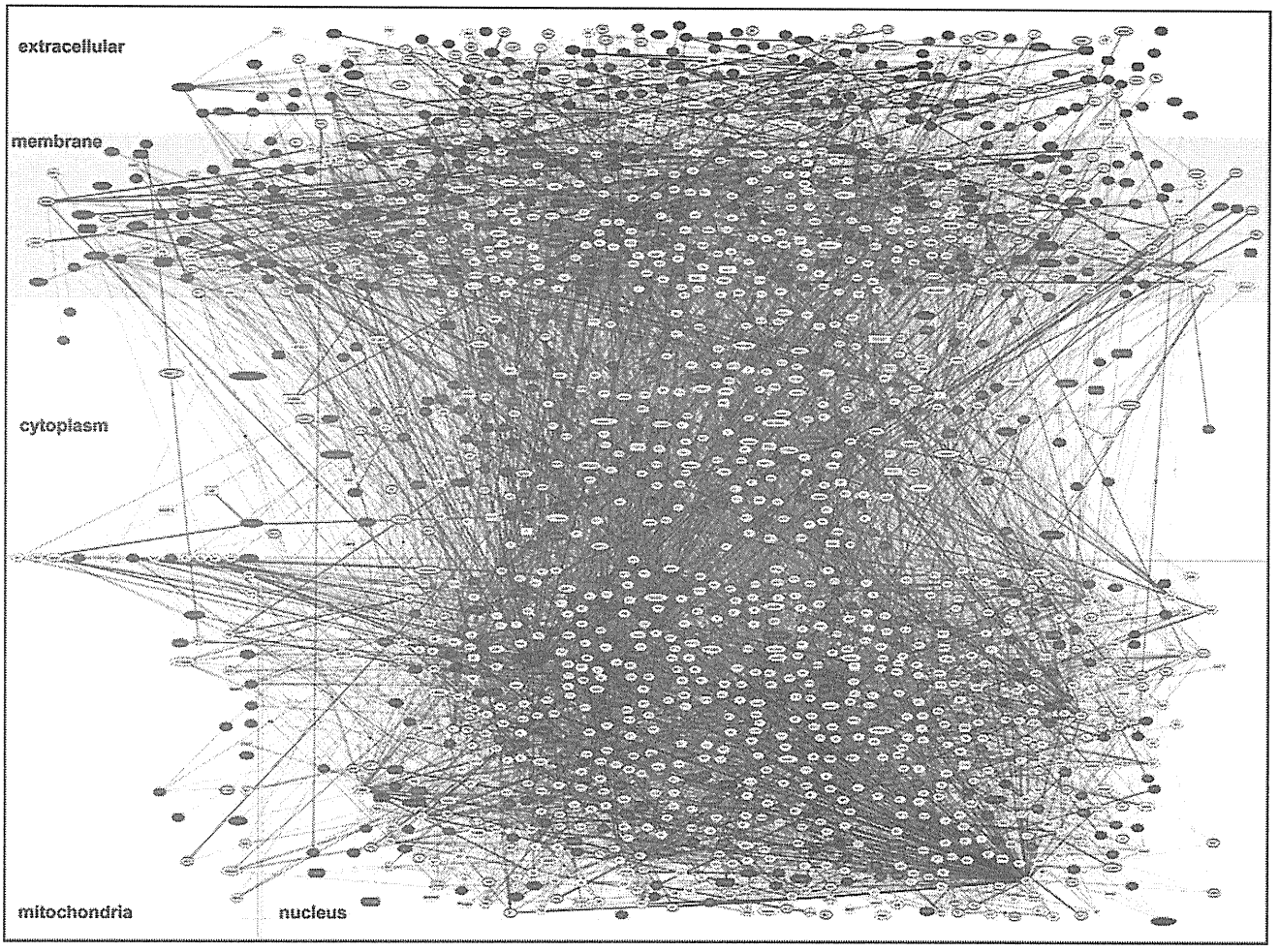

Fig. 6 Molecular network analysis of the genes regulated by stable expression of Ngn1 in F3 cells. The Entrez Gene ID and expression levels of 588 differentially expressed genes (DEG) between F3-WT and F3-Ngn1 cells were imported into KeyMolnet. It extracted 787 genes directly linked to the DEG. The "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

Clinical and Experimental Neuroimmunology 1 (2010) 127–140

REVIEW ARTICLE

Bioinformatics approach to identifying molecular biomarkers and networks in multiple sclerosis

Jun-ichi Satoh

Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, Tokyo, Japan

Keywords
bioinformatics; KeyMolnet; molecular network; multiple sclerosis; systems biology

Correspondence
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
Fax: +81-42-495-8678
Email: satoj@my-pharm.ac.jp

Received: 6 June 2010; accepted: 3 August 2010.

Abstract

Multiple sclerosis (MS) is an inflammatory demyelinating disease of the central nervous system (CNS) white matter mediated by an autoimmune process triggered by a complex interplay between genetic and environmental factors, in which the precise molecular pathogenesis remains to be comprehensively characterized. The global analysis of genome, transcriptome, proteome and metabolome, collectively termed omics, promotes us to characterize the genome-wide molecular basis of MS. However, as omics studies produce high-throughput experimental data at one time, it is often difficult to find out the meaningful biological implications from huge datasets. Recent advances in bioinformatics and systems biology have made major breakthroughs by illustrating the cell-wide map of complex molecular interactions with the aid of the literature-based knowledgebase of molecular pathways. The integration of omics data derived from the disease-affected cells and tissues with underlying molecular networks provides a rational approach not only to identifying the disease-relevant molecular markers and pathways, but also to designing the network-based effective drugs for MS. (Clin. Exp. Neuroimmunol. doi: 10.1111/j.1759-1961.2010.00013.x, September 2010)

Introduction

Multiple sclerosis (MS) is an inflammatory demyelinating disease affecting exclusively the central nervous system (CNS) white matter mediated by an autoimmune process triggered by a complex interplay between genetic and environmental factors.[1] Intravenous administration of interferon-gamma (IFNγ) provoked acute relapses of MS, indicating a pivotal role of proinflammatory T helper type 1 (Th1) lymphocytes. More recent studies proposed the pathogenic role of Th17 lymphocytes in sustained tissue damage in MS.[2] MS shows a great range of phenotypic variability. The disease is classified into relapsing-remitting MS (RRMS), secondary progressive MS (SPMS) or primary progressive MS (PPMS) with respect to the clinical course. Pathologically, MS shows a remarkable heterogeneity in the degree of inflammation, complement activation, antibody deposition, demyelination and remyelination, oligodendrocyte apoptosis, and axonal degeneration.[3] Currently available drugs in clinical practice of MS, including interferon-beta (IFNβ), glatiramer acetate, mitoxantrone, FTY720 and natalizumab, have proven only limited efficacies in subpopulations of the patients.[4] These observations suggest the hypothesis that MS is a kind of neurological syndrome caused by different immunopathological mechanisms leading to the final common pathway that provokes inflammatory demyelination. Therefore, the identification of specific biomarkers relevant to the heterogeneity of MS is highly important to establish the molecular mechanism-based personalized therapy in MS.

After the completion of the Human Genome Project in 2003, the global analysis of genome, transcriptome, proteome and metabolome, collectively termed omics, promotes us to characterize the genome-wide molecular basis of the diseases, and helps us to identify disease-specific molecular signatures

and biomarkers for diagnosis and prediction of prognosis. Actually, the genome-wide association study (GWAS) of MS revealed novel risk alleles for susceptibility of MS.[5] The comprehensive transcriptome and proteome profiling of brain tissues and lymphocytes identified key molecules aberrantly regulated in MS, whose role has not been previously predicted in the pathogenesis of MS.[6,7] Most recently, the application of next-generation sequencing technology to personal genomes has enabled us to investigate the genetic basis of MS at the level of individual patients.[8]

Because omics studies usually produce high-throughput experimental data at one time, it is often difficult to find out the meaningful biological implications from such a huge dataset. Recent advances in bioinformatics and systems biology have made major breakthroughs by showing the cell-wide map of complex molecular interactions with the aid of the literature-based knowledgebase of molecular pathways.[9] The logically arranged molecular networks construct the whole system characterized by robustness, which maintains the proper function of the system in the face of genetic and environmental perturbations.[10] In the scale-free molecular network, targeted disruption of limited numbers of critical components designated the hub, on which the biologically important molecular connections concentrate, could disturb the whole cellular function by destabilizing the network.[11] From the point of these views, the integration of omics data derived from the disease-affected cells and tissues with underlying molecular networks provides a rational approach not only to characterizing the disease-relevant pathways, but also to identifying the network-based effective drug targets.

Increasing numbers of human disease-oriented omics data have been deposited in public databases, such as the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo) and the ArrayExpress archive (http://www.ebi.ac.uk/microarray-as/ae). Most of these are transcriptome datasets. Importantly, they really include the data that have potentially valuable information on molecular biomarkers and networks of the diseases, when they are reanalyzed by appropriate bioinformatics approaches, followed by validation of *in silico* observations with *in vitro* and *in vivo* experiments.[12]

The present review has focused on bioinformatics approaches to identifying MS-associated molecular biomarkers and networks from high-throughput data of omics studies.

Global gene expression analysis

DNA microarray technology is an innovative approach that allows us to systematically monitor the genome-wide gene expression pattern of disease-affected tissues and cells. This approach enables us to illustrate most efficiently a global picture of cellular activity by the messenger RNA (mRNA) expression levels as an indicator, although the levels of mRNA do not always correlate with the levels of proteins directly involved in cellular function. However, the use of DNA microarray is more convenient to collect temporal and spatial snapshots of gene expression than the conventional mass spectrometry, which is often hampered by limited resolution of protein separation. In transcriptome analysis, we could logically assume that a set of coregulated genes might have similar biological functions within the cells.

First of all, I would like to briefly overview the gene expression analysis (Fig. 1). In general, total RNA fractions containing mRNA species are extracted from cells and tissues, individually labeled with fluorescent dyes, and processed for hybridization with thousands of oligonucleotides of known sequences immobilized on the arrays. After washing, they are processed for signal acquisition on a scanner. Various types of microarrays are currently available, although the MicroArray Quality Control (MAQC) project verified that the core results are well reproducible among different platforms used.[13] However, it is recommended that each experiment should contain biological replicates to validate reproducibility of the observations. The raw data are normalized by representative methods, including the quantile normalization method and the Robust MultiChip Average (RMA) method using the R software of the Bioconductor package (cran.r-project.org) or the GeneSpring software (Agilent Technology, Palo Alto, CA, USA).

To identify differentially expressed genes (DEG) among distinct samples, the normalized data are processed for statistical analysis using *t*-test for comparison between two groups or analysis of variance (ANOVA) for comparison among more than three groups, followed by the multiple comparison test with the Bonferroni correction or by controlling false discovery rate (FDR) below 0.05 to adjust *P*-values.

In the next step, the levels of expression of DEG should be validated by quantitative reverse transcription polymerase chain reaction (qRT–PCR). The normalized data are also processed for hierarchical

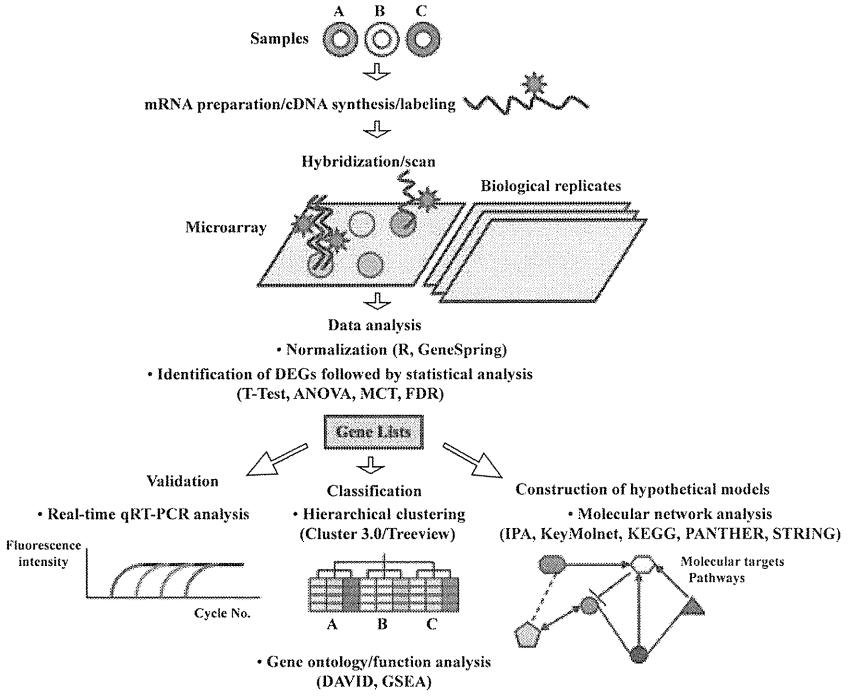

Figure 1 The load map from global gene expression profiling to molecular network analysis. Total RNA samples labeled with fluorescent dyes are processed for hybridization with oligonucleotide probes on the arrays, which should include biological replicates. They are processed for signal acquisition on a scanner. To identify the list of differentially expressed genes (DEG) among the samples, the normalized data are processed for statistical analysis, followed by validation by quantitative reverse transcription polymerase chain reaction (qRT–PCR). They are also processed for hierarchical clustering analysis and gene ontology and function analysis. To identify biologically relevant molecular pathways, the list of DEG is imported into pathway analysis tools endowed with a comprehensive knowledgebase. ANOVA, analysis of variance; DAVID, Database for Annotation, Visualization and Integrated Discovery; FDR, false discovery rate; GSEA, Gene Set Enrichment Analysis; IPA, Ingenuity Pathways Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; MCT, multiple comparison test; PANTHER, Protein Analysis Through Evolutionary Relationships; and STRING; Search Tool for the Retrieval of Interacting Genes/Proteins.

clustering analysis to classify the expression of profile-based groups of genes and samples by using GeneSpring or the open-access resources, such as Cluster 3.0 (bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster) and TreeView (sourceforge.net/projects/jtreeview). The Gene ID Conversion tool of the Database for Annotation, Visualization and Integrated Discovery (DAVID) (david.abcc.ncifcrf.gov)[14] converts the large-scale array-specific probe IDs into the corresponding Entrez Gene IDs, HUGO Gene Symbols, Ensembel Gene IDs or UniProt IDs, being more convenient for application to the downstream analysis. Both the DAVID Functional annotation tool and the Gene Set Enrichment Analysis (GSEA) tool (www.broad.mit.edu/gsea/downloads.jsp)[15] are open-access resources that help us to identify a set of enriched genes with a specified functional annotation in the entire list of genes. Many other approaches for preprocessing microarray data are applicable, and the resources are available elsewhere.

Molecular network analysis

To identify biologically relevant molecular pathways from large-scale data, we could analyze them by using a battery of pathway analysis tools endowed with a comprehensive knowledgebase; that is, Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.kegg.jp), the Protein Analysis Through Evolutionary Relationships (PANTHER) classification system (http://www.pantherdb.org), Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; string.embl.de), Ingenuity Pathways Analysis (IPA; Ingenuity Systems, http://www.ingenuity.com) and KeyMolnet (Institute of Medicinal Molecular Design, http://www.immd.co.jp) (Fig. 1). KEGG, PANTHER and STRING are open-access databases, whereas IPA and KeyMolnet are commercial databases updated regularly. Both transcriptome and proteome data are acceptable for all the databases described here.

KEGG systematically integrates genomic and chemical information to create the whole biological

system *in silico*.[16] KEGG includes manually curated reference pathways that cover a wide range of metabolic, genetic, environmental and cellular processes, and human diseases. Currently, KEGG contains 108 983 pathways generated from 358 reference pathways. PANTHER, operating on the computational algorithms that relate the evolution of protein sequences to the evolution of protein functions and biological roles, provides a structured representation of protein function in the context of biological reaction networks.[17] PANTHER includes the information on 165 regulatory and metabolic pathways, manually curated by expert biologists. By uploading the list of Gene IDs, the PANTHER gene expression data analysis tool identifies the genes in terms of over- or under-representation in canonical pathways, followed by statistical evaluation by multiple comparison test with the Bonferroni correction. STRING is a database that contains physiological and functional protein-protein interactions composed of 2 590 259 proteins from 630 organisms.[18] STRING integrates the information from numerous sources, including experimental repositories, computational prediction methods and public text collections. By uploading the list of UniProt IDs, STRING illustrates the union of all possible association networks.

IPA is a knowledgebase that contains approximately 2 270 000 biological and chemical interactions and functional annotations with definite scientific evidence, curated by expert biologists.[19] By uploading the list of Gene IDs and expression values, 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 content on 123 000 relationships among human genes and proteins, small molecules, diseases, pathways and drugs, curated by expert biologists.[20] They are categorized into the core content 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 Gene ID and expression values, KeyMolnet automatically provides corresponding molecules as a node on networks. The "common upstream" network-search algorithm enables us to extract the most relevant molecular network composed of the genes coordinately regulated by putative common upstream transcription factors. The "neighboring" network-search algorithm selected 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 "N-points to N-points" network-search algorithm identifies the molecular network constructed by the shortest route connecting the start-point molecules and the end-point molecules. The generated network was compared side-by-side with 430 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 is the number of overlapping molecular relations between the extracted network and the canonical pathway, V is the number of molecular relations located in the extracted network, C is the number of molecular relations located in the canonical pathway, T is the number of total molecular relations, and X is the sigma variable that defines coincidence.

$$\text{Score} = -\log_2(\text{Score}[P])$$

$$\text{Score}(P) = \sum_{x=O}^{\text{Min}(C,V)} f(x)$$

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

Biomarkers for predicting MS relapse

Molecular mechanisms underlying acute relapse of MS remain currently unknown. If molecular biomarkers for MS relapse are identified, we could predict the timing of relapses, being invaluable to start the earliest preventive intervention.

By gene expression profiling with Affymetrix Human Genome U133 plus 2.0 arrays, Corvol et al. identified 975 genes that separate clinically isolated syndrome (CIS) into four groups.[21] Surprisingly, 92% of patients in group 1 were characterized by a subset of 108 genes converted to clinically definite MS (CDMS) within 9 months of the first attack. They suggest downregulation of TOB1, a negative regulator of T cell proliferation as a marker predicting the conversion from CIS to CDMS.

By gene expression profiling with Affymetrix Human Genome U133A2 arrays, Achiron et al. showed that 1578 DEG of peripheral blood mononuclear cells (PBMC) of RRMS patients, differentiating

acute relapse from remission, are enriched in the apoptosis-related pathway, in which proapoptotic genes are downregulated, whereas antiapoptotic genes are upregulated during acute relapse.[22] The same group also compared 62 patients with CDMS and 32 patients with CIS by combining gene expression profiling with the support vector machine (SVM)-based prediction of time to the next acute relapse, setting a two stage predictor composed of First Level Predictors (FLP) and Fine Turning Predictors (FTP).[23] They identified three sets of the best 10-gene FLP that predict the next relapse with a resolution of 500 days and four sets of the best 9-gene FTP that predict the forthcoming relapse with a resolution of 50 days. The predictor genes are enriched in the TGFB2-related signaling pathway. More recently, Achiron et al. compared nine subjects who developed MS during a 9-year follow-up period (the preactive stage of MS; MS-to-be) and 11 control subjects unaffected with MS (MS-free) by gene expression profiling.[24] They found downregulation of nuclear receptor NR4A1 in the preactive stage of MS, suggesting that self-reactive T cells are not eliminated in the MS-to-be population, owing to a defect in the NR4A1-dependent apoptotic mechanism.

By gene expression profiling with a custom microarray of the Peter MacCallum Cancer Institute, Arthur et al. showed that a set of dysregulated genes in peripheral blood cells during the relapse and the remission phases of RRMS are enriched in the categories involved in apoptosis and inflammation, when annotated according to the GOstat program.[25] They also found upregulation of TGFB1 during the relapse. These observations support the working hypothesis that MS relapse involves an imbalance between promoting and preventing apoptosis of autoreactive and regulatory T cells. By gene expression analysis with Affymetrix Human Genome U133 plus 2.0 arrays, Brynedal et al. showed that MS relapses reflect the gene expression change in PBMC, but not in cerebrospinal fluid (CSF) lymphocytes, suggesting the importance of initial events triggering relapses occurring outside the CNS.[26]

By gene expression profiling with a custom DNA microarray (Hitachi Life Science, Saitama, Japan), we identified 43 DEG in peripheral blood $CD3^+$ T cells between the peak of acute relapse and the complete remission of RRMS patients.[27] We isolated highly purified $CD3^+$ T cells, because autoreactive pathogenic and regulatory cells, which potentially play a major role in MS relapse and remission, might be enriched in this fraction. By using 43 DEG as a set of discriminators, hierarchical clustering separated the cluster of relapse from that of remission. The molecular network of 43 DEG extracted by the common upstream search of KeyMolnet showed the most significant relationship with transcriptional regulation by the nuclear factor-kappa B (NF-κB). NF-κB is a central regulator of innate and adaptive immune responses, cell proliferation, and apoptosis.[28] A considerable number of NF-κB target genes activate NF-κB itself, providing a positive regulatory loop that amplifies and perpetuates inflammatory responses, leading to persistent activation of autoreactive T cells in MS. These observations support the logical hypothesis that NF-κB plays a central role in triggering molecular events in T cells responsible for induction of acute relapse of MS, and suggest that aberrant gene regulation by NF-κB on T-cell transcriptome serves as a molecular biomarker for monitoring the clinical disease activity of MS. Supporting this hypothesis, increasing evidence has shown that NF-κB represents a central molecular target for MS therapy.[29]

We also studied the gene expression profile of purified $CD3^+$ T cells isolated from four Hungarian monozygotic MS twin pairs with a custom DNA microarray (Hitachi Life Science, Saitama, Japan).[30] By comparing three concordant pairs and one discordant pair, we identified 20 DEG aberrantly regulated between the MS patient and the genetically identical healthy subject. The molecular network of 20 DEG extracted by the common upstream search of KeyMolnet showed the most significant relationship with transcriptional regulation by the Ets transcription factor family. Ets transcription factor proteins, by interacting with various co-regulatory factors, control the expression of a wide range of target genes essential for cell proliferation, differentiation, transformation and apoptosis. Importantly, Ets-1, the prototype of the Ets family members, acts as a negative regulator of Th17 cell differentiation.[31] It is worthy to note that discordant monozygotic MS twin siblings do not show any genetic or epigenetic differences, as validated by whole genome sequencing analysis and genome-scale DNA methylation profiling.[8]

Biomarkers for predicting IFNβ responders

Although recombinant IFNβ therapy is widely used as the gold standard to reduce disease activity of MS, up to 50% of the patients continue to have relapses, followed by progression of disability. If molecular biomarkers for IFNβ responsiveness are identified, we could use the best treatment options depending on the patients, being invaluable to establish the personalized therapy of MS.

By genome-wide screening of single-nucleotide polymorphisms (SNP) with Affymetrix Human 100K SNP arrays, Byun et al. identified allelic differences between IFNβ responders and non-responders of RRMS patients in several genes, including HAPLN1, GPC5, COL25A1, CAST and NPAS3, although odds ratios of SNP differences of individual genes are fairly low.[32]

By gene expression profiling with Affymetrix Human Genome U133A Plus 2.0 arrays, Comabella et al. showed that IFNβ non-responders of RRMS patients after treatment for 2 years are characterized by the overexpression of type I IFN-induced genes in PBMC, associated with increased endogenous production of type I IFN by monocytes at pre-treatment.[33] These observations suggest that a preactivated type I IFN signaling pathway is attributable to IFNβ non-responsiveness in MS. By gene expression profiling with Affymetrix Human Genome Focus arrays, Sellebjerg et al. showed that *in vivo* injection of IFNβ rapidly induces elevation of IFI27, CCL2 and CXCL10 in PBMC of MS patients, even after 6 months of treatment,[34] consistent with previous studies.[35] The induction of IFN-responsive genes is greatly reduced in patients with neutralizing antibodies (NAbs) against IFNβ.[34] In contrast, there exist no global differences in gene expression profiles of PBMC of RRMS patients between NAbs-negative IFNβ non-responders and responders.[36]

By gene expression profiling with Affymetrix Human Genome U133A/B arrays, Goertsches et al. found that IFNβ administration *in vivo* elevates a panel of IFN-responsive genes in PBMC of RRMS patients during a 2-year treatment, but it also downregulates several genes, including CD20, a known target of B-cell depletion therapy in MS.[37] By using the Pathway Architect software (Stratagene, La Jolla, CA, USA), they identified two major gene networks where upregulation of STAT1 and downregulation of ITGA2B act as a central molecule, although they did not further characterize the responder/non-responder-linked gene expression profiles.

By gene expression profiling with a custom array of the National Institutes of Health (NIH)/National Institute of Neurological Disorders and Stroke (NINDS) Microarray Consortium, Fernald et al. showed that a 1-week IFNβ administration *in vivo* induces a set of coregulated genes whose networks are related to immune- and apoptosis-regulatory functions, involving JAK-STAT and NF-κB cascades, whereas the networks of untreated subjects are composed of the genes of cellular housekeeping functions.[38] By combining kinetic RT–PCR analysis of expression of 70 genes in PBMC of RRMS with the integrated Bayesian inference system approach, the same group previously reported that nine sets of gene triplets detected at pretreatment, including a panel of caspases, well predict the response to IFNβ with up to 86% accuracy.[39]

By gene expression profiling with a custom microarary (Hitachi), we previously identified a set of interferon-responsive genes expressed in purified peripheral blood $CD3^+$ T cells of RRMS patients receiving IFNβ treatment.[40] IFNβ immediately induces a burst of expression of chemokine genes with potential relevance to IFNβ-related early adverse effects in MS.[41] The majority of the top 30 most significant DEG in $CD3^+$ T cells between untreated MS patients and healthy subjects are categorized into apoptosis signaling regulators.[42] Furthermore, we found that T cell gene expression profiling classifies a heterogeneous population of Japanese MS patients into four distinct subgroups that differ in the disease activity and therapeutic response to IFNβ.[43] We identified 286 DEG expressed between 72 untreated Japanese MS patients and 22 age- and sex-matched healthy subjects. By importing the list of 286 DEG into the common upstream search of KeyMolnet, the generated network showed the most significant relationship with transcriptional regulation by NF-κB.[30] Although none of the single genes alone serve as a MS-specific biomarker gene, NR4A2 (NURR1), a target of NF-κB acting as a positive regulator of IL-17 and IFNγ production, is highly upregulated in MS T cells.[42,43] It is worthy to note that IFNβ is beneficial in the disease induced by Th1 cells, but detrimental in the disease mediated by Th17 cells in mouse experimental autoimmune encephalomyelitis (EAE), and IFNβ non-responders in RRMS patients show higher serum IL-17F levels, suggesting that IL-17 serves as a biomarker predicting a poor IFNβ response in MS.[44]

Molecular networks of MS brain lesion proteome

Recently, Han et al. investigated a comprehensive proteome of six frozen MS brains.[7] Proteins were prepared from small pieces of brain tissues isolated by laser-captured microdissection (LCM), and they were characterized separately by the standard histological examination, and classified into acute plaques (AP), chronic active plaques (CAP) or chronic plaques (CP) based on the disease activity. The proteins were then separated on one-dimensional SDS-PADE gels, digested in-gel with trypsin, and peptide fragments were processed for mass spectrometric

Table 1 Multiple sclerosis-linked molecules of the KeyMolnet library

KeyMolnet ID	KeyMolnet symbol	Description
KMMC:04422	2,3cnPDE	2',3'-cyclic nucleotide 3'-phosphodiesterase
KMMC:04421	aBcrystallin	Alpha crystallin B chain
KMMC:01024	ADAM17	A disintegrin and metalloproteinase 17
KMMC:04753	AMPAR	AMPA-type glutamate receptor
KMMC:00019	APP	Amyloid beta A4 protein
KMMC:07424	AQP4	Aquaporin 4
KMMC:06672	b-arrestin1	Beta-arrestin 1
KMMC:04017	BAFF	B-cell activating factor
KMMC:00868	Bcl-2	B-cell lymphoma 2
KMMC:00728	Ca	Calcium ion
KMMC:00605	caspase-1	Caspase-1
KMMC:00429	CCL2	Chemokine (C-C motif) ligand 2
KMMC:00425	CCL3	Chemokine (C-C motif) ligand 3
KMMC:00424	CCL5	Chemokine (C-C motif) ligand 5
KMMC:00450	CCR1	Chemokine (C-C motif) receptor 1
KMMC:00454	CCR5	Chemokine (C-C motif) receptor 5
KMMC:03088	CD28	T-cell-specific surface glycoprotein CD28
KMMC:00530	CD80	T-lymphocyte activation antigen CD80
KMMC:03089	CTLA-4	Cytotoxic T-lymphocyte protein 4
KMMC:00418	CXCL10	Chemokine (C-X-C motif) ligand 10
KMMC:00447	CXCR3	Chemokine (C-X-C motif) receptor 3
KMMC:00271	ERa	Estrogen receptor alpha
KMMC:00362	FGF-2	Fibroblast growth factor 2
KMMC:04423	GFAP	Glial fibrillary acidic protein
KMMC:01120	Glu	Glutamic acid
KMMC:00396	glucocorticoid	Glucocorticoid
KMMC:03232	hH1R	Histamine H1 receptor
KMMC:00344	HLA class II	HLA class II histocompatibility antigen
KMMC:09224	HLA-C5	HLA-C5
KMMC:09221	HLA-DQA1*0102	HLA-DQA1*0102
KMMC:06358	HLA-DQA1*0301	HLA-DQA1*0301
KMMC:06359	HLA-DQB1*0302	HLA-DQB1*0302
KMMC:09222	HLA-DQB1*0602	HLA-DQB1*0602
KMMC:06309	HLA-DRB1	HLA-DRB1
KMMC:06315	HLA-DRB1*0301	HLA-DRB1*0301
KMMC:09223	HLA-DRB1*0405	HLA-DRB1*0405
KMMC:09191	HLA-DRB1*11	HLA-DRB1*11
KMMC:07762	HLA-DRB1*15	HLA-DRB1*15
KMMC:06903	HLA-DRB1*1501	HLA-DRB1*1501
KMMC:07763	HLA-DRB1*1503	HLA-DRB1*1503
KMMC:09220	HLA-DRB5*0101	HLA-DRB5*0101
KMMC:04418	HSP105	Heat-shock protein 105 kDa
KMMC:00526	IFNb	Interferon beta
KMMC:00404	IFNg	Interferon gamma
KMMC:00292	IGF1	Insulin-like growth factor 1
KMMC:03611	IgG	Immunoglobulin G
KMMC:00402	IL-10	Interleukin-10
KMMC:03248	IL-12	Interleukin-12
KMMC:04266	IL-12Rb2	Interleukin-12 receptor beta-2 chain
KMMC:03129	IL-17	Interleukin-17
KMMC:03383	IL-18	Interleukin-18
KMMC:00521	IL-1b	Interleukin-1 beta
KMMC:00296	IL-2	Interleukin-2
KMMC:06578	IL-23	Interleukin-23
KMMC:00533	IL-2Rac	Interleukin-2 receptor alpha chain
KMMC:00400	IL-4	Interleukin-4
KMMC:03255	IL-5	Interleukin-5

Table 1 (*Continued*)

KeyMolnet ID	KeyMolnet symbol	Description
KMMC:00108	IL-6	Interleukin-6
KMMC:03257	IL-7Rac	Interleukin-7 receptor alpha chain
KMMC:00523	IL-9	Interleukin-9
KMMC:00555	iNOS	Inducible nitric oxide synthase
KMMC:00982	int-a4/b1	Integrin alpha-4/beta-1
KMMC:00968	int-aM	Integrin alpha-M
KMMC:00970	int-aX	Integrin alpha-X
KMMC:04094	MBP	Myelin basic protein
KMMC:06533	mGluR	Metabotropic glutamate receptor
KMMC:04420	MOG	Myelin-oligodendrocyte glycoprotein
KMMC:04419	MPLP	Myelin proteolipid protein
KMMC:03210	N-VDCC	Voltage dependent N-type calcium channel
KMMC:04712	NCAM	Neural cell adhesion molecule
KMMC:06537	NCE	Na(+)-Ca^{2+} exchanger
KMMC:05576	NeuroF	Neurofilament protein
KMMC:09225	neurofascin	Neurofascin
KMMC:05903	NF-H	Neurofilament triplet H protein
KMMC:05904	NF-L	Neurofilament triplet L protein
KMMC:03785	NMDAR	N-methyl-D-aspartate receptor
KMMC:07764	NMDAR1	N-methyl-D-aspartate receptor subunit NR1
KMMC:07765	NMDAR2C	N-methyl D-aspartate receptor subtype 2C
KMMC:07766	NMDAR3A	N-methyl-D-aspartate receptor subtype NR3A
KMMC:02064	NO	Nitric oxide
KMMC:07767	Olig-1	Oligodendrocyte transcription factor 1
KMMC:01005	OPN	Osteopontin
KMMC:03073	PDGF	Platelet derived growth factor
KMMC:06225	Sema3A	Semaphorin 3A
KMMC:06229	Sema3F	Semaphorin 3F
KMMC:00111	SMAD3	Mothers against decapentaplegic homolog 3
KMMC:03839	tau	Microtubule-associated protein tau
KMMC:00349	TNFa	Tumor necrosis factor alpha
KMMC:00545	VCAM-1	Vascular cell adhesion protein 1
KMMC:03832	VD	Vitamin D
KMMC:03711	VDR	Vitamin D3 receptor

91 multiple sclerosis-linked molecules of the KeyMolnet library are listed in alphabetical order.

analysis. Among 2574 proteins determined with high confidence, the INTERSECT/INTERACT program identified 158, 416 and 236 lesion-specific proteins detected exclusively in AP, CAP and CP, respectively. They found that overproduction of five molecules involved in the coagulation cascade, including tissue factor and protein C inhibitor, plays a central role in molecular events ongoing in CAP. Furthermore, *in vivo* administration of coagulation cascade inhibitors really reduced the clinical severity in EAE, supporting the view that the blockade of the coagulation cascade would be a promising approach for treatment of MS.[43] However, nearly all remaining proteins are uncharacterized in terms of their implications in MS brain lesion development.

We studied molecular networks and pathways of the proteome dataset of Han et al. by using four different bioinformatics tools for molecular network analysis, such as KEGG, PANTHER, KeyMolnet and IPA.[45] KEGG and PANTHER showed the relevance of extracellular matrix (ECM)-mediated focal adhesion and integrin signaling to CAP and CP proteome. KeyMolnet by the N-points to N-points search disclosed a central role of the complex interaction among diverse cytokine signaling pathways in brain lesion development at all disease stages, as well as a role of integrin signaling in CAP and CP. IPA identified the network constructed with a wide range of ECM components, such as COL1A1, COL1A2, COL6A2, COL6A3, FN1, FBLN2, LAMA1, VTN and HSPG2, as one of the networks highly relevant to CAP proteome. Thus, four distinct tools commonly suggested a role of ECM and integrin signaling in development of chronic

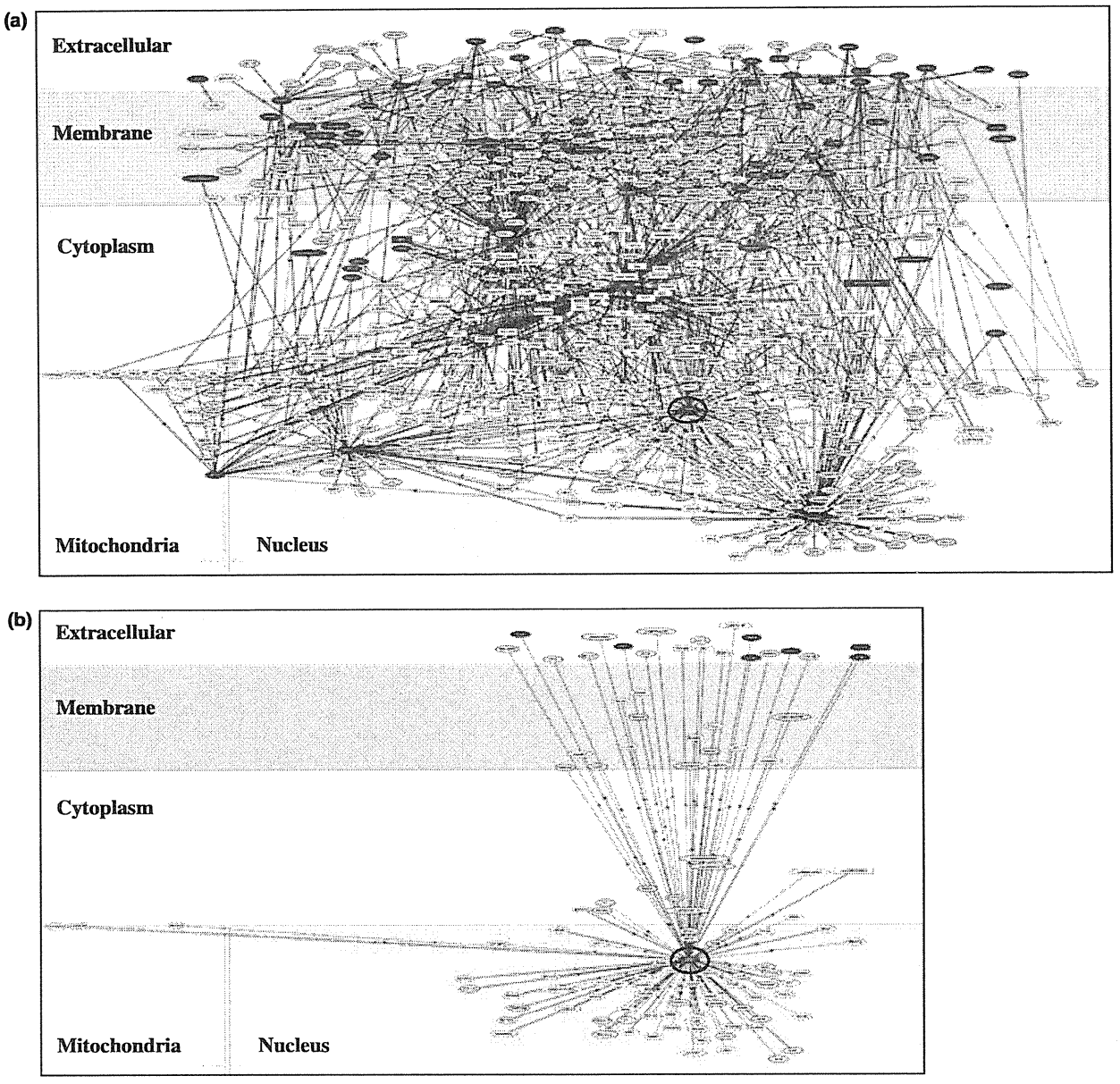

Figure 2 Molecular network of 91 MS-linked molecules. (a) By importing 91 MS-linked molecules into KeyMolnet, the neighboring search within one path from starting points generates the highly complex molecular network composed of 913 molecules and 1005 molecular relations. (b) The extracted network shows the most significant relationship with transcriptional regulation by vitamin D receptor (VDR) that has direct connections with 118 closely related molecules of the extracted network. VDR is indicated by blue circle. Red nodes represent start point molecules, whereas white nodes show additional molecules extracted automatically from core contents to establish molecular connections. The molecular relation is shown by a solid line with an arrow (direct binding or activation), solid line with an arrow and stop (direct inactivation), solid line without an arrow (complex formation), dash line with an arrow (transcriptional activation), and dash line with an arrow and stop (transcriptional repression). Please refer high resolution figures to URL (www.my-pharm.ac.jp/~satoj/sub22.html).

MS lesions, showing that the selective blockade of the interaction between ECM and integrin molecules in brain lesions *in situ* would be a target for therapeutic intervention to terminate ongoing events responsible for the persistence of inflammatory demyelination.

KeyMolnet identifies a candidate of molecular targets for MS therapy

The KeyMolnet library includes 91 MS-linked molecules, collected from selected review articles with the highest reliability (Table 1). By importing the list

Table 2 Molecules constucting the transcriptional regulation by vitamin D receptor network

KeyMolnet ID	KeyMolnet symbol	Description
KMMC:02959	1a,25(OH)2D3	1 alpha, 25-dihydroxyvitamin D3
KMMC:00751	amphiregulin	Amphiregulin
KMMC:03795	ANP	Atrial natriuretic peptide
KMMC:00090	b-catenin	beta-catenin
KMMC:00301	c-Fos	Protooncogene c-fos
KMMC:00183	c-Jun	Protooncogene c-jun
KMMC:00626	c-Myc	Protooncogene c-myc
KMMC:03813	CA-II	Carbonic anhydrase II
KMMC:04105	CalbindinD28K	Vitamin D-dependent calcium-binding protein, avian-type
KMMC:03531	CalbindinD9K	Vitamin D-dependent calcium-binding protein, intestinal
KMMC:00289	caseinK2	Casein kinase 2
KMMC:04195	CaSR	Extracellular calcium-sensing receptor
KMMC:00268	CBP	CREB binding protein
KMMC:00922	CD44	CD44 antigen
KMMC:00136	CDK2	Cyclin dependent kinase 2
KMMC:00135	CDK6	Cyclin dependent kinase 6
KMMC:01008	collagen	Collagen
KMMC:06770	collagenase-I	Type I collagenase
KMMC:04081	CRABP2	Cellular retinoic acid-binding protein II
KMMC:00060	CRT	Calreticulin
KMMC:00401	CXCL8	Chemokine (C-X-C motif) ligand 8 (IL8)
KMMC:00137	cyclinA	Cyclin A
KMMC:00061	cyclinD1	Cyclin D1
KMMC:05926	cyclinD3	Cyclin D3
KMMC:00093	cyclinE	Cyclin E
KMMC:02960	CYP24A1	Cytochrome P450 24A1
KMMC:02958	CYP27B1	Cytochrome P450 27B1
KMMC:04593	CYP3A4	Cytochrome P450 3A4
KMMC:06769	cystatin M	Cystatin M
KMMC:06762	Cytokeratin 13	Keratin, type I cytoskeletal 13
KMMC:06751	Cytokeratin 16	Keratin, type I cytoskeletal 16
KMMC:00053	DHTR	Dihydrotestosterone receptor
KMMC:00928	E-cadherin	E-cadherin
KMMC:00594	ErbB1	Receptor protein-tyrosine kinase erbB-1
KMMC:00068	filamin	Filamin
KMMC:00341	FN1	Fibronectin 1
KMMC:06760	FREAC-1	Forkhead box protein F1
KMMC:06763	G0S2	G0/G1 switch protein 2
KMMC:00617	GM-CSF	Granulocyte macrophage colony stimulating factor
KMMC:06755	Hairless	Hairless protein
KMMC:05978	HOXA10	Homeobox protein Hox-A10
KMMC:06767	HOXB4	Homeobox protein Hox-B4
KMMC:00404	IFNg	Interferon gamma
KMMC:00579	IGF-BP3	Insulin-like growth factor binding protein 3
KMMC:04498	IGF-BP5	Insulin-like growth factor binding protein 5
KMMC:00402	IL-10	Interleukin-10
KMMC:03241	IL-10R	Interleukin-10 receptor
KMMC:03239	IL-10Rac	Interleukin-10 receptor alpha chain
KMMC:03240	IL-10Rbc	Interleukin-10 receptor beta chain
KMMC:03248	IL-12	Interleukin-12
KMMC:03246	IL-12A	Interleukin-12 alpha chain
KMMC:00403	IL-12B	Interleukin-12 beta chain
KMMC:00296	IL-2	Interleukin-2
KMMC:00108	IL-6	Interleukin-6

Table 2 (*Continued*)

KeyMolnet ID	KeyMolnet symbol	Description
KMMC:00973	int-b3	Integrin beta-3
KMMC:03747	IVL	Involucrin
KMMC:00629	JunB	Protooncogene jun-B
KMMC:04334	JunD	Protooncogene jun-D
KMMC:06764	KLK10	Kallikrein-10
KMMC:06765	KLK6	Kallikrein-6
KMMC:04635	Mad1	Max dimerization protein 1
KMMC:06757	Metallothionein	Metallothionein
KMMC:06722	MKP-5	MAP kinase phosphatase 5
KMMC:00595	MMP-2	Matrix metalloproteinase 2
KMMC:03104	MMP-3	Matrix metalloproteinase 3
KMMC:00631	MMP-9	Matrix metalloproteinase 9
KMMC:00556	MnSOD	Manganese superoxide dismutase
KMMC:00927	N-cadherin	N-cadherin
KMMC:00074	NCOA1	Nuclear receptor coactivator 1
KMMC:00075	NCOA2	Nuclear receptor coactivator 2
KMMC:00080	NCOA3	Nuclear receptor coactivator 3
KMMC:00282	NCOR1	Nuclear receptor corepressor 1
KMMC:00270	NCOR2	Nuclear receptor corepressor 2
KMMC:00392	NFAT	Nuclear factor of activated T cells
KMMC:00104	NFkB	Nuclear factor kappa B
KMMC:03120	OPG	Osteoprotegerin
KMMC:01005	OPN	Osteopontin
KMMC:00304	osteocalcin	Osteocalcin
KMMC:00100	p21CIP1	Cyclin dependent kinase inhibitor 1
KMMC:00155	p27KIP1	Cyclin dependent kinase inhibitor 1B
KMMC:00195	p300	E1A binding protein p300
KMMC:03204	PLCb1	Phospholipase C beta 1
KMMC:03295	PLCd1	Phospholipase C delta 1
KMMC:00724	PLCg1	Phospholipase C gamma 1
KMMC:04869	plectin1	Plectin 1
KMMC:06772	PMCA1	Plasma membrane calcium-transporting ATPase 1
KMMC:06766	PP1c	Serine/threonine protein phosphatase PP1 catalytic subunit
KMMC:00786	PP2A	Serine/threonine protein phosphatase 2A
KMMC:03442	PPARd	Peroxisome proliferator activated receptor delta
KMMC:03710	PTH	Parathyroid hormone
KMMC:00346	PTHrP	Parathyroid hormone-related protein
KMMC:03115	RANKL	Receptor activator of NFkB ligand
KMMC:04537	RelB	Transcription factor RelB
KMMC:00091	RIP140	Nuclear factor RIP140
KMMC:00383	RXR	Retinoid X receptor
KMMC:06771	SCCA	Squamous cell carcinoma antigen
KMMC:05340	SKIP	Ski-interacting protein
KMMC:04103	SUG1	26S protease regulatory subunit 8
KMMC:05702	TAFII130	Transcription initiation factor TFIID subunit 4
KMMC:06753	TAFII28	Transcription initiation factor TFIID subunit 11
KMMC:06752	TAFII55	Transcription initiation factor TFIID subunit 7
KMMC:04955	TCF-1	T-cell-specific transcription factor 1
KMMC:03075	TCF-4	T-cell-specific transcription factor 4
KMMC:06754	TFIIA	Transcription initiation factor IIA
KMMC:04089	TFIIB	Transcription initiation factor IIB
KMMC:06768	TGase I	Transglutaminase I
KMMC:04184	TGFb1	Transforming growth factor beta 1
KMMC:05986	TGFb2	Transforming growth factor beta 2
KMMC:04104	TIF1	Transcription intermediary factor 1
KMMC:00349	TNFa	Tumor necrosis factor alpha

Table 2 (*Continued*)

KeyMolnet ID	KeyMolnet symbol	Description
KMMC:00277	TRAP220	Thyroid hormone receptor-associated protein complex component TRAP220
KMMC:06759	TRPV5	TRP vanilloid receptor 5
KMMC:06758	TRPV6	TRP vanilloid receptor 6
KMMC:06756	TRR1	Thioredoxin reductase 1
KMMC:03711	VDR	Vitamin D3 receptor
KMMC:04853	VDUP1	Vitamin D3 up-regulated protein 1
KMMC:06761	ZNF-44	Zinc finger protein 44
KMMC:05147	ZO-1	Tight junction protein ZO-1
KMMC:05811	ZO-2	Tight junction protein ZO-2

118 molecules constucting the transcriptional regulation by VDR network are listed in alphabetical order.

of these molecules into KeyMolnet, the neighboring search within one path from starting points generates the highly complex molecular network composed of 913 molecules and 1005 molecular relations (Fig. 2a). The extracted network shows the most significant relationship with transcriptional regulation by vitamin D receptor (VDR) with *P*-value of the score = 4.415E-242. Thus, VDR, a hub that has direct connections with 118 closely related molecules of the extracted network (Fig. 2b, Table 2), serves as one of the most promising molecular target candidates for MS therapy, because the adequate manipulation of the VDR network capable of producing a great impact on the whole network could efficiently disconnect the pathological network of MS. Indeed, vitamin D plays a protective role in MS by activating VDR, a transcription factor that regulates the expression of as many as 500 genes, although the underlying molecular mechanism remains largely unknown.[46]

Conclusion

MS is a complex disease with remarkable heterogeneity caused by the intricate interplay between various genetic and environmental factors. Recent advances in bioinformatics and systems biology have made major breakthroughs by illustrating the cell-wide map of complex molecular interactions with the aid of the literature-based knowledgebase of molecular pathways. The efficient integration of high-throughput experimental data derived from the disease-affected cells and tissues with underlying molecular networks helps us to characterize the molecular markers and pathways relevant to MS heterogeneity, and promotes us to identify the network-based effective drug targets for personalized therapy of MS.

Acknowledgements

This work was supported by grants from the Research on Intractable Diseases, the Ministry of Health, Labour and Welfare of Japan (H22-Nanchi-Ippan-136), and the High-Tech Research Center Project, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan (S0801043). The author thanks Dr Takashi Yamamura, Department of Immunology, National Institute of Neurosciences, NCNP for his continuous help with our studies.

References

1. Sospedra M, Martin R. Immunology of multiple sclerosis. *Annu Rev Immunol.* 2005; **23**: 683–747.
2. Steinman L. A brief history of T_H17, the first major revision in the T_H1/T_H2 hypothesis of T cell-mediated tissue damage. *Nat Med.* 2007; **13**: 139–45.
3. Lucchinetti C, Brück W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol.* 2000; **47**: 707–17.
4. Rudick RA, Lee JC, Simon J, Ransohoff RM, Fisher E. Defining interferon β response status in multiple sclerosis patients. *Ann Neurol.* 2004; **56**: 548–55.
5. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med.* 2007; **357**: 851–62.
6. Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, et al. Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nature Med.* 2002; **8**: 500–8.
7. Han MH, Hwang SI, Roy DB, Lundgren DH, Price JV, Ousman SS, et al. Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature.* 2008; **451**: 1076–81.

8. Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature*. 2010; **464**: 1351–6.
9. Viswanathan GA, Seto J, Patil S, Nudelman G, Sealfon SC. Getting started in biological pathway construction and analysis. *PLoS Comput Biol*. 2008; **4**: e16.
10. Kitano H. A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov*. 2007; **6**: 202–10.
11. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature*. 2000; **406**: 378–82.
12. Satoh J, Tabunoki H, Arima K. Molecular network analysis suggests aberrant CREB-mediated gene regulation in the Alzheimer disease hippocampus. *Dis Markers*. 2009; **27**: 239–52.
13. MAQC Consortium, Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*. 2006; **24**: 1151–61.
14. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; **4**: 44–57.
15. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; **102**: 15545–50.
16. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 2010; **38**: D355–60.
17. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res*. 2010; **38**: D204–10.
18. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8 – a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 2009; **37**: D412–6.
19. Pospisil P, Iyer LK, Adelstein SJ, Kassis AI. A combined approach to data mining of textual and structured data to identify cancer-related targets. *BMC Bioinformatics*. 2006; **7**: 354.
20. Sato H, Ishida S, Toda K, Matsuda R, Hayashi Y, Shigetaka M, et al. New approaches to mechanism analysis for drug discovery using DNA microarray data combined with KeyMolnet. *Curr Drug Discov Technol*. 2005; **2**: 89–98.
21. Corvol JC, Pelletier D, Henry RG, Caillier SJ, Wang J, Pappas D, et al. Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. *Proc Natl Acad Sci USA*. 2008; **105**: 11839–44.
22. Achiron A, Feldman A, Mandel M, Gurevich M. Impaired expression of peripheral blood apoptotic-related gene transcripts in acute multiple sclerosis relapse. *Ann N Y Acad Sci*. 2007; **1107**: 155–67.
23. Gurevich M, Tuller T, Rubinstein U, Or-Bach R, Achiron A. Prediction of acute multiple sclerosis relapses by transcription levels of peripheral blood cells. *BMC Med Genomics*. 2009; **2**: 46.
24. Achiron A, Grotto I, Balicer R, Magalashvili D, Feldman A, Gurevich M. Microarray analysis identifies altered regulation of nuclear receptor family members in the pre-disease state of multiple sclerosis. *Neurobiol Dis*. 2010; **38**: 201–9.
25. Arthur AT, Armati PJ, Bye C; Southern MS Genetics Consortium, Heard RN, Stewart GJ, et al. Genes implicated in multiple sclerosis pathogenesis from consilience of genotyping and expression profiles in relapse and remission. *BMC Med Genet*. 2008; **9**: 17.
26. Brynedal B, Khademi M, Wallström E, Hillert J, Olsson T, Duvefelt K. Gene expression profiling in multiple sclerosis: a disease of the central nervous system, but with relapses triggered in the periphery? *Neurobiol Dis*. 2010; **37**: 613–21.
27. Satoh J, Misawa T, Tabunoki H, Yamamura T. Molecular network analysis of T-cell transcriptome suggests aberrant regulation of gene expression by NF-κB as a biomarker for relapse of multiple sclerosis. *Dis Markers*. 2008; **25**: 27–35.
28. Barnes PJ, Karin M. Nuclear factor-κB. A pivotal transcription factor in chronic inflammatory diseases. *N Engl J Med*. 1997; **336**: 1066–71.
29. Yan J, Greer JM. NF-κB, a potential therapeutic target for the treatment of multiple sclerosis. *CNS Neurol Disord Drug Targets*. 2008; **7**: 536–57.
30. Satoh J, Illes Z, Peterfalvi A, Tabunoki H, Rozsa C, Yamamura T. Aberrant transcriptional regulatory network in T cells of multiple sclerosis. *Neurosci Lett*. 2007; **422**: 30–3.
31. Du C, Liu C, Kang J, Zhao G, Ye Z, Huang S, et al. MicroRNA miR-326 regulates T_H-17 differentiation and is associated with the pathogenesis of multiple sclerosis. *Nat Immunol*. 2009; **10**: 1252–9.
32. Byun E, Caillier SJ, Montalban X, Villoslada P, Fernández O, Brassat D, et al. Genome-wide pharmacogenomic analysis of the response to interferon beta therapy in multiple sclerosis. *Arch Neurol*. 2008; **65**: 337–44.
33. Comabella M, Lünemann JD, Río J, Sánchez A, López C, Julià E, et al. A type I interferon signature in monocytes is associated with poor response to interferon-β in multiple sclerosis. *Brain*. 2009; **132**: 3353–65.
34. Sellebjerg F, Krakauer M, Hesse D, Ryder LP, Alsing I, Jensen PE, et al. Identification of new sensitive biomarkers for the *in vivo* response to interferon-β treatment in multiple sclerosis using DNA-array evaluation. *Eur J Neurol*. 2009; **16**: 1291–8.

35. Weinstock-Guttman B, Badgett D, Patrick K, Hartrich L, Santos R, Hall D, et al. Genomic effects of IFN-β in multiple sclerosis patients. *J Immunol.* 2003; **171**: 2694–702.
36. Hesse D, Krakauer M, Lund H, Søndergaard HB, Langkilde A, Ryder LP, et al. Breakthrough disease during interferon-β therapy in MS: no signs of impaired biologic response. *Neurology.* 2010; **74**: 1455–62.
37. Goertsches RH, Hecker M, Koczan D, Serrano-Fernandez P, Moeller S, Thiesen HJ, et al. Long-term genome-wide blood RNA expression profiles yield novel molecular response candidates for IFN-β-1b treatment in relapsing remitting MS. *Pharmacogenomics.* 2010; **11**: 147–61.
38. Fernald GH, Knott S, Pachner A, Caillier SJ, Narayan K, Oksenberg JR, et al. Genome-wide network analysis reveals the global properties of IFN-β immediate transcriptional effects in humans. *J Immunol.* 2007; **178**: 5076–85.
39. Baranzini SE, Mousavi P, Rio J, Caillier SJ, Stillman A, Villoslada P, et al. Transcription-based prediction of response to IFNβ using supervised computational methods. *PLoS Biol.* 2005; **3**: e2.
40. Koike F, Satoh J, Miyake S, Yamamoto T, Kawai M, Kikuchi S, et al. Microarray analysis identifies interferon beta-regulated genes in multiple sclerosis. *J Neuroimmunol.* 2003; **139**: 109–18.
41. Satoh J, Nanri Y, Tabunoki H, Yamamura T. Microarray analysis identifies a set of CXCR3 and CCR2 ligand chemokines as early IFNbeta-responsive genes in peripheral blood lymphocytes in vitro: an implication for IFNbeta-related adverse effects in multiple sclerosis. *BMC Neurol.* 2006; **6**: 18.
42. Satoh J, Nakanishi M, Koike F, Miyake S, Yamamoto T, Kawai M, et al. Microarray analysis identifies an aberrant expression of apoptosis and DNA damage-regulatory genes in multiple sclerosis. *Neurobiol Dis.* 2005; **18**: 537–50.
43. Satoh J, Nakanishi M, Koike F, Onoue H, Aranami T, Yamamoto T, et al. T cell gene expression profiling identifies distinct subgroups of Japanese multiple sclerosis patients. *J Neuroimmunol.* 2006; **174**: 108–18.
44. Axtell RC, de Jong BA, Boniface K, van der Voort LF, Bhat R, De Sarno P, et al. T helper type 1 and 17 cells determine efficacy of interferon-β in multiple sclerosis and experimental encephalomyelitis. *Nat Med.* 2010; **16**: 406–12.
45. Satoh JI, Tabunoki H, Yamamura T. Molecular network of the comprehensive multiple sclerosis brain-lesion proteome. *Mult Scler.* 2009; **15**: 531–41.
46. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. *Lancet Neurol.* 2010; **9**: 599–612.

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