

Selection of Neural Differentiation-Specific Genes by Comparing Profiles of Random Differentiation

Min Su Lee, Dae-Hyun Jun, Chang-Il Hwang, Seung Soo Park, Jason Jongho Kang, Hyun-Seok Park, Jihoon Kim, Ju Han Kim, Jeong-Sun Seo and Woong-Yang Park *Stem Cells* 2006;24;1946-1955; originally published online Apr 20, 2006; DOI: 10.1634/stemcells.2005-0325

This information is current as of July 26, 2006

The online version of this article, along with updated information and services, is located on the World Wide Web at: http://www.StemCells.com/cgi/content/full/24/8/1946

STEM CELLS®, an international peer-reviewed journal, covers all aspects of stem cell research: embryonic stem cells; tissue-specific stem cells; cancer stem cells; the stem cell niche; stem cell genetics and genomics; translational and clinical research; technology development.

STEM CELLS® is a monthly publication, it has been published continuously since 1983. The Journal is owned, published, and trademarked by AlphaMed Press, 318 Blackwell Street, Suite 260, Durham, North Carolina, 27701. © 2006 by AlphaMed Press, all rights reserved. Print ISSN: 1066-5099. Online ISSN: 1549-4918.

C AlphaMed Press

Selection of Neural Differentiation-Specific Genes by Comparing Profiles of Random Differentiation

MIN SU LEE,^a DAE-HYUN JUN,^b CHANG-IL HWANG,^b SEUNG SOO PARK,^a JASON JONGHO KANG,^c HYUN-SEOK PARK,^{a,c} JIHOON KIM,^d JU HAN KIM,^d JEONG-SUN SEO,^b WOONG-YANG PARK^b

^aDepartment of Computer Science and Engineering, Ewha Womans University, Seoul, Korea; ^bIlchun Molecular Medicine Institute and Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Korea; ^cMacrogen Inc., Seoul, Korea; ^dSeoul National University Biomedical Informatics, Seoul National University College of Medicine, Seoul, Korea

Key Words. Embryonic stem cells • Microarray • Neural differentiation • Random differentiation • Spearman's correlation

ABSTRACT

Differentiation of embryonic stem cells (ESCs) into neurons requires a high level of transcriptional regulation. To further understand the transcriptional regulation of neural differentiation of ESCs, we used oligonucleotide microarray to examine the gene expressions of the guided differentiation (GD) model for dopaminergic (DA) neurons from mouse ESCs. We also determined the gene expression profiles of the random differentiation (RD) model of mouse ESCs into embryoid bodies. From K-means clustering analysis using the expression patterns of the two models, most of the genes (1,282 of 1,884 genes [68.0%]) overlapped in their expression patterns. Six hundred twenty-two differentially expressed genes (DEGs) from the GD model by random variance F-test were classified by their critical molecular functions in neurogenesis and DNA replication (Gene Ontology analysis). However, 400 genes among GD-DEGs (64.3%) showed a high correlation with RD in Spearman's correlation analysis (Spearman's coefficient $p_s \ge .6$). The genes showing marginal correlation $(-.4 < p_s < .6)$ were present in the early stages of differentiation of both GD and RD, which were non-specific to brain development. Finally, we distinguished 66 GD-specific genes based on p_s \leq -.4, the molecular functions of which were related mainly to vesicle formation, neurogenesis, and transcription factors. From among these GD-specific genes, we confirmed the expression of Serpinil and Rab33a in P19 differentiation models and adult brains. From these results, we identified the specific genes required for neural differentiation by comparing gene expressions of GD with RD; these would potentially be the highly specific candidate genes necessary for differentiation of DA neurons. STEM CELLS 2006;24:1946-1955

INTRODUCTION

Development of the mammalian central nervous system (CNS) is a complex process involving an orchestrated regulation of structural and regulatory genes through differentiation stages of multipotent stem cells into neurons [1]. Mouse embryonic stem cells (ESCs) retain the characteristics of multipotent stem cells, exhibiting the infinite self-renewal activity and the potential of differentiation into various kinds of lineages [2]. Numerous efforts have been made to induce ESCs into neurons by regulating transcriptions of critical genes in the hope of using these cells for therapy for neurological disorders such as Parkinson's disease.

In an early effort to make neurons from P19 embryonic teratocarcinoma cells, retinoic acids were used to induce neural precursor cells [3]. Another efficient differentiation method is

coculturing of the ESCs with the PA-6 feeder cells, producing stromal cell-derived inducing activity [4]. Lee et al. established the five-stage differentiation method to induce the ESCs into dopaminergic (DA) neurons [5]. Recently, transcription factors like Nurr1 [6] and Pitx3 [7] were introduced into stem cells to enhance the efficiency of DA neuron production. Moreover, the methods to make human DA neurons from human ESCs (hESCs) were established by coculturing with PA-6 stromal cells [8, 9].

During the step-wise differentiation into DA neurons, we can follow the levels of differentiation: pluripotent stem cell (ESC, stage I), committed multipotent precursor cells (embryoid body [EB], stage II), neural precursors (stage III), expanded neural precursors (stage IV), and fully differentiated neurons (stage V) [5]. To further dissect this DA neuron differentiation,

Correspondence: Woong-Yang Park, M.D., Ph.D., Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, 28 Yongondong, Chongnogu, Seoul 110-799, Korea. Telephone: +82-2-740-8241; Fax: +82-2-744-4534; e-mail: wypark@snu.ac.kr Received July 17, 2005; accepted for publication April 10, 2006; first published online in STEM CELLS *Express* April 20, 2006. ©AlphaMed Press 1066-5099/2006/\$20.00/0 doi: 10.1634/stemcells.2005-0325

STEM CELLS 2006;2404946ded 955 www.stastemCollsacontekefeller Univ on July 26, 2006

genome-wide gene expression profiling can be used, providing a more comprehensive molecular understanding of neural differentiation using mouse ESCs [10] or hESCs [8]. For example, in-depth examinations of gene expression profiles of ESCs and their differentiated progenies are likely to reveal information about the "stemness" as well as the pathways involved in neural differentiation. These relevant genes, once identified, are good candidates to investigate for their role in neural differentiation. However, the guided differentiation (GD) of ESCs is difficult to control, resulting in a heterogeneous population of fully differentiated neurons. To correctly identify the specific genes related to neural differentiation, we need to minimize the contributions from non-neuronal and undifferentiated cells in the samples.

If the ESCs are induced to enter into the differentiation pathway in vitro, they can form EBs with all types of mesodermal, hematopoietic, endothelial, muscle, and neural lineages [11]. Once we screen the markers for the germ layers or a variety of cell lineages, developmental stages of cells can be estimated in terms of early postimplantation development of mouse embryos [12]. Thus, EB formation is the best in vitro model system for studying early lineage determination and organogenesis in mammals; this system will prove to be a useful tool for identifying developmental genes whose expression is restricted to the particular lineages.

We profiled the transcriptions of two differentiation models: GD into DA neurons and random differentiation (RD) into EBs. We could find the marked correlations between the two models but not the specific expressions of genetic markers of each model. We have compared the sequential expression patterns of GD and RD models by Spearman's correlation analysis in order to find marked overlap between them. Finally, we subtracted those overlapped, common profiles from GD profiles to select GD-specific genes. From this result, we could propose the sizable overlaps between two differentiation models and select the most likely candidate genes for GD for further studies of their roles in neural differentiations.

MATERIALS AND METHODS

Mouse ESC Culture and Differentiation

We induced differentiation of mouse ESCs (R1) as described previously [5]. Briefly, undifferentiated ESCs (stage I) were grown on gelatin-coated tissue-culture plates in knockout (KO)-Dulbecco's modified Eagle's medium (DMEM) media. To induce EB formation (stage II), the cells were dissociated into a single-cell suspension and plated onto nonadherent bacterial culture dishes at a density of 2.5×10^4 cells per cm² in the KO medium. After 4 days, the cells were transferred to the original tissue-culture dish in a serum-free ITSF (insulin/transferrin/ selenium/fibronectin) medium to select the nestin-positive cells (stage III). After 6 days of selection, the cells were expanded (stage IV) by transferring to the plate coated with polyornithine and laminin in N2 medium supplemented with laminin/basic fibroblast growth factor (bFGF)/sonic hedgehog/fibroblast growth factor 8. After 6 days, bFGF was removed to induce the differentiation (stage V) in N2 medium supplemented with laminin and ascorbic acid for 6 days. For RD, EBs were dissociated and plated onto a tissue-culture dish in DMEM with fetal bovine serum and antibiotics for indicated periods.

Oligonucleotide Microarray

Total RNAs from undifferentiated mouse ESCs were used as a reference group in all experiments. Three independent biological replicates were taken at four stages of DA differentiation. For the RD model, three biological replicates were made at days 4, 8, 15, and 21 to extract total RNA. Total RNA was prepared by using TriZol reagent (Invitrogen, Carlsbad, CA, http://www.invitrogen.com). The array used in this experiment was the Macrogen Mouse Oligo 11K Chip (Macrogen Inc., Seoul, Korea, http://www.macrogen.com) as described previously [13, 14]. Cy3 and Cy5 fluorescent intensities were determined using the GenePix scanner (Axon Instruments, Union City, CA, http:// www.axon.com), and images were analyzed using the GenePix Pro to calculate relative ratios and to determine confidence intervals.

Data Analysis

Fluorescence intensities were processed and measured using GenePix Pro software. Intensity data were imported to an inhouse microarray database. The variance stabilizing normalization by Huber et al. was applied with the "vsn" package in Bioconductor using the R statistical package [15]. After performing intensity-dependent global LOWESS (locally-weighted scatterplot smoothing) regression, spatial and intensity-dependent effects were managed by pin-group LOWESS normalization following by the approach of Yang et al. [16].

The gene expression dataset was registered in the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) under the accession numbers GSE3527 (for RD) and GSE3528 (for GD).

Reverse Transcription-Polymerase Chain Reaction (**RT-PCR**) and **Quantitative Real-Time RT-PCR**

The first-strand cDNA was synthesized from 1 μ g of total RNA using reverse transcriptase and 1 μ M olgo-dT primer. Each cDNA sample was amplified by using specific primers (supplemental online data A). Specific bands corresponding to the estimated sizes were analyzed after agarose gel electrophoresis. To quantify the amount of transcripts, real-time RT-PCR based on MyQ system (Bio-Rad Laboratories, Hercules, CA, http://www.bio-rad.com) was performed as described in the manufacturer's recommendations. The relative amount of each transcript was normalized with the level of glyceralde-hyde-3-phosphate dehydrogenase.

P19 Mouse Teratocarcinoma Cells and Mouse Tissues

Undifferentiated P19 mouse teratocarcinoma cells were maintained in a growth medium like α -minimal essential medium supplemented with 10% fetal bovine serum and antibiotics as previously described [17]. We induced the production of neurospheres by adding 1 μ M all-trans-retinoic acid to the media for 4 days. All the suspended EBs were dissociated and plated onto poly-L-lysine-coated dishes with neurobasal media supplemented with B-27 and Ara-C to select neuron-like cells (NLCs). Mouse brain tissues were isolated and pooled from the embryos at days 9, 11, 13, 15, and 17 or from neonates at days 0 and 7 by dissecting out under the stereomicroscope. The whole brain and other tissues from adult mice were also isolated to extract total RNA.

RESULTS AND DISCUSSION

DA Neuron and EB Formation

In an effort to find differentially expressed genes (DEGs) during neural differentiation, we established the five-stage neural differentiation model as described previously [5]. We used this model as a GD of mouse ESCs into DA neurons. First, we checked the expression of sets of markers for the differentiation of DA neurons by RT-PCR in the five-stage GD model (Fig. 1A). Neural markers like Tubb3 (tubulin, beta 3), Gfap (glial fibrillary acidic protein), and Neurod1 (neurogenic differentiation 1) [18] were gradually increased from stage III-V, and Nestin [19] and Pax5 (paired box gene 5) [20] were turned on at stage III and gradually decreased to stage V. In addition, we could detect the expression of DA neuron-specific genes emerging through the GD as in previous reports [6]. The genes related to midbrain development, such as Nr4a2 (nuclear receptor subfamily 4, group A, member 2, Nurr1) [21] and En1 (engrailed 1) [20], started to express at stage II and sustained their expression until stage V. On the other hand, the transcriptions of the genes for DA neuron phenotype, such as Ddc (doapmine decarboxylase, AADC), Slc6a3 (solute carrier family 6, member 3, DAT), and Th (tyrosine hydroxylase) [8], were turned on at stage V. According to the expressions of genetics markers, we could confirm the establishment of GD models as in previous reports [5].

For RD, mouse ESCs were maintained as EBs for 3 weeks by being attached to culture dishes for the indicated period. ESC markers such as Pou5f1 (POU domain, class 5, transcription factor 1, Oct4) and Dppa5 (developmental pluripotency associated 5, Esg1) [22] were gradually decreased in the course of EB formation (Fig. 1B). The differentiation status of EBs could be checked by the expression level of the lineage markers of three layers [12]: T (brachyury) for mesoderm, Nodal for ectoderm, and Gata4 (GATA binding protein 4) for endoderm (Fig. 1B). These results showed that EBs contained all types of three germ layers and their derivatives in quantities large enough to cover any type of cells or tissues.

Overlapping of the Gene Expression Patterns in GD and RD Models

Using GD and RD models of ESCs, we tried to profile the genome-wide gene expression using long oligonucleotide microarray containing 11,376 genes in 13,680 spots. Total RNAs from each stage of differentiating cells (stage II–V for GD and EB day 4–21 for RD) were hybridized with those of mouse ESCs as a reference in biological triplicates for each dataset. From the 24 microarray experiments in total, we could have the gene expression profiles of four stages in two datasets for each differentiation model. The whole dataset of microarray experiments can be browsed in GEO as accession numbers GSE3527 and GSE3528 (http://www.ncbi.nlm.nih.gov/geo).

To capture expression profiles of GD and RD as a whole, K-means clustering analysis was applied to GD and RD normalized log activation fold ratios using the Euclidean distance metric. Genes showing minimal variation across the set of arrays were excluded from the clustering analysis. We took the mean value from the gene expression ratio of each of the three independent experiments and selected genes whose expression levels differed by at least a 1.5-fold change at one or more stages. From all datasets of GD and RD models, 1,884 genes



Figure 1. Gene expression analysis of the differentiation markers in guided differentiation (GD) and random differentiation (RD) models by reverse transcription-polymerase chain reaction (RT-PCR). (A): The levels of transcripts of differentiation markers at indicated stages were analyzed by RT-PCR using total RNA from the GD model. (B): The levels of transcripts of differentiation markers at indicated stages were analyzed by RT-PCR using total RNA from the RD model. (B): The levels of transcripts of differentiation markers at indicated stages were analyzed by RT-PCR using total RNA from the RD model. (B): The levels of transcripts of differentiation markers at indicated stages were analyzed by RT-PCR using total RNA from the RD model. (B): The levels of transcripts of the dearboxylase; Dppa5 (Esg1), developmental pluripotency associated 5; En1, engrailed 1; Gapdh, glyceralde-hyde-3-phosphate dehydrogenase; Gata4, GATA binding protein 4; Gfap, glial fibrillary acidic protein; Nes, Nestin; Neurod1, neurogenic differentiation 1; Nr4a2 (Nurr1), nuclear receptor subfamily 4, group A, member 2; Pax5, paired box gene 5; Pou5f1 (Oct4), POU domain, class 5, transcription factor 1; Slc6a3 (DAT), solute carrier family 6, member 3; T, brachyury; Th, tyrosine hydroxylase; Tubb3, tubulin, beta 3.



Figure 2. K-means cluster analysis of 1,884 genes in GD and RD models. We chose 1,884 genes showing a 1.5-fold change in any one out of eight stages for K-means cluster analysis. Seven clusters were selected to divide global patterns of whole datasets more clearly, which could be visualized and summarized with scales at the bottom. Each cluster could be summarized by plotting average values for each stage. The gene list and log activation fold ratio can be browsed in supplemental online data B. Abbreviations: GD, guided differentiation; RD, random differentiation.

showed more than a 1.5-fold change in any one of eight stages. In the cluster analysis, global patterns of whole datasets could be visualized and summarized (Fig. 2). We could divide all 1,884 genes into seven clusters based on K-means clustering (supplemental online data B) after trying different numbers of clusters to find minimal numbers of RD- and GD-specific clusters. Four out of seven clusters including 1,282 genes (68.0%) showed exactly the same patterns of upregulation (C3 and C5) and downregulation (C6 and C7). The overall pattern of gene expression of two differentiation models reflected a sizable overlapping gene expression between GD and RD. The other two clusters (308 genes) showed the elevation in their gene expression in RD but not in GD. Only one cluster, C4, containing 294 genes (15.6%), showed the upregulated pattern in a GD-specific way.

For the common features, C6 and C7 clusters contained the downregulated genes in both types of differentiation, which were mostly related to development and stem cell markers like Pou5f1 (Oct4) and Nanog (Nanog homeobox). As we expected, any kind of differentiation might lead to a loss of stem cell characteristics such as proliferation and cell cycle. On the other hand, the genes in the C3 and C5 clusters were upregulated in both differentiation models. More interestingly, the genes in C3,

like *Fst* (follistatin) and *Notch3* (Notch gene homolog 3), were related to development, whereas the C5 cluster contained genes, such as *Elavl4* (ELAV [embryonic lethal, abnormal vision, Drosophila]-like 4, HuD) and *Pfn2* (profilin 2), related to the neuronal functions. The genes in the C3 and C5 clusters reflect that a certain part of the mechanism regulating neural differentiation could be overlapped with any other types of differentiation.

DNA Methyltransferases and Imprinted Genes in the Overlapping Clusters

As an example of common features shared between the two models, we found that Dnmt3-â (DNA (cytosine-5-)-methyltransferase 3-B) and Dnmt3l (DNA (cytosine-5-)-methyltransferase 3-like) were downregulated in both the GD and RD models (C6 and C7 clusters in Fig. 2, respectively). Global regulation of transcription through chromatin regulation like DNA methylation seems to be an essential mechanism in the early stages of mammalian development [23]. We confirmed that the expression levels of Dnmt3-â from microarray data (supplemental online data C) and RT-PCR (Fig. 3) clearly showed the downregulated patterns by the induction of any type of differentiation. Because the downregulated methyltransferases could lead to the transcriptional activation of genes imprinted in ESCs, we checked the expressions of the imprinting genes such as Nnat (neuronatin), Igfl (insulin-like growth factor 1), Mest (mesoderm specific transcript), and Ndn (necdin) [23] by RT-PCR in both GD and RD (Fig. 3B and 3D). In a series of analyses on the gene expression of the two differentiation models, DNA methylation and chromatin regulations occurred in the early stage of differentiation. As shown in supplemental online data C, for the genes related to chromatin regulation, the global regulation through chromatin seemed to be a general mechanism for the developmental regulation of gene expression, especially in the early stages of differentiation.

Identification of DEGs

For each RD and GD model, the multivariate permutation test was applied to evaluate the statistical significance of changes in gene expression. Qualite-quantile plot analysis showed that the residual quantiles deviated from the theoretic quantiles (supplemental online data D). Hence, we regarded the distribution of the data as abnormal [24]. Typically, the F-test assumes that the input data hold normal distribution. Therefore, we used the multivariate permutation test to collect the genes at a 90% confidence with a false discovery rate of less than 10% [25]. The test statistics used were random variance versions of F-tests for each gene [26]. Although the F-test was used, the multivariate permutation test is nonparametric and does not require the assumption of normal distributions. By random variance F-test analysis, we selected 622 DEGs (supplemental online data E). Functionally, the genes from GD DEG were classified as being related to DA neurons, neurogenesis, and transcription factors (supplemental online data F). However, 188 genes were differentially expressed in the RD model (supplemental online data G). Because triplicate genes in RD show a larger variance of expression, the number of DEGs in RD is smaller than in GD. Moreover, 69 genes among the 188 RD-related genes were also found in GD DEGs (supplemental online data G).



Figure 3. Gene expression patterns of chromatin regulation-related genes in microarray and reverse transcription-polymerase chain reaction (RT-PCR) (A): The level of transcripts of Dnmt3b, Mest, Igf2, and Nacdin at indicated stages of guided differentiation (GD) was plotted using microarray datasets. (B): The expression of these sets of genes was also confirmed by RT-PCR using total RNA from each sample of GD. (C): Using random differentiation (RD) datasets, the transcripts of the same genes were plotted in each stage. (D): The level of each transcript was validated by RT-PCR in RD samples. Abbreviations: Dnmt3b, DNA (cytosine-5-)-methyltransferase 3 beta; Dnmt3l, DNA (cytosine-5-)-met

To characterize the DEGs of each model, we analyzed gene ontology (GO) categories of the DEGs in GD and RD [27]. This procedure finds gene ontology categories that have higher-thanexpected numbers of genes differentially expressed among the different classes of the samples based on random variance F-tests [26]. By analyzing the GO groups, rather than the individual genes, we were able to reduce the number of tests conducted and enable findings among biologically related genes to reinforce each other (Table 1). In GD, the genes related to neurogenesis, organ development, and DNA replication show a significant variance. On the other hand, the genes related to cell adhesion, cell communication, and structural molecular activity show a dynamic change in RD. The 69 common genes in GD and RD are related to "development (GO: 0007275)" and "organelle organization and biogenesis (GO: 0006996)" categories in GO analysis.

Common Genes for Two Differentiation Models

Spearman's correlation is a nonparametric test for measuring the strength of the association between the two variables. We compared the expression profiles of the 622 DEGs in GD with an RD profile based on the Spearman's correlation coefficient (p_s) . For example, 400 genes from the previously mentioned 622 DEGs (64.3%) showed similar expression patterns in the two models $(p_s \ge .6$, supplemental online data E). As we observed in K-means clustering (Fig. 2), half of the GD-DEGs were also disregulated in the RD model. These common genes could be categorized specifically to cell cycle, DNA replication, and morphogenesis in GO analysis (Table 2).

For example, Sox4 is a member of the SOX (SRY-related HMG-box) family of transcription factors involved in the regulation of embryonic development and in the determination of the cell fate [28]. Sox4 was suggested to be upregulated at stage V of DA neuron differentiation, in which ascorbic acids might induce the expression of Sox4 and other DA neuron-specific genes [10]. We observed that Sox4 was also present in the 622 GD-DEGs in Table 1. However, the expression pattern of Sox4 in GD was similar to that of RD $(p_s = .6, \text{ supplemental online data E})$. To validate the upregulation of Sox4, we quantified Sox4 expressions in another neural differentiation model using P19 teratocarcinoma cells [3]. We used Nestin and Rest (RE1-silencing transcription factor) as controls to monitor the differentiation status of P19 cells. Nestin is an intermediate filament protein that is expressed predominantly in stem cells of the CNS in the neural tube [29]. Upon terminal neural differentiation, nestin is downregulated and replaced by neurofilaments. Rest, which binds a DNA sequence called the neuron-restrictive silencer element (NRSE), was present in the promoter regions of neuron-specific genes. The expression of Rest mRNA was detected in most of the non-neuronal progenitor cells but was absent in differentiated neurons [30]. Whereas Rest and Nestin were downregulated in EB and NLCs, Sox4 could be induced in EB and NLCs (Fig. 4A). We also tried to confirm the expression of genes related to GD-DEGs, using in vivo embryonic mouse brains, by real-time RT-PCR. Like in vitro differentiation models, Sox4 started to express in the brain of embryonic day 11 (E11) mouse embryos, whereas Rest and Nestin were downregulated in embryonic brain after E15 (Fig. 4B). However, in other mouse tissues like spleen

	GO		No. of	
Group	category	GO description	genes	p value
GD-DEG (622 genes)	7,399	Neurogenesis	15	.0272
	6,260	DNA replication	22	.0418
	6,575	Amino acid derivative metabolism	5	.0431
	8,233	Peptidase activity	13	.0458
	48,513	Organ development	34	.0603
	31,012	Extracellular matrix	16	.0710
	43,068	Positive regulation of programmed cell death	5	.0719
	4,601	Peroxidase activity	5	.0736
	902	Cellular morphogenesis	11	.0768
	74	Regulation of cell cycle	19	.0861
RD-DEG (188 genes)	7,155	Cell adhesion	10	.0001
	30,020	Extracellular matrix structural constituent	6	.0005
	7,154	Cell communication	17	.0009
	6,811	Ion transport	5	.0010
	8,026	ATP-dependent helicase activity	5	.0713
	51,234	Establishment of localization	17	.0855
Common to GD- and RD-DEGs (69 genes)	74	regulation of cell cycle	7	.0059
	6,261	DNA-dependent DNA replication	6	.0066
	902	Cellular morphogenesis	6	.0069
	50,874	Organismal physiological process	7	.0200
	5,524	ATP binding	30	.0250
	4,672	Protein kinase activity	6	.0393
	7,275	Development	22	.0430
	40,007	Growth	6	.0590

Table 1. Gene ontology analysis of GD-, RD-, and common DEGs

Abbreviations: DEG, differentially expressed gene; GD, guided differentiation; GO, gene ontology; RD, random differentiation.

Table 2.	GO	analysis	of	GD-sp	pecific,	marginal,	and	common	genes	among	GD-	DEGs
		2										

	GO		No. of	
Group	category	GO description	genes	p value
GD-specific (66 genes)	30,528	Transcription regulator activity	5	.0234
	6,810	Transport	16	.1128
	6,139	Nucleobase, nucleoside, nucleotide, and nucleic acid metabolism	9	.1324
	16,021	Integral to membrane	15	.3223
	3,824	Catalytic activity	18	.3246
	4,871	Signal transducer activity	7	.4469
	7,154	Cell communication	8	.4597
	43,234	Protein complex	6	.4969
Marginal (156 genes)	6,464	Protein modification	13	.0592
	16,491	Oxidoreductase activity	9	.0750
	15,630	Microtubule cytoskeleton	5	.0790
	16,787	Hydrolase activity	10	.1486
	4,713	Protein-tyrosine kinase activity	7	.1907
	6,996	Organelle organization and biogenesis	13	.3057
	48,513	Organ development	7	.3188
	6,397	mRNA processing	6	.3483
Common (400 genes)	6,261	DNA-dependent DNA replication	9	.0019
	74	Regulation of cell cycle	12	.0059
	902	Cellular morphogenesis	8	.0105
	31,012	Extracellular matrix	7	.0166
	7,399	Neurogenesis	9	.0396
	5,198	Structural molecule activity	49	.0556
	4,672	Protein kinase activity	13	.0607
	7,275	Development	42	.0624
	7,046	Ribosome biogenesis	21	.0999

Abbreviations: DEG, differentially expressed gene; GD, guided differentiation; GO, gene ontology.



Figure 4. Quantitative analysis of gene expressions of *Sox4*, *Nes*, and *Rest*. The levels of transcripts of *Sox4*, *Nes*, and *Rest* were analyzed by real-time reverse transcription-polymerase chain reaction using P19 cells (A), mouse brains (B), and adult tissues (C). Each value and error bar was calculated from biological triplicates according to the manufacturer's instructions. Abbreviations: BS, brain stem; CBL, cerebellum; CTX, cerebral cortex; E, embryonic; EB, embryoid body; HT, heart; KID, kidney; LIV, liver; MID, midbrain; Nes, Nestin; NLC, neuron-like cells; P, postnatal; Rest, RE1-silencing transcription factor; SC, spinal cord; Sox4, SRY-box containing gene 4; SPL, spleen; UN, undifferentiated.

and kidney, we were able to detect the expression of Sox4 more than any part of the brains (Fig. 4C). These results revealed that some genes reported to be GD-related could be non-specific to brain and neural differentiation. Although there were variations within neural tissues, Sox4 seems to have an overlapping function in both the neural and non-neural cell types.

GD-Specific Genes

Using Spearman's correlation, we could find the expressions of a large number of genes such as Sox4 in GD-DEGs that showed an overlapping pattern with that of RD. At the other extreme of Spearman's correlation, we could find only 66 highly specific genes upregulated in the GD model ($p_s \leq -.4$). To validate those genes, we analyzed and quantified the expression levels of Rab33a and Serpinil in the P19 neural differentiation model (Fig. 5A). As detected in the microarray experiment, Rab33a (member of RAS family oncogene, $p_s = -.88$) [31] and Serpinil (serine [or cysteine] peptidase inhibitor, clade I, member 1, $p_s = -.4$) [32] were upregulated in NLCs, but not in the EB stage of P19 cells. We also tested the expressions of these genes in embryonic or neonatal brains as well as in a variety of adult tissues. Two genes that are highly specific to adult brain started to express from E15 embryonic brains. Serpinil, a neuroserpin, has been reported to be expressed in the late stages of neurogenesis, during the process of synapse formation corresponding to stage V of the GD model [32]. We also tested four

GD-specific genes, Gng3 ($p_s = -.4$), Uchl1 ($p_s = -1.0$), Vamp8 ($p_s = -.8$), and Sms ($p_s = -.4$), by real-time RT-PCR in the P19 differentiation model (supplemental online data H).

Another group of genes $(-.4 < p_s < .6)$ has been upregulated in the early stages of GD, but also in RD marginally. Like cluster 3 in Figure 2, those genes were also upregulated in RD. However, they were downregulated again at the last stage of GD, and the levels of gene expression were maintained in RD until EB day 21. We checked the expressions of Cdk4 (cyclindependent kinase 4; $p_s = .2$ [33] and $P4h\alpha 2$ (procollagenproline, 2-oxoglutarate 4-dioxygenase [proline 4-hydroxylase], alpha II polypeptide; $p_s = -.2$) [34], as examples of this group, in the P19 differentiation model, embryonic brain, and adult mouse tissues (Fig. 5). The expression of Cdk4 started in EB or E13 embryos, but it was not specific to brain tissue. $P4h\alpha$ also expressed as early as E9 embryos and was also present in heart and liver. These data suggest that genes expressed in the early differentiation stages such as stage III and IV of the GD model were not specific to brain. In another words, genes responsible for the early neural differentiation can be also important to any other type of cells or tissues, even from other germ layers.

After subtracting the group of genes showing a high resemblance to RD, we could finally select the 66 GD-specific genes meeting the criterion of $p_s < -.4$. These genes were related to the vesicle-associated functions and transcriptional regulation in GO analysis (Table 2; GD-specific DEG). Because the majority of these genes were upregulated at stage V of the GD model, the



Figure 5. Quantitative analysis of gene expressions of *Cdk4*, *Ph4a2*, *Serpini1*, and *Rab33a*. The levels of transcripts of *Cdk4*, *Ph4a2*, *Serpini1*, and *Rab33a* were analyzed by real-time reverse transcription-polymerase chain reaction using P19 cells (**A**), embryonic or postnatal mouse brains (**B**), and adult mouse tissues (**C**). Each value and error bar was calculated from biological triplicates according to the manufacturer's instructions. Abbreviations: Cdk4, cyclin-dependent kinase 4; P4ha2, procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase) alpha II polypeptide; Rab33a, member of RAS oncogene family; Serpini1, serine (or cysteine) peptidase inhibitor, clade I, member 1.

early differentiation pathways seem to be shared with the other cell types. Even the downregulated genes were very few, implying that many of the stem cell-related genes were nonspecifically lost during any kind of differentiation. We still need to validate the specificity and functional implications of these genes in neural differentiations. We found that 15 genes might have functions in neurons (*Serpini1*, *Stx4a*, *Stxbp2*, *Scg2*, *Syt1*, *Vamp8*, *Ap3b2*, *Rab33a*, *Lrrn3*, *Dab2*, *Sorcs3*, *Gdap1*, *Resp18*, *Crmp1*, and *Ckb*), and 12 genes were reported in relation to the DA neurons (*Gng5*, *Scg2*, *Apod*, *Slc2a1*, *Kcnip1*, *Gng3*, *Tbrg1*, *Gpx3*, *Ets-1*, *Uchl1*, *Aplp1*, and *Ptgds*) [35–44]. We suggest that those highly selectable genes for neural differentiation by direct comparing might include genes with possible function in neural differentiations.

CONCLUSION

In this study, we set out to find GD-specific genes by comparing GD profiles with RD profiles. First, we used statistical analysis to find DEGs from both groups, which were categorized according to their characteristics of GD or RD. Whereas the genetic markers for GD and RD were exclusively expressed in the two models, the overall expression patterns looked similar to each

REFERENCES

- Sasai Y. Identifying the missing links: Genes that connect neural induction and primary neurogenesis in vertebrate embryos. Neuron 1998;21: 455–458.
- 2 Czyz J, Wobus A. Embryonic stem cell differentiation: The role of extracellular factors. Differentiation 2001;68:167–174.
- 3 Jones-Villeneuve EM, McBurney MW, Rogers KA et al. Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. J Cell Biol 1982;94:253–262.
- 4 Kawasaki H, Mizuseki K, Nishikawa S et al. Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. Neuron 2000;28:31–40.
- 5 Lee SH, Lumelsky N, Studer L et al. Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. Nat Biotechnol 2000;18:675–679.
- 6 Wagner J, Akerud P, Castro DS et al. Induction of a midbrain dopaminergic phenotype in Nurr1-overexpressing neural stem cells by type 1 astrocytes. Nat Biotechnol 1999;17:653–659.
- 7 Chung S, Hedlund E, Hwang M et al. The homeodomain transcription factor Pitx3 facilitates differentiation of mouse embryonic stem cells into AHD2-expressing dopaminergic neurons. Mol Cell Neurosci 2005;28: 241–252.
- 8 Zeng X, Cai J, Chen J et al. Dopaminergic differentiation of human embryonic stem cells. STEM CELLS 2004;22:925–940.
- 9 Park CH, Minn YK, Lee JY et al. In vitro and in vivo analyses of human embryonic stem cell-derived dopamine neurons. J Neurochem 2005;92: 1265–1276.
- 10 Ahn JI, Lee KH, Shin DM et al. Comprehensive transcriptome analysis of differentiation of embryonic stem cells into midbrain and hindbrain neurons. Dev Biol 2004;265:491–501.
- 11 Martin GR, Evans MJ. Differentiation of clonal lines of teratocarcinoma cells: Formation of embryoid bodies in vitro. Proc Natl Acad Sci U S A 1975;72:1441–1445.
- 12 Leahy A, Xiong JW, Kuhnert F et al. Use of developmental marker genes to define temporal and spatial patterns of differentiation during EB formation. J Exp Zool 1999;284:67–81.
- 13 Park WY, Hwang CI, Im CN et al. Identification of radiation-specific responses from gene expression profile. Oncogene 2002;21:8521–8528.

other in K-means clustering. More than half of the GD-DEGs were also found to be similarly changed in RD. We proposed that those overlaps might originate from the heterogeneous cell populations and should be subtracted from the GD-DEGs. The final GD-specific gene list contained only 66 of 11,376 genes in the microarray analysis; these genes were related mainly to transport and transcriptional regulation. Even though we need to further determine the functions of these genes in DA neuron differentiation, this approach has revealed more information about the guided neural differentiation and its role in neural differentiation as a whole.

ACKNOWLEDGMENTS

We thank S.H. Lee and J.Y. Kim for helping us to establish the DA neuron differentiation procedure. This work was supported by grants from the Stem Cell Research Center, KOREA, to W-Y.P. (SC11021) and H-S.P. (SC11022) and also by a grant from Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea, to J.H.K. (0405-BC02-0604-0004).

DISCLOSURES

The authors indicate no potential conflicts of interest.

- 14 Kim JH, Ha IS, Hwang CI et al. Gene expression profiling of anti-GBM glomerulonephritis model: The role of NF-kappaB in immune complex kidney disease. Kidney Int 2004;66:1826–1837.
- 15 Huber W, Von Heydebreck A, Sultmann H et al. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002;18(suppl 1):S96–S104.
- 16 Yang YH, Dudoit S, Luu P et al. Normalization for cDNA microarray data. Technical Report presented at: SPIE BiOS 2001; January 20–26, 2001; San Jose, CA. Available at http://www.stat.berkeley.edu/users/ terry/zarray/Html/normspie.html. Accessed October 31, 2003.
- 17 Wang SY, LaRosa GJ, Gudas LJ. Molecular cloning of gene sequences transcriptionally regulated by retinoic acid and dibutyryl cyclic AMP in cultured mouse teratocarcinoma cells. Dev Biol 1985;107:75–86.
- 18 Przyborski SA, Morton IE, Wood A et al. Developmental regulation of neurogenesis in the pluripotent human embryonal carcinoma cell line NTERA-2. Eur J Neurosci 2000;12:3521–3528.
- 19 Lendahl U, Zimmerman LB, McKay RDG. CNS stem cells express a new class of intermediate filament proteins. Cell 1990;60:585–595.
- 20 Rowitch DH, McMahon AP. Pax-2 expression in the murine neural plate precedes and encompasses the expression domains of Wnt-1 and En-1. Mech Dev 1995;52:3–8.
- 21 Saucedo-Cardenas O, Quintana-Hau JD, Le WD et al. Nurr1 is essential for the induction of the dopaminergic phenotype and the survival of ventral mesencephalic late dopaminergic precursor neurons. Proc Natl Acad Sci U S A 1998;95:4013–4018.
- 22 Western PS, Maldonado-Saldivia J, van den Bergen J et al. Analysis of Esg1 expression in pluripotent cells and the germline reveals similarities with Oct4 and Sox2 and differences between human pluripotent cell lines. STEM CELLS 2005;23:1436–1442.
- 23 Chen T, Li E. Structure and function of eukaryotic DNA methyltransferases. Curr Top Dev Biol 2004;60:55–89.
- 24 Wilk MB. Gnanadesikan. Probability plotting methods for the analysis of data. Biometrika 1968;55:1–17.
- 25 Edward L Korn, James F et al. Controlling the number of false discoveries: Application to high-dimensional genomic data. J Stat Planning Inf 2004;124:379–398.
- 26 Wright GW, Simon R. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 2003;19:2448–2455.

- 27 The Gene Ontology Consortium. Gene Ontology: Tool for the unification of biology. Nat Genet 2000;25:25–29.
- 28 Reppe S, Rian E, Jemtland R et al. Sox-4 messenger RNA is expressed in the embryonic growth plate and regulated via the parathyroid hormone/parathyroid hormone-related protein receptor in osteoblast-like cells. J Bone Miner Res 2000;15:2402–2412.
- 29 Hockfield S, McKay RDG. Identification of major cell classes in the developing mammalian nervous system. J Neurosci 1985;5:3310–3328.
- 30 Schoenherr CJ, Anderson DJ. The neuron-restrictive silencer factor (NRSF): A coordinate repressor of multiple neuron-specific genes. Science 1995;267:1360–1363.
- 31 Zheng JY, Koda T, Arimura Y et al. Structure and expression of the mouse S10 gene. Biochim Biophys Acta 1997;1351:47–50.
- 32 Krueger SR, Ghisu GP, Cinelli P et al. Expression of neuroserpin, an inhibitor of tissue plasminogen activator, in the developing and adult nervous system of the mouse. J Neurosci 1997;17:8984–8996.
- 33 Ferguson KL, Callaghan SM, O'Hare MJ et al. The Rb-CDK4/6 signaling pathway is critical in neural precursor cell cycle regulation. J Biol Chem 2000;275:33593–33600.
- 34 Helaakoski T, Annunen P, Vuori K et al. Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase alphasubunit isoform: Formation of an alpha(2)beta(2) tetramer with the protein disulfide-isomerase/beta subunit. Proc Nat Acad Sci U S A 1995;92:4427–4431.
- 35 Cabrera-Vera TM, Hernandez S, Earls LR et al. RGS9–2 modulates D2 dopamine receptor-mediated Ca2⁺ channel inhibition in rat striatal cholinergic interneurons. Proc Natl Acad Sci U S A 2004;101: 16339–16344.

- 36 Bornstein SR, Tian H, Haidan A et al. Deletion of tyrosine hydroxylase gene reveals functional interdependence of adrenocortical and chromaffin cell system in vivo. Proc Natl Acad Sci U S A 2000;97:14742–14747.
- 37 Thomas EA, Danielson PE, Nelson PA et al. Clozapine increases apolipoprotein D expression in rodent brain: Towards a mechanism for neuroleptic pharmacotherapy. J Neurochem 2001;76:789–796.
- 38 Fernandez C, Nieto O, Fontenla JA et al. Synthesis of glycosyl derivatives as dopamine prodrugs: Interaction with glucose carrier GLUT-1. Org Biomol Chem 2003;1:767–771.
- 39 Liss B, Franz O, Sewing S et al. Tuning pacemaker frequency of individual dopaminergic neurons by Kv4.3L and KChip3.1 transcription. EMBO J 2001;20:5715–5724.
- 40 Jiang M, Spicher K, Boulay G et al. Most central nervous system D2 dopamine receptors are coupled to their effectors by Go. Proc Natl Acad Sci U S A 2001;98:3577–3582.
- 41 Krieglstein K, Suter-Crazzolara C, Fischer WH et al. TGF-beta superfamily members promote survival of midbrain dopaminergic neurons and protect them against MPP⁺ toxicity. EMBO J 1995;14:736–742.
- 42 Avshalumov MV, Rice ME. Activation of ATP-sensitive K⁺ (KATP) channels by H2O2 underlies glutamate-dependent inhibition of striatal dopamine release. Proc Natl Acad Sci U S A 2003;100:11729–11734.
- 43 Shen Y, Yu Y, Guo H et al. Identification and comparative analysis of differentially expressed proteins in rat striatum following 6-hydroxydopamine lesions of the nigrostriatal pathway: Up-regulation of amyloid precursor-like protein 2 expression. Eur J Neurosci 2002;16:896–906.
- 44 Yamaguchi K, Hama H. Separation of periventricular dopaminergic and alpha-adrenergic systems from the vasopressin-secreting mechanisms activated by prostaglandin D2. Brain Res 1991;559:261–266.

See www.StemCells.com for supplemental material available online.

Selection of Neural Differentiation-Specific Genes by Comparing Profiles of Random Differentiation

Min Su Lee, Dae-Hyun Jun, Chang-Il Hwang, Seung Soo Park, Jason Jongho Kang, Hyun-Seok Park, Jihoon Kim, Ju Han Kim, Jeong-Sun Seo and Woong-Yang Park *Stem Cells* 2006;24;1946-1955; originally published online Apr 20, 2006; DOI: 10.1634/stemcells.2005-0325

This information is current as of July 20, 2000					
Updated Information & Services	including high-resolution figures, can be found at: http://www.StemCells.com/cgi/content/full/24/8/1946				
Supplementary Material	Supplementary material can be found at: http://www.StemCells.com/cgi/content/full/24/8/1946/DC1				

This information is current as of July 26, 2006

