Protein Cell
DOI 10.1007/s13238-012-0014-4

Protein & Cell

RESEARCH ARTICLE

System approaches reveal the molecular networks involved in neural stem cell differentiation

Kai Wang[1,5*], Haifeng Wang[2,5*], Jiao Wang[2,5*], Yuqiong Xie[1,5], Jun Chen[3,4], Huang Yan[1,5], Zengrong Liu[5], Tieqiao Wen[1,5] ✉

[1] Laboratory of Molecular Neurobiology, School of Life Sciences, Shanghai University, Shanghai 200444, China
[2] School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China
[3] College of Life Science, Wuhan University, Wuhan 430072, China
[4] College of Chemical Engineering and Technology, Wuhan University of Science and Technology, Wuhan 430081, China
[5] Institute of Systems Biology, Shanghai University, Shanghai 200444, China
✉ Correspondence: tqwen@staff.shu.edu.cn
Received December 20, 2011 Accepted January 12, 2012

ABSTRACT

The self-renewal and multipotent potentials in neural stem cells (NSCs) maintain the normal physiological functions of central nervous system (CNS). The abnormal differentiation of NSCs would lead to CNS disorders. However, the mechanisms of how NSCs differentiate into astrocytes, oligodendrocytes (OLs) and neurons are still unclear, which is mainly due to the complexity of differentiation processes and the limitation of the cell separation method. In this study, we modeled the dynamics of neural cell interactions in a systemic approach by mining the high-throughput genomic and proteomic data, and identified 8615 genes that are involved in various biological processes and functions with significant changes during the differentiation processes. A total of 1559 genes are specifically expressed in neural cells, in which 242 genes are NSC specific, 215 are astrocyte specific, 551 are OL specific, and 563 are neuron specific. In addition, we proposed 57 transcriptional regulators specifically expressed in NSCs may play essential roles in the development courses. These findings provide more comprehensive analysis for better understanding the endogenous mechanisms of NSC fate determination.

KEYWORDS neural stem cell, cell differentiation, molecular networks

INTRODUCTION

Neural stem cells (NSCs) are central nervous system (CNS) originated multipotent stem cells, which possess self-renewal potential and have the ability to differentiate into astrocytes, oligodendrocytes (OLs) and neurons along with the development time course (Temple, 2001). They play essential roles in repairing and maintaining normal CNS cells. Besides, tissue engineering by normal NSCs transplantation also enables the clinical therapeutic applications for CNS regressive diseases, such as Alzheimer's, Parkinson's, Huntington's disease and other CNS disorders (Storch and Schwarz, 2002; Lee et al., 2005; Oliveira and Hodges, 2005; Sanberg, 2007; Tang et al., 2008; Moghadam et al., 2009; Zhongling et al., 2009; Makri et al., 2010).

In recent years, progresses have been made in finding out regulating factors which determine the neural stem cell differentiation orientations. The npBAF complex (neural progenitors-specific chromatin remodeling complex), together with the nBAF complex (neuron-specific chromatin remodeling complex), are required for chromatin remodeling during the transition from NSCs to adult neurons (Lessard et al., 2007). Nestin, a well-known protein marker for NSCs, is critical for NSCs survival, renewal and proliferation (Park et al., 2010). Chrdl1 (Chordin-like protein 1, Neurogenesin-1, Ng1) is reported to prevent glial fate during NSCs differentiation (Ueki et al., 2003). Cxxc5 (CXXC-type zinc finger protein 5) is proposed to be a negative regulator of canonical Wnt signali-

*These authors contributed equally to the work.

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

signaling pathway in response to BMP4 (bone morphogenetic protein 4). Furthermore, it also affects telencephalic forebrain development and telencephalic NSCs differentiation into various cells (Andersson et al., 2009). The extracellular heme-binding protein Cyb5d2 (Cytochrome b5 domain-containing protein 2, Neuferricin), could also promote adult neurogenesis (Kimura et al., 2010). The transcription factor Sox2, along with Sox8 and Sox10, is also required for the maintenance of NSCs multipotent status (Episkopou, 2005; Maka et al., 2005; Favaro et al., 2009). Several recent studies also indicate the Pax6, a transcription factor of Pax family, may have crucial roles in adult neurogenesis (Matsumoto and Osumi, 2008; Osumi et al., 2008). Some other studies demonstrated that the differentiation of NSCs is controlled by both signaling transduction and transcriptional regulation. A recent *in vitro* study also reported many genes and signaling pathways are involved in neural stem cell differentiation (Park et al., 2012). Advances of epigenetic mechanisms regulating neural stem cell differentiation have been reviewed (Namihira et al., 2008; Juliandi et al., 2010). In spite of these efforts, the molecular mechanisms of NSCs differentiation are still unforeseen. And the main obstacle is the complexity of differentiation processes and lack of purification procedures to distinguish different NSCs products (the astrocytes, neurons and OLs) with high specificity and precision (Cahoy et al., 2008).

Recently, with the application of the fluorescent-activated cell sorting (FACS) technology (Cahoy et al., 2008), these technical difficulties have been well resolved. In the present study, we gain an insight into the mechanisms of NSC differentiation in a system approach by employing the public available gene expression data. We show that, many signaling pathways as well as many transcription factors contributed to the differentiation process. Investigation of gene expression of the four neural cell-types (NSCs, astrocytes, neurons and OLs) by comparing with non-CNS cells shows many newly identified specifically expressed genes in CNS. In addition, according to the genes involved in this process, we constructed a molecular interaction network which may determine the NSCs fates. Our study provides a more comprehensive analysis of the mechanisms of NSCs differentiation.

RESULTS

Gene expression profile of neural cells

Besides astrocytes, neurons and OLs, expressions of some well described gene markers of NSCs in mRNA level were also validated. Expressions of these gene markers (Fig. 1A and 1B) are consistent with previously reported data, demonstrating the reliability of the selected samples.

Genes that are not expressed in any of the neural cells are excluded. So MAS 5.0 algorithm and absent/present (A/P) calls were introduced to exclude probes without consistent expression in all of the eight cell types (see MATERIALS AND METHODS). As a result, a collection of 28,175 probes covering 17,327 unique genes were observed that have no less than two-thirds ″P″ signals in at least one cell type. In order to find the genes that are differentially expressed among the neural cells, probe expression levels which were significantly up/down-regulated by at least 1.5-fold comparing with the NSCs after a log2-transformed expression by GCRMA (Wu et al., 2004) were required. Consequently, a total of 11,808 probes representing 8615 unique genes were considered differentially expressed during the process of NSCs differentiation significantly. Among these genes, 4285 were significantly expressed during any astrocyte development stage including 2139 (~49.3%) up-regulated and 2196 (~50.7%) down-regulated; 5943 were significantly expressed during any neuron development stage including 3138 (~52.1%) up-regulated and 2885 (~47.9%) down-regulated; and 5955 were significantly expressed during any OL development stage including 3086 (~49.9%) up-regulated and 3098 (~50.1%) down-regulated.

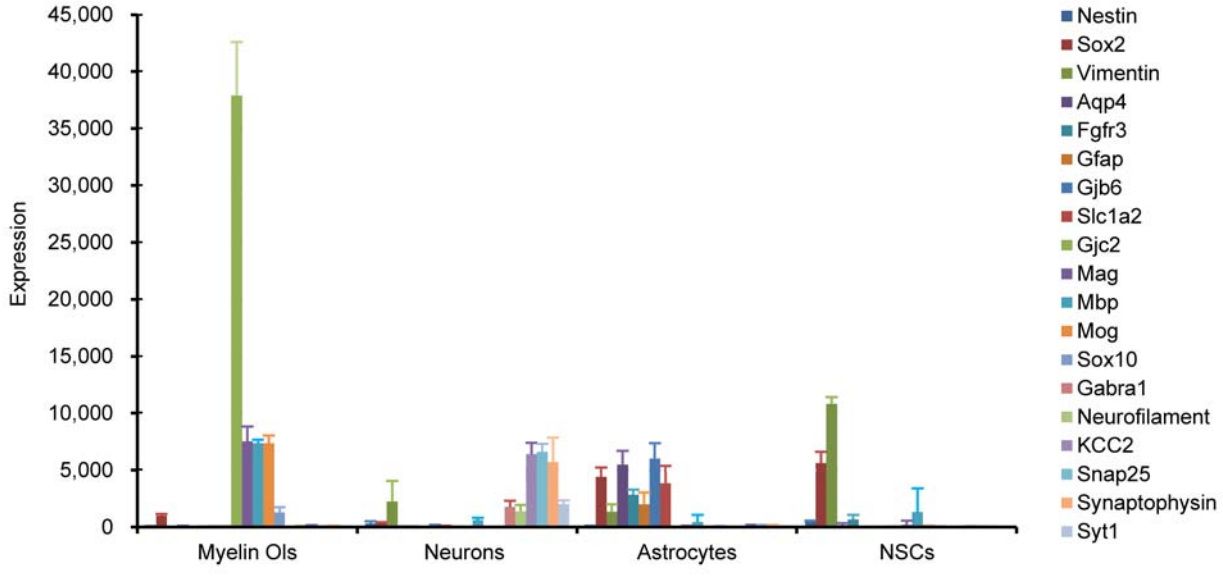
Identifying specifically expressed genes of neural cells

Identifying specifically expressed genes of neural cells will help us distinguish their characteristics from other cell types and further promote their functional mechanism research. It was reported that cell type specific genes were identified by quantitative comparison of different gene expressions among astrocytes, neurons and OLs (Cahoy et al., 2008). However, these quantitative comparisons only restricted to astrocytes, neurons and OLs, which are the three major neural cells. In order to understand the underlying mechanism of NSC differentiation, it is better to distinguish the expression characteristics among all of the NSC and its products as well as other non-neural cell/tissue types. Here, the neural cell, non-CNS and non-multipotent cell samples of 14 different tissues which were used in previous mouse transcriptome analysis (Thorrez et al., 2008) were utilized, including adipose tissue, adrenal gland, bone marrow, diaphragm, eye, heart, kidney, liver, lung, muscle, salivary gland, small intestine, spleen and thymus.

We identified neural cells specifically expressed genes from the 8615 unique genes. As a result, 1559 genes were specifically expressed in at least one of the eight neural cells. In order to further distinguish the eight neural cells, the Maximum Relevance Minimum Redundancy (mRMR) feature selection method (see MATERIALS AND METHODS) (Ding and Peng, 2005; Peng et al., 2007), which had been demonstrated as a very useful supervised machine learning method for discriminative gene selection, was employed to sort these features (probes). Then, leave-one-out cross-validation combined with the supervised SVM predicator (Chang and Lin, 2011) was performed to determine how many features could best represent all the eight neural cell types. As illuminated in the IFS curve, a total of 1758 features were chosen when the

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

A



B

Well described markers	Myelin Ols	Neurons	Astrocytes	NSC
Nestin (NSC)	76.82	314.37	92.83	456.94
Sox2 (NSC)	1005.25	341.36	4414.38	5617.78
Vimentin (NSC)	14.67	2258.68	1360.74	10818.93
Aqp4 (Astrocyte)	81.61	41.16	5471.52	163.72
Fgfr3 (Astrocyte)	29.98	48.24	2840.23	677.47
Gfap (Astrocyte)	38.19	20.10	2001.26	24.41
Gjb6 (Astrocyte)	51.28	118.33	6027.34	14.21
Slc1a2 (Astrocyte)	38.76	76.16	3844.26	23.29
Gjc2 (OL)	37911.64	38.85	28.09	34.38
Mag (OL)	7526.29	28.50	56.14	205.67
Mbp (OL)	7354.93	604.30	454.56	1328.50
Mog (OL)	7357.59	0.51	6.88	47.09
Sox10 (OL)	1295.55	18.86	49.47	39.16
Gabra1 (Neuron)	11.57	1771.69	16.14	6.89
Neurofilament (Neuron)	59.18	1384.81	24.95	20.70
KCC2 (Neuron)	103.44	6417.75	110.42	43.41
Snap25 (Neuron)	59.90	6619.70	134.93	49.84
Synaptophysin (Neuron)	76.05	5707.17	181.22	33.81
Syt1 (Neuron)	31.53	2024.90	38.63	11.13

Figure 1. mRNA level expressions of well described markers for NSCs, astrocytes, OLs and neurons. The y-axis represents gene expression level intensities and error bars represent +SEM (A), and the expression value is shown in (B).

accuracy peaked and contained most features (Fig. 2A), which include 242 NSC specifically expressed genes, 215 astrocyte specifically expressed genes, 551 OL specifically expressed genes, and 563 neuron specifically expressed genes (Fig. 2B; Table 2S). Notably, the most predictive accuracies were among 95% and 100%, indicating that most of the selected probes are specifically expressed. That means a small proportion of those probes could be sufficient to distinguish all neural cells. In addition, unsupervised hierarchical clustering of these samples shows quite similar expression profiles between the same cell types (e.g. the NSCs samples obtained from two different studies show close relationships) and distinct expression profiles between different neural cells (Fig. S1), demonstrating the samples are stabilized and the specifically expressed genes we selected are credible.

General annotation of neural cells specifically expressed genes

To identify the molecular characteristics of each neural cell, the functional enrichment analysis was performed by DAVID (Huang da et al., 2009a, b). These 1559 specifically expressed genes are involved in various functions (Fig. 3; $p \leqslant 0.01$). In the prioritized gene functions, cell cycle, regulation of transcription, DNA metabolic process, DNA damage,

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

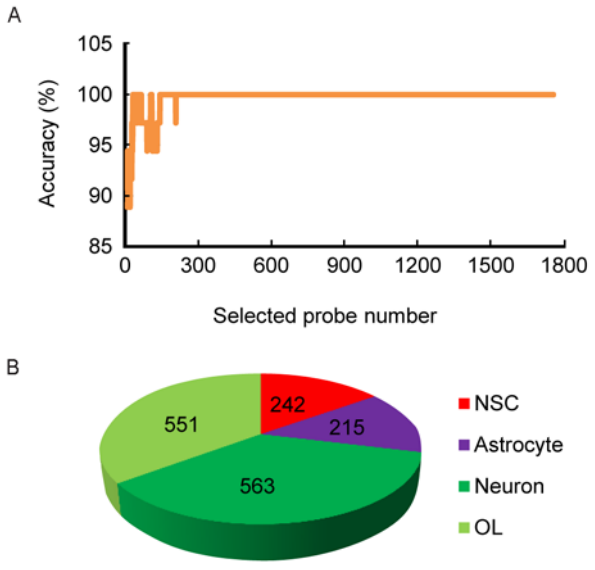


Figure 2. Cell-type-specific genes selection and expression in NSCs, astrocytes, OLs and neurons. (A) IFS curve for cell-type-specific markers selection. The X-axis is the selected probe number; the Y-axis is the predicated accuracy by leave-one-out cross-validation. (B) Destitution of specifically expressed genes in NSCs, astrocytes, OLs and neurons.

DNA repair, ATPase and helicase activity, regulation of microtubule-based process, Wnt receptor signaling pathway and RNA processing were most relevant to NSCs functions (Fig. 3A). Nervous system development and cell differentiation, embryonic morphogenesis and embryonic organ development, activation of caspase activity by cytochrome c and enzyme linked receptor protein signaling pathway were enriched in astrocytes (Fig. 3B). Cell adhesion, nerve impulse and action potential in neuron, hyaluronic acid binding, channel activity and transmembrane transporter activity, neuron differentiation and negative regulation of molecular function were most relevant to ordinary function of OLs (Fig. 3C). While, obviously more functions were observed in neurons, such as channel activity and transmembrane transporter activity, neuron differentiation, development and morphogenesis, serotonin and amine receptor activity, neurotransmitter and neuropeptide receptor activity, somatostatin receptor activity, hormone activity, adult behavior, ephrin receptor activity, behavioral fear and defense response, learning or memory, cAMP biosynthetic process and cAMP-mediated signaling, ion binding, cell adhesion, regulation of neurological system process, regulation of neurogenesis, response to stress and pain andregulation of neuron apoptosis (Fig. 3D). Details of these function clusterings are available in Table S3 ($p \leqslant 0.01$).

Table 1 lists all of the significantly enriched pathways separately ($p \leqslant 0.05$). Although some common pathways were detected (e.g. basal cell carcinoma pathway and neuroactive ligand-receptor interaction pathway), nearly absolutely different pathways are enriched in different cells (Table 1), indicating the differences of their functional mechanisms. To be specific, cell cycle, basal cell carcinoma, glycosylphosphatidylinositol (GPI)-anchor biosynthesis are deposed in NSCs. Basal cell carcinoma, pathways in cancer, colorectal cancer, amyotrophic lateral sclerosis (ALS), dorso-ventral axis formation, MAPK signaling pathway, melanoma, apoptosis, prostate cancer are deposed in astrocytes. Cell adhesion molecules (CAMs), axon guidance, neuroactive ligand-receptor interaction, gap junction are deposed in OLs, and neuroactive ligand-receptor interaction, axon guidance, calcium signaling pathway are deposed in neurons.

These results indicate that NSCs differentiation is a complex biological process, which includes many important signaling pathways and genes with various functions. It's difficult to imagine that a single gene or signaling pathway could initiate the whole differentiation processes. So, a system approach is more powerful to understand the complex differentiation process.

Molecular interaction network models

We particularly paid attention to those genes that specifically expressed in NSCs and function in transcriptional regulation. To detect potential functional associations between the 57 transcriptional regulators (Fig. 3; Table 2) and all of the specifically expressed genes in neural cells, we performed Pearson correlation coefficient (PCC) analysis between them based on the 36 samples. The resulted co-expression network contains 56 transcriptional regulators and 187 other genes with 6645 unique functional associations when PPC $\geqslant 0.80$ (Table S4). As depicted, there were double apices in our dataset compared with only one apex in the randomly generated dataset (Fig. 4A); that is, it was increased other than decreased in number when PCC $\geqslant 0.45$, indicating these genes could be regulated by the transcriptional regulators. Besides, it's obvious that there is an increase both in number and ratio when the PPC $\geqslant 0.80$ (Fig. 4A and 4B), demonstrating the possibility was much more than random.

Literature curation indicated that at least 10 in 56 transcriptional regulators have potential roles in differentiation, development or cell fate determination, including ASCL1, HIP1, MYBL1, POU3F4, RCOR2, SMARCC1, SOX11, SOX3, TEAD2, ZFP2 and ZFP354C (ZNF354C). This suggested that they may play important roles in NSCs differentiation. Furthermore, public protein–protein interaction (PPI) databases, such as BioGRID (Stark et al., 2011), DIP (Salwinski et al., 2004), IntAct (Aranda et al., 2010) and STRING (Szklarczyk et al., 2011) were all used to find additional interactions with the 10 regulators. These databases contain both literature curated associations and computationally

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

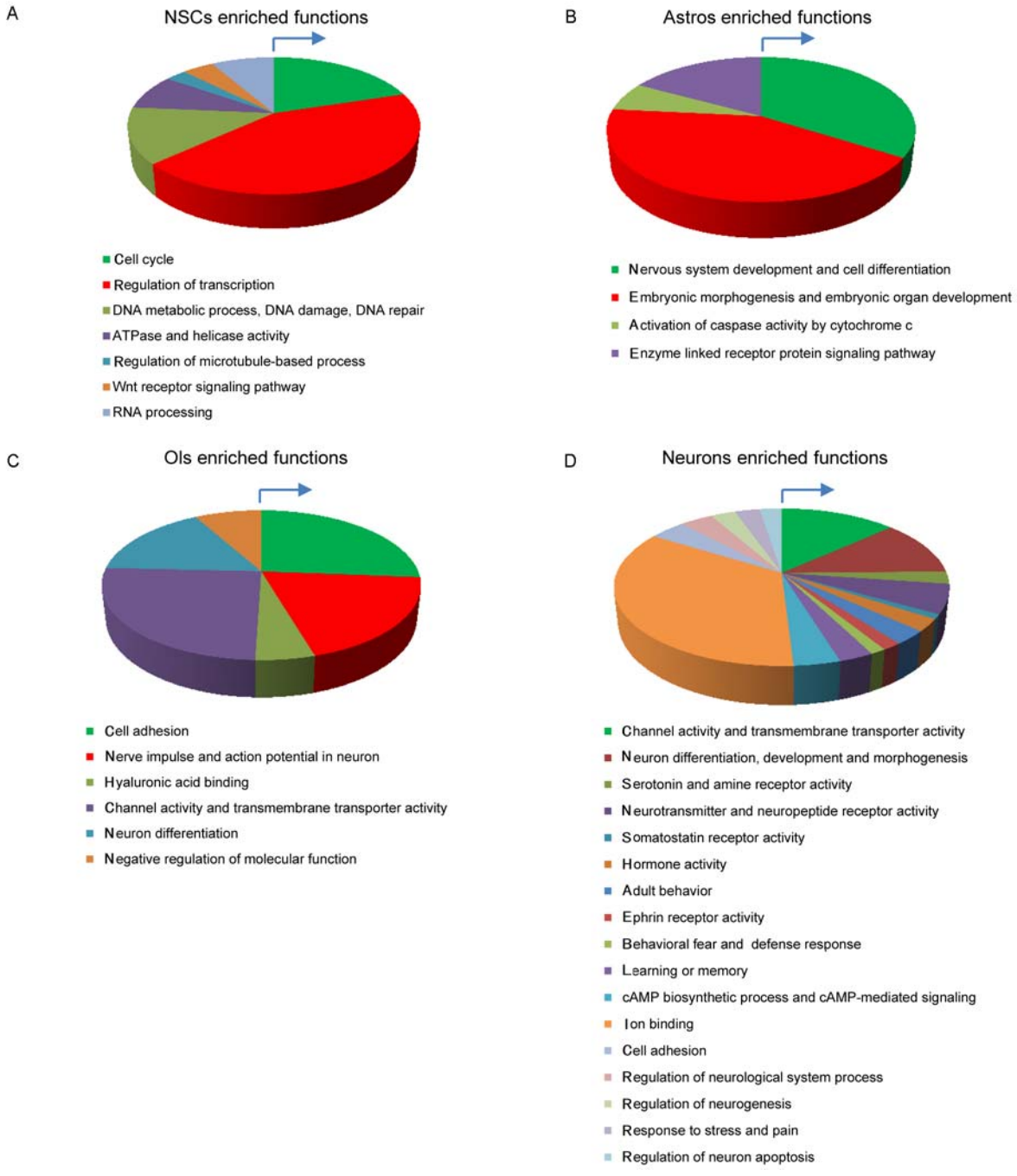


Figure 3. General annotation of most significantly enriched functions of specifically expressed genes in NSCs (A), astrocytes (B), OLs (C) and neurons (D) ($p \leqslant 0.01$).

predicated associations. In order to remove most irrelevant proteins, we use the depth first search (DFS) algorithm to search associations in the first depth, as we demonstrated in our previous study (Wang et al., 2011). Besides, some additional irrelevant proteins from STRING database (Szklarczyk et al., 2011) were excluded by using an integrated score threshold 0.800. The protein–protein interaction network includes 237 nodes and 252 edges, as is shown in Fig. 5. Due to the limitation of the data, no functional association was found about ZFP354C, and only a few associations were found about ZFP2 and RCOR2.

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

Table 1 Pathway enrichment analysis of neural cells specifically expressed genes

Index	Enriched KEGG pathways	*p* value	FDR
NSC enriched pathways			
1	mmu04110:Cell cycle	5.74E-03	5.66
2	mmu05217:Basal cell carcinoma	1.32E-02	12.61
3	mmu00563:Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	2.12E-02	19.47
Astrocyte enriched pathways			
1	mmu05217:Basal cell carcinoma	7.07E-04	0.76
2	mmu05200:Pathways in cancer	2.37E-03	2.53
3	mmu05210:Colorectal cancer	3.72E-03	3.95
4	mmu05014:Amyotrophic lateral sclerosis (ALS)	8.66E-03	8.97
5	mmu04320:Dorso-ventral axis formation	1.15E-02	11.77
6	mmu04010:MAPK signaling pathway	1.33E-02	13.44
7	mmu05218:Melanoma	1.57E-02	15.75
8	mmu04210:Apoptosis	2.69E-02	25.50
9	mmu05215:Prostate cancer	2.93E-02	27.50
Oligodendrocye enriched pathways			
1	mmu04514:Cell adhesion molecules (CAMs)	1.57E-04	0.18
2	mmu04360:Axon guidance	8.47E-04	0.95
3	mmu04080:Neuroactive ligand-receptor interaction	1.07E-02	11.37
4	mmu04540:Gap junction	2.41E-02	24.03
Neuron enriched pathways			
1	mmu04080:Neuroactive ligand-receptor interaction	3.09E-14	0.00
2	mmu04360:Axon guidance	1.32E-06	0.00
3	mmu04020:Calcium signaling pathway	6.35E-05	0.07

$p \leqslant 0.05$

Table 2 Transcriptional regulation related genes enriched in NSCs

Index	Probe name	Gene symbol	Gene name
1	1452334_at	CENPF	centromere protein F
2	1430139_at, 1453361_at	HELLS (ZFM1, ZNF162)	helicase, lymphoid specific
3	1422851_at, 1450780_s_at, 1450781_at	HMGA2	high mobility group AT-hook 2
4	1429557_at	MCM8	minichromosome maintenance deficient 8
5	1416310_at	NUSAP1	nucleolar and spindle associated protein 1
6	1417323_at, 1425416_s_at	PSRC1	proline/serine-rich coiled-coil 1
7	1423026_at, 1438453_at	RAD51C	Rad51 homolog c (*S. cerevisiae*)
8	1422321_a_at	SF1	splicing factor 1
9	1422979_at, 1433996_at	SUV39H2	suppressor of variegation 3-9 homolog 2
10	1437086_at, 1450164_at	ASCL1	achaete-scute complex homolog 1
11	1421149_a_at	ATN1	atrophin 1
12	1428649_at	CAND1	cullin associated and neddylation disassociated 1
13	1421933_at	CBX5	chromobox homolog 5 (Drosophila HP1a)
14	1442576_at	CREB5	cAMP responsive element binding protein 5
15	1432272_a_at	ERCC4	excision repair cross-complementing rodent repair deficiency, complementation group 4

(*to be continued on the next page*)

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

(*Continued*)

Index	Probe name	Gene symbol	Gene name
16	1420998_at	ETV5	ets variant gene 5
17	1448834_at	FOXM1	forkhead box M1
18	1424756_at	HIP1	huntingtin interacting protein 1
19	1420565_at	HOXA1	homeo box A1
20	1422546_at	ILF3	interleukin enhancer binding factor 3
21	1420665_at	ITGB3BP	integrin beta 3 binding protein (beta3-endonexin)
22	1450402_at	MED1	mediator complex subunit 1
23	1452816_at	MLF1IP	myeloid leukemia factor 1 interacting protein
24	1450481_at	MYBL1	myeloblastosis oncogene-like 1
25	1451841_a_at	NCOR2	nuclear receptor co-repressor 2
26	1441325_at	NKRF	NF-kappaB repressing factor
27	1421567_at, 1450287_at	NPAS3	neuronal PAS domain protein 3
28	1450854_at	PA2G4	proliferation-associated 2G4
29	1427550_at	PEG10	paternally expressed 10
30	1453761_at	PHF6	PHD finger protein 6
31	1436738_at	PIF1	PIF1 5'-to-3' DNA helicase homolog
32	1422164_at	POU3F4	POU domain, class 3, transcription factor 4
33	1423318_at	RAD18	RAD18 homolog
34	1421430_at	RAD51L1	RAD51-like 1
35	1417302_at	RCOR2	REST corepressor 2
36	1450489_at	SALL1	sal-like 1
37	1416670_at	SETDB1	SET domain, bifurcated 1
38	1423417_at	SMARCC1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1
39	1436790_a_at	SOX11	SRY-box containing gene 11
40	1435192_at, 1450485_at	SOX3	SRY-box containing gene 3
41	1449370_at	SOX4	SRY-box containing gene 4
42	1449578_at	SUPT16H	suppressor of Ty 16 homolog
43	1442289_at	TAF1D	TATA box binding protein (Tbp)-associated factor, RNA polymerase I, D
44	1457351_at	TAF2	TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor
45	1450100_a_at	TCERG1	transcription elongation regulator 1 (CA150)
46	1448519_at	TEAD2	TEA domain family member 2
47	1431115_at	TGIF2	TGFB-induced factor homeobox 2
48	1430384_at	TLE4	transducin-like enhancer of split 4, homolog of Drosophila E(spl)
49	1452818_at	TTF2	transcription termination factor, RNA polymerase II
50	1452088_at	ZBED3	zinc finger, BED domain containing 3
51	1428887_at	ZFP157 (ZNF157)	zinc finger protein 157
52	1452025_a_at	ZFP2	zinc finger protein 2
53	1450321_at	ZFP354C (ZNF354C)	zinc finger protein 354C
54	1421251_at	ZFP40 (HIVEP1, ZNF40)	zinc finger protein 40
55	1451816_at	ZFP451 (ZNF451)	zinc finger protein 451
56	1458274_at	ZFP69 (ZNF69)	zinc finger protein 69
57	1421395_at	ZIK1	zinc finger protein interacting with K protein 1

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

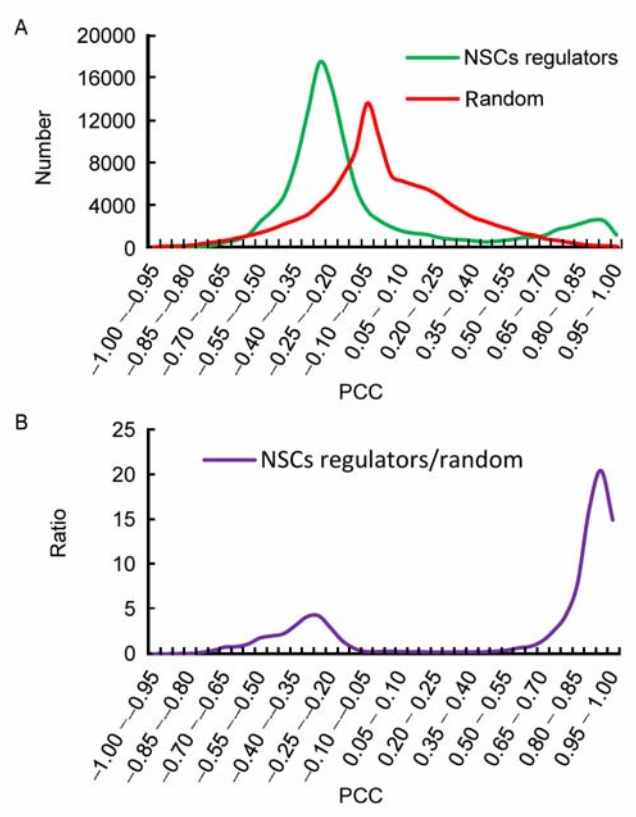


Figure 4. Pearson correlation coefficients distribution curve. (A) The green line represents the numbers of co-expression pairs between our gene set (57 NSC specific regulators and other specifically expressed genes); the red line represents the numbers of co-expression pairs between randomly generated gene set (randomly generated 1559 genes and 57 genes in the randomly generated genes). (B) The purple line represents the ratio of our gene set *versus* randomly generated gene set.

MATERIALS AND METHODS

Neural cells related gene expression data preparation

The raw data of gene expression series GSE12499, GSE18326 and GSE9566 were collected from NCBI Gene Expression Omnibus (GEO) (http: //www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2011). The series GSE12499 was originally reported to investigate the induced multipotent phenomenon by exogenous expression of an embryonic stem cell (ESC) specific transcription factor Oct4 in NSCs (Kim et al., 2009). The series GSE18326 was first reported to study the impact effects of FoxO3 null mutant on NSC maintenance (Renault et al., 2009). The series GSE9566 contains various neural cell samples with high specificity and precision sorted by FACS (Cahoy et al., 2008) except NSCs. All of the three gene expression series are based on the same and widely used Affymetrix Mouse Genome 430 2.0 Array, which contains 45101 probes of ~22,000 unique genes. The wild-type (WT) NSC samples were selected from GSE12499 and GSE18326. All representative samples of astrocytes, OLs and neurons were selected from GSE9566. A summarization of all samples used is listed in Table S1.

Neural cells related data pre-process

We grouped all these samples according to their cell types and development stages (Table S1). To be specific, we grouped the 36 samples as NSCs, immature astrocytes (P1), maturing astrocytes (P7–P8), mature astrocytes (P17–P30), maturing neurons (P7), mature neurons (P16), immature OLs (OPCs), maturing OLs and mature OLs (Myelin OLs). (P represents postnatal day). Obviously, genes that do not express in any of the grouped cell types have no significance. Only those probes that presented in at least two-thirds of the same cell-type, or called consistently expressed genes (probes in actual) in previous study, were kept (Cahoy et al., 2008). The absent/present (A/P) calls were employed to achieve this goal. Probes that were detected as ″P″ usually possess much bigger MAS 5.0 intensities as well as smaller *p*-values compared with those detected as ″A″. And the *p*-value was usually considered as no statistical significance often caused by a low detected intensity value. The GCRMA method in Bioconductor package was conducted for raw data normalization (Wu et al., 2004). GCRMA generally could obtain more accurate gene expression values than RMA method in microarray normalization. Genes with fold changes less than 1.5 were excluded for further process.

Functional and signaling pathway enrichment analysis

Gene ontology (GO) functional enrichment analysis and signaling pathway enrichment analysis of the significantly changed genes were performed by using DAVID (The Database for Annotation, Visualization and Integrated Discovery, http://david.abcc.ncifcrf.gov) (Huang da et al., 2009a, b), which is a powerful tool for biological meaning exploration.

Tissue/cell-type specific expression analysis

We require that the candidate neural cells specifically expressed genes must satisfy the following criteria: (A) the gene expression level no less than the GCRMA normalized mean value plus 3-fold standard deviation among the 22 tissue/cell types (eight neural cells and 14 ordinary cells) and (B) at least 1.5-fold of the mean value was considered as specifically expressed as described in formula (1) and (2):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1)$$

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

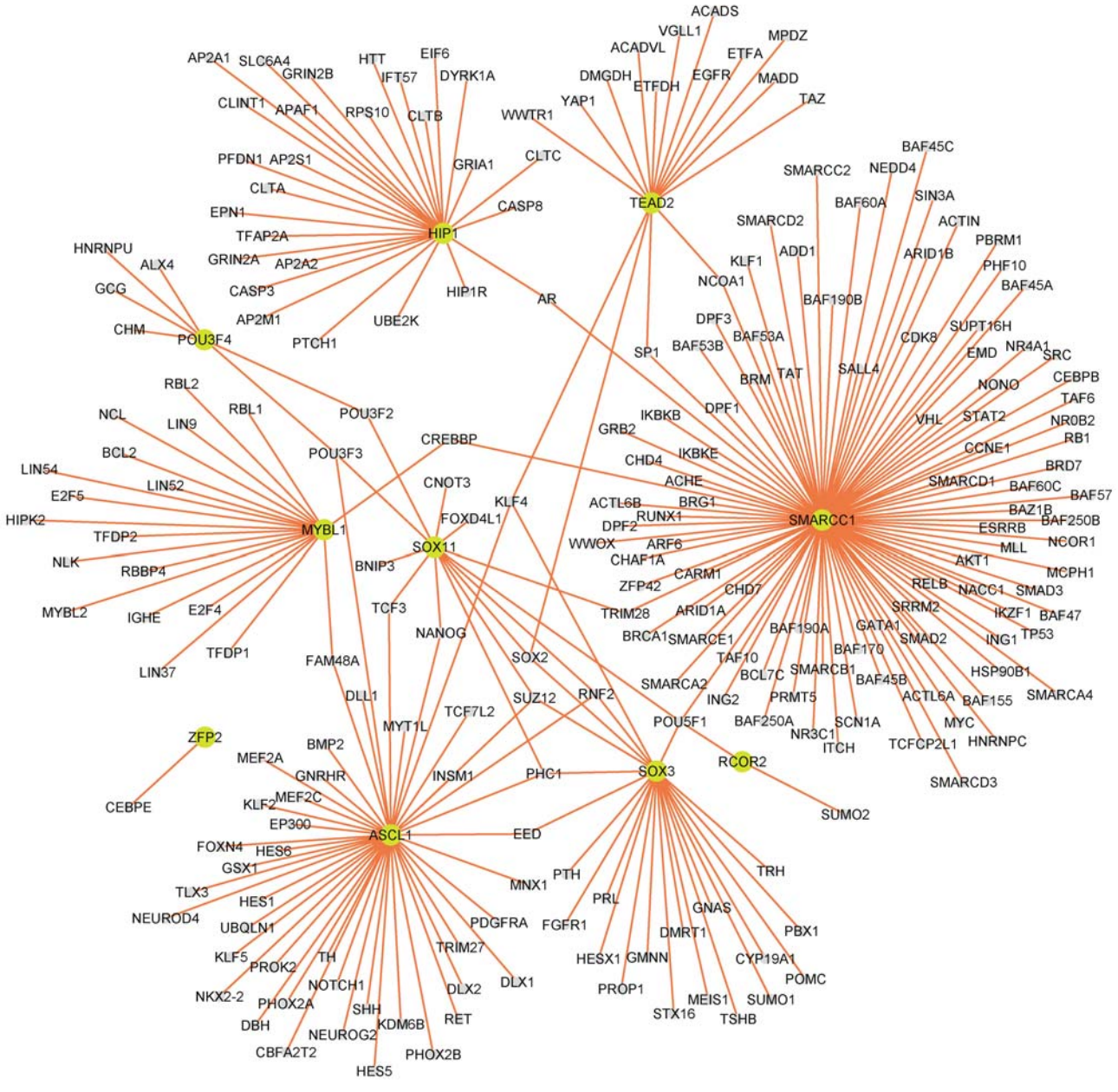


Figure 5. The protein–protein interaction network involved in the master regulators. The master regulators are shown in bigger light green nodes.

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (2)$$

Where n represents the number of tissues; n is 22 here. x_i is the expression leave in the *i-th* tissue of a gene (probe), $\bar{x}$ is the mean expression value among the n tissues of this gene, and s is the standard deviation. For a neural cell specifically expressed gene, it's expression level must be $\geqslant \bar{x}$ + 3s and $\geqslant 1.5\,\bar{x}$.

These genes were then sorted by the Maximum Relevance Minimum Redundancy (mRMR, http://penglab.janelia.org/proj/mRMR/) feature selection method (Ding and Peng, 2005; Peng et al., 2005). The mRMR method balances the relevant features and redundant features to generate a better classification performance. In other words, it requires the selected features (probes) are highly related to the target categories (tissue/cell type), and tries to reduce the redundancy between all selected features as much as possible. Here, the redundant features (probes) should not be ex-

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

cluded. So, only Maximum Relevance feature selection method was performed. Firstly, mRMR will generate a list of sorted genes (probes in actual) according to the maximum relevance minimum redundancy criterion. However, not all of these genes (probes) will be considered specifically expressed in a certain cell type. Therefore, Incremental Feature Selection (IFS) method was then performed to select the top N genes (probes) which could better characterize all the neural cells. To achieve this goal, we should construct a series of sub-feature collections from the total listed probes, as shown in formula (3):

$$S_i = \{p_1, p_2, ..., p_{10}, ..., p_i\} (10 \leqslant i \leqslant 1758) \quad (3)$$

where p_i is the *i-th* probe in the sorted probe listed, S_i is sub-feature collection containing i probes (the total probe number is 1758). Combing with the SVM predicator (Chang and Lin, 2011), leave-one-out cross-validation was employed to evaluate the performance (only for neural cells) of each sub-feature collection. If the predicated accuracy is maximum, the corresponding i genes (probes) would be considered as neural cells specifically expressed. The detailed method was also described in our former work (Wang et al., 2012).

DISCUSSION AND CONCLUSION

We conducted a systemic analysis by employing the highly purified gene expression data, combined with protein–protein interaction data. The dynamics of NSCs differentiation were reconstructed by computational modeling from the aspect of both specifically expressed transcriptomic data from gene expression arrays and proteomic data containing various interactions. The gene specific expression strategy is especially suitable for resolving this complex differentiation problem with multi-cell types. It allows us to better describe their own properties as well as discriminate their differences. This systemic approach is the basis of globally detecting key regulators.

The primary goal of our work is to find out which regulators dominate the differentiation processes. It is known that NSCs could differentiate into three major neural cells: astrocytes, neurons and OLs. Therefore we assumed that there may be some factors that drive the NSCs exit to the cell cycle of multipotent status and control the different differentiation directions. Indeed, we identified 29 genes that mediate the neural cell mitosis and chromosome remodeling, including 4632434I11RIK, APC, ASPM, CCNB1, CDC25C, CDC7, CENPF, CEP55, ERCC6L, EXO1, FANCD2, FANCI, HELLS, HMGA2, INCENP, KIF20B, LFNG, MCM8, NASP, NCAPD3, NUSAP1, PSRC1, RAD51C, SEPT2, SF1, SKP2, SMC2, SPDYA and SUV39H2, as well as 57 potential regulating genes (Table 2) which may play essential roles in the differentiation processes. We predicated their downstream target genes from gene co-expression (Suzuki et al., 2009) samples. The transcriptional regulation interactions were shown in Table S4. By manual curation, 10 transcription factors were identified as key regulators. These genes are all enriched in the early stage (NSCs) of the neural cell differentiation processes, which suggested that they may change the direction of cell cycle and initiate the NSC differentiation. In conclusion, these findings describe some key regulators that we suggest could determine the NSC fates.

In addition, the promoter binding situations of the 1559 genes were detected by employing JASPAR CORE database (Portales-Casamar et al., 2010) analysis. Specifically, –400 bp – +100 bp from the transcription start sites (TSS) of the 1559 genes were selected from NCBI Mouse Genome Assembly Build 37.1, and were subsequently sent to JASPAR CORE database for binding analysis with default parameters. However, no binding was found by the 57 regulators, which was primarily due to the limited motif information. And it was reported that the accuracy of this motif based method is generally no more than 15%. So more biological experiments such as ChIP-chip or ChIP-seq will be needed in the future research.

Besides, signal transduction and microRNA regulation are also under our consideration. And many relevant methods, tools and databases are publicly available, such as the microRNA database miRBase (Kozomara and Griffiths-Jones, 2011), the signal transduction network mining methods CASCADE_SCAN (Wang et al., 2011), NetSearch (Steffen et al., 2002) and integer linear programming (Zhao et al., 2008). However, in this study we only focused on transcriptional regulations and protein–protein interactions.

In summary, our studies provide a more comprehensive analysis of the dynamics during NSC differentiation in a system approach based on neural cells specifically expressed genes. It suggests that some regulators may play essential roles during the development stages. These findings provide more theoretic evidence for further studies.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31070954, 11172158 and 10832006), the Shanghai Commission of Science and Technology Basic Research Fund (No. 09JC1406600), and the Shanghai Education Committee and Shanghai Leading Academic Discipline Project (J50103; key discipline "Molecular Physiology").

REFERENCES

Andersson, T., Sodersten, E., Duckworth, J.K., Cascante, A., Fritz, N., Sacchetti, P., Cervenka, I., Bryja, V., and Hermanson, O. (2009). CXXC5 is a novel BMP4-regulated modulator of Wnt signaling in neural stem cells. J Biol Chem 284, 3672–3681.

Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., Feuermann, M., Ghanbarian, A.T., Kerrien, S., Khadake, J., et al. (2010). The IntAct molecular interaction database in 2010. Nucleic Acids Res 38, D525–531.

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., et al. (2011). NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic Acids Res 39, D1005–1010.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci 28, 264–278.

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2, 1–27.

Ding, C., and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol 3, 185–205.

Episkopou, V. (2005). SOX2 functions in adult neural stem cells. Trends Neurosci 28, 219–221.

Favaro, R., Valotta, M., Ferri, A.L., Latorre, E., Mariani, J., Giachino, C., Lancini, C., Tosetti, V., Ottolenghi, S., Taylor, V., et al. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. Nat Neurosci 12, 1248–1256.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1–13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44–57.

Juliandi, B., Abematsu, M., and Nakashima, K. (2010). Epigenetic regulation in neural stem cell differentiation. Dev Growth Differ 52, 493–504.

Kim, J.B., Sebastiano, V., Wu, G., Arauzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., et al. (2009). Oct4-induced pluripotency in adult neural stem cells. Cell 136, 411–419.

Kimura, I., Nakayama, Y., Konishi, M., Kobayashi, T., Mori, M., Ito, M., Hirasawa, A., Tsujimoto, G., Ohta, M., Itoh, N., et al. (2010). Neuferricin, a novel extracellular heme-binding protein, promotes neurogenesis. J Neurochem 112, 1156–1167.

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39, D152–157.

Lee, S.T., Chu, K., Park, J.E., Lee, K., Kang, L., Kim, S.U., and Kim, M. (2005). Intravenous administration of human neural stem cells induces functional recovery in Huntington's disease rat model. Neurosci Res 52, 243–249.

Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron 55, 201–215.

Maka, M., Stolt, C.C., and Wegner, M. (2005). Identification of Sox8 as a modifier gene in a mouse model of Hirschsprung disease reveals underlying molecular defect. Dev Biol 277, 155–169.

Makri, G., Lavdas, A.A., Katsimpardi, L., Charneau, P., Thomaidou, D., and Matsas, R. (2010). Transplantation of embryonic neural stem/precursor cells overexpressing BM88/Cend1 enhances the generation of neuronal cells in the injured mouse cortex. Stem Cells 28, 127–139.

Matsumoto, Y., and Osumi, N. (2008). [Role of Pax6 in the developing central nervous system]. Brain Nerve 60, 365–374.

Moghadam, F.H., Alaie, H., Karbalaie, K., Tanhaei, S., Nasr Esfahani, M.H., and Baharvand, H. (2009). Transplantation of primed or unprimed mouse embryonic stem cell-derived neural precursor cells improves cognitive function in Alzheimerian rats. Differentiation 78, 59–68.

Namihira, M., Kohyama, J., Abematsu, M., and Nakashima, K. (2008). Epigenetic mechanisms regulating fate specification of neural stem cells. Philos Trans R Soc Lond B Biol Sci 363, 2099–2109.

Oliveira, A.A., Jr., and Hodges, H.M. (2005). Alzheimer's disease and neural transplantation as prospective cell therapy. Curr Alzheimer Res 2, 79–95.

Osumi, N., Shinohara, H., Numayama-Tsuruta, K., and Maekawa, M. (2008). Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. Stem Cells 26, 1663–1672.

Park, D., Xiang, A.P., Mao, F.F., Zhang, L., Di, C.G., Liu, X.M., Shao, Y., Ma, B.F., Lee, J.H., Ha, K.S., et al. (2010). Nestin is required for the proper self-renewal of neural stem cells. Stem Cells 28, 2162–2171.

Park, J.H., Choi, M.R., Park, K.S., Kim, S.H., Jung, K.H., and Chai, Y.G. (2012). The characterization of gene expression during mouse neural stem cell differentiation in vitro. Neurosci Lett 506, 50–54.

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27, 1226–1238.

Peng, H., Long, F., Zhou, J., Leung, G., Eisen, M.B., and Myers, E.W. (2007). Automatic image analysis for gene expression patterns of fly embryos. BMC Cell Biol 8 Suppl 1, S7.

Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res 38, D105–110.

Renault, V.M., Rafalski, V.A., Morgan, A.A., Salih, D.A., Brett, J.O., Webb, A.E., Villeda, S.A., Thekkat, P.U., Guillerey, C., Denko, N.C., et al. (2009). FoxO3 regulates neural stem cell homeostasis. Cell Stem Cell 5, 527–539.

Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. Nucleic Acids Res 32, D449–451.

Sanberg, P.R. (2007). Neural stem cells for Parkinson's disease: to protect and repair. Proc Natl Acad Sci U S A 104, 11869–11870.

Stark, C., Breitkreutz, B.J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M.S., Nixon, J., Van Auken, K., Wang, X., Shi, X., et al. (2011). The BioGRID Interaction Database: 2011 update. Nucleic Acids Res 39, D698–704.

Steffen, M., Petti, A., Aach, J., D'Haeseleer, P., and Church, G. (2002). Automated modelling of signal transduction networks. BMC Bioinformatics 3, 34.

Storch, A., and Schwarz, J. (2002). Neural stem cells and Parkinson's disease. J Neurol 249 Suppl 3, III/30–32.

Suzuki, H., Forrest, A.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M.J., et al. (2009). The transcriptional network that controls

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2012

growth arrest and differentiation in a human myeloid leukemia cell line. Nat Genet 41, 553–562.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 39, D561–568.

Tang, J., Xu, H., Fan, X., Li, D., Rancourt, D., Zhou, G., Li, Z., and Yang, L. (2008). Embryonic stem cell-derived neural precursor cells improve memory dysfunction in Abeta(1–40) injured rats. Neurosci Res 62, 86–96.

Temple, S. (2001). The development of neural stem cells. Nature 414, 112–117.

Thorrez, L., Van Deun, K., Tranchevent, L.C., Van Lommel, L., Engelen, K., Marchal, K., Moreau, Y., Van Mechelen, I., and Schuit, F. (2008). Using ribosomal protein genes as reference: a tale of caution. PLoS One 3, e1854.

Ueki, T., Tanaka, M., Yamashita, K., Mikawa, S., Qiu, Z., Maragakis, N.J., Hevner, R.F., Miura, N., Sugimura, H., and Sato, K. (2003). A novel secretory factor, Neurogenesin-1, provides neurogenic environmental cues for neural stem cells in the adult hippocampus. J Neurosci 23, 11732–11740.

Wang, K., Hu, F., Xu, K., Cheng, H., Jiang, M., Feng, R., Li, J., and Wen, T. (2011). CASCADE_SCAN: mining signal transduction network from high-throughput data based on steepest descent method. BMC Bioinformatics 12, 164.

Wang, K., Hu, L., Shi, X., Dong, Y., Li, H., and Wen, T. (2012). PSCL: Predicting protein subcellular localization based on optimal functional domains. Protein Pept Lett 19, 15–22.

Wu, Z., Irizarry, R., Gentleman, R., Martinez-Murillo, F., and Spencer, F. (2004). A model-based background adjustment for oligonucleotide expression arrays. J Am Stat Assoc. 99, 909–917.

Zhao, X.M., Wang, R.S., Chen, L., and Aihara, K. (2008). Uncovering signal transduction networks from high-throughput data by integer linear programming. Nucleic Acids Res 36, e48.

Zhongling, F., Gang, Z., and Lei, Y. (2009). Neural stem cells and Alzheimer's disease: challenges and hope. Am J Alzheimers Dis Other Demen 24, 52–57.

 © Higher Education Press and Springer-Verlag Berlin Heidelberg 2012