子の分子ネットワークを同定した．彼らはRORγt欠損(knockout：KO)および野生型(wild-type：WT)C57BL/6マウスのリンパ節と脾臓から細胞自動解析-分離装置(fluorescence activated cell sorter：FACS)で分離した$CD4^+CD8^-CD19^-CD25^-CD44^{low/int}CD62L^+$ナイーブT細胞を，プレートコートした抗CD3ε抗体と抗CD28抗体で刺激し，さらにIL-6，TGF-β，抗IFN-γ抗体と抗IL-4抗体を添加した培地(Th17-inducing condition)で48時間培養して，Th17細胞の分化を誘導した[11]．また培養系にdimethyl sulfoxide(DMSO)またはDMSOに溶解したdigoxin(DIG：10 μM)を添加し，Th17細胞分化に対する抑制効果を調べた．GSE27241は，各条件の細胞からRNAを精製し，Mouse Genome 430 2.0 Array(Affymetrix)で解析し，robust multiarray average(RMA)法で正規化したデータセットである(WT-DMSO，WT-DIG，KO-DMSO，KO-DIG各2サンプル)．

われわれはHuhらのデータに関して，はじめにWT-DMSO群とKO-DMSO群を比較し，Th17-inducing conditionにおいて前者で2倍以上発現上昇した57遺伝子(Th17細胞分化関連遺伝子群)を同定した(**表1**)．つぎにWT-DIG群とWT-DMSO群を比較し，前者で0.5倍以下に発現低下した12遺伝子を同定した(DIG応答性遺伝子群)(**表1下線**)．また57遺伝子を指標として階層クラスター解析をおこなったところ，WT-DMSO，KO-DMSO，WT-DIG，KO-DIGの各群はおのおの独立したクラスターを形成し，WT-DIG群はKO-DMSO群に近接したクラスターに分類された(**図2**)．DIGで発現抑制された12遺伝子のうち11遺伝子は，ほかの遺伝子から独立したクラスターを形成し，11遺伝子を含む16遺伝子は転写が共調節(co-regulation)されている可能性がある(**図2点線**)．KeyMolnetの共通上流検索により，これらの16遺伝子に関連する分子ネットワークを解析したところ，TGF-β受容体シグナル伝達系で中心的な役割を果たしている転写因子SMADによる発現調節の関与が最も強く示唆された(p＝1.194E-060)(**図3**)．

TGF-β受容体の活性化に伴ってリン酸化されたSMAD2とSMAD3は，SMAD4と複合体を形成して核へ運ばれ，種々のコアクチベーターやコリプレッサーと協調して，標的遺伝子の転写を制御する．SMAD2はTh17細胞の分化に必須であると報告されている[12]．DIGは，Huhらが4812種類の低分子化合物に関して，RORγt転写活性化抑制を指標にスクリーニングし，RORγtと結合してTh17細胞の分化を抑制することが判明した化合物である[11]．またRORγt，RORαと結合してTh17細胞の分化を抑制する合成リガンドSR1001も報告されている[13]．DIGは転写因子SMADが制御している分子ネットワーク上のTh17細胞分化関連遺伝子群(**表1下線**)の発現を共抑制している可能性がある．SMAD系転写因子は分子ネットワークのハブに位置し，MSにおけるTh17細胞分化制御薬の標的分子と成りうる．しかしながら，TGF-β受容体シグナル系を全般的に抑制すると，induced $Foxp3^+$ regulatory T(iTreg)細胞の分化まで抑制してしまう可能性があること，SMAD非依存性のTh17細胞分化経路も存在すること[14]，SMAD2，SMAD3は機能的に冗長にはたらくこと，SMAD3それ自体はRORγtと結合してその活性を抑制し，Th17細胞の分化を負に制御していること[15]などを考慮し，SMAD系転写因子を選択的かつ部分的に抑制できるような新薬の開発が望ましい．

おわりに

膨大なオミックスデータに関与する分子ネットワークを手際よく解析するためには，精査された文献情報にもとづく解析ツールを使う必要がある．解析ツールはいまだ発展途上かつ日進月歩であり，現時点では，どのツールもスプライスバリアントや翻訳後修飾，細胞特異的発現，細胞内局在化，

表 1. Th17 細胞分化関連 57 遺伝子

Entrez Gene ID	Gene Symbol	Gene Name	Ratio
70337	iyd	iodotyrosine deiodinase	11.06
16171	IL17A	interleukin 17A	4.30
76142	ppp1r14c	protein phosphatase 1, regulatory(inhibitor)subunit 14c	4.25
193740	Hspa1a	heat shock protein 1A	3.47
50929	il22	interleukin 22	3.39
15511	Hspa1b	heat shock protein 1B	3.38
56312	nupr1	nuclear protein 1	3.23
14538	GCNT2	glucosaminyl(N-acetyl)transferase 2, I-branching enzyme	2.79
74103	Nebl	nebulette	2.76
75573	2310007L24Rik	RIKEN cDNA 2310007L24 gene	2.55
68549	SGOL2	shugoshin-like 2(S. pombe)	2.53
237436	GAS2L3	growth arrest-specific 2 like 3	2.47
76131	depdc1a	DEP domain containing 1a	2.43
100043766	Gm14057	predicted gene 14057	2.37
14235	FOXM1	forkhead box M1	2.37
230098	E130306D19Rik	RIKEN cDNA E130306D19 gene	2.30
171284	Timd2	T-cell immunoglobulin and mucin domain containing 2	2.29
12235	BUB1	budding uninhibited by benzimidazoles 1 homolog(S. cerevisiae)	2.26
51944	D2Ertd750e	DNA segment, Chr 2, ERATO Doi 750, expressed	2.24
17863	myb	myeloblastosis oncogene	2.24
229841	CENPE	centromere protein E	2.22
270906	PRR11	proline rich 11	2.19
12316	ASPM	asp(abnormal spindle)-like, microcephaly associated(Drosophila)	2.18
108000	CENPF	centromere protein F	2.18
17345	MKI67	antigen identified by monoclonal antibody Ki 67	2.16
14432	gap43	growth associated protein 43	2.15
105988	ESPL1	extra spindle poles-like 1(S. cerevisiae)	2.15
15366	HMMR	hyaluronan mediated motility receptor(RHAMM)	2.15
27053	asnS	asparagine synthetase	2.15
52276	CDCA8	cell division cycle associated 8	2.15
18005	NEK2	NIMA(never in mitosis gene a)-related expressed kinase 2	2.15
72080	2010317E24Rik	RIKEN cDNA 2010317E24 gene	2.15
74107	CEP55	centrosomal protein 55	2.13
29817	igfbp7	insulin-like growth factor binding protein 7	2.13
71819	KIF23	kinesin family member 23	2.10
75317	4930547N16Rik	RIKEN cDNA 4930547N16 gene	2.10
12704	CIT	citron	2.10
72140	CCDC123	coiled-coil domain containing 123	2.08
234258	Neil3	nei like 3(E. coli)	2.08
12442	CCNB2	cyclin B2	2.07
72119	Tpx2	TPX2, microtubule-associated protein homolog(Xenopus laevis)	2.07
68743	Anln	anillin, actin binding protein	2.06
20419	SHCBP1	Shc SH2-domain binding protein 1	2.05
208084	PIF1	PIF1 5'-to-3'DNA helicase homolog(S. cerevisiae)	2.04
17279	Melk	maternal embryonic leucine zipper kinase	2.04
19348	kif20a	kinesin family member 20A	2.04
21335	TACC3	transforming, acidic coiled-coil containing protein 3	2.03
208628	KNTC1	kinetochore associated 1	2.02
19659	Rbp1	retinol binding protein 1, cellular	2.02
72155	CENPN	centromere protein N	2.02
257630	Il17f	interleukin 17F	2.02
215819	nhsl1	NHS-like 1	2.02
54141	SPAG5	sperm associated antigen 5	2.01
12189	BRCA1	breast cancer 1	2.01
19362	RAD51AP1	RAD51 associated protein 1	2.01
110033	Kif22	kinesin family member 22	2.00
69534	AVPI1	arginine vasopressin-induced 1	2.00

GSE27241 は RORγt 欠損(knockout：KO)および野生型(wild-type：WT)マウスナイーブ T 細胞を，Th17-inducing condition で 48 時間培養して Th17 細胞の分化を誘導し，培養系に DMSO に溶解した digoxin(DIG)を添加した細胞のトランクリプトームデータである．GSE27241 に関して，はじめに WT-DMSO と KO-DMSO を比較し，Th17-inducing condition において前者で 2 倍以上発現上昇した Th17 細胞分化関連 57 遺伝子を同定した．つぎに WT-DIG と WT-DMSO を比較し，前者で 0.5 倍以下に発現低下した DIG 応答性 12 遺伝子を同定した(下線)．

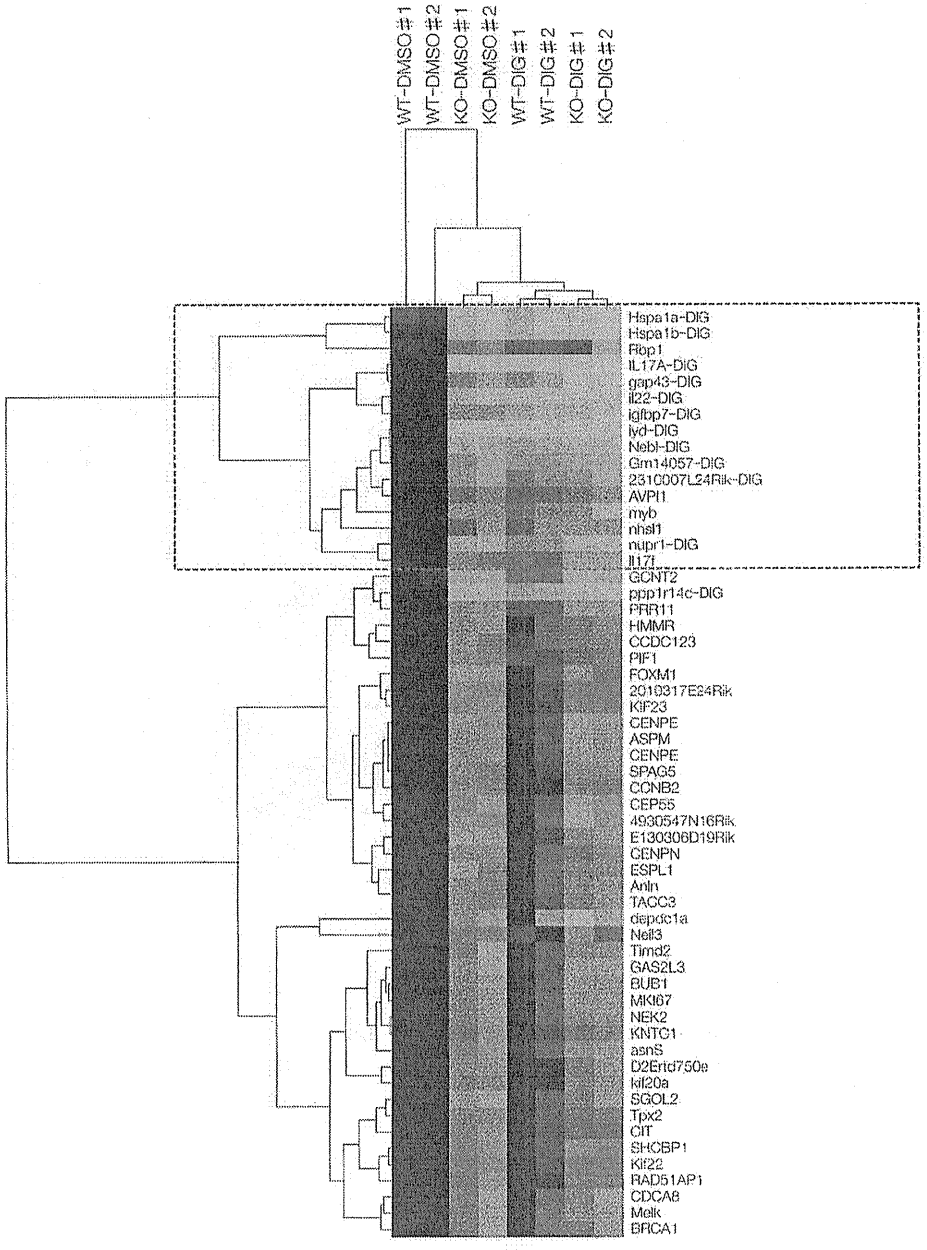

図 2. Th17 細胞分化関連 57 遺伝子の階層クラスター解析

GSE27241 は，RORγt 欠損(knockout：KO)および野生型(wild-type：WT)マウスナイーブ T 細胞を，Th17-inducing condition で 48 時間培養し，Th17 細胞の分化を誘導．培養系に DMSO に溶解した digoxin(DIG)を添加した細胞のトランクリプトームデータである．はじめに WT-DMSO と KO-DMSO を比較し，Th17 細胞分化関連 57 遺伝子を同定し，階層クラスター解析をおこなった．つぎに WT-DIG と WT-DMSO を比較し，DIG 応答性 12 遺伝子を同定した(図の遺伝子-DIG)．DIG 応答性 11 遺伝子は独立したクラスターを形成し，これらを含む 16 遺伝子は転写が共調節(co-regulation)されている可能性がある(点線)．

動的な特性に関しては十分対応できていない．しかしながら生命現象を複雑なシステムとして捉えるシステム生物学的観点からすると，分子ネットワークを詳細に解析することにより，はじめて論理的な仮説に裏づけられた創薬標的分子を効率的に同定することができる．

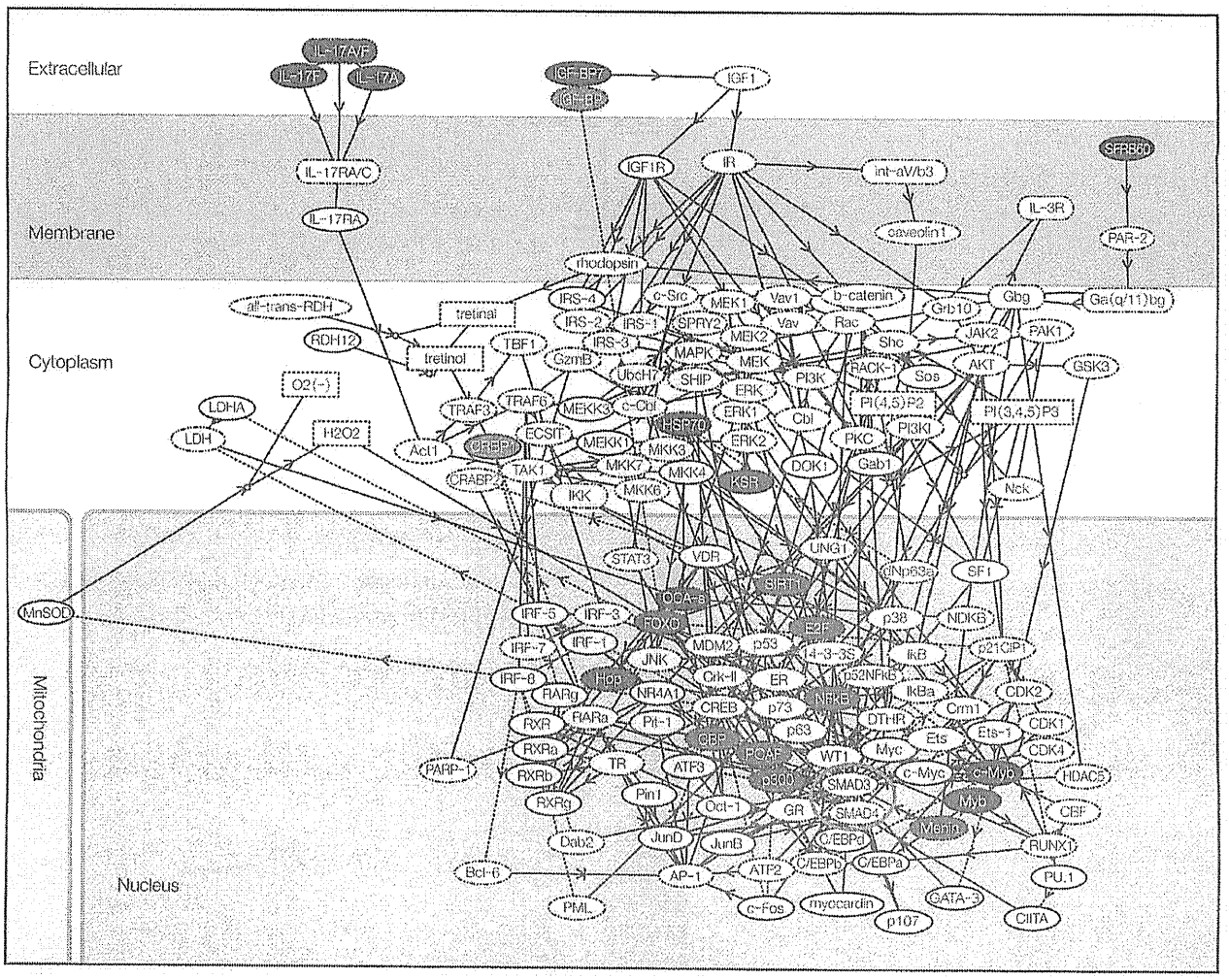

図 3. DIG 応答性 Th17 細胞分化関連遺伝子の分子ネットワーク解析
KeyMolnet 共通上流検索(コアコンテンツ：発現制御・直接結合・複合体形成)により，DIG 応答性 11 遺伝子のクラスターの 16 遺伝子(図 2)に関連する分子ネットワークを解析した．転写因子 SMAD(楕円)による発現調節の関与が示唆された．

謝辞

本稿で紹介した研究は，国立精神・神経医療研究センター神経研究所免疫研究部 山村隆部長，明治薬科大学バイオインフォマティクス 天竺桂弘子助教との共同研究でなされ，文部科学省基盤研究(C22500322)と私立大学戦略的研究基盤形成支援事業明治薬科大学ハイテクリサーチセンター研究事業(S0801043)および厚生労働科学難治性疾患克服研究事業(H21-難治-一般-201：H22-難治-一般-136)の補助を受けた．

文 献

1) Satoh J：Bioinformatics approach to identifying molecular biomarkers and networks in multiple sclerosis. *Clin Exp Neuroimmunol* 1：127-140, 2010
2) Han MH *et al*：Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature* **451**：1076-1081, 2008
3) Satoh JI *et al*：Molecular network of the comprehensive multiple sclerosis brain-lesion proteome. *Mult Scler* **15**：531-541, 2009
4) van Horssen J *et al*：The extracellular matrix in multiple sclerosis pathology. *J Neurochem* **103**：1293-1301, 2007
5) Milner R *et al*：Fibronectin- and vitronectin-induced microglial activation and matrix metallo-

proteinase-9 expression is mediated by integrins $\alpha_5\beta_1$ and $\alpha_v\beta_5$. *J Immunol* **178** : 8158-8167, 2007
6) Liu TJ *et al* : Inhibition of both focal adhesion kinase and insulin-like growth factor-I receptor kinase suppresses glioma proliferation in vitro and in vivo. *Mol Cancer Ther* **6** : 1357-1367, 2007
7) McFarland HF *et al* : Multiple sclerosis : a complicated picture of autoimmunity. *Nat Immunol* **8** : 913-919, 2007
8) Durelli L *et al* : T-helper 17 cells expand in multiple sclerosis and are inhibited by interferon-β. *Ann Neurol* **65** : 499-509, 2009
9) Tzartos JS *et al* : Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. *Am J Pathol* **172** : 146-155, 2008
10) Axtell RC *et al* : T helper type 1 and 17 cells determine efficacy of interferon-β in multiple sclerosis and experimental encephalomyelitis. *Nat Med* **16** : 406-412, 2010
11) Huh JR *et al* : Digoxin and its derivatives suppress T_H17 cell differentiation by antagonizing RORγt activity. *Nature* **472** : 486-490, 2011
12) Malhotra N *et al* : SMAD2 is essential for TGFβ-mediated Th17 cell generation. *J Biol Chem* **285** : 29044-29048, 2010
13) Solt LA *et al* : Suppression of T_H17 differentiation and autoimmunity by a synthetic ROR ligand. *Nature* **472** : 491-494, 2011
14) Lu L *et al* : Role of SMAD and non-SMAD signals in the development of Th17 and regulatory T cells. *J Immunol* **184** : 4295-4306, 2010
15) Martinez GJ *et al* : Smad3 differentially regulates the induction of regulatory and inflammatory T cell differentiation. *J Biol Chem* **284** : 35283-35286, 2009

Second Version, Submitted on October 11, 2011

The Book Entitled "Biomedical Engineering and Cognitive Neuroscience for Healthcare: Interdisciplinary Applications" Published by IGI Global

The Section: Complex Bioinformatics and Healthcare

Title:

Molecular network analysis of target RNAs and interacting proteins of TDP-43, a causative gene for neurodegenerative diseases ALS/FTLD

Jun-ichi Satoh

Correspondence to Jun-ichi Satoh, MD, PhD, Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, 2-522-1 Noshio, Kiyose, Tokyo 204-8588, Japan.

Tel.: +81-42-495-8678

e-mail: satoj@my-pharm.ac.jp

The manuscript is composed of the text (33 pages), four figures, and two tables.

Total number of the words is 6,755.

I recommend that Figure 1 and Figure 2 are suitable for those occupying the half page, and Figure 3 and Figure 4 are suitable for those occupying the full page, individually, in view of their resolution of the images.

ABSTRACT

TAR DNA-binding protein-43 (TDP-43) is an evolutionarily conserved nuclear protein that regulates gene expression by forming a multimolecular complex with a wide variety of target RNAs and interacting proteins. Abnormally phosphorylated, ubiquitinated, and aggregated TDP-43 proteins constitute a principal component of neuronal and glial cytoplasmic and nuclear inclusions in the brains of amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD), establishing a novel clinical entity designated as TDP-43 proteinopathy. Although increasing evidence suggests that the neurodegenerative process underlying ALS and FTLD is attributable to a toxic gain of function or a loss of cellular function of TDP-43, the precise molecular mechanisms remain largely unknown. Recent advances in systems biology enable us to characterize the global molecular network extracted from large-scale data of the genome, transcriptome and proteome by using pathway analysis tools of bioinformatics endowed with comprehensive knowledgebase. The present study is conducted to characterize the comprehensive molecular network of TDP-43 target RNAs and interacting proteins, recently identified by deep sequencing with next-generation sequencers and mass spectrometric analysis. Our results propose the systems biological view that TDP-43 serves as a molecular coordinator of RNA-dependent regulation of gene transcription and translation pivotal for achievement of diverse neuronal functions, and the disruption of TDP-43-mediated molecular coordination induces neurodegeneration in ALS and FTLD.

INDEX WORDS: ALS; FTLD; KEGG; KeyMolnet; IPA; Molecular Network; Systems Biology; TDP-43

INTRODUCTION

TDP-43 is a nuclear RNA/DNA-binding protein highly conserved through evolution, 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/TG repeat stretches of target RNAs/DNAs, plays a key role in regulation of transcription, alternative splicing, mRNA stability and transport, and microRNA biogenesis (Buratti and Baralle, 2010; Lagier-Tourenne et al., 2010). Structurally, it 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 (Lagier-Tourenne and Cleveland, 2009). The RRM1 domain is necessary and sufficient for recognition of UG repeats of target RNAs, while the C-terminal domain that mediates the protein-protein interaction plays an essential role in regulation of splicing (Ayala et al., 2005; Buratti et al., 2005). In normal cells under physiological conditions, the vast majority of TDP-43 proteins are located in the nucleus, highly enriched in nuclear bodies.

Accumulating evidence indicates that abnormally phosphorylated, ubiquitinated, and aggregated TDP-43 proteins constitute a principal component of neuronal and glial cytoplasmic and nuclear inclusions in the brains of amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration with ubiquitin-positive inclusions (FTLD-TDP) (Arai et al., 2006; Neumann et al., 2006). ALS and FTLD share substantial clinical and pathological manifestations (Mackenzie et al., 2010). ALS patients show generalized skeletal and bulbar muscle atrophy owing to progressive loss of cortical and spinal motor neurons. Up to 10% of ALS cases are caused by inheritable genetic mutations. FTLD is a cause of the second most common early-onset dementia. FTLD patients show behavioral changes with progressive decline in cognitive function caused by neuronal loss chiefly affecting the frontotemporal cortex. Importantly, substantial numbers of ALS patients show cognitive impairment, while significant populations of FTLD patients develop symptoms of motor neuron disease. Approximately 50% of the cases of clinical FTLD exhibit TDP-43 pathology (Chen-Plotkin et al., 2010).

Furthermore, TDP-43-immunoreactive inclusions are accumulated, to a lessor extent, in the brains of various neurodegenerative diseases, such as Alzheimer disease (AD), dementia with Lewy bodies (DLB), Pick disease (PiD), and the Guam parkinsonism-dementia complex (G-PDC) (Geser et al., 2009).

Because ALS and FTLD show clinicopathologically overlapping features, they are recently categorized into a novel disease entity named TDP-43 proteinopathy (Geser et al., 2009). In TDP-43 proteinopathy, TDP-43 often translocates from the nucleus to the cytoplasm where it forms detergent-insoluble urea-soluble aggregates. The accumulated proteins are hyperphosphorylated, polyubiquitinated, and proteolytically cleaved to produce 25-kDa and 35-kDa C-terminal fragments (Hasegawa et al., 2008). Following exposure to stressful insults, TDP-43 promptly accumulates in cytoplasmic stress granules (SGs) that regulate stress-induced translational arrest (Colombrita et al., 2009). Impaired dynein-mediated microtubule transport promotes aggregation of the C-terminal fragments (Pesiridis et al., 2011). The C-terminal domain of TDP-43 contains multiple phosphorylation consensus sites, among which phosphorylation of Ser-409/410 on TDP-43 serves as the pathological hallmark of sporadic and familial ALS cases (Neumann et al., 2009). Hyperphosphorylation of TDP-43 promotes oligomerization and fibril formation *in vitro* (Hasegawa et al., 2008). Overexpression of the TDP-43 C-terminal fragment is sufficient to generate hyperphosphorylated and ubiquitinated cytoplasmic aggregates that alter the exon-splicing pattern (Zhang et al., 2009). Importantly, various missense mutations encoding 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 and FTLD (Mackenzie et al., 2010). TDP-43 A315T mutant peptides form prion protein-like proteinase K-resistant amyloid fibrils with profound neurotoxicity (Guo et al., 2011). The gene named fused in sarcoma/translocated in liposarcoma (FUS/TLS) encodes another RNA/DNA-binding protein with the structural relationship to TDP-43. Genetic mutations of the FUS/TLS gene are also causative for ALS (Kwiatkowski et al., 2009). TDP-43 and FUS/TLS are located in the common disease-related pathway involving RNA metabolism (Lagier-Tourenne et al., 2010; Mackenzie et al., 2010).

Transgenic mice overexpressing wild-type and ALS-linked mutant TDP-43 proteins develop motor dysfunction and cognitive impairment with cytoplasmic ubiquitinated TDP-43 inclusions (Wils et al., 2010). In contrast, TDP-43 knockout mice are embryonically lethal, indicating that TDP-43 is indispensable for early embryonic development (Kraemer et al., 2010). However, conditional knockout mice revealed an unexpected role of TDP-43 related to fat metabolism and embryonic stem (ES) cell survival (Chiang et al., 2010). All of these observations, taken together, suggest that the neurodegenerative process underlying ALS and FTLD is attributable to a toxic gain of function or a loss of essential function of TDP-43 (Kabashi et al., 2010; Janssens et al., 2011). However, at present, the precise molecular mechanisms remain largely unknown.

Recent advances in systems biology have made major breakthroughs by illustrating the cell-wide map of complex molecular interactions extracted from large-scale data of the genome, transcriptome and proteome with the aid of the literature-based knowledgebase of molecular pathways (Satoh, 2010). The logically arranged molecular networks construct the whole system characterized by robustness that maintains the proper function of the system in the face of genetic and environmental perturbations (Kitano, 2007). In the scale-free molecular network, targeted disruption of several critical components designated hubs, on which the biologically important molecular interactions concentrate, efficiently disturbs the whole cellular function by destabilizing the network (Albert et al., 2000). Therefore, the intensive characterization of large-scale molecular networks and pathways that specifically regulate biologically relevant functions *in vivo* would help us to understand the disease-linked molecular mechanisms. By analyzing molecular networks, we recently clarified key roles of transcriptional regulation by NF-κB in acute relapse of multiple sclerosis (MS) (Satoh et al., 2008), integrin and extracellular matrix proteins in development of chronic active plaques in MS brains (Satoh et al., 2009c), Akt signaling pathway in the cellular prion protein interactome (Satoh et al., 2009a), transcriptional regulation by CREB (Satoh et al., 2009b) and microRNA-controlled cell cycle regulators (Satoh, 2011) in the pathology of AD brains.

To achieve proper biological functions, TDP-43 forms a multimolecular complex with

various target RNAs and interacting proteins, exceeding the size of 500-kDa in human cells (Sephton et al., 2011). Previously, we found that substantial amounts of TDP-43 proteins form a dimer in human cells and brain tissues (Shiina et al., 2010), being consistent with the self-assembling capacity (Kuo et al., 2009) and the aggregation-prone propensity of TDP-43 (Johnson et al., 2009). Several recent studies identified numerous TDP-43 target RNAs and interacting proteins by deep sequencing with next-generation sequencers and mass spectrometric analysis (Freibaum et al., 2010; Sephton et al., 2011). The present study is conducted to characterize the comprehensive molecular network of TDP-43 target RNAs and interacting proteins derived from publicly available datasets by using pathway analysis tools of bioinformatics endowed with comprehensive knowledgebase.

METHODS

Dataset of TDP-43 Target RNAs

By RNA immunoprecipiation followed by deep sequencing (RIP-seq), a recent study identified thousands of TDP-43 targets RNAs in cultured rat cortical neurons (Sephton et al., 2011). The cell lysate of rat cortical neurons was processed for immunoprecipitation with anti-TDP-43 antibody or nonspecific rabbit IgG. Then, RNA was purified from the immunoprecipitates, and treated with DNAse I to remove genomic DNA contamination. cDNA libraries were constructed and sequenced on the Illumina GA IIx Genome Analyzer. The short reads were mapped onto the rat genome (build rn4) using the Bowtie software. The numbers of short sequence fragments enriched in the TDP-43 library versus the control library were counted. The study identified top 25% enriched 4,352 TDP-43 target RNAs as the most reliable set. We converted individual RefSeq accession numbers of the genes listed in this dataset into corresponding Entrez Gene IDs by using the Gene ID Conversion tool of the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.7 (david.abcc.ncifcrf.gov) (Huang et al., 2009). Non-annotated IDs and overlapping IDs were deleted, resulting in the selection of 4,163 rat TDP-43 target RNAs.

Dataset of TDP-43 Interacting Proteins

By mass spectrometry analysis, a recent study identified 261 TDP-43-interacting proteins in cultured human cells (Freibaum et al., 2010). The cell lysate was prepared from HEK 293 cells following overexpression of FLAG-tagged TDP-43 by transfection of the expression vector, and purified by immunoprecipitation with anti-FLAG M2 affinity gel. Then, the immunoprecipitates were gel-separated, digested into peptides, and processed for LC-MS/MS analysis, followed by database search using the Mascot software. The cell lysate of nontransfected cells processed in parallel was used as a control to determine nonspecific interactions. Among 261 TDP-43

interacting proteins they identified, we selected 227 proteins that exhibit a two-fold or greater increase in TDP-43 immunoprecipitates versus the control as the most reliable set. We converted individual UniProt accession numbers listed in this dataset into corresponding Entrez Gene IDs by using the Gene ID Conversion tool of DAVID.

Molecular Network Analysis

Entrez Gene IDs of TDP-43 target RNAs (Sephton et al., 2011) and interacting proteins (Freibaum et al., 2010) described above were imported into three distinct pathway analysis tools of bioinformatics endowed with comprehensive knowledgebase, including Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.kegg.jp), Ingenuity Pathways Analysis (IPA) (Ingenuity Systems; www.ingenuity.com), and KeyMolnet (Institute of Medicinal Molecular Design; www.immd.co.jp). KEGG is a public database, while both IPA and KeyMolnet are commercial ones updated regularly.

KEGG systematically integrates genomic and chemical information to create the whole biological system *in silico*. KEGG includes manually curated reference pathways that cover a wide range of metabolic, genetic, environmental, and cellular processes, and human diseases and drugs. Currently, KEGG contains 148,769 pathways generated from 410 reference pathways. By importing the list of Entrez Gene IDs into the Functional Annotation tool of DAVID, it identifies KEGG pathways and Gene Ontology (GO) categories composed of the genes enriched in the given set with statistical significance evaluated by the modified Fisher's exact test.

IPA is a knowledgebase that contains approximately 2,500,000 biological and chemical interactions and functional annotations with definite scientific evidence, curated by expert biologists. By uploading the list of Entrez Gene IDs, the network-generation algorithm identifies focused genes integrated in a global molecular network. IPA calculates the score p-value, the statistical significance of association between the genes and the networks by the Fisher's exact test.

KeyMolnet contains knowledge-based contents on 137,300 relationships among human genes and proteins, small molecules, diseases, pathways and drugs, curated by expert biologists (Satoh, 2010). They are categorized into the core contents collected from selected review articles with the highest reliability or the secondary contents extracted from abstracts of PubMed and Human Reference Protein database (HPRD). By importing the list of Entrez Gene IDs, KeyMolnet automatically provides corresponding molecules as a node on networks. The "neighboring" network-search algorithm selects one or more molecules as starting points to generate the network of all kinds of molecular interactions around starting molecules, including direct activation/inactivation, transcriptional activation/repression, and the complex formation within the designated number of paths from starting points. The generated network was compared side by side with 459 human canonical pathways installed in 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 statistically significant contribution to the extracted network, as described previously (Satoh, 2010).

RESULTS

Molecular Network of TDP-43 Target RNAs

By importing Entrez Gene IDs of 4,163 TDP-43 target RNAs detected in cultured rat cortical neurons (Sephton et al., 2011) into the Functional Annotation tool of DAVID, it identified 43 KEGG pathways with statistically significant relevance under the Expression Analysis Systematic Explorer (EASE) Score threshold of 0.01, a modified Fisher Exact test p-value. The top three KEGG pathways included "Axon guidance" (p = 7.474E-13), "ErbB signaling pathway" (p = 2.507E-07), and "Spliceosome" (p = 3.495E-07) (Fig. 1; **Fig. 1. Molecular Network of TDP-43 Target RNAs Illustrated by KEGG.** By importing Entrez Gene IDs of 4,163 rat TDP-43 target RNAs (Sephton et al., 2011) into the Functional Annotation tool of DAVID, it identified KEGG pathways with statistically significant relevance. The top KEGG pathway termed "Axon guidance" is shown, where the genes highlighted by orange are included in the set of TDP-43 target RNAs.). The top three GO categories consisted of "GO:0000166~nucleotide binding " (p = 7.846E-29), "GO:0008104~protein localization" (p = 1.403E-25), and "GO:0005829~cytosol" (p = 1.942E-25) (Table 1). These results suggest that TDP-43 target RNAs are chiefly involved in regulation of neuronal signaling pathways via RNA-protein interactions.

By importing Entrez Gene IDs of 4,163 TDP-43 target RNAs into IPA, it extracted the top three molecular networks with statistical significance, composed of "RNA Post-Transcriptional Modification, Genetic Disorder, Neurological Disease" (p = 1E-104), "Genetic Disorder, Metabolic Disease, RNA Damage and Repair" (p = 1E-104), and "Gene Expression, Cancer, DNA Replication, Recombination, and Repair" (p = 1E-98). The category of "RNA Post-Transcriptional Modification" network includes modification, processing, and splicing of various classes of RNAs. These results suggest a key role of TDP-43 in RNA-dependent regulation of gene expression. By importing Entrez Gene IDs of 4,163 TDP-43 target RNAs into KeyMolnet, the neighboring network-search algorithm based on the core contents extracted

the highly complex molecular network composed of 4,430 molecules and 8,999 molecular relations, which exhibited the most significant relationship to the canonical pathways termed "Transcriptional regulation by p53" (p = 2.082E-256), "Transcriptional regulation by RB/E2F" (p = 1.041E-194), and "Integrin family" (p = 3.760E-171). It is of particular interest that p53, RB, and E2F act as a principal transcriptional factor that regulates cell cycle progression and apoptosis.

Molecular Network of TDP-43 Interacting Proteins

By importing Entrez Gene IDs of 227 TDP-43 interacting proteins detected in HEK293 cells overexpressing TDP-43 (Freibaum et al., 2010) into the Functional Annotation tool of DAVID, it identified only 2 KEGG pathways with statistically significant relevance named "Ribosome" (p = 7.438E-56) and "Spliceosome" (p = 2.182E-11) under the EASE Score threshold of 0.01 (Fig. 2; **Fig. 2. Molecular Network of TDP-43 Interacting Proteins Illustrated by KEGG.** By importing Entrez Gene IDs of 227 human TDP-43 interacting proteins (Freibaum et al., 2010) into the Functional Annotation tool of DAVID, it identified KEGG pathways with statistically significant relevance. The second rank KEGG pathway termed "Spliceosome" is shown, where the genes highlighted by orange are included in the set of TDP-43 interacting proteins.). The top three GO categories included "GO:0030529~ribonucleoprotein complex" (p = 9.329E-101), "GO:0003723~RNA binding" (p = 2.460E-98), and "GO:0006414~translational elongation" (1.252E-60) (Table 1). These results suggest that TDP-43 and its binding partners play a key role in regulation of spliceosome and ribosome functions.

By importing Entrez Gene IDs of 227 TDP-43-interacting proteins into IPA, it extracted the top three molecular networks with statistical significance, composed of "Protein Synthesis, RNA Post-Transcriptional Modification, Cell Cycle" (p = 1E-69), "RNA Post-Transcriptional Modification, Cell Morphology, Cellular Function and Maintenance" (p = 1E-69), and "RNA Post-Transcriptional Modification, Cancer, Gene Expression" (p = 1E-63), suggesting again a central role of TDP-43 in RNA-dependent regulation of gene expression. By importing Entrez

Gene IDs of 227 TDP-43 interacting proteins into KeyMolnet, the neighboring network-search algorithm based on the core contents extracted the complex molecular network comprised of 517 molecules and 552 molecular relations, which exhibited the most significant relationship to the canonical pathways, composed of "spliceosome assembly" (4.149E-086), being consistent with the results of KEGG, and in addition, "Kinesin family signaling pathway" (1.806E-081) and "14-3-3 signaling pathway" (3.042E-073) (Fig. 3; **Fig. 3. Molecular Network of TDP-43 Interacting Proteins Illustrated by KeyMolnet.** By importing Entrez Gene IDs of 227 human TDP-43 interacting proteins (Freibaum et al., 2010) into KeyMolnet, it extracted the complex molecular network composed of 517 molecules and 552 molecular relations, exhibiting the most significant relationship to the canonical pathways composed of "spliceosome assembly". Red nodes reflect the molecules included in the set of TDP-43 interacting proteins, while white nodes exhibit additional molecules extracted automatically from the core contents to establish molecular connections. The molecular relation is indicated by solid line with arrow (direct binding or activation), solid line with arrow and stop (direct inactivation), solid line without arrow (complex formation), dash line with arrow (transcriptional activation), and dash line with arrow and stop (transcriptional repression).).

The Integrated Molecular Network of TDP-43 Target RNAs and Interacting Proteins

Finally, we converted Entrez Gene IDs of 4,163 rat TDP-43 target RNAs (Sephton et al., 2011) into the corresponding human Entrez Gene IDs by using the Gene ID Conversion tool of DAVID, resulting in the identification of 4,063 presumptive human TDP-43 target RNAs. Among them, we identified the set of 106 genes shared between human TDP-43 target RNAs and interacting proteins (Table 2), in which the components of Ribosome and Splicesome pathways of KEGG were enriched significantly (data not shown). Then, we imported Entrez Gene IDs of the combined set of total 4,063 human TDP-43 RNA targets and 227 TDP43 interacting proteins into the Functional Annotation tool of DAVID, IPA and KeyMolnet.

DAVID identified 32 KEGG pathways with statistically significant relevance under the EASE Score threshold of 0.01. The top three KEGG pathways included "Ribosome" (4.249E-10), "Axon guidance" (p = 1.247E-09), and "Spliceosome" (p = 1.240E-08). The top three GO categories included "GO:0003723~RNA binding" (p = 3.995E-38), "GO:0005829~cytosol" (p = 5.094E-34), and "GO:0000166~nucleotide binding" (p = 1.365E-27) (Table 1). IPA extracted the top three molecular networks with statistical significance, composed of "RNA Post-Transcriptional Modification, Genetic Disorder, Neurological Disease" (p = 1E-104), "Protein Synthesis, Gene Expression, RNA Trafficking" (p = 1E-90), and "Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry" (p = 1E-88) (Fig. 4; **Fig. 4. The Integrated Molecular Network of Human TDP-43 RNA Targets and Interacting Proteins Illustrated by IPA.** Entrez Gene IDs of the combined set of total 4,063 human TDP-43 RNA targets and 227 TDP-43 interacting proteins were imported into IPA. The top molecular network termed "RNA Post-Transcriptional Modification, Genetic Disorder, Neurological Disease" is shown. Dark red nodes indicate the molecules that concurrently serve as both TDP-43 target RNAs and interacting proteins, while bright red nodes represent the molecues that act only as TDP-43 interacting proteins, and green nodes indicate the molecules that act only as TDP-43 target RNAs. The molecular relation is indicated by solid line (direct interaction), dash line (indirect interaction), line only (binding), line with arrowhead (acts on), line with stop (inhibits), and line with stop and arrowhead (inhibits and acts on).). These results validated a crucial role of TDP-43 in RNA-dependent regulation of gene expression. By the neighboring network-search algorithm based on the core contents, KeyMolnet extracted the highly complex molecular network comprised of 4,515 molecules and 9,362 molecular relations, which exhibited the most significant relationship to the canonical pathways, composed of "Transcriptional regulation by p53" (p = 1.383E-252), "Integrin family" (p = 2.121E-192), and "Transcriptional regulation by RB/E2F" (p = 1.823E-189), being almost consistent with the molecular network of rat TDP-43 target RNAs.

DISCUSSION

Abnormally phosphorylated, ubiquitinated, and aggregated TDP-43 proteins constitute a principal component of neuronal and glial cytoplasmic and nuclear inclusions in the brains of ALS and FTLD (Arai et al., 2006; Neumann et al., 2006). These intractable neurodegenerative diseases share substantial clinical and pathological features, establishing a novel clinical entity designated as TDP-43 proteinopathy (Geser et al., 2009). Although increasing evidence suggests that the neurodegenerative process underlying ALS and FTLD is attributable to a toxic gain of function or a loss of cellular function of TDP-43 (Kabashi et al., 2010; Janssens et al., 2011), the precise molecular mechanisms remain largely unknown. To investigate molecular mechansims responsible for development of TDP-43 proteinopathy, we attempted to characterize the comprehensive molecular network of TDP-43 target RNAs and interacting proteins, which are identified by recent studies using deep sequencing with next-generation sequencers and mass spectrometric analysis (Freibaum et al., 2010; Sephton et al., 2011).

By using RIP-seq, a recent study identified thousands of TDP-43 target RNAs in cultured rat cortical neurons (Sephton et al., 2011). The study showed that the binding sites consisting of (UG)n and (UG)nUA(UG)m are enriched in TDP-43 target RNAs. They are categorized into three groups composed of the genes having the binding sites in exons, introns, and both exons and introns. The study showed that functional categories related to synaptic function, RNA metabolism, and neuronal development are enriched in TDP-43 RNA targets. By importing their dataset into three different pathway analysis tools, we found that TDP-43 RNA targets are most closely associated with "Axon guidance" pathway of KEGG, "RNA Post-Transcriptional Modification, Genetic Disorder, Neurological Disease" network of IPA, and "Transcriptional regulation by p53" pathway of KeyMolnet. Our results suggest a pivotal role of TDP-43 in RNA-dependent regulation of neuronal gene expression. Most importantly, axonal guidance function is markedly disturbed in the brains of ALS patients and mouse models, characterized by accumulation of neurofilamentous swellings named spheroids within the axons of motor neurons (King et al., 2011). A recent genome-wide association study (GWAS) showed that

TMEM106B variants on chromosome 7p21 serve as a risk factor for development of FTLD-TDP (Van Deerlin et al., 2010). Interestingly, we identified TMEM106B as one of 4,163 TDP-43 target RNAs.

More recently, by using ultraviolet cross-linking and immunoprecipitation (CLIP) and deep sequencing (CLIP-seq), a different study identified 6,304 TDP-43 target RNAs in the adult mouse brain (Polymenidou et al., 2011). The study showed that the vast majority of TDP-43-binding sites are located in introns of the target genes regardless of the presence or absence of UG-rich repeats. Among them, 2,672 genes (42.4%) are identical to the previous results derived from rat cortical neurons (Sephton et al., 2011). By using RNA sequencing (RNA-seq) and splicing-sensitive microarrays, they found that depletion of TDP-43 by local administration of TDP-43-specific antisense oligonucelotides in the striatum of the mouse brain downregulated the expression of 239 genes involved chiefly in synaptic activity, and altered splicing events of 965 mRNAs (Polymenidou et al., 2011).

By using individual nucleotide resolution CLIP (iCLIP) and deep sequencing, another study characterized the global picture of TDP-43 target RNAs isolated from human brain tissues of healthy subjects, FTLD-U patients, and cultured cells (Tollervey et al., 2011). The study verified that the majority of TDP-43-binding sites are located in introns of target genes having long clusters of UG-rich repeats. Importantly, small populations of iCLIP cDNAs are mapped on long non-coding RNAs (ncRNAs), including small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), ribosomal RNAs (rRNAs), and telomeric ncRNAs. They found that TDP-43-binding sites located on NEAT1 and NEAT2 ncRNAs are overrepresented in FTLD-U brain samples. Knockdown of TDP-43 in SH-SY5Y cells affected 158 alternative splicing events related to regulation of the genes involved in neuronal development and survival (Tollervey et al., 2011). They also confirmed that TDP-43 binds RNA by forming a homodimer, supporting our previous observations (Shiina et al., 2010). These results suggest that TDP-43, by binding to various classes of RNAs, regulates diverse neuronal functions.

By using immunoaffinity purification and mass spectrometric analysis, a recent study identified over 200 TDP-43 interacting proteins in HEK293 cells (Freibaum et al., 2010). The