NPC	Neural progenitor cells
NSC	Neural stem cells
ORF	Open-reading frame
RMA	Robust multiarray average
RT-PCR	Reverse transcription-polymerase chain reaction
SHH	Sonic hedgehog homolog
Wnt	Wingless-type MMTV integration site family

Introduction

Neural stem cells (NSC) with self-renewal and multipotent properties serve as an ideal cell source for transplantation to treat spinal cord injury, stroke, and neurodegenerative diseases (Kim 2004; Kim and de Vellis 2009). To efficiently induce neuronal lineage cells from NSC for neuron replacement therapy, we should clarify the intrinsic genetic programs involved in a time- and place-specific regulation of human NSC differentiation. Previously, we found that primary cultures of human neural progenitor cells (NPC) exhibit an intrinsic capacity to differentiate into astrocytes in response to bone morphogenic protein 4 (BMP4) included in the serum (Obayashi et al. 2009). This might be a major hindrance against the proper commitment to neuronal linage cells following transplantation of NSC in vivo. Recently, we established an immortalized human NSC clone HB1.F3 by retroviral vector-mediated v-myc gene transfer into fetal human telencephalon cell cultures (Kim 2004). HB1.F3 cells could provide an unlimited NSC source applicable to genetic manipulation ex vivo for cell-based therapy. Actually, F3 cells stably expressing therapeutic genes migrate and integrate into target brain tissues upon transplantation in animal models of Parkinson disease, Huntington disease, and amyotrophic lateral sclerosis, and they differentiate into neurons, followed by an enhanced functional recovery (Kim et al. 2006; Kim and de Vellis 2009).

Neurogenin-1 (NEUROG1, Ngn1) is a member of proneural basic helix-loop-helix (bHLH) transcription factors that promote neurogenesis by activating a battery of target genes, including the NeuroD family of bHLH transcription factors (Morrison 2001). During embryogenesis, Ngn1 is expressed in NPC distributed in dorsal root ganglia (DRG), dorsal and ventral regions of the neural tube, dorsal telencephalon, and specific regions within the midbrain and hindbrain (Sommer et al. 1996). Although there exists a functional redundancy among Ngn1, Ngn2, and Ngn3, Ngn1-deficient mice failed to generate a $TrkA^+$ subset of cervical DRG neurons (Ma et al. 1999). Overexpression of Ngn1 induces neurite outgrowth in F11 rat DRG and mouse neuroblastoma hybrid cells (Kim et al. 2002). Stable expression of Ngn1 induces neuronal differentiation of pluripotent mouse embryonal carcinoma P19 cells (Kim et al. 2004). Ngn1 inhibits differentiation of rat NSC into astrocytes by sequestering a transcriptional coactivator complex composed of CBP and SMAD1 and blocking activation of STAT transcription factors (Sun et al. 2001).

In the present study, to investigate the role of Ngn1 in human NSC differentiation, we established a clonal cell line stably overexpressing Ngn1 by retroviral vector-mediated gene transfer into HB1.F3 cells. Then, we studied genome-wide gene expression profiles of Ngn1-overexpressing F3 cells (F3-Ngn1) and wild-type F3 cells (F3-WT) by using whole genome DNA microarrays. As a result, we unexpectedly found that stable expression of a single gene Ngn1 in F3 cells induced a robust upregulation of leucine-rich repeat-containing G protein-coupled receptor 5 (LGR5), a recently identified marker for intestine and hair follicle stem cells (Barker et al. 2007; Jaks et al. 2008; Sato et al. 2009). Our results suggested that stable expression of Ngn1 in human NSC cells induces not only simply neurogenic but also multifunctional changes that potentially affect the differentiation of NSC via a reorganization of complex gene regulatory networks.

Methods

Human Neural Stem Cell Clone HB1.F3 and Its Derivative HB1.F3-Ngn1

Primary cultures of fetal human telencephalon cells were transformed with a retroviral vector pLSNmyc carrying the v-myc oncogene and the neomycin resistance gene. Following selection with G418, a single continuously dividing clone with a capacity to self-renew and differentiate into neurons and glial cells both in vitro and in vivo was isolated and designated HB1.F3 (Kim 2004). It carried normal human karyotype of 46 XX. After transducing a retroviral vector pBabePNgn1 carrying the open-reading frame (ORF) of the human Ngn1 gene and the puromycin resistance gene into HB1.F3 cells, a single puromycin-resistant clone was selected, expanded, and designated HB1.F3-Ngn1. In the present study, the wild-type HB1.F3 cells and the HB1.F3-Ngn1 cells are abbreviated as F3-WT and F3-Ngn1. They were incubated in the feeding medium composed of Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin and 100 μg/ml streptomycin. The medium was renewed every 3 days.

Microarray Analysis

Total RNA was isolated from subconfluent cells by using the TRIZOL Plus RNA Purification kit (Invitrogen). The

quality of total RNA was evaluated on Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). One hundred nanograms of total RNA was processed for cRNA synthesis, fragmentation, and terminal labeling with the GeneChip Whole Transcript Sense Target Labeling and Control Reagents (Affymetrix, Santa Clara, CA, USA). Then, it was processed for hybridization at 45°C for 17 h with Human Gene 1.0 ST Array (Affymetrix) containing 28,869 genes with approximately 26 probes per each gene that spread across the full length of the gene. The arrays were washed in the GeneChip Fluidic Station 450 (Affymetrix), and scanned by the GeneChip Scanner 3000 7G (Affymetrix). The data expressed as CEL files were normalized by the robust multiarray average (RMA) method with the Expression Console software version 1.1 (Affymetrix). By comparing the signal intensity levels between F3-WT and F3-Ngn1, the genes exhibiting either greater than twofold upregulation or smaller than 0.5-fold downregulation are considered as differentially expressed genes (DEG). To perform unsupervised clustering analysis of gene expression profiles, the CEL file-based data were imported to GeneSpring GX10 (Agilent).

Molecular Network Analysis

KeyMolnet is a comprehensive knowledgebase, originally established by the Institute of Medicinal Molecular Design (IMMD), Tokyo, Japan (Sato et al. 2005). It contains numerous contents of human genes, molecules and molecular relations, diseases, pathways, and drugs, all of which are manually collected, carefully curated, and regularly updated by expert biologists. The database is categorized into the core contents collected from selected review articles with the highest reliability or the secondary contents extracted from abstracts of PubMed database and Human Protein Reference database (HPRD). By importing the list of Entrez Gene ID and signal intensity data, KeyMolnet automatically provides corresponding molecules as a node on networks. Among various network-searching algorithms, the "N-points to N-points" search extracts the molecular network with the shortest route connecting the starting point molecules and the end point molecules. The generated network was compared side by side with 403 human canonical pathways of the KeyMolnet library. The algorithm counting the number of overlapping molecular relations between the extracted network and the canonical pathway makes it possible to identify the canonical pathway showing the most significant contribution to the extracted network. The significance in the similarity between both is scored following the formula, where $O =$ the number of overlapping molecular relations between the extracted network and the canonical pathway, $V =$ the number of molecular relations located in the extracted network, $C =$ the number of molecular relations located in the canonical pathway, $T =$ the number of total molecular relations composed of approximately 110,000 sets, and the $X =$ the sigma variable that defines coincidence.

$$\text{Score} = -\log_2(\text{Score (p)}) \quad \text{Score(p)} = \sum_{x=O}^{\text{Min}(C,V)} f(x)$$

$$f(x) = {}_C C_x \cdot {}_{T-C} C_{V-x} / {}_T C_V$$

Gene Annotation Analysis

Functional annotation of differentially expressed genes was searched by the web-accessible program named Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 2008, National Institute of Allergy and Infectious Diseases (NIAID), NIH (david.abcc.ncifcrf.gov) (Huang da et al. 2009). It covers more than 40 annotation categories, including Gene ontology (GO) terms, protein–protein interactions, protein functional domains, disease associations, biological pathways, sequence general features, homologies, gene functional summaries, and tissue expressions. By importing the list of Entrez Gene ID, this program creates the functional annotation chart, an annotation-term-focused view that lists annotation terms and their associated genes under study. To avoid excessive counting of duplicated genes, the Fisher Exact statistics is calculated based on corresponding DAVID gene IDs by which all redundancies in original IDs are removed.

Real-Time RT-PCR Analysis

DNase-treated total cellular RNA was processed for cDNA synthesis using oligo(dT)$_{12-18}$ primers and SuperScript II reverse transcriptase (Invitrogen). Then, cDNA was amplified by PCR in LightCycler ST300 (Roche Diagnostics, Tokyo, Japan) using SYBR Green I and primer sets listed in Table 1. The expression levels of target genes were standardized against those of the glyceraldehyde-3-phosphate dehydrogenase (G3PDH) gene detected in parallel in identical cDNA samples. All the assays were performed in triplicate.

In some experiments, the ORF of Ngn1 was amplified by PCR using PfuTurbo DNA polymerase (Stratagene, La Jolla, CA, USA) and primer sets listed in Table 1. It was then cloned into the mammalian expression vector p3XFLAG-CMV7.1 (Sigma, St. Louis, MO, USA) to express a fusion protein with an N-terminal Flag tag. At 48 h after transfection of the vector in F3-WT cells by Lipofectamine 2000 reagent (Invitrogen), the cells were processed for real-time RT-PCR analysis of LGR5 and Western blot analysis of a Flag-fusion protein with anti-Flag M2 antibody (Sigma).

Table 1 Primers for RT-PCR and cloning utilized in the present study

Genes	GenBank accession no.	Sense primers	Antisense primers
NES	NM_006617	5′ctgctcaggagcagcactcttaac3′	5′cttagcctatgagatggagcaggc3′
LGR5	NM_003667	5′aacagtcctgtgactcaactcaag3′	5′ttagagacatgggacaaatgccac3′
GAS2	NM_005256	5′acaaacatgtcatggtccgtgtgg3′	5′aactggcagagaccaccaagtagt3′
HAS2	NM_005328	5′gccagctgccttagaggaaatatc3′	5′atggttttccttcctgatgtgccc3′
MMP9	NM_004994	5′tcttccagtaccgagagaaagcct3′	5′ctgcaggatgtcataggtcacgta3′
NEUROG1	NM_006161	5′ttcctcaccgacgaggaagactgt3′	5′tcaagttgtgcatgcggttgcgct3′
NEUROG1 for cloning	NM_006161	5′cggaattccccagcccgccttgagacctgc3′	5′cgggatcccgctagtggtaaggaatgaaac3′
G3PDH	NM_002046	5′ccatgttcgtcatgggtgtgaacca3′	5′gccagtagaggcagggatgatgttc3′

NES Nestin, *LGR5* leucine-rich repeat-containing G protein-coupled receptor 5, *GAS2* growth arrest-specific 2, *HAS2* hyaluronan synthase 2, *MMP9* matrix metallopeptidase 9, *NEUROG1* neurogenin 1, and *G3PDH* glyceraldehyde-3-phosphate dehydrogenase

Western Blot Analysis

To prepare total protein extract, the cells were homogenized in RIPA buffer containing a cocktail of protease inhibitors (Sigma). After separation on a 12% SDS-PAGE gel, the protein was transferred onto a nitrocellulose membrane, and the blot was incubated with rabbit polyclonal anti-LGR5 antibody (AP2745d) (ABGENT, Flanders Court, San Diego, CA, USA). Then, it was labeled with HRP-conjugated anti-rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA, USA). The specific reaction was visualized by exposing of the blot to a chemiluminescence substrate (Pierce, Rockford, IL, USA). After the antibodies were stripped by incubating the membrane at 50°C for 30 min in stripping buffer, composed of 62.5 mM Tris–HCl, pH 6.7, 2% SDS, and 100 mM 2-mercaptoethanol, it was processed for relabeling with anti-Hsp60 antibody (N-20; Santa Cruz Biotechnology), an internal control for protein loading.

Results

Overexpression of Neurogenin 1 in F3-Ngn1 Cells

When incubated in the feeding medium, both F3-WT and F3-Ngn1 cells proliferated continuously with a doubling time ranging from 3 to 7 days. Although they were morphologically different, i.e. F3-WT exhibited a fusiform morphology, while F3-Ngn1 exhibited a cuboidal appearance (Fig. 1A, panels a and b), both of them expressed nestin but did not form a neurosphere when cultured in the feeding medium. The levels of expression of nestin mRNA were higher in F3-Ngn1 than F3-WT (Fig. 1B, panel a, lanes 1 and 2). Importantly, only F3-Ngn1 expressed Ngn1 mRNA (Fig. 1B, panel b, lanes 1 and 2).

We conducted genome-wide gene expression profiling of F3-WT and F3-Ngn1 by using two sets of Human Gene 1.0 ST Array for each, followed by two comparisons composed of F3-WT array-1 (WT-1) versus F3-Ngn1 array-1 (NGN-1) and F3-WT array-2 (WT-2) versus F3-Ngn1 array-2 (NGN-2). Unsupervised clustering analysis of these data clearly separated the cluster of F3-Ngn1 from that of F3-WT, based on gene expression profiles of 59 genes differentially expressed between both cell types (Fig. 2). The gene expression profile of WT-1 was similar to that of WT-2, while the gene expression profiles of NGN-1 and NGN-2 were almost identical, supporting the reproducibility among the results of repeated microarray analysis (Fig. 2). The analysis of individual probe data identified significant upregulation of Ngn1 ORF expression

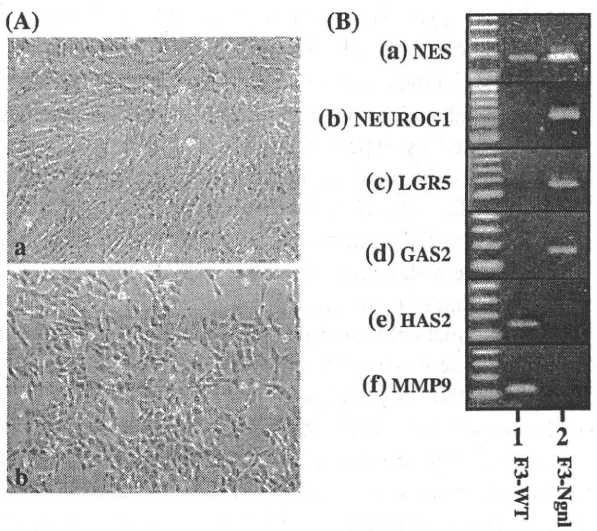

Fig. 1 Characterization of phenotypes of F3-WT and F3-Ngn1 cells. **A** Phase contrast photomicrograph. Both F3-WT cells (panel *a*) and F3-Ngn1 cells (panel *b*) were incubated in the feeding medium at a subconfluent density. **B** RT-PCR analysis. cDNA prepared from F3-WT cells (lane 1) and F3-Ngn1 cells (lane 2) was amplified by PCR for 30 cycles using primer sets listed in Table 1. The panels (*a–f*) represent (*a*) nestin (NES), (*b*) neurogenin 1 (NEUROG1), (*c*) leucine-rich repeat-containing G protein-coupled receptor 5 (LGR5), (*d*) growth arrest-specific 2 (GAS2), (*e*) hyaluronan synthase 2 (HAS2), and (*f*) matrix metallopeptidase 9 (MMP9)

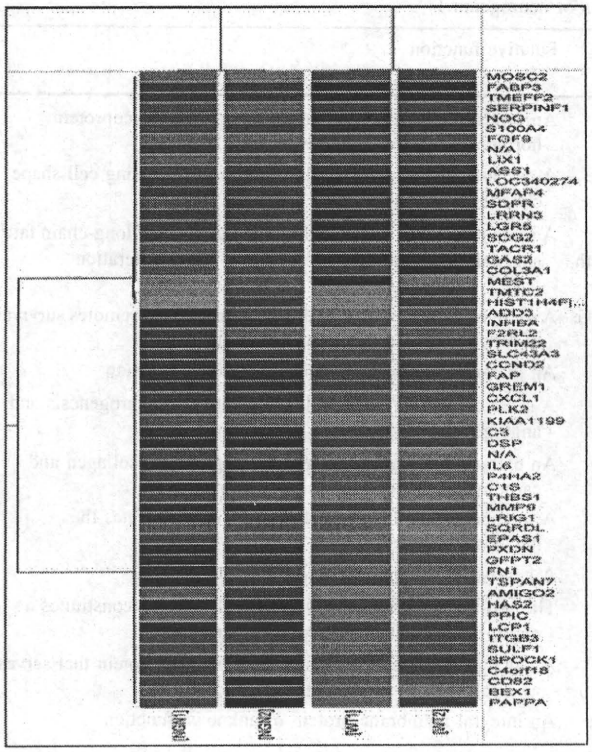

Fig. 2 Clustering analysis of gene expression profiles of F3-WT and F3-Ngn1 cells. Genome-wide gene expression profiling of F3-WT and F3-Ngn1 was performed by using two sets of Human Gene 1.0 ST Array for each, followed by two comparisons composed of WT array-1 (WT1) versus Ngn1 array-1 (NGN1) and WT array-2 (WT2) versus Ngn1 array-2 (NGN2). The microarray data are processed for unsupervised clustering analysis on GeneSpring GX10. A set of 59 differentially expressed genes between both cell types separated the cluster of F3-Ngn1 from that of F3-WT. The heat map represents upregulated genes (*orange*) and downregulated genes (*blue*)

in F3-Ngn1 cells (Supplementary Fig. 1). However, the Affymetrix GeneChip Command Console (AGCC) algorithm that calculates the cumulative gene expression levels excluded Ngn1 from the group of upregulated genes in F3-Ngn1 owing to low baseline expression of Ngn1 in the set of probes distributed outside its ORF on Human Gene 1.0 ST Array.

Microarray Analysis Identifies a Robust Induction of LGR5 in F3-Ngn1 Cells

Microarray analysis identified total 588 differentially expressed genes (DEG), composed of 250 upregulated genes and 338 downregulated genes in F3-Ngn1 versus F3-WT (see Supplementary Tables 1 and 2 for the complete lists). Top 20 upregulated genes are shown in Table 2. Notably, LGR5, a novel stem cell marker (Barker et al. 2007; Jaks et al. 2008; Sato et al. 2009), showed an 167-fold increase in F3-Ngn1 (Table 2; Fig. 3).

In view of cell type-specific markers for NSC, neurons, and glial cells, nestin (NES) exhibited a 3.1-fold increase in F3-Ngn1 (Supplementary Table 1; Fig. 3), consistent with RT-PCR results (Fig. 1B, panel a, lanes 1 and 2). However, the expression of other NSC-specific markers, such as musashi homolog 1 (MSI1) and ATP-binding cassette sub-family G member 2 (ABCG2), was not elevated in F3-Ngn1 (Fig. 3). Although neurofilament medium polypeptide (NEFM) showed a 2.1-fold increase, the expression of other neuron-specific markers, such as neurofilament heavy polypeptide (NEFH), enolase 2 (ENO2), and tubulin beta 3 (TUBB3), was not substantially upregulated in F3-Ngn1 (Fig. 3). The expression of astroglial (GFAP), oligodendroglial (MBP, MOG, and CNP), and microglial (CD68) markers remained unaltered (Fig. 3). Furthermore, NEUROD1, a putative Ngn-1 target gene,[5,10] was not upregulated in F3-Ngn1 (Fig. 3).

Top 20 downregulated genes are shown in Table 3. It is worthy to note that the great majority of top 20 downregulated genes are categorized as extracellular matrix-associated proteins.

RT-PCR and Western Blot Analysis Validated the Results of Microarray Analysis

Both the conventional RT-PCR and real-time RT-PCR analysis validated marked upregulation of LGR5 and GAS2, and remarkable downregulation of HAS2 and MMP9 in F3-Ngn1 (Fig. 1B, panels c–f, lanes 1 and 2; Fig. 4, panels a–d). Western blot analysis verified LGR5 protein expression exclusively in F3-Ngn1 (Fig. 4, panel e, lane 2).

To address the question whether LGR5 is a direct target for Ngn1, an expression vector of either Ngn1 or green fluorescent protein (GFP) was transfected in F3-WT cells (Fig. 5a, upper panel, lanes 1 and 2). At 48 h after transfection, the cells were processed for real-time RT-PCR analysis. Transient overexpression of Ngn1 did not induce LGR5 expression in F3-WT, suggesting that LGR5 is not a direct transcriptional target of Ngn1 (Fig. 5b).

An Involvement of the Complex Interaction of Networks Regulated by Multiple Transcription Factors in Development of F3-Ngn1 Cells

To clarify the molecular network of the genes differentially expressed between F3-WT and F3-Ngn1, we imported microarray data into KeyMolnet, a bioinformatics tool for analyzing molecular relations on a comprehensive knowledgebase. When Entrez Gene ID and expression levels of 588 DEG were imported, KeyMolnet recognized a set of 51 non-annotated genes to be removed. Then, it extracted 787

Table 2 Top 20 upregulated genes in F3 cells following stable expression of neurogenin 1

No.	Gene symbol	Fold change	Entrez gene ID	Gene name	Putative function
1	LGR5	166.623	8549	Leucine-rich repeat-containing G protein-coupled receptor 5	An orphan G protein-coupled receptor of the glycoprotein hormone receptor subfamily
2	GAS2	32.861	2620	Growth arrest-specific 2	A caspase-3 substrate that plays a role in regulating cell shape changes during apoptosis
3	FABP3	32.739	2170	Fatty acid-binding protein 3, muscle and heart (mammary-derived growth inhibitor)	A protein involved in intracellular metabolism of long-chain fatty acids and modulation of cell growth and proliferation
4	TMEFF2	28.233	23671	Transmembrane protein with EGF-like and two follistatin-like domains 2	A secreted protein with a EGF-like domain that promotes survival of hippocampal and mesencephalic neurons
5	LRRN3	24.284	54674	Leucine-rich repeat neuronal 3	An integral membrane protein of unknown function
6	SCG2	13.841	7857	Secretogranin II (chromogranin C)	A secretory protein involved in regulation of neurogenesis and angiogenesis
7	MFAP4	13.429	4239	Microfibrillar-associated protein 4	An extracellular matrix protein binding to both collagen and carbohydrate involved in cell adhesion
8	HIST1H4F	11.875	8361	H4 histone, family 2	A member of the histone H4 family that constitutes the nucleosome structure
9	TACR1	11.35	6869	Tachykinin receptor 1	A neurokinin receptor selective for substance P
10	COL3A1	10.828	1281	Collagen, type III, alpha 1	The pro-alpha 1 chains of type III collagen that constitutes a major component of the extracellular matrix
11	SDPR	9.975	8436	Serum deprivation response (phosphatidylserine-binding protein)	A calcium-independent phospholipid-binding protein that serves as a substrate of protein kinase C
12	TMTC2	9.744	160335	Transmembrane and tetratricopeptide repeat containing 2	An integral membrane protein of unknown function
13	FGF9	9.38	2254	Fibroblast growth factor 9 (glia-activating factor)	A member of the FGF family whose expression is dependent on Sonic hedgehog signaling
14	ASS1	8.813	445	Argininosuccinate synthetase 1	The enzyme that catalyzes the penultimate step of the arginine biosynthetic pathway
15	S100A4	8.555	6275	S100 calcium binding protein A4	A member of the S100 family of proteins involved in motility, invasion, and tubulin polymerization
16	LIX1	7.99	167410	Lix1 homolog (chicken)	A protein involved in RNA metabolism that has an essential function for motor neuron survival
17	FAM65B	7.974	9750	Family with sequence similarity 65, member B	a protein involved in myogenic cell differentiation
18	NOG	7.65	9241	Noggin	A secreted protein that plays a principal role in creating morphogenic gradients by antagonizing bone morphogenetic proteins
19	C1orf115	7.571	79762	Chromosome 1 open reading frame 115	An integral membrane protein of unknown function
20	CYSLTR2	7.23	57105	Cysteinyl leukotriene receptor 2	A G protein-coupled receptor for cysteinyl leukotrienes

Genome-wide gene expression profiling of F3-WT and F3-Ngn1 was performed by using two sets of Human Gene 1.0 ST Array for each, followed by two comparisons composed of WT array-1 (F3-WT-1) versus Ngn1 array-1 (F3-Ngn1-1) and WT array-2 (F3-WT-2) versus Ngn1 array-2 (F3-Ngn1-2). Top 20 upregulated genes in F3-Ngn1 cells are shown with fold change derived from the comparison between F3-WT-2 and F3-Ngn1-2

genes directly linked to the 537 genes. Subsequently, we performed the "N-points to N-points" search by starting from Ngn1 and ending with the set of 787 genes via the shortest route connecting starting and ending points. It generated a highly complex molecular network composed of 1,816 fundamental nodes and 7,238 molecular relations (Fig. 6). When the network was referred to the canonical pathways of the KeyMolnet library, the generated network has the most significant relationship with transcriptional regulation by nuclear factor kappa-B (NF-κB) with the score of 59.9 and score (p) = 9.467E−019. This was followed by transcriptional regulation by cyclic AMP-response element-binding protein (CREB) in the second rank with the score of 52.3 and score (p) = 1.771E−016, transcriptional regulation by vitamin D receptor (VDR) in the third rank with the score of 45.8 and score (p) = 1.582E−014, transcriptional

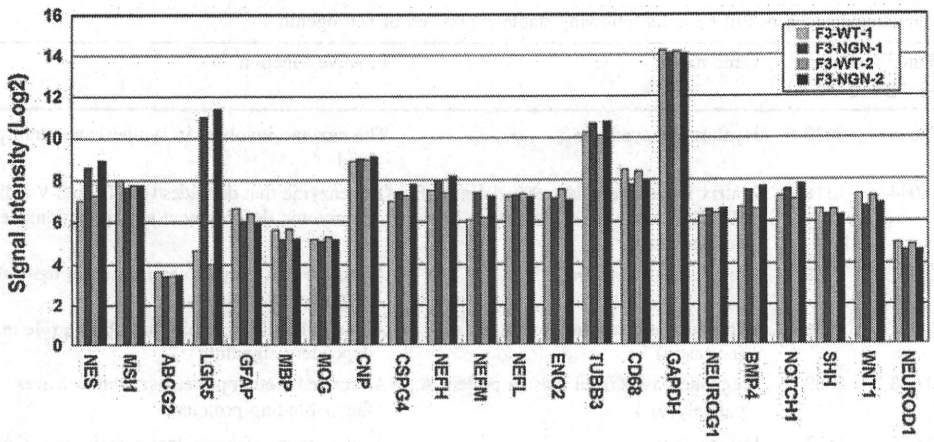

Fig. 3 Gene expression profiles of NSC, neuronal and glial markers. Genome-wide gene expression profiling of F3-WT and F3-Ngn1 was performed by using two sets of Human Gene 1.0 ST Array for each, followed by two comparisons composed of WT array-1 (F3-WT-1; the first column) versus Ngn1 array-1 (F3-Ngn1-1; the second column) and WT array-2 (F3-WT-2; the third column) versus Ngn1 array-2 (F3-Ngn1-2; the fourth column). Signal intensities of NSC, neuronal and glial marker genes are expressed as log 2 after normalization. *NES* nestin, *MSI1* musashi homolog 1, *ABCG2* ATP-binding cassette, subfamily G member 2, *LGR5* leucine-rich repeat-containing G protein-coupled receptor 5, *GFAP* glial fibrillary acidic protein, *MBP* myelin basic protein, *MOG* myelin oligodendrocyte glycoprotein, *CNP*, 2′3′-cyclic nucleotide 3′ phosphodiesterase, *CSPG4* chondroitin sulfate proteoglycan 4 (NG2), *NEFH* neurofilament heavy polypeptide, *NEFM* neurofilament medium polypeptide, *NEFL* neurofilament light polypeptide, *ENO2* enolase 2 (NSE), *TUBB3* tubulin beta 3, *GAPDH* glyceraldehyde-3-phosphate dehydrogenase (G3PDH), *NEUROG1* neurogenin 1, *BMP4* bone morphogenic protein 4, *NOTCH1* notch homolog 1, *SHH* sonic hedgehog homolog, *WNT1* wingless-type MMTV integration site family member 1, and *NEUROD1* neurogenic differentiation 1. A robust upregulation of LGR5 is evident in both F3-Ngn1-1 and F3-Ngn1-2

regulation by hypoxia-inducible factor (HIF) in the fourth rank with the score of 35.7 and score (p) = 1.781E−011, transcriptional regulation by glucocorticoid receptor (GR) in the fifth rank with the score of 31.0 and score (p) = 4.779E−010, and the complement activation pathway in the sixth rank with the score of 20.5 and score (p) = 6.589E−007. Thus, the molecular network of the genes differentially expressed between F3-WT and F3-Ngn1 involves the complex interaction of networks regulated by multiple transcription factors.

Gene Annotation Analysis Suggested Multifunctional Changes in F3-Ngn1 Cells

We studied functional annotation terms overrepresented in 588 DEG by using the web-accessible program named DAVID. By importing the list of Entrez Gene ID, DAVID identified top 20 enriched gene ontology (GO) terms in the list of 250 upregulated genes, most of which are related to development and morphogenesis (Table 4). In contrast, top 20 enriched GO terms in the list of 338 downregulated genes were chiefly composed of the molecules closely associated with extracellular matrix and adhesion (Table 4). Thus, gene annotation analysis suggested that stable expression of a single gene Ngn1 in F3 cells induces multifunctional changes that potentially affect the differentiation of human NSC.

Discussion

Recently, we established an immortalized human NSC clone named HB1.F3, which could serve as an unlimited source for cell replacement therapy of various neurological diseases (Kim 2004; Kim and de Vellis 2009). Ngn1 is a proneural bHLH transcription factor that promotes neuronal differentiation but inhibits glial differentiation of rodent NSC and NPC (Morrison 2001; Sun et al. 2001). In the present study, to investigate a role of Ngn1 in human NSC differentiation, we established a clone derived from F3 stably overexpressing Ngn1. Genome-wide gene expression profiling identified 250 upregulated genes and 338 downregulated genes in F3-Ngn1 versus F3-WT cells. Notably, the expression of LGR5, a recently identified marker for intestine and hair follicle stem cells (Barker et al. 2007; Jaks et al. 2008; Sato et al. 2009), was greatly elevated in F3-Ngn1 cells at both mRNA and protein levels. However, transient overexpression of Ngn1 did not induce upregulation of LGR5 in F3-WT cells, suggesting that LGR5 is not a direct transcriptional target of Ngn1. KeyMolnet, a bioinformatics tool for analyzing molecular relations on a comprehensive knowledgebase, indicated that the molecular network of differentially expressed genes involves the complex interaction of networks regulated by multiple transcription factors, such as NF-κB, CREB, VDR, HIF, and GR. Gene annotation analysis

Table 3 Top 20 downregulated genes in F3 cells following stable expression of neurogenin 1

No.	Gene symbol	Fold change	Entrez gene ID	Gene name	Putative function
1	HAS2	0.024	3037	Hyaluronan synthase 2	The enzyme involved in synthesis and transport of hyaluronic acid
2	MMP9	0.044	4318	Matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)	The enzyme that degrades type IV and V collagens involved in embryonic development and tissue remodeling
3	C3	0.05	718	Complement component 3	A protein that plays a central role in the activation of complement system
4	LCP1	0.05	3936	Lymphocyte cytosolic protein 1 (L-plastin)	An actin-binding protein that plays a role in cell adhesion-dependent signaling
5	PAPPA	0.068	5069	Pregnancy-associated plasma protein A, pappalysin 1	A secreted metalloproteinase which cleaves insulin-like growth factor binding proteins
6	DSP	0.072	1832	Desmoplakin	A component of functional desmosomes that anchors intermediate filaments to desmosomal plaques
7	SPOCK1	0.075	6695	Sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1	A chondroitin sulfate/heparan sulfate proteoglycan expressed in the postsynaptic region of hippocampal pyramidal neurons
8	TRIM22	0.076	10346	Tripartite motif-containing 22	A member of the tripartite motif family induced by interferon and mediates interferon's antiviral effects
9	CCND2	0.087	894	Cyclin D2	A protein that forms a complex with CDK kinases involved in cell cycle G1/S transition
10	IL6	0.088	3569	Interleukin 6 (interferon, beta 2)	An immunoregulatory cytokine that functions in inflammation and the maturation of B cells
11	CD82	0.092	3732	CD82 molecule	A membrane glycoprotein activated by p53 involved in suppression of metastasis
12	SLC43A3	0.093	29015	Solute carrier family 43, member 3	An integral membrane protein of the SLC43A transporter family
13	GREM1	0.093	26585	Gremlin 1, cysteine knot superfamily, homolog (*Xenopus laevis*)	A member of bone morphogenic protein antagonist family expressed in the neural crest
14	INHBA	0.095	3624	Inhibin, beta A	A growth/differentiation factor for various cell types by acting as a homodimer (activin A) or a heterodimer (activin A-B)
15	ITGB3	0.095	3690	Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	A subunit of integrins involved in cell adhesion and cell-surface-mediated signaling
16	FAP	0.097	2191	Fibroblast activation protein, alpha	A homodimeric integral membrane gelatinase involved in epithelial–mesenchymal interactions during development
17	C1S	0.101	716	Complement component 1, s subcomponent	A major constituent of the human complement subcomponent C1 that associates with C1r and C1q to yield the first component of the serum complement system
18	PXDN	0.103	7837	Peroxidasin homolog (Drosophila)	An extracellular matrix-associated peroxidase involved in extracellular matrix consolidation
19	C4orf18	0.103	51313	Chromosome 4 open reading frame 18	A Golgi apparatus membrane of unknown function
20	CFH	0.104	3075	Complement factor H	Q serum glycoprotein that regulates the function of the alternative complement pathway

Genome-wide gene expression profiling of F3-WT and F3-Ngn1 was performed by using two sets of Human Gene 1.0 ST Array for each, followed by two comparisons composed of WT array-1 (F3-WT-1) versus Ngn1 array-1 (F3-Ngn1-1) and WT array-2 (F3-WT-2) versus Ngn1 array-2 (F3-Ngn1-2). Top 20 downregulated genes in F3-Ngn1 cells are shown with fold change derived from the comparison between F3-WT-2 and F3-Ngn1-2

suggested that GO terms of development and morphogenesis are enriched in upregulated genes, while those of extracellular matrix and adhesion are enriched in downregulated genes. These results suggest that stable expression of a single gene Ngn1 in F3 cells induces not simply neurogenic but multifunctional changes that potentially affect the differentiation of human NSC via a reorganization of complex gene regulatory networks.

LGR5, an orphan G protein-coupled receptor alternatively named GRP49 with structural similarity to the

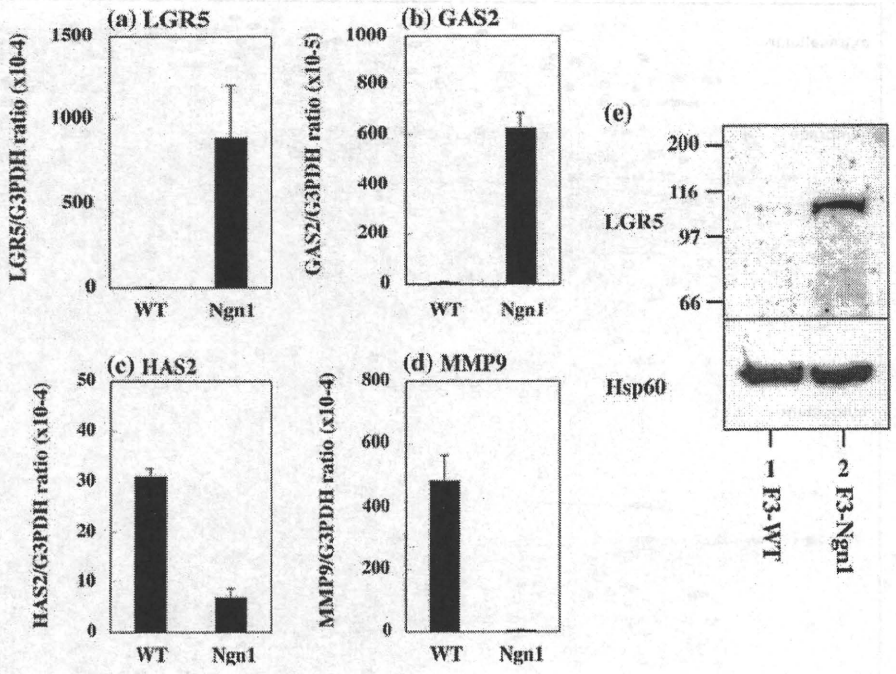

Fig. 4 Real-time RT-PCR and Western blot analysis. cDNA prepared from F3-WT and F3-Ngn1 cells was processed for real-time RT-PCR using primer sets listed in Table 1. Total protein extract was processed for western blot with anti-LGR5 antibody. The panels (**a–e**) represent real-time RT-PCR of **a** LGR5, **b** GAS2, **c** HAS2, and **d** MMP9, and western blot of (**e**, *upper panel*) LGR5 and (**e**, *lower panel*) Hsp60, an internal control for protein loading

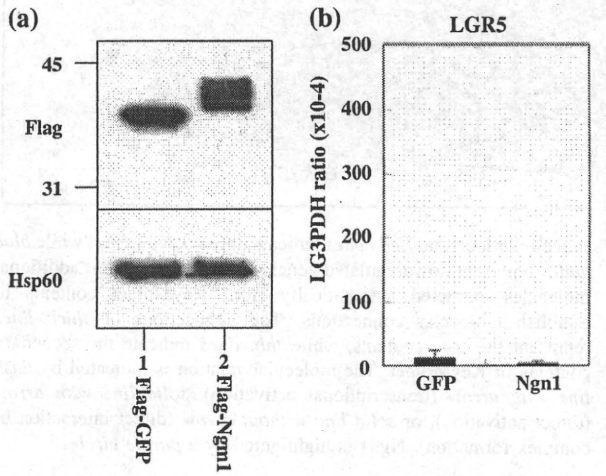

Fig. 5 Transient overexpression of Ngn1 did not induce upregulation of LGR5 in F3-WT cells. Expression vectors of Flag-tagged Ngn1 or GFP were transfected in F3-WT cells. At 48 h after transfection, the cells were processed for Western blot analysis of Flag and real-time RT-PCR analysis of LGR5. **a** Western blot analysis. The lanes (1, 2) represent *1* Flag-tagged GFP and *2* Flag-tagged Ngn1. The *upper panel* indicates Flag-tagged proteins, while the *lower panel* indicates Hsp60, an internal control for protein loading. **b** Real-time RT-PCR analysis. The *left bar* represents F3-WT cells with transient overexpression of Flag-tagged GFP, while the *right bar* represents those with transient overexpression of Flag-tagged Ngn1

glycoprotein hormone receptor family, is recently identified as a marker of adult intestinal stem cells and hair follicle stem cells by lineage-tracing studies (Barker et al. 2007; Jaks et al. 2008; Sato et al. 2009). LGR5 expression is also identified in the adult human spinal cord and brain at least at mRNA levels (Hsu et al. 1998). At present, the precise physiological function of LGR5 and downstream signaling pathways remain unknown owing to the lack of an identified natural ligand. LGR gene knockout mice showed neonatal lethality due to a breast-feeding defect caused by ankyloglossia, suggesting an involvement of LGR5 in craniofacial development (Morita et al. 2004). A more recent study showed that LGR5 deficiency induces premature differentiation of Paneth cells in the small intestine, accompanied by overactivation of the Wnt pathway, indicating that LGR5 acts as a negative regulator of Wnt (Garcia et al. 2009). A different study revealed that LGR5 is a marker for the sublineage of intestinal stem cells that are responsive to Wnt signals derived from stem cell niche (Ootani et al. 2009). In the populations of intestinal stem cells, LGR5 labels cycling cells, while doublecortin-like kinase-1 (DCLK1) marks quiescent cells (May et al. 2009). Interestingly, the expression of DCLK2, a putative paralog of DCLK1, is elevated with a 3.38-fold increase in F3-Ngn1 cells (Supplementary Table 1).

The interaction between Wnt proteins and Frizzled receptors on the cell surface transduces the signals to β-catenin by inactivating glycogen synthase kinase 3β (GSK3β), and stabilized β-catenin is translocated into the nucleus and forms a complex with T-cell factor (TCF) transcription factors to activate transcription of Wnt target genes. Importantly, LGR5 is identified as one of Wnt target genes (Yamamoto et al. 2003), suggesting a key role of LGR5 in establishment of a negative feedback loop in the

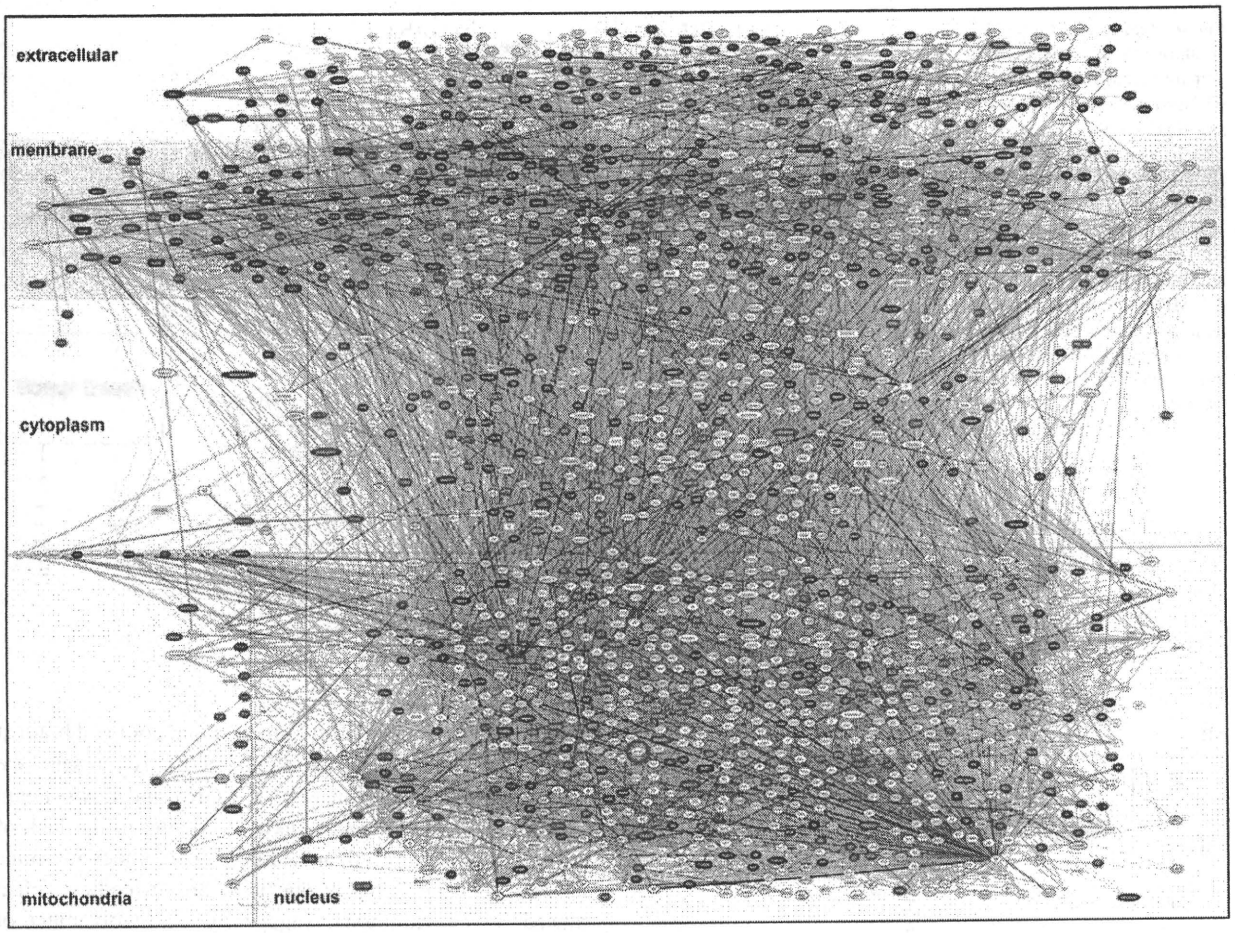

Fig. 6 Molecular network analysis of the genes regulated by stable expression of Ngn1 in F3 cells. The Entrez Gene ID and expression levels of 588 differentially expressed genes (DEG) between F3-WT and F3-Ngn1 cells were imported into KeyMolnet. It extracted 787 genes directly linked to the DEG. The "N-points to N-points" search was performed by starting from Ngn1 and ending with the set of 787 genes via the shortest route connecting starting and ending points. It generated a complex molecular network composed of 1,816 fundamental nodes and 7,238 molecular relations, arranged according to the subcellular location. *Red nodes* indicate upregulated genes, while *blue nodes* represent downregulated genes. *White nodes* exhibit additional molecules extracted automatically from KeyMolnet contents to establish molecular connections. The connections of *thick lines* represent the core contents, while *thin lines* indicate the secondary contents of KeyMolnet. The molecular relation is indicated by *dash line with arrow* (transcriptional activation), *solid line with arrow* (direct activation), or *solid line without arrow* (direct interaction or complex formation). Ngn1 is highlighted by a *purple circle*

Wnt pathway. In the present study, several Wnt target genes, such as FGF9 (Hendrix et al. 2006) and Dick homolog 1 (DKK1) (Niida et al. 2004), are coordinately upregulated, whereas MMP9 (Wu et al. 2007) is markedly downregulated in F3-Ngn1 (Tables 2 and 3; Supplementary Tables 1 and 2). FGF9 inhibits astrocyte differentiation of adult mouse NPC (Lum et al. 2009). One of us (SUK) recently found that DKK1 is a negative regulator of Wnt signaling in HB1.F3 cells (Ahn et al. 2008). MMP9 plays a central role in migration of adult NSC and NPC (Barkho et al. 2008). Interestingly, Ngn1 is also identified as a target of Wnt signaling, and it inhibits the self-renewal capacity of mouse cortical neural precursor cells (Hirabayashi et al. 2004). The expression of LGR5 is also controlled by the sonic hedgehog (SHH) signaling pathway (Tanese et al. 2008). SHH promotes Ngn1 expression in trigeminal neural crest cells (Ota and Ito 2003). Importantly, both Wnt and SHH signaling pathways play a central role in NSC development and differentiation (Prakash and Wurst 2007). Therefore, F3-Ngn1 cells might serve as a valuable tool for screening natural ligands of LGR5 that potentially affect the human NSC differentiation via Wnt and SHH signaling pathways, although it remains to be investigated whether a specialized subset of $LGR5^+$ NSC exists in vivo in the adult human central nervous system (CNS).

Table 4 Functional annotation terms of upregulated and downregulated genes in F3-Ngn1 cells

No.	Top 20 enriched GO terms in upregulated genes	*P* value	Top 20 enriched GO terms in downregulated genes	*P* value
1	GO:0048513 ~ organ development	4.71E−08	GO:0044421 ~ extracellular region part	1.66E−24
2	GO:0048731 ~ system development	5.15E−08	GO:0005576 ~ extracellular region	4.49E−22
3	GO:0048856 ~ anatomical structure development	1.32E−07	GO:0005578 ~ proteinaceous extracellular matrix	5.78E−15
4	GO:0007275 ~ multicellular organismal development	1.02E−06	GO:0031012 ~ extracellular matrix	9.50E−15
5	GO:0032502 ~ developmental process	1.73E−06	GO:0009605 ~ response to external stimulus	9.65E−14
6	GO:0009887 ~ organ morphogenesis	1.96E−06	GO:0005615 ~ extracellular space	1.42E−12
7	GO:0001501 ~ skeletal development	4.29E−05	GO:0009611 ~ response to wounding	2.42E−12
8	GO:0009653 ~ anatomical structure morphogenesis	5.80E−05	GO:0022610 ~ biological adhesion	3.08E−12
9	GO:0050793 ~ regulation of developmental process	1.46E−04	GO:0007155 ~ cell adhesion	3.08E−12
10	GO:0051216 ~ cartilage development	1.78E−04	GO:0048731 ~ system development	7.30E−10
11	GO:0007399 ~ nervous system development	1.97E−04	GO:0006954 ~ inflammatory response	1.24E−09
12	GO:0048754 ~ branching morphogenesis of a tube	2.23E−04	GO:0048856 ~ anatomical structure development	1.38E−09
13	GO:0001657 ~ ureteric bud development	2.37E−04	GO:0032502 ~ developmental process	1.87E−09
14	GO:0032501 ~ multicellular organismal process	2.47E−04	GO:0044420 ~ extracellular matrix part	2.45E−09
15	GO:0048598 ~ embryonic morphogenesis	2.48E−04	GO:0005581 ~ collagen	3.15E−09
16	GO:0030154 ~ cell differentiation	3.15E−04	GO:0007275 ~ multicellular organismal development	1.32E−08
17	GO:0048869 ~ cellular developmental process	3.15E−04	GO:0048513 ~ organ development	2.45E−08
18	GO:0000786 ~ nucleosome	3.33E−04	GO:0005509 ~ calcium ion binding	4.38E−08
19	GO:0043583 ~ ear development	3.42E−04	GO:0005125 ~ cytokine activity	6.33E−08
20	GO:0001763 ~ morphogenesis of a branching structure	3.42E−04	GO:0006950 ~ response to stress	1.21E−07

Functional annotation terms overrepresented in the list of 588 genes differentially expressed between F3-WT and F3-Ngn1 cells were searched on the web-accessible program named DAVID. Top 20 enriched gene ontology (GO) terms in 250 upregulated genes and top 20 enriched GO terms in 338 downregulated genes are shown with GO ID and *P* value

Acknowledgments This work was supported by a research grant to J-IS from the High-Tech Research Center Project, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan (S0801043), and from Research on Intractable Diseases, the Ministry of Health, Labour and Welfare of Japan. The microarray data are available from Gene Expression Omnibus (GEO) under the accession number GSE18296.

References

Ahn SM, Byun K, Kim D, Lee K, Yoo JS, Kim SU, Jho EH, Simpson RJ, Lee B (2008) Olig2-induced neural stem cell differentiation involves downregulation of Wnt signaling and induction of Dickkopf-1 expression. PLoS One 3:e3917

Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, Clevers H (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449:1003–1007

Barkho BZ, Munoz AE, Li X, Li L, Cunningham LA, Zhao X (2008) Endogenous matrix metalloproteinase (MMP)-3 and MMP-9 promote the differentiation and migration of adult neural progenitor cells in response to chemokines. Stem Cells 26:3139–3149

Garcia MI, Ghiani M, Lefort A, Libert F, Strollo S, Vassart G (2009) LGR5 deficiency deregulates Wnt signaling and leads to precocious Paneth cell differentiation in the fetal intestine. Dev Biol 331:58–67

Hendrix ND, Wu R, Kuick R, Schwartz DR, Fearon ER, Cho KR (2006) Fibroblast growth factor 9 has oncogenic activity and is a downstream target of Wnt signaling in ovarian endometrioid adenocarcinomas. Cancer Res 66:1354–1362

Hirabayashi Y, Itoh Y, Tabata H, Nakajima K, Akiyama T, Masuyama N, Gotoh Y (2004) The Wnt/β-catenin pathway directs neuronal differentiation of cortical neural precursor cells. Development 131:2791–2801

Hsu SY, Liang SG, Hsueh AJ (1998) Characterization of two LGR genes homologous to gonadotropin and thyrotropin receptors with extracellular leucine-rich repeats and a G protein-coupled, seven-transmembrane region. Mol Endocrinol 12:1830–1845

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57

Jaks V, Barker N, Kasper M, van Es JH, Snippert HJ, Clevers H, Toftgård R (2008) Lgr5 marks cycling, yet long-lived, hair follicle stem cells. Nat Genet 40:1291–1299

Kim SU (2004) Human neural stem cells genetically modified for brain repair in neurological disorders. Neuropathology 24: 159–174

Kim SU, de Vellis J (2009) Stem cell-based cell therapy in neurological diseases: a review. J Neurosci Res 87:2183–2200

Kim S, Ghil SH, Kim SS, Myeong HH, Lee YD, Suh-Kim H (2002) Overexpression of neurogenin1 induces neurite outgrowth in F11 neuroblastoma cells. Exp Mol Med 34:469–475

Kim S, Yoon YS, Kim JW, Jung M, Kim SU, Lee YD, Suh-Kim H (2004) Neurogenin1 is sufficient to induce neuronal differentiation of embryonal carcinoma P19 cells in the absence of retinoic acid. Cell Mol Neurobiol 24:343–356

Kim SU, Park IH, Kim TH, Kim KS, Choi HB, Hong SH, Bang JH, Lee MA, Joo IS, Lee CS, Kim YS (2006) Brain transplantation of human neural stem cells transduced with tyrosine hydroxylase and GTP cyclohydrolase 1 provides functional improvement in animal models of Parkinson disease. Neuropathology 26: 129–140

Lum M, Turbic A, Mitrovic B, Turnley AM (2009) Fibroblast growth factor-9 inhibits astrocyte differentiation of adult mouse neural progenitor cells. J Neurosci Res 87:2201–2210

Ma Q, Fode C, Guillemot F, Anderson DJ (1999) Neurogenin1 and neurogenin2 control two distinct waves of neurogenesis in developing dorsal root ganglia. Genes Dev 13:1717–1728

May R, Sureban SM, Hoang N, Riehl TE, Lightfoot SA, Ramanujam R, Wyche JH, Anant S, Houchen CW (2009) DCAMKL-1 and LGR5 mark quiescent and cycling intestinal stem cells respectively. Stem Cells. doi:10.1002/stem.193

Morita H, Mazerbourg S, Bouley DM, Luo CW, Kawamura K, Kuwabara Y, Baribault H, Tian H, Hsueh AJ (2004) Neonatal lethality of LGR5 null mice is associated with ankyloglossia and gastrointestinal distension. Mol Cell Biol 24:9736–9743

Morrison SJ (2001) Neuronal differentiation: proneural genes inhibit gliogenesis. Curr Biol 11:R349–R351

Niida A, Hiroko T, Kasai M, Furukawa Y, Nakamura Y, Suzuki Y, Sugano S, Akiyama T (2004) DKK1, a negative regulator of Wnt signaling, is a target of the β-catenin/TCF pathway. Oncogene 23:8520–8526

Obayashi S, Tabunoki H, Kim SU, Satoh J (2009) Gene expression profiling of human neural progenitor cells following the serum-induced astrocyte differentiation. Cell Mol Neurobiol 29: 423–438

Ootani A, Li X, Sangiorgi E, Ho QT, Ueno H, Toda S, Sugihara H, Fujimoto K, Weissman IL, Capecchi MR, Kuo CJ (2009) Sustained in vitro intestinal epithelial culture within a Wnt-dependent stem cell niche. Nat Med 15:701–706

Ota M, Ito K (2003) Induction of neurogenin-1 expression by sonic hedgehog: its role in development of trigeminal sensory neurons. Dev Dyn 227:544–551

Prakash N, Wurst W (2007) A Wnt signal regulates stem cell fate and differentiation in vivo. Neurodegener Dis 4:333–338

Sato H, Ishida S, Toda K, Matsuda R, Hayashi Y, Shigetaka M, Fukuda M, Wakamatsu Y, Itai A (2005) New approaches to mechanism analysis for drug discovery using DNA microarray data combined with KeyMolnet. Curr Drug Discov Technol 2:89–98

Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, Clevers H (2009) Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature 459:262–265

Sommer L, Ma Q, Anderson DJ (1996) Neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. Mol Cell Neurosci 8:221–241

Sun Y, Nadal-Vicens M, Misono S, Lin MZ, Zubiaga A, Hua X, Fan G, Greenberg ME (2001) Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. Cell 104:365–376

Tanese K, Fukuma M, Yamada T, Mori T, Yoshikawa T, Watanabe W, Ishiko A, Amagai M, Nishikawa T, Sakamoto M (2008) G-protein-coupled receptor GPR49 is up-regulated in basal cell carcinoma and promotes cell proliferation and tumor formation. Am J Pathol 173:835–843

Wu B, Crampton SP, Hughes CC (2007) Wnt signaling induces matrix metalloproteinase expression and regulates T cell transmigration. Immunity 26:227–239

Yamamoto Y, Sakamoto M, Fujii G, Tsuiji H, Kenetaka K, Asaka M, Hirohashi S (2003) Overexpression of orphan G-protein-coupled receptor, Gpr49, in human hepatocellular carcinomas with β-catenin mutations. Hepatology 37:528–533

Cell Mol Neurobiol (2010) 30:641–652
DOI 10.1007/s10571-009-9489-9

ORIGINAL RESEARCH

TDP-43 Dimerizes in Human Cells in Culture

Yuki Shiina · Kunimasa Arima · Hiroko Tabunoki · Jun-ichi Satoh

Received: 8 October 2009 / Accepted: 16 December 2009 / Published online: 31 December 2009

Abstract TAR DNA-binding protein-43 (TDP-43) is a 43-kDa nuclear protein involved in regulation of gene expression. Abnormally, phosphorylated, ubiquitinated, and aggregated TDP-43 constitute a principal component of neuronal and glial cytoplasmic and nuclear inclusions in the brains of frontotemporal lobar degeneration with ubiquitin-positive inclusions (FTLD-U) and amyotrophic lateral sclerosis (ALS), although the molecular mechanism that triggers aggregate formation remains unknown. By Western blot analysis using anti-TDP-43 antibodies, we identified a band with an apparent molecular mass of 86-kDa in HEK293, HeLa, and SK-N-SH cells in culture. It was labeled with both N-terminal-specific and C-terminal-specific TDP-43 antibodies, enriched in the cytosolic fraction, and the expression levels were reduced by TDP-43 siRNA but unaltered by treatment with MG-132 or by expression of ubiqulin-1 or casein kinase-1. By immunoprecipitation analysis, we found the interaction between the endogenous full-length TDP-43 and the exogenous Flag-tagged TDP-43, and identified the N-terminal half of TDP-43 spanning amino acid residues 3–183 as an intermolecular interaction domain. When the tagged 86-kDa tandemly connected dimer of TDP-43 was overexpressed in HEK293, it was sequestered in the cytoplasm and promoted an accumulation of high-molecular-mass TDP-43-immunoreactive proteins. Furthermore, the 86-kDa band was identified in the immunoblot of human brain tissues, including those of ALS. These results suggest that the 86-kDa band represents dimerized TDP-43 expressed constitutively in normal cells under physiological conditions.

Keywords Dimerization · Immunoprecipitation · Seed · TDP-43

Abbreviations

TDP-43	TAR DNA-binding protein-43
FTLD-U	Frontotemporal lobar degeneration with ubiquitin-positive inclusions
ALS	Amyotrophic lateral sclerosis
RRM	RNA-recognition motif
CSNK1A1	Casein kinase-1 alpha-1
UBQLN1	Ubiqulin-1
PARP	PolyADP ribose-polymerase

Introduction

TAR DNA-binding protein-43 (TDP-43) is a 43-kDa nuclear protein encoded by the TARDBP gene on chromosome 1p36.22, originally identified as a transcriptional repressor of the human immunodeficiency virus (HIV) gene (Ou et al. 1995). TDP-43, capable of interacting with UG and TG repeat stretches of RNA and DNA (Buratti and Baralle 2008). It plays a role in regulation of exon exclusion and inclusion of target genes during alternative splicing events, thereby being involved in cell division, apoptosis, mRNA stability, and microRNA biogenesis (Wang et al. 2008).

Y. Shiina · H. Tabunoki · J. Satoh (✉)
Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, 2-522-1 Noshio Kiyose, Tokyo 204-8588, Japan
e-mail: satoj@my-pharm.ac.jp

K. Arima
Department of Psychiatry, National Center Hospital, National Center of Neurology and Psychiatry, Tokyo 187-8551, Japan

TDP-43 is highly conserved through evolution from human to *Caenorhabditis elegans*, suggesting a phylogenetically pivotal role (Ayala et al. 2005). In its protein structure, TDP-43 is composed of an N-terminal domain and two highly conserved RNA-recognition motifs named RRM1 and RRM2, followed by a glycine-rich C-terminal domain that mediates the interaction of TDP-43 with heterogeneous ribonucleoproteins (Buratti and Baralle 2008; Wang et al. 2008). The RRM1 domain is necessary and sufficient for recognition of UG and TG repeat stretches of nucleic acids, while the C-terminal domain plays an essential role in regulation of splicing (Ayala et al. 2005; Buratti et al. 2005). In normal cells under physiological conditions, more than 90% of total TDP-43 proteins are accumulated in the nucleus, enriched in nuclear bodies, where TDP-43 coexists with survival motor neuron (SMN) and fragile X mental retardation (FMR) proteins, whereas very small amounts are located in the cytoplasm (Wang et al. 2008).

Abnormally, phosphorylated and ubiquitinated TDP-43 constitutes a principal component of neuronal cytoplasmic inclusions (NCIs), dystrophic neurites (DNs), neuronal intranuclear inclusions (NIIs), and glial cytoplasmic inclusions (GCIs), in the brains of frontotemporal lobar degeneration with ubiquitin-positive inclusions (FTLD-U) and amyotrophic lateral sclerosis (ALS) (Arai et al. 2006; Neumann et al. 2006). In view of overlapping clinicopathological features, both FTLD-U and ALS are categorized into a novel disease entity named TDP-43 proteinopathy (Geser et al. 2009). In TDP-43 proteinopathy, TDP-43 protein often translocates from the nucleus to the cytoplasm by forming detergent-insoluble urea-soluble aggregates, where it is hyperphosphorylated, polyubiquitinated, and proteolytically cleaved to produce 25- and 35-kDa C-terminal fragments (Neumann et al. 2006; Zhang et al. 2007; Hasegawa et al. 2008). Furthermore, abnormal TDP-43 immunoreactivity is occasionally found in the brains of Alzheimer disease (AD), dementia with Lewy bodies (DLB), Pick disease (PiD), corticobasal degeneration (CBD), argyrophilic grain disease (AGD), the Guam parkinsonism-dementia complex (G-PDC), and Huntington disease (HD) (Geser et al. 2009). The C-terminal domain of TDP-43 contains multiple phosphorylation consensus sites, among which the major phosphorylated epitopes are created by casein kinase-1 (CK1) (Kametani et al. 2009). Out of them, phosphorylation of Ser409/410 on TDP-43 is the pathological hallmark of certain sporadic and familial ALS cases (Neumann et al. 2009). Hyperphosphorylation of TDP-43 promotes oligomerization and fibril formation in vitro (Hasegawa et al. 2008). Importantly, missense mutations expressing mutant proteins with an increased aggregation property are clustered in the C-terminal domain of the TDP-43 gene in the patients with sporadic and familial ALS (Kabashi et al. 2008).

At present, the precise molecular events that trigger aggregate formation of TDP-43 remain to be characterized. In this study, by Western blot analysis, we identified a small amount of the TDP-43-immunoreactive 86-kDa protein constitutively expressed in HEK293, HeLa, and SK-N-SH cells in culture and human brain tissues in vivo. We suppose that this 86-kDa protein represents dimerized TDP-43.

Methods

Human Cell Lines and Brain Tissues

Human cell lines, such as SK-N-SH neuroblastoma, HeLa cervical carcinoma, and HEK293 embryonic kidney cells, were maintained in the culture medium consisting of DMEM (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS), 100-U/ml penicillin, and 100-μg/ml streptomycin. In some experiments, the cells were exposed for 24 h to 1-μM MG-132 (Calbiochem, San Diego, CA), a proteasome inhibitor. Human brain tissues of the cerebrum (CBR) and the cerebellum (CBL) were provided by Research Resource Network (RRN), Japan. They include a 29-year-old woman with secondary progressive MS (MS#1), a 40-year-old woman with secondary progressive MS (MS#2), a 43-year-old woman with primary progressive MS (MS#3), a 76-year-old woman with PD (PD#1), a 61-year-old woman with ALS (ALS#1), a 74-year-old woman with ALS (ALS#2), a 61-year-old man with ALS (ALS#3), a 66-year-old man with ALS (ALS#4), a 73-year-old man with schizophrenia (SCH#1), and a 77-year-old woman with depression (DEP#1). The postmortem interval of the cases ranges from 1.5 to 10 h prior to freezing the brain tissues. All autopsies were performed at the National Center Hospital, National Center of Neurology and Psychiatry (NCNP), Tokyo, Japan. This study was approved by the Ethics Committee of NCNP. Written informed consent was obtained form all the autopsy cases examined.

Western Blot Analysis

To prepare total protein extract, the cells and tissues were homogenized in either M-PER protein extraction buffer (Pierce, Rockford, IL) or RIPA buffer (Sigma, St. Louis, MO), supplemented with a cocktail of protease inhibitors (Sigma). The cell and tissue lysate were centrifuged at 12,000 rpm for 5 min at room temperature (RT). The mixture of the supernatant and a ×2 Lammeli loading buffer was boiled and separated on a 10 or 12% SDS–PAGE gel. The molecular weight of the proteins was calculated by the position of a broad-range SDS–PAGE standard (BioRad,

Hercules, CA). The protein concentration was determined by a Bradford assay kit (BioRad). After gel electrophoresis, the protein was transferred onto nitrocellulose membranes, and immunolabeled at RT overnight with rabbit polyclonal anti-TDP-43 antibody that recognizes amino acid residues 1–260 located at the N-terminal half of the human TDP-43 protein (1:10,000; 10782-2-AP; Proteintech Group, Chicago, IL) or rabbit polyclonal anti-TDP-43 antibody that recognizes amino acid residues 350–414 located at the C-terminus of the human TDP-43 protein (1:500; NB110-55376; Novus Biologicals, Littleton, CO). Then, the membranes were incubated at RT for 30 min with HRP-conjugated anti-rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA). The specific reaction was visualized by exposing the membranes to a chemiluminescent substrate (Pierce). In some experiments, the antibodies were stripped by incubating the membranes at 50°C for 30 min in stripping buffer, composed of 62.5-mM Tris–HCl, pH 6.7, 2% SDS, and 100-mM 2-mercaptoethanol. Then, the membranes were processed for relabeling with mouse monoclonal anti-ubiquitin antibody (P4D1; Santa Cruz Biotechnology), rabbit polyclonal anti-polyADP ribose-polymerase (PARP) antibody (Roche Diagnostics, Tokyo, Japan), rabbit anti-Halo tag antibody (Promega, Madison, WI), goat polyclonal anti-HSP60 antibody (N-20; Santa Cruz Biotechnology) for an internal control of protein loading, anti-Xpress antibody (Invitrogen), or anti-V5 antibody (Invitrogen).

Fractionation of Cellular Proteins

To determine subcellular location of dimeric TDP-43 proteins, we performed the differential extraction of native proteins using the ProteoExtract subcellular Proteome Extraction kit (Calbiochem, San Diego, CA). Then, the fractionated proteins from cytosol, membrane, nuclear, and cytoskeletal compartments were processed for Western blot with anti-TDP-43 antibody (10782-2-AP). The blots were relabeled with goat polyclonal anti-HSC70 antibody (K19; Santa Cruz Biotechnology), rabbit polyclonal anti-pan-cadherin antibody (RB-9036; Thermo Fisher Scientific, Fremont, CA), mouse monoclonal anti-vimentin antibody (V9; Santa Cruz Biotechnology), and mouse monoclonal anti-histone H1 antibody (SPM256; AnaSpec, San Jose, CA).

Vector Construction

To study the molecular interaction between TDP-43 proteins, the genes coding for the full-length (FL) TDP-43 (GenBank Accession No. NM_007375) and a panel of truncated forms of TDP43 (Fig. 1) or GFP were amplified by PCR using PfuTurbo DNA polymerase (Stratagene, La Jolla, CA) and the sense and the antisense primer sets listed in Table 1. After digesting the PCR products with restriction enzymes KpnI, XbaI, XhoI, and NotI (New England BioLabs, Beverly, MA), they were cloned in the expression vector p3XFLAG-CMV7.1 (Sigma), pFN21A-CMV Flexi (Promega), or pCMV-Myc (Clontech, Mountain View, CA) to express a fusion protein with an N-terminal Flag, Halo, or Myc tag. The vector containing the tandemly-connected dimer of TDP-43, tentatively named as TDP-43 tandem dimer, was constructed by tail-to-head ligation of two TDP-43 PCR products with distinct restriction-enzyme-digested ends, one having a C-terminal KpnI site and the other having an N-terminal KpnI site (Table 1). The siRNA vector constructs targeted to TDP-43 and a scrambled sequence (Table 1) were generated using GeneClip U1 Hairpin cloning system (Promega) following the manufacturer's instruction.

To investigate the role of casein kinase-1 alpha-1 (CSNK1A1) and ubiqulin-1 (UBQLN1) in dimer formation of TDP-43, the genes coding for CSNK1A1 (NM_001892), and UBQLN1 (NM_013438), were amplified by PCR using PfuTurbo DNA polymerase and the sense and the antisense primer sets listed in Table 1. Then, the PCR products were cloned in the expression vector pEF6/V5/His-TOPO (Invitrogen) or pCDNA4/HisMax-TOPO (Invitrogen) to express a fusion protein with a C-terminal V5 tag or an N-terminal Xpress tag.

All the vectors were transfected in the cells using Lipofectamine 2000 reagent (Invitrogen).

Immunoprecipitation

Immunoprecipitation analysis was performed according to the methods described previously (Satoh et al. 2009). 24–48 h after transfection of the vectors, HEK293 cells were homogenized in M-PER protein extraction buffer supplemented with a cocktail of protease inhibitors (Sigma). The protein extract was incubated at 4°C overnight with mouse monoclonal anti-Flag M2 affinity gel (Sigma). After several washes, the anti-Flag M2 affinity gel-binding proteins were eluted by incubating the gel with an exceeding amount of Flag peptide (Sigma). Then, the eluted proteins were precipitated by cold acetone. The immunoprecipitates were processed for Western blot with mouse monoclonal anti-Flag M2 antibody (Sigma) or rabbit polyclonal anti-TDP-43 antibody (10782-2-AP). Reciprocal coimmunoprecipitation analysis was performed according to the methods described previously (Satoh et al. 2009).

Cell Imaging

To visualize the subcellular location of TDP-43 in cultured cells, the genes encoding the monomer or the tandem dimer

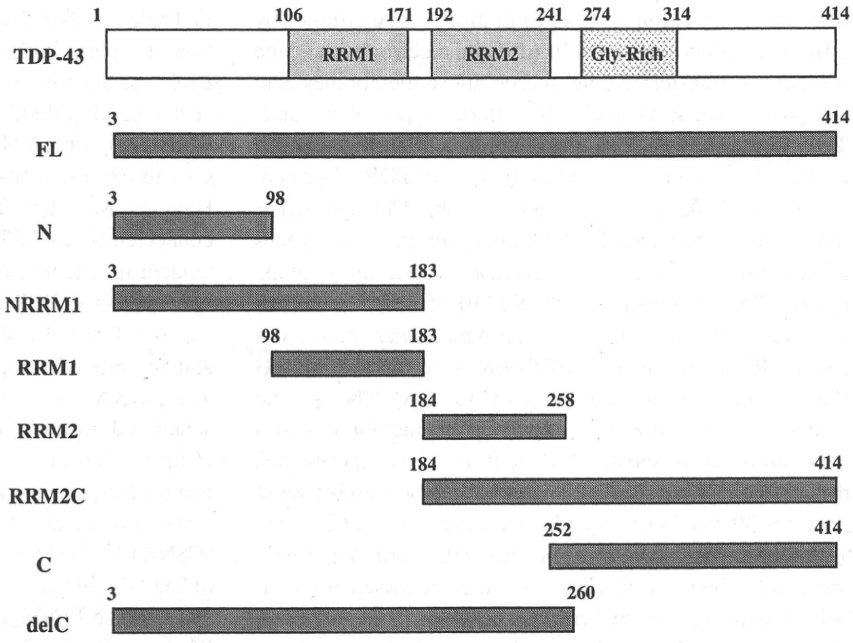

Fig. 1 The structural domains of the human TDP-43 protein. The domain structure of the human TDP-43 protein is shown on the *top* with the position of amino acid residues. The genes encoding the full-length (*FL*) TDP-43 protein and a panel of truncated forms, such as the C-terminal domain-cleaved protein (*delC*), the C-terminal domain fragment (*C*), the N-terminal domain fragment (*N*), *RRM1*, *RRM2*, the N-terminal domain plus RRM1 (*NRRM1*), and the C-terminal domain plus RRM2 (*RRM2C*), were cloned in the expression vectors as listed in Table 1

of TDP-43 were cloned in the vector pFN21A-CMV Flexi (Table 1), and transfected in HEK293 cells. 24 h after transfection, the live cells were exposed for 15 min to Oregon Green (Promega), a fluorochrome specifically bound to the Halo tag protein, fixed in 4% paraformaldehyde, and exposed to 4′,6′-diamidino-2-phenylindole (DAPI). They were mounted with glycerol-polyvinyl alcohol and examined under the Olympus BX51 universal microscope.

Results

Constitutive Expression of TDP-43 Dimer in Human Cell Cultures

By Western blot of the detergent-soluble protein extract with anti-TDP-43 antibody (10782-2-AP), HEK293, HeLa, and SK-N-SH cells in culture expressed constitutively a major 43-kDa protein representing the full-length (FL) TDP-43 protein, variable levels of 36- and 27-kDa proteins representing the N-terminally cleaved fragments, and a small but discernible amount of the 86-kDa protein (Fig. 2, panels a–c, lane 1). By densitometric analysis, the 86-kDa protein consisted of approximately 4% of total TDP-43-immunoreactive proteins in HEK293 cells, where minimum 50 μg of protein is required for loading on the gel to detect the 86-kDa protein. A 24-h treatment of the cells with MG-132, an inhibitor of proteasome function, did not alter the expression levels of the 86-kDa protein (Fig. 2, panels a–c, lane 2).

Because the 86-kDa TDP-43-immunoreactive protein has not been reported previously, we assumed that it represents a fairly small fraction of the endogenous FL TDP-43 dimer, based on its molecular weight. First, to characterize this protein, we concentrated both 86- and 43-kDa proteins extracted from the corresponding SDS–PAGE gel bands. They were then processed for Western blot analysis using anti-TDP-43 antibody (10782-2-AP) that reacts with amino acid residues 1–260 located at the N-terminal half or anti-TDP-43 antibody (NB110-55376) that reacts with amino acid residues 350–414 located at the C-terminus of the human TDP-43 protein. Both antibodies recognized the 86-kDa protein, along with the 43-kDa protein (Fig. 3A, panels a and b, lanes 1 and 2). In contrast, the exclusion of the primary antibodies did not react with either (Fig. 3A, panel c, lanes 1 and 2). These results argue against the possibility that the 86-kDa protein reflects a protein of non-TDP43 origin that simply exhibits cross-reactivity to TDP-43 on the blot. Furthermore, transient expression of the siRNA vector targeted specifically to TDP-43 but not to a scrambled sequence substantially reduced the expression of both 43- and 86-kDa TDP-43 immunoreactive proteins, suggesting that the 86-kDa protein is composed of TDP-43 (Fig. 3B, panels a and b, lanes 2 and 3).

Overexpression of ubiqulin-1 (UBQLN1), a proteasome-targeting factor that promotes the formation of cytoplasmic aggregations of TDP-43 (Kim et al. 2008) or casein kinase-1 alpha-1 (CSNK1A1), a protein kinase that mediates phosphorylation of major serine epitopes in the C-terminal domain of TDP-43 (Kametani et al. 2009), did not change the levels of expression of the 86-kDa protein in HEK293

Table 1 Primers utilized for PCR-based cloning in this study

Genes	Cloning vectors	Restriction enzyme sites for cloning	Sense primers	Antisense primers
TDP-43 (the full-length: amino acid residues 3-414; FL)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtacccctgtctgaatatattcgggta3′	5′gctctagagcctacattccccagccagaag3′
TDP-43 (deletion of the C-terminus: amino acid residues 3-260; delC)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtacccctgtctgaatatattcgggta3′	5′gctctagagcctaggcattggatatatgaacgct3′
TDP-43 (the C-terminus: amino acid residues 252–414; C)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtaccccaaggaatcagcgttcatatatcc3′	5′gctctagagcctacattccccagccagaag3′
TDP-43 (the N-terminus: amino acid residues 3–98; N)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtacccctgtctgaatatattcgggta3′	5′gctctagagcctatcttttcactttcactgctga3′
TDP-43 (the RRM1 domain: amino acid residues 98–183; RRM1)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtaccccgagcagtccagaaaacatcc3′	5′gctctagagcctagctttgcttagaatttg3′
TDP-43 (the RRM2 domain: amino acid residues 184–258; RRM2)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtaccccgccaagatgagcctttgaga3′	5′gctctagagcctaggatatatgaacgctga3′
TDP-43 (the N-terminus plus the RRM1 domain: amino acid residues 3–183; NRRM1)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtacccctgtctgaatatattcgggta3′	5′gctctagagcctagctttgcttagaatttg3′
TDP-43 (the RRM2 domain plus the C-terminus: amino acid residues 184–414; RRM2C)	p3XFLAG-CMV7.1	KpnI, XbaI	5′ggggtaccccgccaagatgagcctttgaga3′	5′gctctagagcctacattccccagccagaag3′
TDP-43 (the full-length monomer: amino acid residues 1–414; MM)	pFN21A-CMV Flexi	SgfI, PmeI	5′cgaagcgatcgccatgtctgaatatattcgggtaa3′	5′tgtcgtttaaaccattccccagccagaagact3′
TDP-43 (the full-length tandem dimer: amino acid residues 1–414 ×2; DM)	pFN21A-CMV Flexi	SgfI, KpnI	5′cgaagcgatcgccatgtctgaatatattcgggtaa3′	5′cggggtaccccgcattccccagccagaagactt3′
		KpnI, PmeI	5′cggggtaccccgtctgaatatattcgggtaacc3′	5′tgtcgtttaaaccattccccagccagaagact3′
TDP-43 (the full-length: amino acid residues 2–414; FL)	pCMV-Myc	XhoI, NotI	5′ccgctcgagcggtgtctgaatatattcgggta3′	5′atagtttagcggccgcattcttatctacattccccagccagaaga3′
GFP	p3XFLAG-CMV7.1	EcoRI, KpnI	5′cggaattccgccagcaaaggagaagaactt3′	5′ggggtacccctttatttgtagagctcatcca3′
GFP	pFN21A-CMV Flexi	SgfI, PmeI	5′taaagcgatcgccatggccagcaaaggagaaga3′	5′cgaggtttaaactttgtagagctcatccatgcca3′
CSNK1A1	pEF6/V5/His-TOPO	Not necessary	5′gggatggcgagtagcagcggctccaag3′	5′gaaacctttcatgttactcttggt3′
UBQLN1	pCDNA4/HisMax-TOPO	Not necessary	5′gccgagagtggtgaaagcggcggt3′	5′ctatgatggctgggagcccagtaa3′

Genes	Cloning vectors	Restriction enzyme sites for cloning	Sense oligonucleotides	Antisense oligonucleotides
siRNA of TDP-43	pGeneClip Hygromycin	Not necessary	5′tctcggaatcagcgttcatatatcttcctgtcaatatatgaacgctgattccct3′	5′ctgcagggaatcagcgttcatatattgacaggaagatatatgaacgctgattcc3′
Scrambled sequence	pGeneClip Hygromycin	Not necessary	5′tctcgctaggtacctttgaaaatcttcctgtcaattttcaaaggtacctagcct3′	5′ctgcaggctaggtacctttgaaaattgacaggaagattttcaaaggtacctagc3′

See Fig. 1 for the structure of TDP-43 domains cloned in the vectors

TDP-43 TAR DNA-binding protein-43, *RRM* RNA recognition motif, *GFP* green fluorescent protein, *CSNKA1* casein kinase-1 alpha-1, *UBQLN1* ubiquilin-1

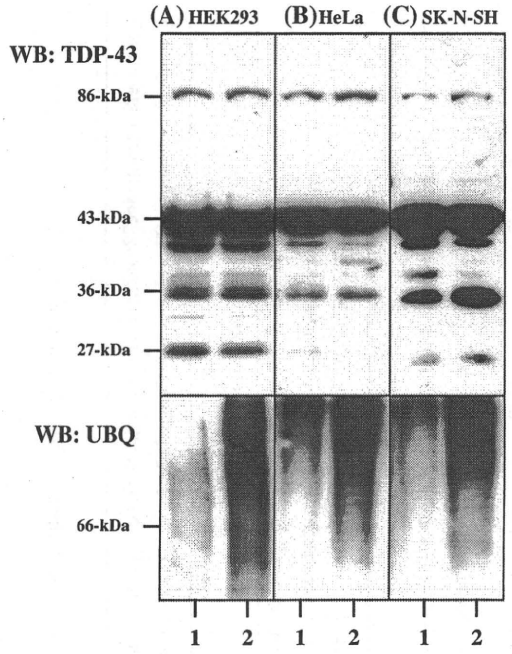

Fig. 2 The constitutive expression of an 86-kDa TDP-43-immunoreactive protein in human cell lines in culture. The detergent-soluble protein extract of HEK293, HeLa, and SK-N-SH cells exposed for 24 h to the vehicle (dimethylsulfoxide) (*lane 1*) or 1-μM MG-132 (*lane 2*) was processed for Western blot with anti-TDP43 antibody (*upper panels*) or with anti-ubiquitin antibody (*lower panels*). A small but discernible amount of the 86-kDa TDP-43-immunoreactive band is expressed in all the cells examined

cells (Fig. 3C, upper panel a, lane 3; upper panel b, lane 6). Furthermore, either a 48-h exposure of the cells to 10-nM okadaic acid, an inhibitor of protein phosphatases 1 and 2A, or a 48-h treatment with 20-μM hydrogen peroxide, an oxidative stress inducer did not alter the levels of expression of the 86-kDa protein (not shown). The use of different extraction buffers, either the M-PER protein extraction buffer or the RIPA buffer for protein extraction or the inclusion of 20-mM dithiothreitol (DTT) or an increased amount of 2-mercaptoethanol (2-ME) that reduces a typical disulfide bond in the sample-loading buffer did not affect the levels of expression of the 86-kDa protein, suggesting that the intermolecular interaction is fairly a detergent-resistant (not shown).

Next, we conducted immunoprecipitation analysis. After Flag-tagged FL TDP-43 or Flag-tagged GFP was expressed in HEK293, HeLa, and SK-N-SH cells, the protein extract was processed for immunoprecipitation with anti-Flag affinity gel. Then, the precipitates were eluted and processed for Western blot with anti-TDP-43 antibody (10782-2-AP). The proteins binding to Flag-tagged TDP-43 always included the endogenous FL TDP-43 protein in addition to several truncated species (Fig. 4A, upper panels a–c, lane 1). We identified multimeric forms of Flag-tagged TDP-43 proteins, when the immunoprecipitates were overloaded on the gel (Fig. 4B, lanes 1 and 2). In contrast, the precipitates of Flag-tagged GFP-binding proteins were completely devoid of the endogenous TDP-43 protein (Fig. 4A, upper panels a–c, lane 2). Furthermore, reciprocal immunoprecipitation analysis verified that Flag-tagged TPD-43 interacts with Myc-tagged TDP-43, excluding non-specific binding of TDP-43 to immunoglobulins (Fig. 4C, upper and lower panels, lanes 1–3). These results indicate that a substantial part of endogenous TDP-43 protein interacts with the exogenous Flag-tagged TDP-43 but not with the exogenous Flag-tagged GFP, suggesting that TDP-43 intrinsically forms the dimer, and TDP-43 forms a specific molecular interaction with TDP-43. Taken all these results together, we concluded that the 86-kDa protein represents a dimer of TDP-43.

By fractionation analysis of total cellular proteins of HEK293, the 86-kDa protein was highly enriched in the cytosolic fraction but absent in the nuclear fraction, while the 43-kDa FL endogenous TDP-43 protein was distributed in all the compartments, including cytosol, membrane, nuclear, and cytoskeletal fractions (Fig. 5, panel a, lanes 1–5).

Involvement of N-Terminal Half in Intermolecular Interaction of TDP-43 Proteins

To identify the domains involved in the intermolecular interaction of TDP43 proteins, a panel of Flag-tagged truncated proteins (Fig. 1), including the C-terminal fragment spanning amino acid residues 252–414 (C), the N-terminal fragment spanning amino acid residues 3–98 (N), the RRM1 domain spanning amino acid residues 98–183 (RRM1), the RRM2 domain spanning amino acid residues 184–258 (RRM2), the N-terminal domain plus RRM1 spanning amino acid residues 3–183 (NRRM1), the C-terminal domain plus RRM2 spanning amino acid residues 184–414 (RRM2C), and the C-terminal domain-cleaved protein spanning amino acid residues 3–260 (delC) were individually expressed in HEK293 cells. Then, the protein extract was processed for immunoprecipitation with anti-Flag affinity gel, followed by Western blot with anti-TDP-43 antibody (10782-2-AP). The anti-TDP-43 antibody (10782-2-AP) reacted with FL TDP-43, delC, N, NRRM1, RRM2, and RRM2C, but not with RRM1 or C (Fig. 6, lanes 1, 3, 5, 9, 11, and 15). These results indicate that the 10782-2-AP antibody recognizes at least two separate epitopes located in N and RRM2 in the N-terminal half of TDP-43. After immunoprecipitation, the endogenous FL TDP-43 protein was detected exclusively in the precipitates of Flag-tagged FL TDP-43, delC, and NRRM1 fragments (Fig. 6, lanes 1, 3, and 9). These results suggest that the continuous N-terminal half domain spanning amino acid

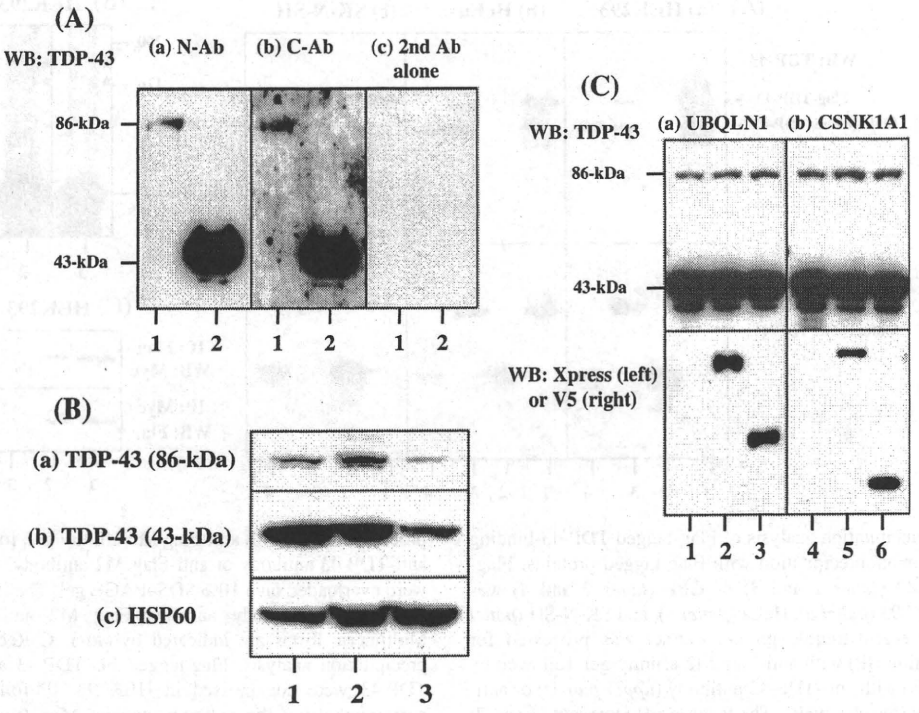

Fig. 3 The characterization of the 86-kDa protein. **A** The immunolabeling of the protein with N- or C-terminal-specific anti-TDP-43 antibodies. Both 86- and 43-kDa proteins were extracted from the corresponding SDS-PAGE gel bands, concentrated, and processed for Western blot with the N-terminal-specific anti-TDP-43 antibody (*panel a*), the C-terminal-specific anti-TDP-43 antibody (*panel b*), or the secondary antibody alone (*panel c*). The lanes (*1* and *2*) represent the 86- and the 43-kDa protein, respectively. **B** The effects of transient expression of TDP-43 siRNA. SK-N-SH cells were transfected with the siRNA expression vector targeted to TDP-43 or a scrambled sequence listed in Table 1. 96 h after transfection, the protein extract was processed for Western blot with anti-TDP-43 antibody (*panels a* and *b*) or anti-HSP60 antibody, an internal control for protein loading (*panel c*). The lanes (*1–3*) represent non-transfected cells, the cells transfected with the vector of a scrambled sequence, and the cells transfected with the vector of TDP-43 siRNA, respectively. **C** The effects of expression of ubiqulin-1 (UBQLN1) and casein kinase-1 alpha-1 (CSNK1A1). HEK293 cells were untransfected (*lanes 1* and *4*) or transfected with the expression vectors of Xpress-tagged LacZ (*lane 2*), Xpress-tagged UBQLN1 (*lane 3*), V5-tagged LacZ (*lane 5*), or V5-tagged CSNK1A1 (*lane 6*). The detergent-soluble protein extract was processed for Western blot with anti-TDP-43 antibody (*upper panels*), anti-Xpress antibody (*the left lower panel*), or anti-V5 antibody (*the right lower panel*)

residues 3–183, possibly by constructing a conformationally ordered interacting domain, but neither N nor RRM1 alone, is sufficient for the intermolecular interaction of TDP-43, while the C-terminal domain is dispensable for this.

TDP-43 Dimer Acts as a Seed for Promoting Aggregation of TDP-43 Proteins

To investigate a role of the TDP-43 dimer in TDP-43 protein aggregation, Halo-tagged TDP-43 FL monomer, tandem dimer, or GFP was expressed in HEK293 cells. Then, the blot was labeled with anti-TDP43 antibody (10782-2-AP), anti-Halo tag antibody, anti-PARP antibody, anti-ubiquitin antibody, or anti-HSP60 antibody. The expression of the TDP-43 tandem dimer but neither the monomer nor GFP greatly enhanced an accumulation of TDP-43-immunoreactive proteins with higher molecular mass ranging from 70- to 200-kDa (Fig. 7A, panel a, lane 5), while the levels of expression of Halo-tagged proteins were identical among GFP, the monomer, and the dimer (Fig. 7A, panel b, lanes 3–5). These results suggest that TDP-43 dimer by acting as a seed promotes protein aggregation that involves various endogenous species of TDP-43, including N-terminally cleaved fragments. Unexpectedly, untransfected and untreated HEK293 cells expressed constitutively both uncleaved (116-kDa) and cleaved (85-kDa) forms of PARP without morphological features of apoptosis (Fig. 7A, panel c, lanes 1–5). The expression of TDP-43 tandem dimer did not alter the levels of the cleaved PARP (Fig. 7A, panel c, lane 5) or did not elevate substantially the levels of ubiquitinated proteins, compared with the cells with expression of GFP or the monomer (Fig. 7A, panel d, lanes 3–5). In contrast, a 24-h treatment with MG-132 enhanced an accumulation of ubiquitinated proteins (Fig. 7A, panel d, lane 2). Importantly, the numbers of the cells presenting with nuclear

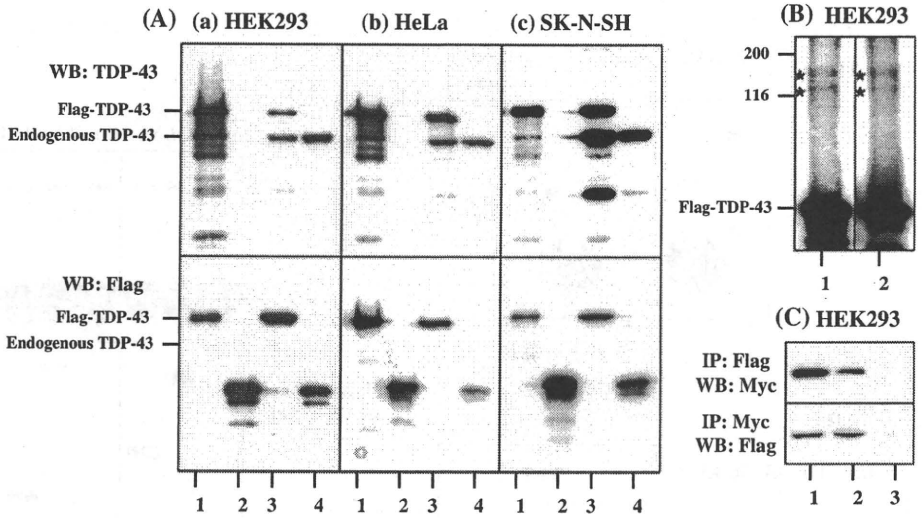

Fig. 4 Immunoprecipitation analysis of Flag-tagged TDP-43-binding proteins. **A** Coimmunoprecipitation with Flag-tagged proteins. Flag-tagged FL TDP-43 (*lanes 1* and *3*) or GFP (*lanes 2* and *4*) was expressed in HEK293 (*panel a*), HeLa (*panel b*), and SK-N-SH (*panel c*) cells. The detergent-soluble protein extract was processed for immunoprecipitation (IP) with anti-Flag M2 affinity gel, followed by Western blot (WB) with anti-TDP-43 antibody (*upper panels*) or anti-Flag M2 antibody (*lower panels*). The lanes (*1–4*) represent (*1* and *2*) the immunoprecipitates and (*3* and *4*) the corresponding input controls. **B** Flag-tagged TDP-43 constitutes multimeric forms. Flag-tagged FL TDP-43 was expressed in HEK293 cells, and then processed for IP with anti-Flag M2 affinity gel, followed by WB with anti-TDP-43 antibody or anti-Flag M2 antibody. Immunoprecipitates were overloaded on a 10% SDS–PAGE gel. The lanes (*1*, *2*) represent anti-TDP-43 antibody and anti-Flag M2 antibody, respectively. Multimeric forms are indicated by *stars*. **C** Reciprocal coimmunoprecipitation analysis. Flag-tagged FL TDP-43 and Myc-tagged FL TDP-43 were coexpressed in HEK293. IP followed by WB was performed using the antibody against Myc (*upper panel*) or Flag (*lower panel*), reciprocally. The lanes (*1–3*) represent input control, IP with anti-Flag M2 or anti-Myc antibody, and IP with normal mouse or rabbit IgG, respectively

accumulation of Halo tag immunoreactivity was significantly smaller in the tandem dimer-expressing HEK293 cells than those expressing the monomer ($P = 0.00008$, two-tailed Student's t test) (Fig. 7B, panels a–d and Fig. 7C).

Identification of TDP-43 Dimer in Human Brain Tissues

Finally, we conducted Western blot analysis of the detergent-soluble protein extract of human brain tissues with anti-TDP43 antibody (10782-2-AP). The expression of the 86-kDa TDP-43 dimer, along with the 43-kDa FL TDP-43 monomer and several cleaved fragments, was identified at varying levels in both the cerebrum (CBR) and the cerebellum (CBL) of the patients with PD, ALS, MS, schizophrenia, or depression (Fig. 8, panel a, lanes 1–15). Although the brain tissues derived from ALS patients constantly expressed higher levels of the TDP-43 dimer, we could not obtain the definite conclusion because of the limitation of the samples examined.

Discussion

By Western blot analysis, we identified the expression of the 86-kDa TDP-43-immunoreactive protein in HEK293, HeLa, and SK-N-SH cells in culture. We considered this protein as a dimer of TDP-43 for the following reasons. First, the molecular weight of the protein corresponds exactly to that of the dimer. Second, two different anti-TDP-43 antibodies that recognize distinct antigenic epitopes located in the N- or the C-terminal domain equally labeled the 86-kDa protein, excluding the possibility of immunological cross reactivity on a non-TDP-43 protein. Third, the levels of expression of the 86-kDa protein were greatly reduced by treatment with TDP-43 siRNA, suggesting that it is composed of TDP-43. Fourth, by immunoprecipitation analysis, we verified the interaction between the endogenous FL TDP-43 and the exogenous Flag-tagged TDP-43. We excluded non-specific binding of TDP-43 to immunoglobulins. We identified the N-terminal half spanning amino acid residues 3–183 as an intermolecular interacting domain. Finally, the 86-kDa protein was identified in human brain tissues of various neurological diseases, arguing against the possibility of an in vitro artificial byproduct. Nevertheless, we could not completely exclude the possibility that the 86-kDa protein represents a complex of TDP-43 with an unidentified 43-kDa protein. It is of particular interest that SDS-resistant dimers of DJ-1, whose loss of function causes PD, are accumulated in PD and AD brains (Choi et al. 2006).

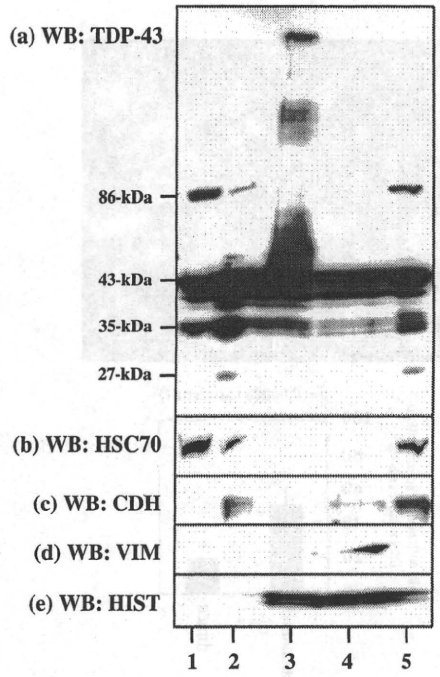

Fig. 5 Fractionation analysis of cellular proteins of HEK293. The cellular proteins of HEK293 separated into cytosol (*lane 1*), membrane (*lane 2*), nuclear (*lane 3*), and cytoskeletal (*lane 4*) fractions were processed for Western blot with anti-TDP-43 antibody (*panel a*), anti-HSC70 antibody (*panel b*), anti-pan-cadherin antibody (*panel c*), anti-vimentin antibody (*panel d*), and anti-histone H1 antibody (*panel e*). The lane (*5*) represents the unfractionated input control

Supporting our observations, a recent crystal structural study has clarified the self-assembling capacity of TDP-43 (Kuo et al. 2009). Another study showed that Flag-tagged FL TDP-43 protein binds to GFP-tagged C-terminally deleted products of TDP-43, suggesting a role of the N-terminal domain in the intermolecular interaction (Zhang et al. 2009). In contrast, a different study disclosed that the wild-type human TDP-43 protein has an intrinsically aggregation-prone propensity that requires the C-terminal domain (Johnson et al. 2009). Furthermore, DsRed-tagged FL TDP-43 is colocalized in cytoplasmic inclusions with GFP-tagged C-terminal fragments (Nonaka et al. 2009a). These observations suggest the possibility that the dimer formation of TDP-43 requires the N-terminal half, while afterward the aggregate formation involves the C-terminal domain as well.

In contrast to the normal nuclear location of TDP-43 under physiological conditions, the pathogenic TDP-43 is often redistributed in the neuronal and glial cytoplasm and dystrophic neuritis by forming insoluble aggregates (Arai et al. 2006; Neumann et al. 2006). Importantly, we found that the 86-kDa TDP-43 dimer was enriched in the cytosolic fraction in HEK293 cells. When the 86-kDa tandemly connected dimer of TDP-43 was overexpressed in HEK293, it also showed a trend for being distributed outside the nucleus, and promoted an accumulation of high-molecular-mass TDP-43-immunoreactive proteins. These observations propose the working hypothesis that a very small amount of TDP-43 dimer expressed constitutively in normal cells under physiological conditions might serve as a starting seed that triggers aggregation of posttranslationally modified TDP-43 species under pathological conditions of TDP-43 proteinopathy.

A previous study showed that TDP-43 continuously shuttles between the nucleus and the cytoplasm in a transcription-dependent manner (Ayala et al. 2008). Because TDP-43 plays a key role in shuttling of mRNA species in response to the neuronal injury (Moisse et al. 2009), even a trivial perturbation affecting the nucleocytoplasmic

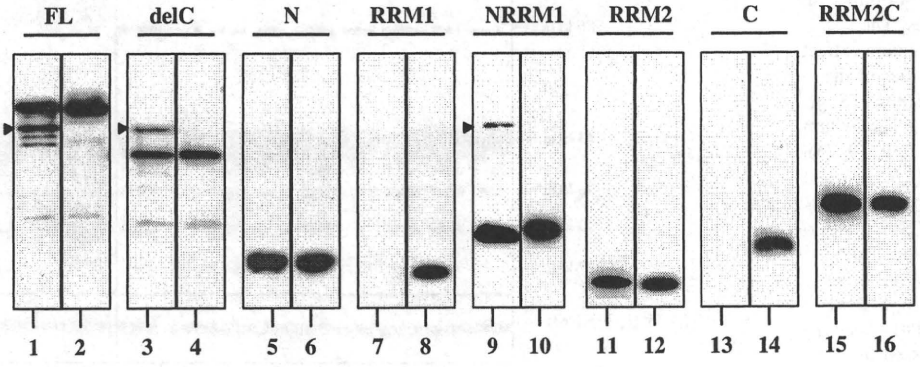

Fig. 6 The domains involved in the intermolecular interaction of TDP-43 proteins. Flag-tagged FL TDP-43 (*lanes 1* and *2*) and a panel of truncated proteins (see the details in Fig. 1), including delC (*lanes 3* and *4*), N (*lanes 5* and *6*), RRM1 (*lanes 7* and *8*), NRRM1 (*lanes 9* and *10*), RRM2 (*lanes 11* and *12*), C (*lanes 13* and *14*), and RRM2C (*lanes 15* and *16*) were individually expressed in HEK293 cells. The protein extract was processed for IP with anti-Flag affinity gel, followed by Western blot with anti-TDP-43 antibody (*lanes 1, 3, 5, 7, 9, 11, 13,* and *15*) or anti-Flag M2 antibody (*lanes 2, 4, 6, 8, 10, 12, 14,* and *16*). The position (43-kDa) of the immunoprecipitated endogenous FL TDP-43 is indicated by *arrow heads*