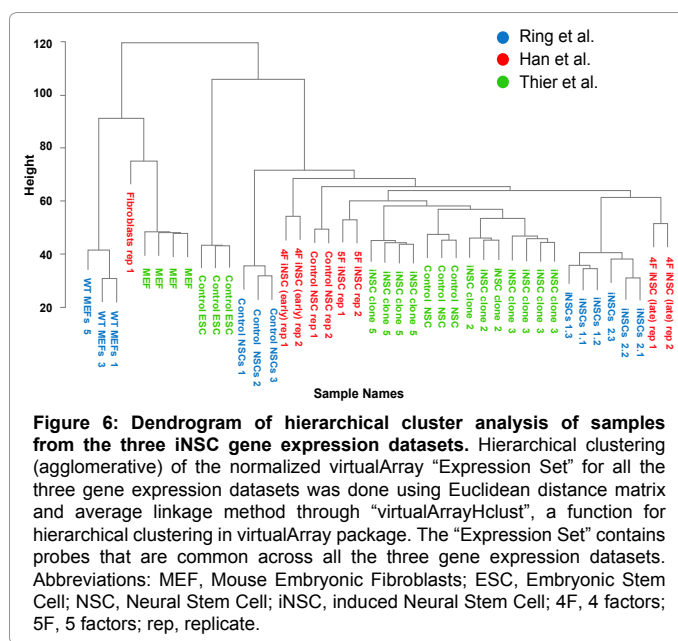
different microarray datasets generated by these groups on different platforms. The microarray datasets of the three studies were integrated using virtualArray and then compared using hierarchical clustering analysis and PCA to determine the similarity in gene expression profiling of iNSCs. First, the quality of the samples in all the three datasets before and after transformation was verified using Box plots (Figure S1; 5). Each box represents individual samples in the datasets and all the boxes were in the same range indicating that all the samples were of good quality (Figure 5). Hierarchical clustering analysis (Figure 6) of the combined expression data revealed that fibroblasts from Han et al., Ring et al. and Thier et al. cluster together while ESCs cluster separately. The control NSCs of Thier et al and Han et al, as well as iNSCs clones of [12], 4F iNSCs (early) replicates and 5F iNSC replicates cluster in between NSCs and iNSC replicates of [11]. Thier et al. NSCs and iNSCs resemble closely 5F iNSCs of Han et al. Notably, the iNSCs of Ring et al. cluster closely with replicates of 4F iNSCs (late) of Han et al. and by that are quite far away from control NSCs of Ring et al.

For PCA analysis, the combined expression datasets generated from virtualArray package were exported to Partek Genomics suite [32]. The high dimensionalities of three datasets were reduced and their patterns were identified using Principal Components (PCs) [33]. PCs are the set of new variables, created by linear transformation of original variables (gene expression). Rather looking into gene expression levels, PCA plot shows set of principal components calculated based upon actual gene expression values. Through this approach, the high dimensional complex data can be simplified and visualized, by retaining as much variance possible. The PCA plot explained in this study shows only first three PCs, as they contain the most information. The first 3 PCs were able to retain a total of 40.6% of variation among the data. Each PC explains the variance in the data and they are stacked in a decreasing order i.e. PC#1=20.18 %, PC#2=13.23%, PC#3=7.14% and so on (Figure S2). The list of total PCs and the cumulative % of variance proportion accumulating to 100% is shown in Figure S2. 3D scatterplot (Figure 7) generated from Partek illustrates that fibroblasts from Han et al. (Red), Ring et al. (Violet) and Thier et al. (Pink) have been projected along the PC#2. Control NSCs (Blue), iNSCs (Dark green) of Han et al, control NSCs (Grey), iNSCs (Light green) of Thier et al and control NSCs (Orange), iNSCs (Cyan) have been projected along PC#3. The figure also clearly visualizes that Ring iNSCs (Cyan) were distant from iNSCs of Han et al. and Thier et al. iNSCs of Han et al. and Thier et al. were grouped closer among each other suggesting that iNSCs reported by Han et al. and Thier et al. were more similar in their expression levels when compared to those of Ring et al.

Therefore, using both hierarchical clustering analysis and PCA we demonstrate that iNSCs reported by Thier et al. are similar to iNSCs from Han et al. Notably, hierarchical clustering analysis also revealed that iNSCs generated by Ring et al. are relatively similar to 4F iNSC (late) of the study by Han et al and NSCs and iNSCs from Thier et al. resemble more closely to 5F iNSCs of Han et al. Therefore, this study clearly demonstrates that cross platform comparison is feasible and could be used to investigate the transcriptional state of newly generated multipotent cells from different groups. Recently, several data handling/organizational tools are presented that might serve as meta tools for integrating cell profiling data [34-37]. Muller et al. [38] designed a new computational strategy for classifying the stem cells by designing a database, "stem cell matrix" of global gene expression profiles of pluripotent, multipotent and other differentiated cells. Using unsupervised clustering method, they were able to classify the pluripotent stem cells from the other cell lines based on their gene

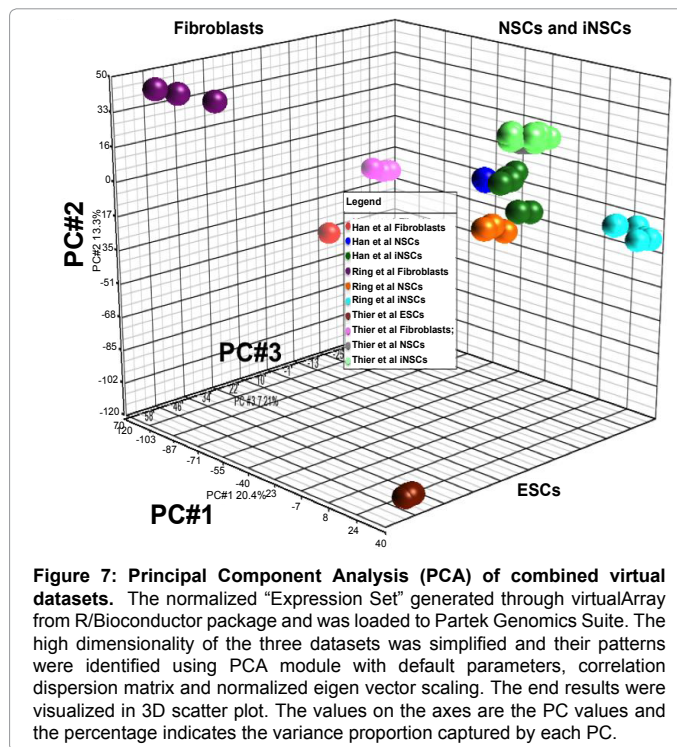


**Figure 5: Box Plot analysis of the three datasets after transformation, integration and normalization.** The three heterogeneous datasets (Han et al., Ring et al., and Thier et al. [10-12]) generated from different platforms were log2 transformed, integrated and quantile normalized using standard R functions and Bioconductor packages. The X-axis shows the list of samples used to transform for each datasets and the Y-axis shows the expression values in log2 scale.



**Figure 6: Dendrogram of hierarchical cluster analysis of samples from the three iNSC gene expression datasets.** Hierarchical clustering (agglomerative) of the normalized virtualArray "Expression Set" for all the three gene expression datasets was done using Euclidean distance matrix and average linkage method through "virtualArrayHclust", a function for hierarchical clustering in virtualArray package. The "Expression Set" contains probes that are common across all the three gene expression datasets. Abbreviations: MEF, Mouse Embryonic Fibroblasts; ESC, Embryonic Stem Cell; NSC, Neural Stem Cell; iNSC, induced Neural Stem Cell; 4F, 4 factors; 5F, 5 factors; rep, replicate.





expression. They constructed a pluripotency network called PluriNet based on MATISSE (Module Analysis via Topology of interactions and similarity sets). With visualization and data analysis tools embedded within IntNetDB [39], it provides the researchers a good framework to analyze, visualize and interpret the results. Recently, PluriTest [16], a robust bioinformatics assay for assessing pluripotency, has been developed which can be used to upload and analyze gene expression data of human pluripotent cells based on Illumina platform. This tool compares the known pluripotent gene expression signature with a newly generated reprogrammed cell line to determine the pluripotency status of the candidate cell line and by that avoid teratoma assay in mice. However, this bioinformatics assay has a limitation in assessing pluripotency of human cells from Illumina platform only. More recently, a similar web portal Stemformatics has been generated to assist stem cell biologists to identify and assess relevant datasets, gene sets and pathways for easy visualization and comparison of gene expression generated across platforms, laboratories and cell models [18]. Similar advances using novel stem cell specific bioinformatics tools could replace cumbersome and time-consuming approaches like teratoma assays or in vitro differentiation assays to characterize cell lines to determine their identity [40].

Our study involving cross platform comparison could further help us in identifying biomarkers from various data handling/organizational tools, which then can be used to compare published cell types with novel cell types. A web interface for iNSCs should be established to easily upload gene expression data for different species and platforms to determine the transcriptional status of newly derived iNSCs and other multipotent stem cell lines. The analysis can be extended in identifying the differentially expressed genes between cell lines, which could further shed light on the key genes involved in the cell conversion process. Therefore, with already existing algorithms and development of new tools and databases, the new experimental datasets can be applied to accessible databases and will infer new

results based on known biological processes to understand more about stem cell fate and cellular reprogramming. Hence, combination of both, experimental and computational approaches will help biologists to understand cell fate determination. In conclusion, our study substantiates that bioinformatics approaches are essential to robustly determine the transcriptional state of reprogrammed cells and envisage their functionality.

## Future Perspectives

The availability of different platforms for microarray experiments impedes comparison of expression data from the different platforms. This limitation also restricts the extraction of valuable information hidden in the datasets uploaded online in multiple data handling/organizational tools. Therefore, a tool capable of performing cross platform comparison without losing much information is required and will provide infinite opportunities for researchers to carry out the comparative analysis. The mouse and human stem cells at the system level remain poorly explored; hence, more bioinformatics approaches have to be deployed. As the application of bioinformatics in cellular reprogramming is at an infant stage, therefore integrating the experimental data along with the other "OMICS" data will definitely provide more details of biological properties of the reprogrammome at a broader spectrum.

## Acknowledgments

We are grateful to Andreas Heider for providing valuable guidance regarding the "virtualArray" package. We also thank Marc Thier and Philipp Wörsdörfer and all the members of the Stem Cell Engineering Group for their support and stimulating discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft, DFG, (SPP1356-ED79/1-2), EU FP7 funding for project CADMAD (265505), BMBF (NewNeurons; 01GN1009B) and Merck KGaA, Darmstadt, Germany.

## References

- Amemori T, Romanyuk N, Jendelova P, Herynek V, Turnovcova K, et al. (2013) Human conditionally immortalized neural stem cells improve locomotor function after spinal cord injury in the rat. *Stem Cell Res Ther* 4: 68.
- Kumar AR, Kumar SD, Nishanth T (2011) A Review on Stem cell Research & Their Role in Various Diseases. *J Stem Cell Res Ther* 1: 3.
- Neirinckx V, Coste C, Rogister B, Wislet-Gendebien S (2013) Concise review: adult mesenchymal stem cells, adult neural crest stem cells, and therapy of neurological pathologies: a state of play. *Stem Cells Transl Med* 2: 284-296.
- Okano H, Nakamura M, Yoshida K, Okada Y, Tsuji O, et al. (2013) Steps toward safe cell therapy using induced pluripotent stem cells. *Circ Res* 112: 523-533.
- Ramachandran RP, Yelledahalli (2011) Exploring the Recent Advances in Stem Cell Research. *J Stem Cell Res Ther* 1: 3.
- Siddiqui A, Jeevani T, Anusha N, Bindu H (2011) Stem Cell Therapy for Liver Diseases. *J Stem Cell Res Ther* 1: 3.
- Simara P, Motl JA, Kaufman DS (2013) Pluripotent stem cells and gene therapy. *Transl Res* 161: 284-292.
- van Gorp S, Leerink M, Kakinohana O, Platoshyn O, Santucci C, et al. (2013) Amelioration of motor/sensory dysfunction and spasticity in a rat model of acute lumbar spinal cord injury by human neural stem cell transplantation. *Stem Cell Res Ther* 4: 57.
- Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126: 663-676.
- Han DW, Tapia N, Hermann A, Hemmer K, Hoing S, et al. (2012) Direct reprogramming of fibroblasts into neural stem cells by defined factors. *Cell Stem Cell* 10: 465-472.
- Ring KL, Tong LM, Balestra ME, Javier R, Andrews-Zwilling Y, et al. (2012) Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor. *Cell Stem Cell* 11: 100-109.
- Thier M, Worsdorfer P, Lakes YB, Gorris R, Herms S, et al. (2012) Direct

- conversion of fibroblasts into stably expandable neural stem cells. *Cell Stem Cell* 10: 473-479.
13. Fong CY, Gauthaman K, Bongso A (2010) Teratomas from pluripotent stem cells: A clinical hurdle. *J Cell Biochem* 111: 769-781.
14. Miura K, Okada Y, Aoi T, Okada A, Takahashi K, et al. (2009) Variation in the safety of induced pluripotent stem cell lines. *Nat Biotechnol* 27: 743-745.
15. Yamanaka S (2009) A fresh look at iPS cells. *Cell* 137: 13-17.
16. Muller FJ, Schultdt BM, Williams R, Mason D, Altun G, et al. (2011) A bioinformatic assay for pluripotency in human cells. *Nat Methods* 8: 315-317.
17. Nestor MW, Noggle SA (2013) Standardization of human stem cell pluripotency using bioinformatics. *Stem Cell Res Ther* 4: 37.
18. Wells CA, Mosbergen R, Korn O, Choi J, Seidenman N, et al. (2013) Stemformatics: visualisation and sharing of stem cell gene expression. *Stem Cell Res* 10: 387-395.
19. Yong W (2013) Bioinformatics Studies on Induced Pluripotent Stem Cell. *Current Bioinformatics* 8: 80-86.
20. Massie CE, Mills IG (2008) ChIPping away at gene regulation. *EMBO Rep* 9: 337-343.
21. Panda S, Sato TK, Hampton GM, Hogenesch JB (2003) An array of insights: application of DNA chip technology in the study of cell biology. *Trends Cell Biol* 13: 151-156.
22. Maojo V, Martin-Sanchez F (2004) Bioinformatics: towards new directions for public health. *Methods Inf Med* 43: 208-214.
23. Perez-Iratxeta C, Palidwor G, Porter CJ, Sanche NA, Huska MR, et al. (2005) Study of stem cell function using microarray experiments. *FEBS Lett* 579: 1795-1801.
24. Chang HY, Thomson JA, Chen X (2006) Microarray analysis of stem cells and differentiation. *Methods Enzymol* 420: 225-254.
25. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA (2002) "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science* 298: 597-600.
26. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863-14868.
27. Dudoit S, Fridlyand J (2002) A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biol* 3: RESEARCH0036.
28. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207-210.
29. Heider A, Alt R (2013) virtualArray: a R/bioconductor package to merge raw data from different microarray platforms. *BMC Bioinformatics* 14: 75.
30. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19: 185-193.
31. Shannon W, Culverhouse R, Duncan J (2003) Analyzing microarray data using cluster analysis. *Pharmacogenomics* 4: 41-52.
32. Downey T (2006) Analysis of a multifactor microarray study using Partek genomics solution. *Methods Enzymol* 411: 256-270.
33. Li L (2010) Dimension reduction for high-dimensional data. *Methods Mol Biol* 620: 417-434.
34. Mallon BS, Chenoweth JG, Johnson KR, Hamilton RS, Tesar PJ, et al. (2013) StemCellDB: the human pluripotent stem cell database at the National Institutes of Health. *Stem Cell Res* 10: 57-66.
35. Sandie R, Palidwor GA, Huska MR, Porter CJ, Krzyzanowski PM, et al. (2009) Recent developments in StemBase: a tool to study gene expression in human and murine stem cells. *BMC Res Notes* 2: 39.
36. Som A, Harder C, Greber B, Siatkowski M, Paudel Y, et al. (2010) The PluriNetWork: an electronic representation of the network underlying pluripotency in mouse, and its applications. *PLoS One* 5: e15165.
37. Wohlers I, Stachelscheid H, Borstlap J, Zeilinger K, Gerlach JC (2009) The Characterization Tool: A knowledge-based stem cell, differentiated cell, and tissue database with a web-based analysis front-end. *Stem Cell Res* 3: 88-95.
38. Muller FJ, Laurent LC, Kostka D, Ulitsky I, Williams R, et al. (2008) Regulatory networks define phenotypic classes of human stem cell lines. *Nature* 455: 401-405.
39. Xia K, Dong D, Han JD (2006) IntNetDB v1.0: an integrated protein-protein interaction network database generated by a probabilistic model. *BMC Bioinformatics* 7: 508.
40. Buta C, David R, Dressel R, Emgard M, Fuchs C, et al. (2013) Reconsidering pluripotency tests: Do we still need teratoma assays? *Stem Cell Res* 11: 552-562.