OPEN ACCESS Freely available online

PLOS ONE

An Efficient Weighted Graph Strategy to Identify Differentiation Associated Genes in Embryonic Stem Cells

Jie Zhang[1]*, Li Li[2], Luying Peng[2], Yingxian Sun[3], Jue Li[1]*

1 Department of Prevention, Tongji University School of Medicine, Shanghai, China, **2** Key Laboratory of Arrhythmias, Ministry of Education, Tongji University School of Medicine, Shanghai, China, **3** Department of Cardiology, The First Hospital of China Medical University, Shenyang, China

Abstract

In the past few decades, embryonic stem cells (ESCs) were of great interest as a model system for studying early developmental processes and because of their potential therapeutic applications in regenerative medicine. However, the underlying mechanisms of ESC differentiation remain unclear, which limits our exploration of the therapeutic potential of stem cells. Fortunately, the increasing quantity and diversity of biological datasets can provide us with opportunities to explore the biological secrets. However, taking advantage of diverse biological information to facilitate the advancement of ESC research still remains a challenge. Here, we propose a scalable, efficient and flexible function prediction framework that integrates diverse biological information using a simple weighted strategy, for uncovering the genetic determinants of mouse ESC differentiation. The advantage of this approach is that it can make predictions based on dynamic information fusion, owing to the simple weighted strategy. With this approach, we identified 30 genes that had been reported to be associated with differentiation of stem cells, which we regard to be associated with differentiation or pluripotency in embryonic stem cells. We also predicted 70 genes as candidates for contributing to differentiation, which requires further confirmation. As a whole, our results showed that this strategy could be applied as a useful tool for ESC research.

Citation: Zhang J, Li L, Peng L, Sun Y, Li J (2013) An Efficient Weighted Graph Strategy to Identify Differentiation Associated Genes in Embryonic Stem Cells. PLoS ONE 8(4): e62716. doi:10.1371/journal.pone.0062716

Editor: Jordi Garcia-Ojalvo, Universitat Politecnica de Catalunya, Spain

Received September 27, 2012; **Accepted** March 25, 2013; **Published** April 26, 2013

Funding: Research was supported by the Young Talents Scheme of Tongji University, China (To Zhang Jie, 1500219046); Shanghai Municipal Health Bureau Project (To Zhang Jie, z0124y166); National Natural Science Foundation of China (30971621, 81270231 and 31170791); and International Science & Technology Cooperation Program of China (2011DFB30010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jiezhang@tongji.edu.cn (JZ); jueli@tongji.edu.cn (JL)

Introduction

Embryonic stem cells (ESCs) are unspecialized cells that have the ability of self-renewal, producing daughter cells with equivalent developmental potential, or to differentiate into more specialized cells. Experiments performed several decades ago showed that dormant gene expression programs can be awakened in differentiated cells by the fusion of different pairs of cell types [1]. Different cell fates can be induced by the defined transcription factors [2]. However, the global transcription activities in ESCs are not well understood, and the set of differentiation associated genes, i.e. the genes which are active in the pluripotent state and become inactive upon differentiation (and vice versa), is still unknown.

Rapid increase of high throughput biological data supplies us both opportunities and challenges to explore mechanisms in ESCs differentiation. In fact, initial approaches derive predictions based on specific information such as gene expression profile [3] and protein-protein interactions [2]. Also, it has been shown that the use of global optimization may not actually yield significant improvement over simpler local prediction methods [4,5,6].Here, we propose an intuitive method, which uses a unified framework for combining multiple sources, including mRNA expression profile dataset, sequence dataset and protein-protein interaction dataset. Our method involves three steps. Firstly, each evidence source is assessed with a reliability score based on their functional correlation. According to the data characteristics, a weighted value is defined. Secondly, undirected graphs are constructed based on each data source respectively, with genes as vertices and functional relationships between gene pairs as edges. Finally, these undirected graphs are integrated into a weighted functional linked network. The genes are predicted to be differentiation associated genes based on their degrees in the final network, which are regarded to be associated with differentiation or pluripotency in embryonic stem cells.

Our results showed that despite the simplicity of its formulation, our method performed relatively well on the prediction ability of identifying the differentiation associated genes. It was also shown that our method could involve a large amount of datasets, including cross genome information, in order to make much better predictions.

Materials and Methods

Datasets Preprocessing and Normalization

Four different types of datasets were analyzed. The Affymetrix mouse stem cell microarray data (GSE7506) consisted of 36 samples, which were used for prediction and testing of novel

networks regulating ESCs self-renewal and commitment [7]. It was pre-processed by Robust Multi-array Analysis (RMA) followed by median normalization between arrays [8]. The protein sequences were downloaded from RefSeq database containing a total of 38129 distinct sequences (June 11, 2010). Functional annotations were taken from Gene Ontology (GO) (June 20, 2010). The annotations were arranged in a hierarchical manner and compiled using up-to-date information from GO's three ontology divisions, including Molecular Function (MF), Biological Process (BP) and Cellular Component (CC). The mouse protein-protein interaction (PPI) datasets (October 10, 2010) were downloaded from APID [9], BIND [10], iRefIndex [11], MINT [12] and STRING [13], which contained 12026, 8164, 19727, 4333 and 207211 PPIs respectively. To increase the coverage of the PPI network, the five datasets were pooled together as previously done in Lage et al. [14].

Selection of Differentially Expressed Genes

We used the popular SAM (Significance Analysis of Microarrays, samr R package) method [15] to select differentially expressed genes (DEGs). Multiple statistical tests were controlled by false discovery rate (FDR) defined as the expected percentage of false positives among the claimed DEGs [16]. Because the FDR estimation of SAM might be overly conservative [17,18], we also applied the FDR estimation method suggested by Zhang [18] using the idea of Xie [17], and refer to it as the modified SAM method.

Scoring Functions

The weighted graph strategy utilized different weighted scores as inputs. According to the character of different dataset, we applied a simple weighted strategy similar to the weighted averages method [19].

(1) Gene expression profiles. Relationship between gene i and j was scored based on Pearson Correlation Coefficient (PCC) by the expression profiles, denoted as p_{ij}. An average of d_i and d_j which are SAM statistics [18] is assigned as the weight of p_{ij}. The weighted score was defined as formula 1. To scale the score between 0 and 1, the score of each gene pair was divided by $Max(X^{\exp})$ which was the highest value of all gene pairs.

$$X_{ij}^{Exp} = \frac{|d_i| + |d_j|}{2} \cdot |p_{ij}| \tag{1}$$

$$S_{ij}^{Exp} = \frac{X_{ij}^{Exp}}{Max(X^{Exp})} \tag{2}$$

(2) Sequence analysis. Each mouse sequence was aligned with all other sequences using software ClustalX [20]. The identity matrix was applied to calculate scores S_{ij}^{Blast} which represented the sequence similarities of each two amino acids. And the score was automatically adjusted to positive values, scaled between 0 and 1.

(3) GO functional analysis. The pair-wise functional similarities of the DEGs were computed and analyzed. Each gene was represented by a feature vector containing the gene's similarities to predefined prototype genes. The scores between gene i and j were calculated in molecular function, biological process and cellular component respectively. For scaling the score between 0 and 1, S_{ij}^{GO}, representing the semantic similarity between gene i and j [21,22,23], was calculated as formula 3.

$$S_{ij}^{GO} = \frac{X_{ij}^{MF} + X_{ij}^{BP} + X_{ij}^{CC}}{3} \tag{3}$$

$$\begin{aligned} X_{ij}^{MF} &= \frac{\left|GO_i^{MF} \cap GO_j^{MF}\right|}{\left|GO_i^{MF} \cup GO_j^{MF}\right|}, \\ X_{ij}^{BP} &= \frac{\left|GO_i^{BP} \cap GO_j^{BP}\right|}{\left|GO_i^{BP} \cup GO_j^{BP}\right|}, \\ X_{ij}^{CC} &= \frac{\left|GO_i^{CC} \cap GO_j^{CC}\right|}{\left|GO_i^{CC} \cup GO_j^{CC}\right|}, \end{aligned} \tag{4}$$

$$GO_i = \{go_{i1}, go_{i2}, ..., go_{im}\}, GO_j = \{go_{j1}, go_{j2}, ..., go_{jn}\} \tag{5}$$

X_{ij}^{MF}, X_{ij}^{BP} and X_{ij}^{CC} measured the functional similarities of three basic ontology divisions between gene i and j (Formula 4). GO_i and GO_j were two sets of GO terms that annotated with gene i and j respectively (Formula 5).

(4) Protein-protein interactions. The FSWeight [24] has been shown to provide a good estimate of functional similarity between the interacting protein pairs (direct interactions), as well as between the protein pairs that do not interact, but share common interaction partners (indirect interactions). To keep our comparison simple, we only used direct interaction pairs. Each interacting protein pair was scored using a simplified variant of the FSWeight measurement (Formula 6), where N_p referred to the set that contains p and its interaction neighbors.

$$\begin{aligned} S_{ij}^{PPI} = &\frac{2\left|N_i \cap N_j\right|}{\left|N_i - N_j\right| + 2\left|N_i \cap N_j\right| + 1} \\ &\times \frac{2\left|N_j \cap N_i\right|}{\left|N_j - N_i\right| + 2\left|N_i \cap N_j\right| + 1} \end{aligned} \tag{6}$$

Combination of Different Datasets

The initial four score matrixes of four datasets just included DEGs respectively. Lee et al. [25] used a unified log-likelihood scoring function to combine several sources of binary gene relationship data into a graph, which could be clustered into groups that show strong similarity in function. It has been illustrated that different data sources have different degrees of correlation with function similarity. Here, we adapted a simple model in our approach to integrate the four datasets. Each dataset can be modeled as an undirected graph, where each vertex represents a protein and each edge represents the functional relationship between proteins. The edges in different graphs have different scoring schemes as previously described. Different graphs

derived from four score matrixes were combined to form a larger and presumably more complete graph. The confidence relationship of each edge in the last complete graph can be estimated by an integrated score, which represents a particular function shared between two genes. The score of the two proteins in the final integrated graph can be calculated as formula 7.

$$S_{ij} = (S_{ij}^{Exp} + S_{ij}^{Blast} + S_{ij}^{GO} + S_{ij}^{PPI})/4 \quad (7)$$

Generation of Differentiation Associated Genes

The final network was built with the gene pairs if their scores were larger than the 75% quantile of the whole score values (Formula 7), since when the threshold was higher than 75%, some differentiation associated genes would not be selected and when it was lower than 75%, too many redundant genes would be selected. In a network, nodes with high connectivity were more important than low connectivity. They were named as "hubs". A line graph showed the relationship between degree and gene number. According to the chart, genes with most of higher degrees were selected, which were considered as differentiation associated genes.

Validation Method

For comparison, we ran three separate methods, SAM (using the mRNA expression data set) [18], decision tree (DT) [26] and normal graph strategy (NGS). In normal graph strategy, scores were calculated just based on Pearson Correlation Coefficients, blast scores, GO scores (same as S_{ij}^{GO}) and the PPI scores (1 representing that the protein i interacts with protein j, 0 representing non-interacting proteins). The selected differentiation associated genes were predicted as positive gene set using three repetitions of 5-fold cross-validation. The area under Receiver Operating Characteristics [27] graph was computed for each class (associated or not associated with differentiation) and the average was obtained based on the predictions 15 times in total.

Results

Differentially Expressed Genes Selection

Current FDR control procedures, including the one adopted in SAM [15], may be unstable in small samples especially in the presence of correlated expression changes. Hence, we evaluated the actual FDR of a DEG list detected in simulated small samples, according to the predefined DEGs. Based on the simulated results, using SAM with 0.05% FDR control, we tentatively defined the DEGs obtained from the full samples as a nominal gold standard set [28]. The procedure outputs totaled 3277 DEGs. Although, there were false positives in the selected DEGs, this was just a preliminary procedure which was prepared for the subsequent functional analysis of various data source integration.

Generation of ESC Differentiation Associated Genes

Different kinds of datasets can supply us different information, which can improve the prediction performance. In our method, each of the four score matrixes had been scaled between 0 and 1, and their combination was a merging process based on the previous DEGs selection result. That is, each dataset contained 3277 dimensions. The score between two genes which had no relation was denoted by 0. Next, we selected the final network based on the combination result. The final network was built with the genes whose scores were larger than the 75% quantile of the total score values (Formula 7). A line chart showed the relationship between the degree and the gene numbers (Figure 1). An increase in gene number resulted in a significant decrease in degree. A significant drop in degree in the graph threshold was selected for analysis. The 100th gene, Bmi1 had a degree of 1369, while the 101th gene, Tmcc3had a degree of 1100. Tmcc3 is not associated with differentiation; hence we selected the top 100 genes with the highest degrees as differentiation associated candidate genes

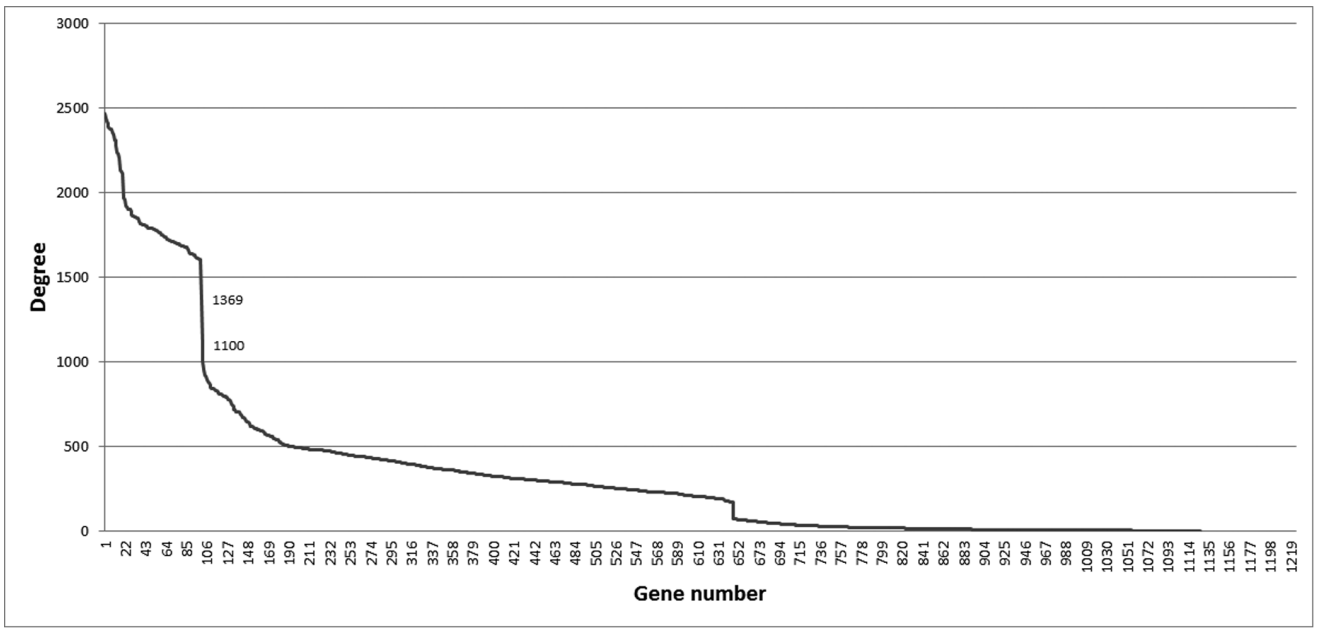

Figure 1. The relationship between degree and gene number. A line chart can show the relationship between the degree and gene number. Abscissa represents the gene number, and ordinate represents the degree. Bmi1 has the lowest degree, which is at the corner of the line chart. With the increase in gene number, there is a decrease in degree.
doi:10.1371/journal.pone.0062716.g001

(Table S1). The cutoff for selecting differentiation genes in the integrated network is set as 75%. 75% was the highest cutoff that included all Nanog, Poutf1 and Sox2 in the selected group. If the cutoff value was raised from this, however, Pou5f1 and Sox2 were excluded from the selected group.

Among the 100 candidate genes, 30 genes had been reported to be associated with stem cell differentiation processes (Table 1). Briefly, 17 genes tended to be significantly active in the pluripotent state and became inactive or repressed during differentiation. 13 genes tended to be significantly inactive in pluripotent state and became active or expressed during differentiation. The other 70 genes were listed in Table S2.

Comparisons with Three Other Different Methods

The SAM (only using the mRNA expression data set), normal graph strategy (NGS), decision tree (DT) and weighted graph strategy (WGS) were compared using the 30 differentiation associated genes as a positive gene set. Figure 2 showed the averaged Receiver Operating Characteristics (ROC) for the 30 differentiation associated genes predicted using SAM, NGS, DT and WGS. WGS took less time to make a better performance than the other three methods, and was especially easy to be understood and accepted.

Table 1. 30 differentiation associated genes selected by weighted graph strategy.

Gene	Degree	Roles	Expression	Tissues/cells	PMID/References
Aire	1889	*+	↓	endoderm	20226168 [29]
App	1776	+	↓	neuron	18535156 [30];17908039 [31]
Bmi1	1369	+	↓	mammary stem cells	18635350 [32]
Brca1	1689	+	↓;↑	ESCs; mammary stem cells	19340312 [33];18230721 [34]
Carm1	1786	*+	↓	ESCs	19544422 [35]
Cd24a	1606	+	↑	hepatic progenitor cells; ESCs->brain,liver	17641245 [36];19998061 [37]
Cdh1	2233	+	↑	ESCs; neural stem cells->neuron	20473026 [38];19918205 [39]
Cdx2	1782	*+	↑	trophectoderm	16325584 [40]
Cyr61	1698	+	↑	neuronal differentiation; endoderm/mesoderm differentiation	9832196 [41]; 19544440 [42]
Eed	2310	*+	↓	ESCs	11803473 [43];21540835 [44]
Ids	2382	+	↑	epithelial cells	9737997 [45]
Ilk	2468	*+	↑	ESCs->cardiomyogenic differentiation	21344393 [46];22666394 [47]
Irs1	2407	*+	↓	ESCs	17620314 [48]
Irx3	1898	+	↑	ESCs->neuronal cells	21710438 [49];15611653 [50]
Klf4	2312	+	↓	monocyte differentiation	17762869 [51]
Lrp4	2275	+	↑	cardiovascular formation	15699019 [52]
Nanog	2443	*+	↓	ESCs->embryonic ectoderm	19544440 [42];22482508 [53]
Nr0b1	1642	*+	↓	individual germ layer fates	16466956 [54]
Npdc1	2241	+	↑	neural and glial precursors	9181131 [55]
Pin4	2110	+	↑	plant embryogenesis	19000164 [56]
Pou5f1	2353	*+	↑;↓	ESCs->mesoderm, ectoderm; neuronal differentiation	10742100 [57];15615706 [58]
Prc1	2122	+	↓	three germ layers	20123906 [59]
Prnp	2389	+	↓	Neuronal differentiation	10617928 [60]
Psen1	2375	+; *+	↓	ESCs->endothelial cell lineage; neuronal lineage	16376112 [61];20484632 [62]
Ptk7	1898	+	↓	expressed in un-differentiated ESC	17671748 [63]
Rap1gds1	2237	*+	↑	colony formation	20039365 [64]
Satb1	1792	*+	↑	early erythroid differentiation	15618465 [65];19933152 [66]
Sfrp2	2343	*+	↓	mesenchymal stem cells; ESCs-> dopamine neuron;ESCs->mesoderm	20826809 [67];22290867 [68];17462603 [69]
Sox2	2421	*+	↑ ↓	neuronal differentiation;ESCs->mesoderm	21663792 [70]
Stat3	2375	+	↓	mesoderm and endoderm differentiation	19544440 [42]

Gene: gene symbols; Degree: the degree of aim gene in the final network;Roles: the role of aim gene in the stem cell,
*represents the aim gene plays a role in maintaining stem cell pluripotency,
+represents the aim gene plays a role in stem cell differentiation process; Expression: the trend of expression level of aim gene,
↓ represents a decreasing expression in differentiation,
↑ represents an increasing expression in differentiation; Tissues/cells: the tissue or cells where the differentiation occurs; PMID/References: the pubmed ID of supporting published works (www.pubmed.org) and the references means the citation number in this work.
doi:10.1371/journal.pone.0062716.t001

The Evaluation of Our Weighted Graph Strategy (WGS)

Differentiation associated genes were selected based on their high connectivities. The selection rule in WGS was based on the degree rather than the integrated score. This could avoid the score bias of specific datasets. There were less than 30 differentiation associated genes if the selection was based on integrated score.

Discussion

WGS supplied a simple but reliable method to search for differentiation associated genes. Although some genes had different expression styles in different cells, the 30 genes we listed were associated with differentiation occurring not only in ESCs but also other stem cells, such as hepatic progenitor cells, plant stem cells, and neural stem cells.

We found the GO function similarity scores were higher than the sequence similarity scores, but lower than the expression scores. That was because different types of data source reflect different nature of functional relevance. As a whole, the scores of mRNA expression were always higher than others. However, the expression data might not have a higher reliability than other data sources. In order to get four balanced score matrixes, a simple weighted strategy was applied here. Firstly, the scores must be scaled between 0 and 1. Secondly, a coefficient was added into the formula. Because the scores of the other three datasets were generally lower than the expression similarity scores, a different coefficient was added in different scores, which was based on the character of dataset. For example, an average of d_i and d_j was assigned as the weight of p_{ij}. The weighted coefficient for sequence similarities was assigned as 1. Our results showed that this treatment could balance the scores, and reduced the data bias.

Weighted graph strategy based on our analysis is more efficient than SAM, DT and NGS. Firstly, weighted strategy could avoid the experimental technical biases of the derivation of different datasets according to the data character (Figure 2). Secondly, the integrated scores were used for constructing the integrated network, and the differentiation associated genes were selected based on the rank of degree in the final network.

Our weighted graph strategy was a simple but reliable method to search for differentiation associated genes. Moreover, it provided a novel way to discover candidate features associated with cell fates. Our strategy was intuitive and could be easily scaled up to for both diverse and large quantities of rapidly growing information. It could also utilize the cross genome information to further improve prediction performance. In addition, the candidate features identified in our work will be helpful in understanding the physiological processes of stem cell differentiation.

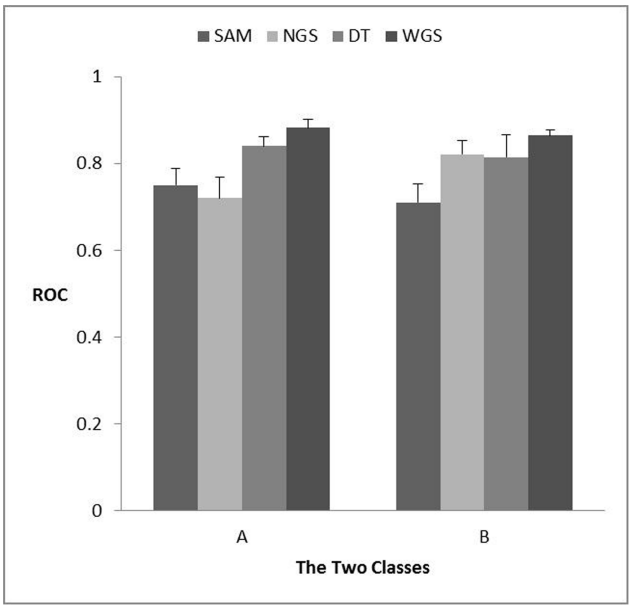

Figure 2. Average ROC scores for predicting the 2 classes. (Class A: the selected 30 genes; Class B: the other genes) using the 4 different approaches: (1) SAM; (2) Normal graph strategy (NGS); (3) Decision Tree (DT); (4) weighted graph strategy (WGS).
doi:10.1371/journal.pone.0062716.g002

Supporting Information

Table S1 List of the top 100 genes selected by weighted graph strategy.
(DOCX)

Table S2 Functions of 70 ''differentiation candidate genes'' in stem cells.
(DOCX)

Acknowledgments

We appreciated Da Wo for his help in language revision.

Author Contributions

Conceived and designed the experiments: JZ JL. Performed the experiments: JZ JL. Analyzed the data: JZ. Contributed reagents/materials/analysis tools: LP. Wrote the paper: JZ JL. Helped to analyse the data: YS.

References

1. Blau HM (1989) How fixed is the differentiated state? Lessons from heterokaryons. Trends Genet 5: 268–272.
2. Wang J, Rao S, Chu J, Shen X, Levasseur DN, et al. (2006) A protein interaction network for pluripotency of embryonic stem cells. Nature 444: 364–368.
3. Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, et al. (2004) Gene expression in human embryonic stem cell lines: unique molecular signature. Blood 103: 2956–2964.
4. Murali TM, Wu CJ, Kasif S (2006) The art of gene function prediction. Nat Biotechnol 24: 1474–1475; author reply 1475–1476.
5. Nishiyama A, Xin L, Sharov AA, Thomas M, Mowrer G, et al. (2009) Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell Stem Cell 5: 420–433.
6. Hailesellasse Sene K, Porter CJ, Palidwor G, Perez-Iratxeta C, Muro EM, et al. (2007) Gene function in early mouse embryonic stem cell differentiation. BMC Genomics 8: 85.
7. Walker E, Ohishi M, Davey RE, Zhang W, Cassar PA, et al. (2007) Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. Cell Stem Cell 1: 71–86.
8. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, et al. (2005) Multiple-laboratory comparison of microarray platforms. Nat Methods 2: 345–350.
9. Prieto C, De Las Rivas J (2006) APID: Agile Protein Interaction DataAnalyzer. Nucleic Acids Res 34: W298–302.
10. Bader GD, Betel D, Hogue CW (2003) BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res 31: 248–250.
11. Razick S, Magklaras G, Donaldson I (2008) iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics 9: 1471–2105.
12. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, et al. (2007) MINT: the Molecular INTeraction database. Nucleic Acids Res 35: D572–574.
13. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, et al. (2003) STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 31: 258–261.
14. Lage K, Karlberg EO, Storling ZM, Olason PI, Pedersen AG, et al. (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 25: 309–316.

15. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98: 5116–5121.
16. Klipper-Aurbach Y, Wasserman M, Braunspiegel-Weintrob N, Borstein D, Peleg S, et al. (1995) Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. Med Hypotheses 45: 486–490.
17. Xie Y, Pan W, Khodursky AB (2005) A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. Bioinformatics 21: 4280–4288.
18. Zhang S (2007) A comprehensive evaluation of SAM, the SAM R-package and a simple modification to improve its performance. BMC Bioinformatics 8: 230.
19. Jane Grossman MG, Robert Katz (1980) The First Systems of Weighted Differential and Integral Calculus.
20. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
21. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF (2007) A new method to measure the semantic similarity of GO terms. Bioinformatics 23: 1274–1281.
22. Wu H, Su Z, Mao F, Olman V, Xu Y (2005) Prediction of functional modules based on comparative genome analysis and Gene Ontology application. Nucleic Acids Res 33: 2822–2837.
23. Langaas Mea (2005) Statistical hypothesis testing of association between two reporter lists within the GO-hierarchy. Technical report Department of Mathematical Sciences, Norwegian University of Science and Technology.
24. Chua HN, Sung WK, Wong L (2006) Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. Bioinformatics 22: 1623–1630.
25. Lee I, Date SV, Adai AT, Marcotte EM (2004) A probabilistic functional network of yeast genes. Science 306: 1555–1558.
26. Zhang J, Li X, Jiang W, Wang YQ, Li CX, et al. (2005) A novel ensemble decision tree approach for mining genes coding ion channels for cardiopathy subtype. Fuzzy Systems and Knowledge Discovery, Pt 2, Proceedings 3614: 852–860.
27. Gribskov M, Robinson NL (1996) Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. Comput Chem 20: 25–33.
28. Pavlidis P, Li Q, Noble WS (2003) The effect of replication on gene expression microarray experiments. Bioinformatics 19: 1620–1627.
29. Gu B, Zhang J, Chen Q, Tao B, Wang W, et al. Aire regulates the expression of differentiation-associated genes and self-renewal of embryonic stem cells. Biochem Biophys Res Commun 394: 418–423.
30. Schrenk-Siemens K, Perez-Alcala S, Richter J, Lacroix E, Rahuel J, et al. (2008) Embryonic stem cell-derived neurons as a cellular system to study gene function: lack of amyloid precursor proteins APP and APLP2 leads to defective synaptic transmission. Stem Cells 26: 2153–2163.
31. Sugaya K, Kwak YD, Ohmitsu O, Marutle A, Greig NH, et al. (2007) Practical issues in stem cell therapy for Alzheimer's disease. Curr Alzheimer Res 4: 370–377.
32. Pietersen AM, Evers B, Prasad AA, Tanger E, Cornelissen-Steijger P, et al. (2008) Bmi1 regulates stem cells and proliferation and differentiation of committed cells in mammary epithelium. Curr Biol 18: 1094–1099.
33. Amleh A, Nair SJ, Sun J, Sutherland A, Hasty P, et al. (2009) Mouse cofactor of BRCA1 (Cobra1) is required for early embryogenesis. PLoS One 4: e5034.
34. Liu S, Ginestier C, Charafe-Jauffret E, Foco H, Kleer CG, et al. (2008) BRCA1 regulates human mammary stem/progenitor cell fate. Proc Natl Acad Sci U S A 105: 1680–1685.
35. Wu Q, Bruce AW, Jedrusik A, Ellis PD, Andrews RM, et al. (2009) CARM1 is required in embryonic stem cells to maintain pluripotency and resist differentiation. Stem Cells 27: 2637–2645.
36. Ochsner SA, Strick-Marchand H, Qiu Q, Venable S, Dean A, et al. (2007) Transcriptional profiling of bipotential embryonic liver cells to identify liver progenitor cell surface markers. Stem Cells 25: 2476–2487.
37. Shaw L, Johnson PA, Kimber SJ Gene expression profiling of the developing mouse kidney and embryo. In Vitro Cell Dev Biol Anim 46: 155–165.
38. Bar-On O, Shapira M, Skorecki K, Hershko A, Hershko DD (2010) Regulation of APC/C (Cdh1) ubiquitin ligase in differentiation of human embryonic stem cells. Cell Cycle 9: 1986–1989.
39. Yao W, Qian W, Zhu C, Gui L, Qiu J, et al. (2010) APC is involved in the differentiation of neural stem cells into neurons. Neuroreport 21: 39–44.
40. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, et al. (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123: 917–929.
41. Chung KC, Ahn YS (1998) Expression of immediate early gene cyr61 during the differentiation of immortalized embryonic hippocampal neuronal cells. Neurosci Lett 255: 155–158.
42. Bourillot PY, Aksoy I, Schreiber V, Wianny F, Schulz H, et al. (2009) Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. Stem Cells 27: 1760–1771.
43. Richie ER, Schumacher A, Angel JM, Holloway M, Rinchik EM, et al. (2002) The Polycomb-group gene eed regulates thymocyte differentiation and suppresses the development of carcinogen-induced T-cell lymphomas. Oncogene 21: 299–306.
44. Ura H, Murakami K, Akagi T, Kinoshita K, Yamaguchi S, et al. (2011) Eed/Sox2 regulatory loop controls ES cell self-renewal through histone methylation and acetylation. EMBO J 30: 2190–2204.
45. Wice BM, Gordon JI (1998) Forced expression of Id-1 in the adult mouse small intestinal epithelium is associated with development of adenomas. J Biol Chem 273: 25310–25319.
46. Suh HN, Han HJ Collagen I regulates the self-renewal of mouse embryonic stem cells through alpha2beta1 integrin- and DDR1-dependent Bmi-1. J Cell Physiol.
47. Traister A, Aafaqi S, Masse S, Dai X, Li M, et al. (2012) ILK induces cardiomyogenesis in the human heart. PLoS One 7: e37802.
48. Rubin R, Arzumanyan A, Soliera AR, Ross B, Peruzzi F, et al. (2007) Insulin receptor substrate (IRS)-1 regulates murine embryonic stem (mES) cells self-renewal. J Cell Physiol 213: 445–453.
49. Salehi H, Karbalaie K, Razavi S, Tanhaee S, Nematollahi M, et al. (2011) Neuronal induction and regional identity by co-culture of adherent human embryonic stem cells with chicken notochords and somites. Int J Dev Biol 55: 321–326.
50. Perry P, Sauer S, Billon N, Richardson WD, Spivakov M, et al. (2004) A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction. Cell Cycle 3: 1645–1650.
51. Feinberg MW, Wara AK, Cao Z, Lebedeva MA, Rosenbauer F, et al. (2007) The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. EMBO J 26: 4138–4148.
52. Zhang SX, Garcia-Gras E, Wycuff DR, Marriot SJ, Kadeer N, et al. (2005) Identification of direct serum-response factor gene targets during Me2SO-induced P19 cardiac cell differentiation. J Biol Chem 280: 19115–19126.
53. Wang Z, Oron E, Nelson B, Razis S, Ivanova N (2012) Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell 10: 440–454.
54. Niakan KK, Davis EC, Clipsham RC, Jiang M, Dehart DB, et al. (2006) Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. Mol Genet Metab 88: 261–271.
55. Dupont E, Sansal I, Toru D, Evrard C, Rouget P (1997) [Identification of NPDC-1, gene involved in the control of proliferation and differentiation of neural and glial precursors]. C R Seances Soc Biol Fil 191: 95–104.
56. Casson SA, Topping JF, Lindsey K (2009) MERISTEM-DEFECTIVE, an RS domain protein, is required for the correct meristem patterning and function in Arabidopsis. Plant J 57: 857–869.
57. Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24: 372–376.
58. Deb-Rinker P, Ly D, Jezierski A, Sikorska M, Walker PR (2005) Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. J Biol Chem 280: 6257–6260.
59. Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, et al. Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev 24: 265–276.
60. Mouillet-Richard S, Laurendeau I, Vidaud M, Kellermann O, Laplanche JL (1999) Prion protein and neuronal differentiation: quantitative analysis of prnp gene expression in a murine inducible neuroectodermal progenitor. Microbes Infect 1: 969–976.
61. Nakajima M, Ogawa M, Shimoda Y, Hiraoka S, Iida M, et al. (2006) Presenilin-1 controls the growth and differentiation of endothelial progenitor cells through its beta-catenin-binding region. Cell Biol Int 30: 239–243.
62. Veeraraghavalu K, Choi SH, Zhang X, Sisodia SS (2010) Presenilin 1 mutants impair the self-renewal and differentiation of adult murine subventricular zone-neuronal progenitors via cell-autonomous mechanisms involving notch signaling. J Neurosci 30: 6903–6915.
63. Katoh M (2007) Comparative integromics on non-canonical WNT or planar cell polarity signaling molecules: transcriptional mechanism of PTK7 in colorectal cancer and that of SEMA6A in undifferentiated ES cells. Int J Mol Med 20: 405–409.
64. Li L, Wang S, Jezierski A, Moalim-Nour L, Mohib K, et al. (2010) A unique interplay between Rap1 and E-cadherin in the endocytic pathway regulates self-renewal of human embryonic stem cells. Stem Cells 28: 247–257.
65. Wen J, Huang S, Rogers H, Dickinson LA, Kohwi-Shigematsu T, et al. (2005) SATB1 family protein expressed during early erythroid differentiation modifies globin gene expression. Blood 105: 3330–3339.
66. Savarese F, Davila A, Nechanitzky R, De La Rosa-Velazquez I, Pereira CF, et al. (2009) Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev 23: 2625–2638.
67. Alfaro MP, Vincent A, Saraswati S, Thorne CA, Hong CC, et al. (2010) sFRP2 suppression of bone morphogenic protein (BMP) and Wnt signaling mediates mesenchymal stem cell (MSC) self-renewal promoting engraftment and myocardial repair. J Biol Chem 285: 35645–35653.
68. Kele J, Andersson ER, Villaescusa JC, Cajanek L, Parish CL, et al. (2012) SFRP1 and SFRP2 dose-dependently regulate midbrain dopamine neuron development in vivo and in embryonic stem cells. Stem Cells 30: 865–875.
69. Wawrzak D, Metioui M, Willems E, Hendrickx M, de Genst E, et al. (2007) Wnt3a binds to several sFRPs in the nanomolar range. Biochem Biophys Res Commun 357: 1119–1123.
70. Thomson M, Liu SJ, Zou LN, Smith Z, Meissner A, et al. (2011) Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 145: 875–889.