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

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

127

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.
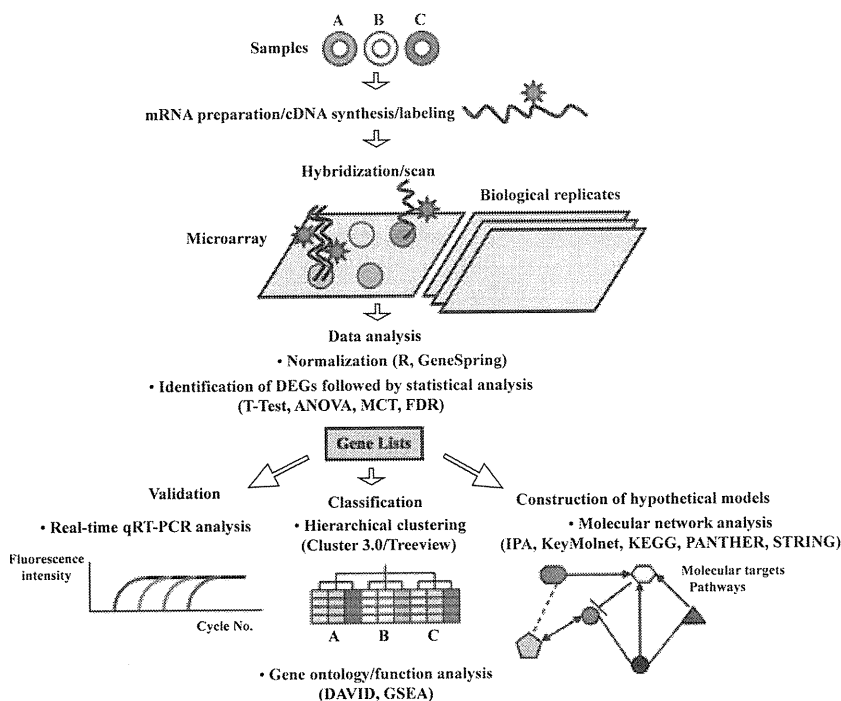
Global gene expression analysis

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

128

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

129

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.

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

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

$$f(x) = {}_CC_x \cdot {}_{T-C}C_{V-x}/{}_TC_V$$

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

130

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.

131

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 down-regulates 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

132

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

133

| 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
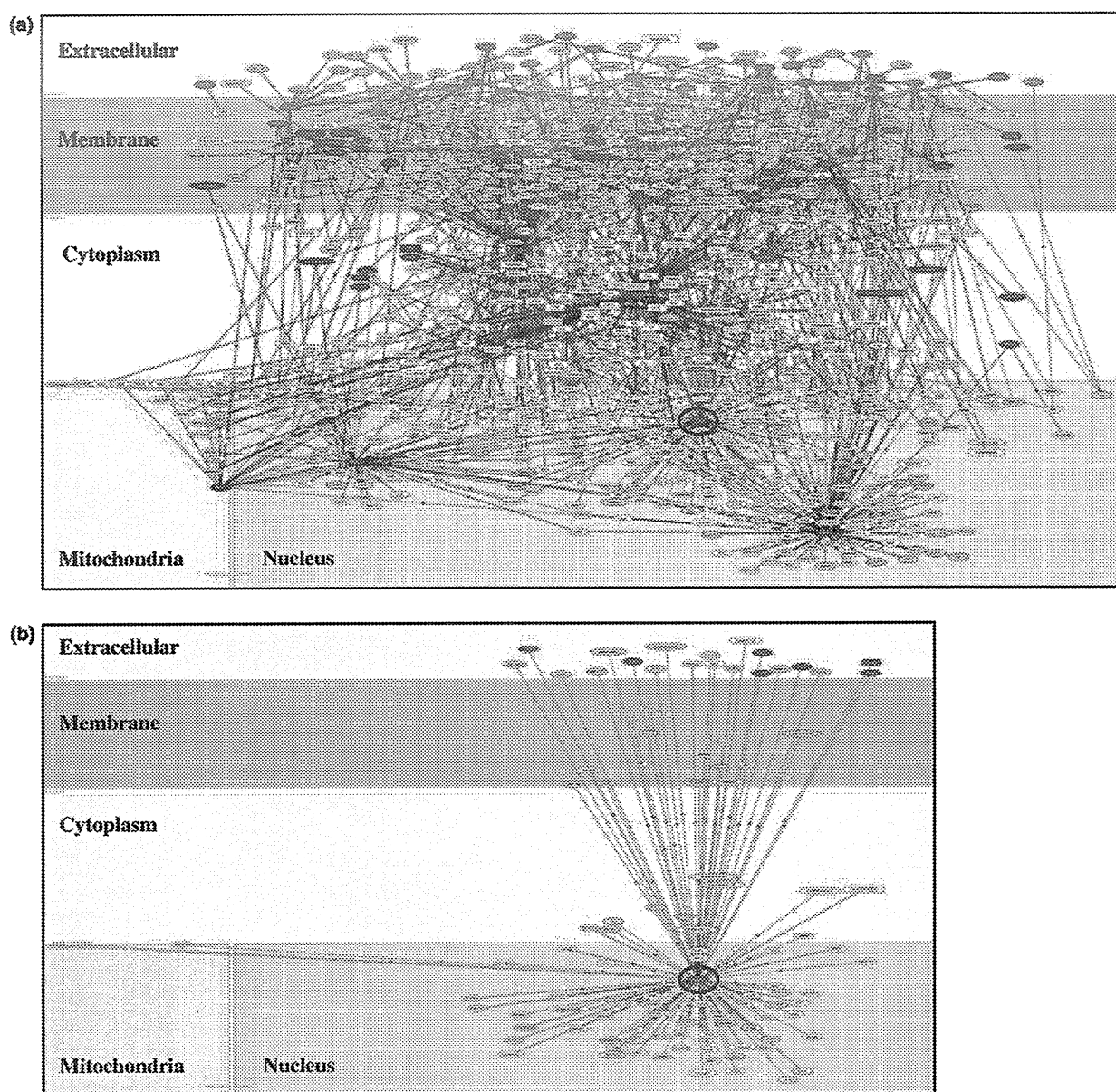
Figure 2 Molecular network of 91 MS-linked molecules. (a) By importing 91 MS-linked molecules into KeyMolnet, the neighboring search within one path from starting points generates the highly complex molecular network composed of 913 molecules and 1005 molecular relations. (b) The extracted network shows 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

135

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 Molecules constucting the transcriptional regulation by vitamin D receptor network

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

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.

138

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.

140

総　説
特集：ヒト免疫疾患研究の新展開—From clinic to bench

多発性硬化症病変分子のネットワーク解析

佐 藤 準 一

Molecular Network Analysis of Multiple Sclerosis Brain Lesion Proteome

Jun-ichi SATOH

Department of Bioinformatics, Meiji Pharmaceutical University, 2-522-1 Noshio, Kiyose, Tokyo 204-8588, Japan

summary

A recent proteomics study of multiple sclerosis (MS) brain lesion–specific proteome profiling clearly revealed a pivotal role of coagulation cascade proteins in chronic active demyelination (Han MH et al. *Nature* 451 : 1076–1081, 2008). However, among thousands of proteins identified, nearly all of remaining proteins were left behind to be characterized in terms of their implications in MS brain lesion development. By the systems biology approach using four different pathway analysis tools of bioinformatics, we studied molecular networks and pathways of the proteome dataset of acute plaque (AP), chronic active plaque (CAP), and chronic plaque (CP). The database search on KEGG and PANTHER indicated the relevance of extracellular matrix (ECM)–mediated focal adhesion and integrin signaling to CAP and CP proteome. IPA identified the network constructed with a wide range of ECM components as one of the networks highly relevant to CAP proteome. KeyMolnet 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. Although four distinct platforms produced diverse results, they commonly suggested a role of ECM and integrin signaling in development of chronic lesions of MS. These observations indicate that the selective blockade of the interaction between ECM and integrins would be a rational approach for designing inhibitors of chronic inflammatory demyelination in MS brain lesions.

Key words——KeyMolnet; molecular network; multiple sclerosis; proteome; systems biology

抄　録

多発性硬化症（multiple sclerosis；MS）は，中枢神経系白質に炎症性脱髄巣が多発し，様々な神経症状が再発を繰り返す難病である．MS では，自己反応性 Th1 細胞や Th17 細胞が血液脳関門を通過して脳や脊髄に浸潤し，マクロファージやミクログリアを活性化して，脱髄を惹起する．炎症が遷延化すると軸索傷害を来して不可逆的機能障害が残存する．現在まで，髄鞘や軸索の再生に有効な治療薬は開発されていない．最近，MS 脳病巣の網羅的プロテオーム解析データが報告された（Han MH et al. *Nature* 451 : 1076-1081, 2008）．Han らはステージを確認した病巣から laser microdissection で分離したサンプルを質量分析で解析して，4324 種類のタンパク質を同定した．彼らは慢性活動性脱髄巣（chronic active plaque；CAP）における血液凝固系の亢進を見出し，抗凝固薬を用いて MS 動物モデル自己免疫性脳脊髄炎の治療に成功した．しかしながら，凝固系以外の多くのタンパク質に関しては，MS 脳分子病態における意義は明らかではない．われわれは，Han らのデータセットを分子ネットワーク解析ツール KEGG, PANTHER, IPA, KeyMolnet を用いて再解析し，MS 脳病巣プロテオームの主要分子ネットワークを調べた．その結果，CAP における extracellular matrix (ECM)-integrin ネットワークの中心的役割を発見した．すなわちシステム生物学の観点からは，ECM-integrin シグナル伝達系は，MS における炎症性脱髄遷延化抑制のための創薬標的パスウェイとなる可能性がある．

I. はじめに

多発性硬化症（multiple sclerosis；MS）は，中枢

神経系白質に炎症性脱髄巣が多発し，様々な神経症状が再発を繰り返して進行する難病である．MS では，遺伝的要因と環境因子の複雑な相互作用を背景に出現した活性化自己反応性 Th1 細胞や Th17 細胞が，血液脳関門（blood-brain barrier；BBB）を

明治薬科大学バイオインフォマティクス

通過して脳や脊髄に浸潤し，マクロファージやミクログリアを活性化して，TNFα などの proinflammatory mediator の産生を誘導し，脱髄を惹起すると考えられている[1]．回復期には髄鞘再生を認めるが，炎症が遷延化すると軸索傷害を来して不可逆的機能障害が残存する．MS では，急性増悪期に intravenous methylprednisolone pulse (IVMP) を行い，寛解期に IFNβ の継続的投与を行う方法が，最も一般的な治療法として選択されている．しかし IFNβ に対する nonresponder や副作用のため投与出来ない症例も存在する．MS は臨床経過から再発寛解型 (relapsing-remitting MS；RRMS)，2 次進行型 (secondary-progressive MS；SPMS)，1 次進行型 (primary-progressive MS；PPMS) に分類され，病理学的には T 細胞浸潤，抗体沈着，オリゴデンドロサイトアポトーシスの所見により 4 型に分類されており，このような不均一性 (heterogeneity) の存在も治療難航の一因となっている[2]．現在まで，髄鞘や軸索の再生を促進する治療薬はなく，新規の標的分子に対する画期的な創薬が待望されている．

2003 年に全ヒトゲノムの解読が完了し，個々の細胞における全遺伝子やタンパク質の発現情報をルーチンに解析可能なポストゲノム時代が到来し，創薬研究の中心はゲノム創薬へとパラダイムシフトした．このようなオミックス研究により，癌や神経難病の診断バイオマーカーや治療標的分子が次々明らかにされた．さらに薬理ゲノミクスの分野は急成長を遂げ，薬物応答性個人差をある程度予測可能となり，テーラメイド医療 (personalized medicine) の樹立に道が開かれた．またシステム生物学 (systems biology) の観点からは，ヒトは大規模な分子ネットワークで精密に構築された複雑系であり，多くの難病がシステムの持つロバストネスの破綻に起因すると考えられている[3]．従って難病の病態解明のためには，オミックス解析に直結したゲノムワイドの分子ネットワーク解析が必須の研究手段となりつつある[4]．

最近，MS 脳病巣の網羅的プロテオーム解析データが報告された[5]．この研究では，種々のステージの MS 脳病巣から 4324 種類のタンパク質を同定した．彼らはその中から慢性活動性脱髄巣における血液凝固系の亢進を見出した．その所見に基づき，抗凝固薬を用いて，MS 動物モデルである自己免疫性脳脊髄炎 (experimental autoimmune encephalomye-

litis；EAE) の治療に成功し，膨大なプロテオミクスデータから新規創薬標的候補を同定出来た．しかしながら，大多数を占める凝固系以外のタンパク質に関しては，MS 脳分子病態における意義は明らかにされていない．われわれは彼らのデータセットを利用して，分子ネットワーク解析ツール KEGG，Panther，KeyMolnet，IPA を用いて再解析し，MS 脳病巣プロテオームの主要分子ネットワークを調べ，システム生物学の観点から創薬標的の探索を試みた[6]．

II. MS 脳病巣の網羅的プロテオーム解析

2008 年，Han らは 6 例の MS 凍結脳を用いて，病理学的にステージを確認した脳病巣から laser microdissection で採取したサンプルを SDS-PAGE で分離後に，タンパク質を抽出し，トリプシン消化ペプチド断片を質量分析で解析した[5]．ステージは，炎症性細胞浸潤と浮腫を主徴とする急性脱髄巣 (active plaque；AP)，脱髄巣辺縁部に炎症が限局している慢性活動性脱髄巣 (chronic active plaque；CAP)，炎症所見に乏しくアストログリアの瘢痕形成を主徴とする慢性非活動性脱髄巣 (chronic plaque；CP) に分類した．同時に 2 例の健常脳に関しても質量分析で解析した．その結果，AP から 1082，CAP から 1728，CP から 1514，合計 4324 種類のタンパク質を同定した．さらに INTERSECT プログラムを用いて，健常脳では検出されずかつステージ特異的なタンパク質を選び出し，AP 158，CAP 416，CP 236 種類のタンパク質データを公開した．彼らは PROTEOME-3D を用いてアノテーションを調べた結果，CAP において血液凝固系タンパク質 protein C inhibitor，tissue factor，thrombospondin 1，fibronectin 1，vitronectin の発現を認めた．この所見に基づいて，抗凝固薬である thrombin inhibitor hirudin および activated protein C を用いて，MS 動物モデルである PLP139-151 ペプチド誘導性 SJL/J マウス EAE を治療した．どちらの抗凝固薬も，脾細胞やリンパ節細胞の増殖と IL-17，TNFα 産生を抑制した．以上の結果より，血液凝固系タンパク質は新規創薬標的分子となることが明らかになった．しかしながらどのような経緯で，膨大なプロテオミクスデータから上記 5 種類のタンパク質を選出し，焦点を絞ったのかに関しては記載がない．また大多数を占める凝固系以外のタンパク質に関しては，MS 脳分子病態における意義は明らかにされていない．

III. MS 脳病巣プロテオームデータの分子ネットワーク解析

生体では，タンパク質は複雑なシステムを構築しているので，病態の解明には個々のタンパク質の機能解析のみならず，タンパク質が構築する分子ネットワークやパスウェイの同定も重要である．タンパク質間相互作用 (protein-protein interaction ; PPI) には，単純な直接的結合関係のみならず，活性化，抑制，運搬，酵素反応，複合体形成など多彩な相互作用様式が存在する．複雑多岐のオミックスデータに関連している分子ネットワークを解析するためには，精査された文献情報に裏付けられた専用の解析ツールを使う必要がある．すなわち，膨大な文献情報から様々な分子間相互作用を抽出し，信頼性の高い知識を整理してコンテンツとして収録したデータベース (knowledgebase) を利用して，既知のどのネットワークやパスウェイに最も高い類似性を呈しているかについて調べる方法である．Web 上でフリーに利用出来る代表的な knowledgebase としては，Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.kegg.jp)[7]，the Protein Analysis Through Evolutionary Relationships (PANTHER) classification system (www.pantherdb.org)[8]，Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (string.embl.de)[9] などがある．特に KEGG と PANTHER は，curator と呼ばれる専門家によって精査された遺伝子や代謝物などの情報を統合しており，2010 年 4 月現在，KEGG PATHWAY には 352 reference pathways から構成される 104,520 パスウェイが登録されており，search objects in pathways ボックスに目的分子の KEGG ID を入力することにより，該当するパスウェイを検索出来る．PANTHER では reference set との比較により，類似性に関する統計学的有意差を多重検定で評価出来る．STRING では KEGG，HPRD，BIND，IntAct の情報も統合して収録しており，PubMed アブストラクトからは自然言語処理 (natural language processing) によるテキストマイニングを介して，分子間相互作用に関する情報を収集している．有償ツールとしては，Ingenuity Pathways Analysis (IPA) (Ingenuity Systems, Redwood City, CA) (www.ingenuity.com) や KeyMolnet (Institute of Medicinal Molecular Design, Tokyo) (www.immd.co.jp) などがある．どちらも専門家

が精選された文献を精読して，信頼性の高い分子間相互作用に関する情報を収集しており，定期的にアップデートされている．KeyMolnet は日本語入力にも対応しており，結合・発現制御・複合体形成を包括的に調べる周辺検索 (neighboring search)，発現制御に関与する転写因子群を調べる共通上流検索 (common upstream search)，始点と終点間のネットワークを調べる始点終点検索 (N-points to N-points search)，複数の端点を始点として，最多数の始点を含む最小の分子ネットワークを調べる相互関係検索 (interrelation search) を，検索法として選択出来る[10]．

著者らは，Han らの MS 脳病巣 AP 158, CAP 416, CP 236 プロテオームデータ[5]に該当する UniProt ID を，Entrez Gene ID および KEGG ID に変換して，KEGG, PANTHER, IPA, KeyMolnet に入力し，それぞれのステージ特異的プロテオームデータを最も良く反映している分子ネットワークを同定した[6]．ID 変換には UniProt (www.uniprot.org) と KEGG Identifiers を用いたが，DAVID Bioinformatics Resources (david.abcc.ncifcrf.gov)[11]の Gene ID conversion ツールを利用しても，Entrez Gene ID への一括変換は容易に行える．なお DAVID はプロテオミクスデータのアノテーション解析の際にも非常に有用なツールである[12]．

KEGG による解析では，CAP プロテオームと focal adhesion (hsa04510), cell communication (hsa01430), ECM-receptor interaction (hsa04512), CP プロテオームと focal adhesion (hsa04510) との関連性が示唆された．CAP プロテオームの focal adhesion (hsa04510) は，COL1A1, COL1A2, COL5A2, COL6A2, COL6A3, FN1, LAMA1, MYLK, SHC3, PPP1CA, PARVA, PRKCB1, MYL7, RAC3, SPP1, SRC, THBS1, VTN から構成され，CP プロテオームの focal adhesion (hsa04510) は，COL4A2, COL6A1, CRK, FYN, ITGA6, LAMB2, LAMC1, PIK3CA, ZYX から構成されていた．PANTHER による解析では，CAP プロテオームと inflammation mediated by chemokine and cytokine signaling pathway (p=2.63E-03)，integrin signaling pathway (p=3.55E-03) (図 1)，CP プロテオームと integrin signaling pathway (p=4.33E-02) との関連性が検出された．すなわち KEGG と PANTHER の解析から，MS 慢性病巣における ECM-integrin シグナル伝達系の中心的役割
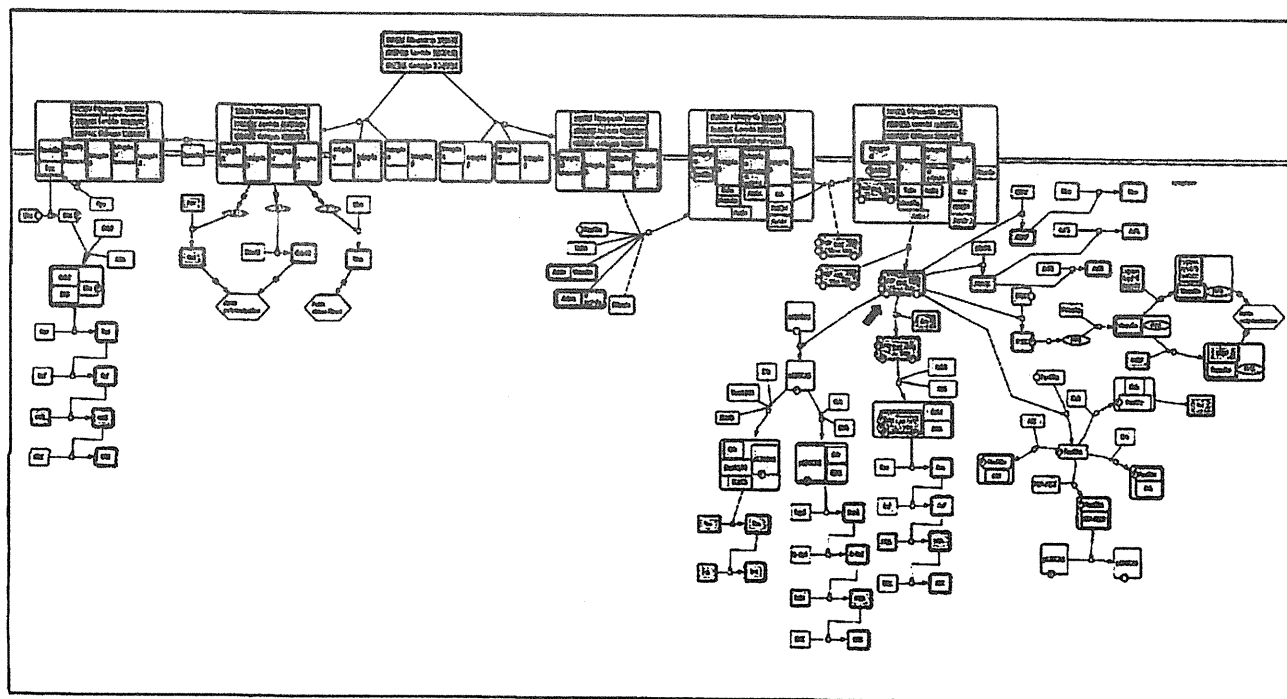
図1　PANTHER による MS 脳病巣 CAP プロテオームの分子ネットワーク解析

MS 脳病巣 CAP プロテオームの PANTHER による解析では，integrin signaling pathway との関連性が示唆された（p=3.55E-03）．Reference pathway 上の分子とヒットしたタンパク質を濃いシャドウで示す．Focal adhesion kinase (FAK) が，ネットワークのハブ（矢印）となることがわかる．文献6）より改変．

が示唆された．KEGG と PANTHER の解析では，AP プロテオームと密接に関連するパスウェイは検出されなかった．

一方 IPA core analysis による解析では，AP プロテオームは cellular assembly and organization, cancer, and cellular movement (p=1.00E-49), CAP プロテオームは dermatological diseases and conditions, connective tissue disorders, and inflammatory disease (p=1.00E-47), lipid metabolism, molecular transport, and small molecule biochemistry (p=1.00E-47), CP プロテオームは cell cycle, cell morphology, and cell-to-cell signaling and interaction (p=1.00E-50) との関連性を認めた．CAP プロテオームの dermatological diseases and conditions, connective tissue disorders, and inflammatory disease ネットワークは，BGN, CHI3L1, CNN2, COL1A1, COL1A2, COL6A2, COL6A3, CXCL11, ENTPD1, ERK, FBLN2, FERMT2, FN1, GBP1, HSPG2, Ifn gamma, INPP5D, Integrin, LAMA1, LUM, Mlc, MYL7, MYL6B, NES, P4HA1, Pak, PARVA, POSTN, PRELP, SERPINA5, SERPINH1, Tgf beta, TGFBR3, THBS1, VTN から構成されており，ECM-integrin 相互作用の関与を強く示唆している．最後に KeyMolnet に収録されている MS 関連

75 分子を始点，ステージ特異的プロテオームの各分子を終点として，最短経路で始点終点検索を施行したところ，非常に複雑な分子ネットワークが抽出された[6]．AP プロテオームは IL-4 signaling pathway (p=1.79E-13)，CAP プロテオームは PI3K signaling pathway (p=7.25E-18)，CP プロテオームは IL-4 signaling pathway (p=1.04E-16) と最も密接に関連していた．また CAP と CP は integrin signaling pathway との関連性も認め（p=2.13E-12 および p=2.57E-12），他にも様々なサイトカインシグナル伝達系との関連性も見られた．

IV.　MS 脳病巣プロテオームネットワークの創薬標的分子

上述のように，MS 脳病巣網羅的プロテオームデータセット[5]に関して，4種類の異なる分子ネットワーク解析ツール KEGG, PANTHER, IPA, KeyMolnet は，様々な独自の分子ネットワークを抽出したが，共通して CAP, CP における ECM-integrin シグナル伝達系の中心的役割が示唆された[6]．さらに CAP プロテオームデータを STRING で解析したところ，描画された複雑な分子ネットワーク中に，ECM のクラスターを同定することが出来た（図2）．
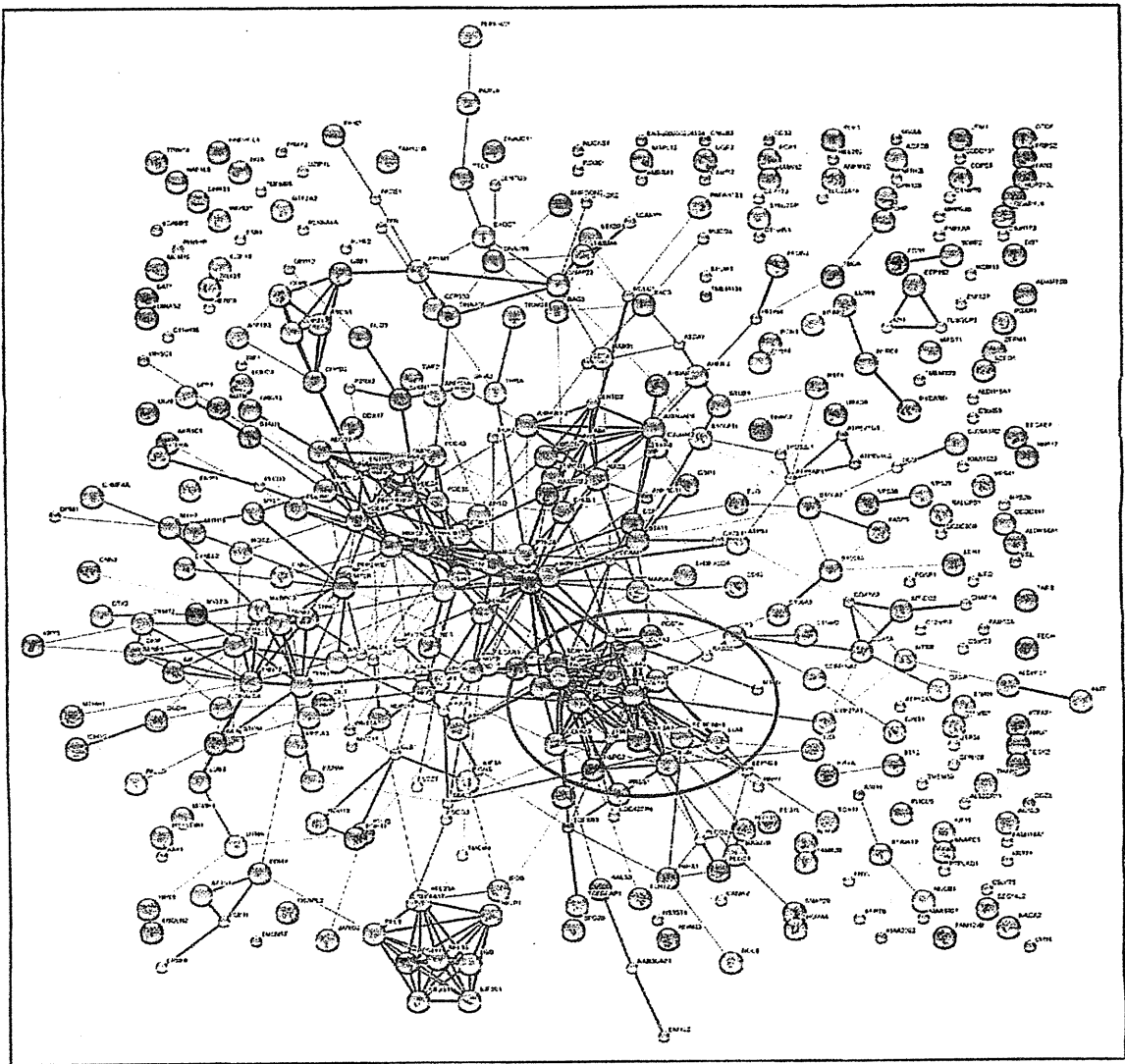
図2　STRING による MS 脳病巣 CAP プロテオームの分子ネットワーク解析
　MS 脳病巣 CAP プロテオームの STRING による解析では，複雑な分子ネットワーク中に ECM のクラスター（赤楕円）が存在するのがわかる．

　Integrin は複数の α, β サブユニットから構成される 24 種類のヘテロダイマータンパク質であり，細胞外基質 ECM のリガンドである．例えば β1 integrin ファミリーは collagen, fibronectin, laminin と結合し，αv integrin ファミリーは vitronectin と結合する．細胞骨格動態制御を介する細胞の接着，遊走，分化，増殖には，ECM-integrin 間の相互作用を介する outside-in, inside-out シグナルが必須である[13]．MS 脳病巣に集積を認めた fibronectin や vitronectin は，主として破綻した BBB を通過して脳実質に浸透した血漿成分に由来する．ECM, integrin が著増している慢性病巣において髄鞘や軸索の再生が乏しい理由として，グリア瘢痕に含まれている ECM タンパク質が再生阻害因子として働く可能性や，活性化マクロファージやミクログリアが産生

する種々のタンパク質分解酵素が ECM に結合して長期に保持され，髄鞘崩壊が遷延化している可能性が挙げられている[14,15]．また ECM-integrin 間の相互作用は，リンパ球のホーミングや血管外遊出，アストログリア・ミクログリアの活性化，オリゴデンドログリア前駆細胞の分化の抑制を介して，脱髄と軸索傷害を増強する[16,17]．
　MS における臨床試験では，α4β1 integrin（VLA4）に対するヒト化モノクローナル抗体 natalizumab が再発抑制に著効を呈した．しかしながら，natalizumab は進行性多巣性白質脳症（progressive multifocal leukoencephalopathy；PML）を惹起するので，より安全な治療薬の登場が待望されている[18]．分子ネットワークから創薬標的分子を探索する場合は，ハブ（hub）と呼ばれる，多くの分子か

らのリレーションが集中しているネットワークの中心分子を同定することが重要である．ハブの抑制薬または活性化薬は，ネットワーク全体すなわちシステムの維持に多大な影響をもたらす[19]．システム生物学の観点からは，ECM-integrin シグナル伝達系は，MS における炎症性脱髄遷延化抑制のための創薬標的パスウェイとなる可能性がある．PANTHER による CAP プロテオームの解析で同定された integrin signaling pathway においては，focal adhesion kinase (FAK) がハブとなることが明らかである（図 1 矢印）．低分子化合物 TAE226 は，ECM による FAK の自己リン酸化を選択的に抑制し，in vivo モデル系で経口投与により腫瘍細胞の増殖と血管新生を抑制する[20]．従って分子ネットワークから見ると，TAE226 は FAK を分子標的とする MS 慢性炎症性脱髄抑制薬候補となる可能性があり，EAE における前臨床試験の実施が待たれる．

V. おわりに

ポストゲノム時代の膨大なオミックスデータに関する分子ネットワーク解析のためには，精査された文献情報（knowledgebase）に基づく解析ツールを使う必要がある．解析ツールは未だ発展途上・日進日歩であり，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・組織特異的発現，細胞内局在化，動的な特性に関しては十分対応出来ていない．しかしながら生体をシステムとして捉える見方から，分子ネットワークを解析することにより，初めて論理的な仮説に裏付けられた創薬標的分子・分子ネットワークを効率的に同定することが出来る．

文　献

1) McFarland HF, Martin R. : Multiple sclerosis : a complicated picture of autoimmunity. *Nat Immunol* 8 : 913-919. 2007.

2) Lassmann H, et al. : The immunopathology of multiple sclerosis : an overview. *Brain Pathol* 17 : 210-218, 2007.

3) Kitano H. : A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* 6 : 202-210, 2007.

4) Quintana FJ, et al. : Systems biology approaches for the study of multiple sclerosis. *J Cell Mol Med* 12 : 1087-1093, 2008.

5) Han MH, et al. : Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature* 451 : 1076-1081, 2008.

6) Satoh JI, et al. : Molecular network of the comprehensive multiple sclerosis brain-lesion proteome. *Mult Scler* 15 : 531-541, 2009.

7) Kanehisa M, et al. : KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38 : D355-D360, 2010.

8) Mi H, et al. : PANTHER version 7 : improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res* 38 : D204-D210, 2010.

9) Jensen LJ, et al. : STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37 : D412-D416, 2009.

10) Sato H, et al. : New approaches to mechanism analysis for drug discovery using DNA microarray data combined with KeyMolnet. *Curr Drug Discov Technol* 2 : 89-98, 2005.

11) Huang da W, et al. : Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4 : 44-57, 2009.

12) Satoh J, et al. : Protein microarray analysis identifies human cellular prion protein interactors. *Neuropathol Appl Neurobiol* 35 : 16-35, 2009.

13) Luo BH, et al. : Structural basis of integrin regulation and signaling. *Annu Rev Immunol* 25 : 619-647, 2007.

14) Sobel RA. : The extracellular matrix in multiple sclerosis lesions. *J Neuropathol Exp Neurol* 57 : 205-217, 1998.

15) van Horssen J, et al. : The extracellular matrix in multiple sclerosis pathology. *J Neurochem* 103 : 1293-1301, 2007.

16) Sisková Z, et al. : Fibronectin impedes "myelin" sheet-directed flow in oligodendrocytes : a