Form 3-3-2

Japan-US Brain Research Cooperation Program
The Report of Information Exchange Seminar in 2004 fiscal year
[field :③ Research on the neural mechanisms of emotion and memory]

1. The Seminar Title: Bioinformatic Analysis of Brain Function

2. The Term : From Y. 2004 M. 9 D. 29 To Y. 2004 M. 10 D. 2

3. The Location: Hilton Waikoloa Village, Hawaii (USA)

4. The Representative's Name, Title and Affiliation:
   Japanese Coordinator:
   Kazuhiro Ikenaka, Ph.D. Professor National Institute for Physiological Sciences
   US Coordinator:
   Karoly Mirnics, M.D. Assistant Professor, University of Pittsburgh

5. The Participants:
   Japan: The Invited participants _10 people The others _0 people
   Name, Title and Affiliation of the Invited participants
   As stated on another sheet

   US : The Invited participants _15 people The others _2 people
   Name, Title and Affiliation of the Invited participants
   As stated on another sheet

6. The Abstract and the Significance of this seminar (300 words):

The central theme of this workshop was integrating technologies and knowledge from "-omics" fields and sharing bioinformatics approaches with other disciplines of neuroscience such as imaging and neurophysiology. Speaker presentations of 30 minutes were followed by lively 15 minute discussions which were centered around the most critical aspects of follow-up, sharing, integration and interpretation of the emerging datasets. The presentations were divided in three thematic sessions.

Session I. **Linking the Neural Transcriptome with Existing Databases to Uncover New Facets of Brain Disease (Chair: Dr. Lynn Hudson).**

Session II. **Integration of Neuroimaging and Behavioral Data into Molecular Maps of the Brain. (Chair: Dr. Karoly Mirnics)**

Session III. **Superimposing the Glycome and Proteome onto the "-omics" Palette to Move from Molecules to Function in the Nervous System. (Chair: Dr. Kazuhiro Ikenaka)**

We are planning to submit a scientific report from the workshop to one of the eminent neuroscience journals. (see the attached file)

Beyond the formal scientific exchanges there was a significant interaction between speakers and participants that occurred outside the meeting room. These discussions were also an important aspect of the workshop as they may lead to establishing novel collaborations and build bridges between the US and Japanese neuroscience communities. On the organizational note, Capital Management did a superb job in planning, coordinating and seamlessly running the workshop.

In summary, the workshop was a great success.

7. The Result of this seminar and the results expected (300 words):

1. Emerging databases must become transparent, interactive, easily searchable and must integrate archival data. Retrieval of data from multiple linked databases simultaneously is essential for interpretation of the "omic" data. These databases must allow searches based on virtually any criteria, including (but not limited to) cell type, electrophysiological property, brain region or cellular function. Integration of archival anatomical, physiological, biochemical and other data accumulated over years must become one of the top priorities in building the new databases.

2. Careful experimental design is critical and should be tailored in the context of the scientific questions asked. It is important to realize that "one shoe does not fit all", especially in "omics" research. Enforcing strict standards on the end of experimental design, sampling dynamics, platforms used or analytical approaches implemented may impede our ability to ask the scientific questions the right way. However, *reporting* information obtained in the experiments should be standardized, and further evolvement of the MIAME/MGED standards should proceed in parallel with the development of the "omics" platforms.

3. Imaging, neurophysiology and "omics" fields are facing many of the similar biostatistic challenges. These involve datasets that are inherently noisy and often require complex pre-processing procedures to uncover the meaningful data. As a result, the potential for obtaining false positive data is high in all high-throughput methods, and appropriate methods must be implemented to estimate the false discovery ratio. We need to develop innovative, open source statistical approaches to correlate the data from multiple fields of neuroscience. These tools should be simple enough to be used by neuroscientists with a basic statistical background, and should be downloadable from the data repositories themselves. A big difference between neuroimaging data and "omics" data is that the latter does not typically have cell-type information in each voxel, which makes higher order analysis of the raw data difficult. Dissection of individual neural cell types (e.g. oligodendrocytes) or structures (e.g. myelin sheath), completion of the protein-protein interaction analysis project, and detailed analysis of transcription factor expression in the brain should be useful adjuncts for defining the neural transcriptome or proteome and supplying second order information on regulatory networks.

4. Metabolomics, lipidomics and glycomics must evolve in parallel with the transcriptome and proteome studies. The classic dogma of DNA-RNA-protein does not fully address the physiological processes of the brain. Metabolic events, glycoprotein synthesis, lipid dynamics are susceptible to environmental influences and play a critical

role in normal homeostasis and pathological states. At present, the expression pattern of the sugar chains (glycomic data) is difficult to predict solely by transcriptomic or proteomic data.

5. Reference "omics" datasets are needed for a meaningful interpretation of the experimental data. Brain regional mapping of transcriptome factors, "omics" and electrophysiological signatures of individual cell types are amongst the many useful tools that will allow more meaningful data mining, which ultimately leads to testable hypotheses.

6. Use of data from repositories should be governed by a set of guidelines. While the imaging and electrophysiology community has already established some guidelines, there is much less clear consensus in the "omics" fields about the rules of data sharing. We need to start a broad discussion in the neuroscience community about a number of critical questions such as " Does reanalysis of my dataset by other scientists entitle me for authorship?"..."Can I combine raw data from different data repositories to create a new manuscript?"... "Is combining raw data from different experimental series confounded to the degree that it does not represent good science?"... "My whole raw dataset is used in a meta study by a group of scientists. Do they need my permission for analysis and publishing the data?"

7. Biomedical and pharmaceutical research should participate in the data integration process. Industrial entities are routinely generating, managing, mining and integrating datasets from multiple sources. Discovery tools that are not commercialized in industrial research should become part of a public domain, where they can be further developed as an open source. Furthermore, industrial profiling data that is not used for patent applications (for example, experiments on control tissue) should be included in publicly available databases.

8. "Omics" data must be obtained more systematically. The molecular signatures of different subcellular compartments, cell types, layers, nuclei and brain regions are unique. For meaningful interpretation of the "omics" data the resolution of the technologies must increase and go beyond profiling the bulk brain tissue. In this context, efforts to obtain molecular signatures at cellular and subcellular levels should be promoted and refined.

9. It is all about biology. It is important to emphasize that in today's science bioinformatics is an integral part of any scientific discovery: classic molecular biology - often combined with reductionist approaches - can generate valuable findings and follow-up with simple informatics tools (e.g. sequence analysis or prediction of domains) to formulate biological hypotheses about the new discovery. In contrast, neuroimaging and "omics" analysis of human brain use complex statistical procedures to extract biological discovery from the data sets. As a result, the novel data-driven and the classic hypothesis-driven research are complementary approaches that depend on each other. They should exchange information, and the nature of the questions asked and the available *a priori* knowledge will define which approach may be more appropriate.

8. The Others (Practical Issues, Special Mention Matters):

Interactions between the Japanese and US neuroscientists must further evolve.

Although there has been an improvement in the contacts and collaboration between the

scientists of the two countries over the last several years, there is much left to accomplish. This must be a true, equal commitment partnership, and the BRCP program is the ideal framework to ensure the further evolvement of joint collaborative projects, information exchange and scientist exchange between the two neuroscience communities. The size of the meeting was ideal to facilitate intensive discussion and initiation of the future collaborations.

The Participants (Name, Title and Affiliation of the Invited participants)

Japan:
National Institute for Physiological Sciences Professor Kazuhiro Ikenaka
National Institute for Physiological Sciences Professor Norihiro Sadato
Tokyo Metropolitan Institute of Gerontology Group Leader Tosifusa Toda
Riken Brain Science Institute Research Scientist Hiroyuki Nakahara
Riken Genomic Sciences Center Team Leader Harukazu Suzuki
Institute of Statistical Mathematics Professor Tohru Ozaki
National Institute of Advanced Industrial Science and Technology Team Leader Tohru Natsume
National Institute of Advanced Industrial Science and Technology Team Leader Hisashi Narimatsu
Faculty of Engineering Kagawa University Research Associate Takanori Kochiyama
Nara Institute of Science and Technology Graduate Student Ryoko Morioka

US :
School of Medicine USU Denes V. Agoston
McLean Hospital Director David Borsook
National Institute of Medical Health, NIH Robert W. Cox
University of Pennsylvania Medical Center Professor Jim Eberwine
Weill Medical College of Cornell University Head Daniel Gardner
NYU School of Medicine Professor Esther P. Gardner
Reed Neurological Research Center UCLA Daniel Geschwind
National Institute of Neurological Disorders and Stroke NIH Chief Lynn Hudson
University of Pittsburgh Professor Davis A. Lewis
University of Pittsburgh Assistant Professor and Director Karoly Mirnics
University of Connecticut Medical School Professor Steve Pfeiffer
University of British Columbia Research Associate Julie Pongrac
UCLA School of Medicine Desmond J. Smith
The Scripps Research Institute Professor J. Gregor Sutcliffe
Dartmouth College Associate Professor John Van Horn

MEETING REPORT
US-Japan Brain Research Cooperation Program (BRCP) 2004
"Bioinformatic Analysis of Brain Function" Workshop
September 29 – October 3, 2004
Hilton Waikoloa Village, Hawaii, USA

I. Overview

The central theme of this workshop was integrating technologies and knowledge from "-omics" fields and sharing bioinformatics approaches with other disciplines of neuroscience such as imaging and neurophysiology. Speaker presentations of 30 minutes were followed by lively 15 minute discussions which were centered around the most critical aspects of follow-up, sharing, integration and interpretation of the emerging datasets. The presentations were divided in three thematic sessions (see enclosed program) chaired by Drs. Lynn Hudson, Kazuhiro Ikenaka and Karoly Mirnics. We are planning to submit a scientific report from the workshop to one of the eminent neuroscience journals.

Beyond the formal scientific exchanges there was a significant interaction between speakers and participants that occurred outside the meeting room. These discussions were also an important aspect of the workshop as they may lead to establishing novel collaborations and build bridges between the US and Japanese neuroscience communities. On the organizational note, Capital Management did a superb job in planning, coordinating and seamlessly running the workshop.

In summary, the workshop was a great success.

II. Presentations and discussion

Although there were some last minute speaker cancellations (Drs. Barlow, Lockhart and Storey) both the Japanese and US speaker lineups were outstanding. A brief summary of the presentations is enclosed as follows:

Session I. **Linking the Neural Transcriptome with Existing Databases to Uncover New Facets of Brain Disease (Chair: Dr. Lynn Hudson)**

1. Dan Gardner, Weill Medical College of Cornell University

"Overview: What the Human Brain Project and SFN's 'Neuroscience Gateway' Contribute to Neuroscience Bioinformatics"

Dr. Gardner gave an overview of the goals and achievement of the Human Brain Project to date. In his talk he emphasized the importance of defining common descriptors across different databases in order to achieve the desired integration of the data. The 'Neuroscience Gateway' introduced by the Society for Neuroscience (SFN) is another response to the challenge of the "4 ½ dimensions" of interoperability. Discussion focused on the usefulness of implementing common experimental standards and the importance of integrating archival data into the newly created databases.

2. Jim Eberwine, University of Pennsylvania

"Neuronal Gene Expression – Integrative Activity of RNA Binding Proteins"

This talk focused on the diversity of the transcriptome in individual neuronal compartments (e.g. dendrites), pointing out that splicing and protein synthesis also occurs in dendrites, which makes the interpretation of the profiling experiments even more challenging. Furthermore, Dr. Eberwine suggested that from all the potential datasets, currently the transcriptome is the best descriptor of the cell phenotype. The discussion was focused on the need to obtain this data systematically; moreover, databases must be populated with this information for the proper interpretation of the "omics" datasets.

3. Hiroyuki Nakahara, Riken Brain Science Institute

"Information Geometric Approach to Decipher the Functions of Many 'Variables': Neural Population Network and Gene Regulatory Network"

Dr. Nakahara pointed out that the number of variables in both electrophysiological and microarray analysis is high and that reducing the dataset to two populations will lead to loss of critical data. He presented a comprehensive approach to deciphering electrophysiological properties of neural populations using higher order interaction modeling. As the information geometric measures can be successfully used to systematically examine spike firing patterns of diverse neuronal populations, they can be also very useful for analyzing information obtained in the "omics" platforms, where information is also derived from a number of different cellular phenotypes. The discussion focused on the causality vs. correlative nature of changes that are observed across the "omics" and physiology/imaging data sets. The participants agreed that these novel approaches are critical for generating novel hypotheses that can be tested in follow-up experiments.

4. Lynn Hudson, NINDS, NIH

"Dissecting out the Glial Contribution to the Neural Transcriptome"

Following up on Dr. Eberwine's approach of tackling transcriptome diversity at a subcellular level, this talk focused on parsing out individual contributions at the cellular level. Cells of the oligodendrocyte lineage were sorted by flow cytometry prior to expression profiling. Varying stages of development of a given lineage as well as distinct neural lineages can be directly purified using either antibodies directed against surface antigens or GFP-labeled cells from transgenic mice, animals which are being made publicly available through the GENSAT project initiated by NINDS. Discussion centered on how such approaches can address the tremendous cellular heterogeneity of the nervous system.

5. Greg Sutcliffe, Scripps Research Institute

"Discovery and characterization of Hcrt1 in the brain"

Dr. Sutcliffe presented an example how to identify and characterize differentially expressed genes by subtractive hybridization. Along with dozen of other transcripts, *Hcrt1* was identified as a promising candidate for follow-up based on it's predicted structure. The discussion centered around importance of following up findings at a functional level and prioritizing of gene lists for further investigation using already existing knowledge bases. Views were also exchanged as to what added benefits bioinformatics can bring to traditional approaches for establishing gene function.

6. Karoly Mirnics, University of Pittsburgh

"Transcriptome changes in schizophrenia: deficits across multiple functional systems"

This presentation reviewed human postmortem transcription profiling efforts in psychiatric disorders. Furthermore, it was pointed out that many of the observed transcriptome changes might be part of a wide, co-regulated transcript network that is defined by common

regulatory elements in the DNA sequence. The discussion was centered around specificity of the identified expression changes vis a vis other disorders and effects of drug treatment on the human brain.

7. David A. Lewis, University of Pittsburgh

"Transcriptome Abnormalities in Schizophrenia: What Do They Mean?"

The presentation of Dr. Lewis highlighted how to use anatomical knowledge to attribute transcriptome changes in the context of the neuronal phenotype in subjects of schizophrenia. The presentation and following discussion focused on the tremendous phenotypic variability of cortical neurons and discussed the GAD67-GABA receptor-Parvalbumin expression deficits in the context of the connectivity of interneurons and their inhibitory action on the pyramidal cells.

8. Daniel Geschwind, UCLA

"From Gene Expression Patterns to Function"

Dr. Geschwind used a combination of microarray screening and *in vitro* and *in vivo* confirmation in animal models to discover and characterize key genes in biological pathways of interest. He presented a regional analysis of wild type and P301L mutant mice, as well as transcriptome profiling signatures of stem cells. Finally, he directed attention to gene expression networks in human brain and pointed out that microarray findings follow a scale free topology and are organized into modules that can be readily identified. Analysis of gene transcript networks can identify biology that is relevant to the disease and is not detected by traditional microarray analysis. He pointed out that disease genes may represent molecular hubs, and as such they are appealing drug targets, a topic that was further explored in the following discussion.

9. Harukazu Suzuki, Riken Yokohama Institute, Genomic Sciences Center

"From Dynamic Transcriptome Analysis to Future Life Science"

Dr. Suzuki informed the participants about the Japanese Genome Network Project, which is a coordinated, 5-year effort focusing on identification of transcriptional regulatory networks. This effort will identify the expression profile of 1,500 mouse and 2,000 human transcription factors across the distinct anatomical structures of the brain. These efforts already revealed examples of functional tissue-specific gene networks. The results are accessible at http://genome.gsc.riken.jp/qRT-PCR/. The discussion focused on genes with low level of expression, which are hard to detect and properly quantify with both microarrays and qRT-PCR. It was also pointed out that these resources are necessary for understanding expression networks where transcript levels are interdependent.

Session II. **Integration of Neuroimaging and Behavioral Data into Molecular Maps of the Brain. (Chair: Dr. Karoly Mirnics)**

1. Desmond Smith, UCLA

"Neurogenome Makes Neurotranscriptome Makes Neuroproteome"

Dr. Smith described investigations at three levels to understand how the genome constructs the brain. At the DNA level, a new set of genome-wide congenic mouse strains is being used to map genetic loci involved in behavior. At the RNA level, voxelation and gene expression tomography are being employed to map gene expression patterns in normal and abnormal human and mouse brains at a genomic scale. At the protein level, voxelation and mass spectrometry are being used for high-throughput mapping of protein expression patterns in the brain. Discussion focused on additional approaches that can combine imaging and 3D gene expression data, and on what added value the voxelation approach may bring to the "traditional" approach that is based on neuronal circuits and/or functional grouping of neural cells.

2. Robert W. Cox, NIMH, NIH

"Processing, Analyzing, and Displaying Functional MRI Data"

Dr. Cox gave an overview of statistical and interpretational challenges in the files of functional brain imaging. He pointed out that, just like the microarray data, these images must be preprocessed and normalized, and that raw images contain little usable data. Both microarray and imaging datasets are noisy, with confounds that are hard to control for. Statistical analyses often include various assumptions, and these assumptions must be continuously re-evaluated based on newly obtained knowledge. Comparing and sharing datasets is a continuous challenge due to differences in cohorts, experimental design and equipment used. The discussion centered around the fact that special types of questions may require special experimental designs and unique analytical approaches.

3. Takanori Kochiyama, Kagawa University

"Functional Neuroimaging: Principles of Analysis"

This presentation pointed out that the aim of any imaging method is to identify brain regions that are significantly activated, while keeping the probability of both types 1 and 2 error low. As the two errors are opposite in nature, this is a significant challenge, and critical experimental design (e.g. Fisher's approach of replication, randomization and local control) is necessary to generate data that keeps true discovery to a maximum and false discovery to the minimum. He pointed out that errors are often not identically, independently and normally distributed. Dr. Kochiyama also discussed the advantages and drawbacks of the multivariate and univariate approaches, and argued that noise can be important and interesting, as it may convey specific, novel information about the system. The discussion focused on the similarities and differences in sources of variation between the microarray and imaging experiments.

4. Norihiro Sadato, National Institute for Physiological Sciences

"Functional neuroimaging: Applications to brain plasticity"

The presentation of Dr. Sadato focused on functional network reorganization due to visual or hearing loss in human. We learned that with a task-specific experimental design and analysis, one successfully answer questions regarding the activity of visual cortex in blind people during tactile tasks. Furthermore, one can assess whether the tactile activation of visual cortex in blind is due a learning effect, if it shows age-dependency, if it is functionally relevant and how it is achieved. In the discussion it was pointed out that "-omics" technologies, together with functional imaging during plasticity, might lead to significant discoveries in the future. Correlation of the two datasets may identify molecular correlates of brain function, which can be tested in follow-up experiments.

5. David Borsook, Harvard University

"What Imaging of Pain Can Contribute to Understanding Brain Function"

Dr. Borsook focused on imaging approaches to make sense of neural circuits. Assessing neuropathic pain in a focused model using a top-down approach, he found that in the trigeminal system, pain activated a number of brain structures not classically thought to be involved in conveying noxious stimuli, including the accumbens, hippocampus and amygdala. The imaging brain maps can be used as powerful tools for novel approaches of drug development, where the drug effects are assessed in relationship to the circuitry they act upon. This data, used together with animal models, genetic and "omics" information provides a solid foundation for developing and assessing the usefulness of future therapeutic approaches. Discussion focused on combining the imaging data sets with other sources of knowlege and the need for greater industrial involvement in developing and sharing bioinformatics tools.

6. Tohru Ozaki, Institute of Statistical Mathematics

"Innovation Approach to the Identification of Dynamic Causal Model for Microarray Time Series"

Dr Ozaki's presentation focused on imaging and expression profiling data as dynamic events. The dynamics of expression changes or brain activation sequences carry information about causality, and with innovative approaches one can form data-driven hypotheses about previously unknown causal relationships. He discussed prediction models such as AR, State space, neural network and chaos models, and pointed out that the dimensions of spatial dynamics are infinite. Finally, he discussed implications of dynamic noise and observation noise in relation to continuous and discrete time. Discussion focused on causality vs. correlation of the observed changes in such models, and it was pointed out that sampling rate both in microarrays (e.g. time course experiments) and imaging datasets is critical for establishing causal relationships.

7. Tohru Natsume, National Institute of Advanced Industrial Science and Technology

"Bioinformatics for Large-scale Protein-Protein Interaction Analysis"

To identify proteins that interact with a protein is crucial in understanding the function of a protein. Dr. Natsume's group is planning to express 42,000 independent full-length human cDNA clones and identify all the proteins that bind to each of the protein. For this purpose they optimized the immunoprecipitation protocol and developed nanoflow LC-MSMS method to identify immunoprecipitated proteins. So far they have finished analyzing 1,500 cDNA clones, most of which were signaling molecules. They have already been able to draw a map presenting a precise interaction of intracellular signaling molecules. When this database is accomplished, it would enable us to analyze the proteomics data at a higher order.

8. Ryoko Morioka, Nara Institute of Science and Technology

"A multivariate analysis for estimating cellular adaptations using gene expression"

This analysis emphasized detection of cellular state transition points and characterization of individual stages based on time-series global gene expression profiles in cells. The study assumed a variable regulatory system, where critical transition points defined the transcriptome as a function of time. State transitions were estimated using a Linear Dynamical System (LDS) and self-organizing maps in a multidimensional expression space. With this approach, they were able to reliably establish points of transition in B. subtilis growth and sporulation that were associated with distinct gene expression profiles. The discussion focused on how this approach can be applied to mammalian systems.

9. Dan Gardner, Weill Medical College of Cornell University

"Neuroinformatics and Neural Information Processing"

Dr. Gardner pointed out that information is encoded and transduced across the whole neural system via spike activity. The pattern of spiking is unique to cell type, and the temporal sequence of the electrophysiological events encodes critical information. As a result, databasing of these recordings should be performed at the highest possible resolution. Representative recording traces in native resolution are an ideal source of data that can subsequently be used in meta-analysis. He pointed out that open-source analytical tools are also warehoused. Discussion focused on sharing datasets and the principles governing use of data from the repositories. Furthermore, it was stressed by multiple participants that data repositories must combine information from multiple sources that is readily searchable for virtually any possible parameter.

10. Esther Gardner, New York University School of Medicine

"Bioinformatic Analyses Reveal Neurophysiologic Correlates of Behavior"

Neurons respond to both simple stimuli and complex behaviors. Correlating motor behavior with the simultaneously recorded firing patterns of neurons in parietal cortex of monkeys revealed that activity of PPC neurons preceded that in S-I, and was often shape-selective for particular objects, suggesting that they play an important role in motor planning of prehension. In S-I, distinct groups of neurons responded to specific behaviors such as grasping, lifting, holding or releasing objects. Discussion focused on heterogeneity of the neuronal population and specialization of their fuctions, with brainstorming about ways to correlate the eletrophysiological phenotypes with the "omics" phenotypes.

Session III. **Superimposing the Glycome and Proteome onto the "-omics" Palette to Move from Molecules to Function in the Nervous System. (Chair: Dr. Kazuhiro Ikenaka)**

1. Kazuhiro Ikenaka, National Institute for Physiological Sciences

"Link between Glycome and Transcriptome"

Glycosylation is essential for protein function. Most proteins are glycosylated, and a glycoprotein is a final functional product synthesized by the coordinated actions of >100 genes. Glycogenes are dramatically regulated during cell differentiation & carcinogenesis. However, it is still an open question whether the transcriptomic or proteomic data can predict or define the structure of sugar chains expressed in cells or tissues. Both the structure of the major N-linked sugar chains and their expression levels in the brain have been determined, an undertaking which revealed very little variation among individuals, even in pathological conditions. A macroarray system to analyze the expression levels of "glycogenes" has been developed and the correlation between glycogene expression and the levels of N-linked sugar chains was investigated in various brain-derived cell lines. However, only limited correlation was observed, suggesting that the regulation of sugar chain expression is not only determined by glycogene expression, but also by various intrinsic as well as extrinsic factors.

2. Hisashi Narimatsu, National Institute of Advanced Industrial Science and Technology

"A System for Automated Rapid Identification of Oligosaccharide Structures using Observational MSn Database"

Following Dr. Ikenaka's presentation, a rapid method to determine the sugar chain structure was introduced. In this glycoproteomic study of the *Glycogene Project*, Dr. Narimatsu and colleagues used a high-throughput method to determine which amino acids are glycosylated in a glycoprotein. In addition, they were able to determine the carbohydrate structure and analyze its biological role. Using database searches and recombinant expression, they identified a number of new, high specificity glycosyltransferases. In addition, they found that one can determine the mode of linkage of the sugar chains as well as its sugar composition using MS^n (MS/MS/MS). This method enables an analysis of sugar chains in normal and diseased states. In the future, this can be performed as a diagnostic procedure from a drop of blood. The discussion pointed out that glycoproteomics is currently an under-appreciated field that should develop in parallel with other "omics" technologies.

3. Tosifusa Toda, Tokyo Metropolitan Institute of Gerontology

"Proteome Profiling of Brain Tissues and Isolated Cells for Research on Brain Aging and Neurological Disorders"

The Human Brain Proteome Project (JHUPO) is a collaborative project focusing on determining proteomic profiles in brain tissue and isolated cell types. Diversity in the cellular composition of the brain gives rise to a compartmentalized transcriptome that is specific to

neurons, oligodendrocytes, Schwann cells and other cell types. Progenitor populations of cellular phenotypes can be identified in development, purified, differentiated and analyzed for proteomic signatures. The results are reported as quantitative and comparative 2-D gel image analysis of protein expression in neuron, oligodendrocytes, astrocytes and primary culture cells, allowing easy data mining and linking to various proteome databases. This database serves as a solid foundation for interpretation of expression changes in various brain disease states.

4. Steve Pfeiffer, University of Connecticut

"Proteomic Analysis of the Myelin-Axon Complex"

Oligodendrocytes and neurons extensively communicate, and this communication is happening at the myelin-axolemmal complex. In an attempt to identify all of the protein elements in myelin and place them into functional context, Dr. Pfeiffer's research is focused on profiling oligodendrocytes, myelin and myelin subdomains using a variety of discovery approaches, including proteomic, glycomic, phosphoproteomic and functional analyses. These analyses lead to a number of functionally important discoveries. For example, NSF is a novel component of Schwann cell microvillae, while Nm23A and B are dramatically reduced in CGT-null myelin mice. Furthermore, oligodendrocyte progenitors and mature oligodendrocytes show distinct proteomic signatures. Further analyses revealed that myelin has a specific glycome profile, with enriched BA1 and BA2 expression. In addition, phosphoproteomic approaches revealed many functional facts about MOG, which may be relevant for our understanding of multiple brain disorders. The discussion focused around the value of reductionist approaches in high-throughput discovery methods and the distinction between the "hypothesis-driven" and "data-driven" approaches.

III. Conclusions

As part of the final round-table, the workshop participants made a set of critical bioinformatic, administrative and scientific observations and suggestions:

1. Emerging databases must become transparent, interactive, easily searchable and must integrate archival data. Retrieval of data from multiple linked databases simultaneously is essential for interpretation of the "omic" data. These databases must allow searches based on virtually any criteria, including (but not limited to) cell type, electrophysiological property, brain region or cellular function. Integration of archival anatomical, physiological, biochemical and other data accumulated over years must become one of the top priorities in building the new databases.

2. Careful experimental design is critical and should be tailored in the context of the scientific questions asked. It is important to realize that "one shoe does not fit all", especially in "omics" research. Enforcing strict standards on the end of experimental design, sampling dynamics, platforms used or analytical approaches implemented may impede our ability to ask the scientific questions the right way. However, *reporting* information obtained in the experiments should be standardized, and further evolvement of the MIAME/MGED standards should proceed in parallel with the development of the "omics" platforms.

3. Imaging, neurophysiology and "omics" fields are facing many of the similar biostatistic challenges. These involve datasets that are inherently noisy and often require complex pre-processing procedures to uncover the meaningful data. As a result, the potential for

obtaining false positive data is high in all high-throughput methods, and appropriate methods must be implemented to estimate the false discovery ratio. We need to develop innovative, open source statistical approaches to correlate the data from multiple fields of neuroscience. These tools should be simple enough to be used by neuroscientists with a basic statistical background, and should be downloadable from the data repositories themselves. A big difference between neuroimaging data and "omics" data is that the latter does not typically have cell-type information in each voxel, which makes higher order analysis of the raw data difficult. Dissection of individual neural cell types (e.g. oligodendrocytes) or structures (e.g. myelin sheath), completion of the protein-protein interaction analysis project, and detailed analysis of transcription factor expression in the brain should be useful adjuncts for defining the neural transcriptome or proteome and supplying second order information on regulatory networks.

4. Metabolomics, lipidomics and glycomics must evolve in parallel with the transcriptome and proteome studies. The classic dogma of DNA-RNA-protein does not fully address the physiological processes of the brain. Metabolic events, glycoprotein synthesis, lipid dynamics are susceptible to environmental influences and play a critical role in normal homeostasis and pathological states. At present, the expression pattern of the sugar chains (glycomic data) is difficult to predict solely by transcriptomic or proteomic data.

5. Reference "omics" datasets are needed for a meaningful interpretation of the experimental data. Brain regional mapping of transcriptome factors, "omics" and electrophysiological signatures of individual cell types are amongst the many useful tools that will allow more meaningful data mining, which ultimately leads to testable hypotheses.

6. Use of data from repositories should be governed by a set of guidelines. While the imaging and electrophysiology community has already established some guidelines, there is much less clear consensus in the "omics" fields about the rules of data sharing. We need to start a broad discussion in the neuroscience community about a number of critical questions such as " Does reanalysis of my dataset by other scientists entitle me for authorship?"..."Can I combine raw data from different data repositories to create a new manuscript?"... "Is combining raw data from different experimental series confounded to the degree that it does not represent good science?"... "My whole raw dataset is used in a meta study by a group of scientists. Do they need my permission for analysis and publishing the data?"

7. Biomedical and pharmaceutical research should participate in the data integration process. Industrial entities are routinely generating, managing, mining and integrating datasets from multiple sources. Discovery tools that are not commercialized in industrial research should become part of a public domain, where they can be further developed as an open source. Furthermore, industrial profiling data that is not used for patent applications (for example, experiments on control tissue) should be included in publicly available databases.

8. "Omics" data must be obtained more systematically. The molecular signatures of different subcellular compartments, cell types, layers, nuclei and brain regions are unique. For meaningful interpretation of the "omics" data the resolution of the technologies must increase and go beyond profiling the bulk brain tissue. In this context, efforts to obtain molecular signatures at cellular and subcellular levels should be promoted and refined.

9. It is all about biology. It is important to emphasize that in today's science bioinformatics is an integral part of any scientific discovery: classic molecular biology - often combined with reductionist approaches - can generate valuable findings and follow-up with simple informatics tools (e.g. sequence analysis or prediction of domains) to formulate biological hypotheses about the new discovery. In contrast, neuroimaging and "omics" analysis of human

brain use complex statistical procedures to extract biological discovery from the data sets. As a result, the novel data-driven and the classic hypothesis-driven research are complementary approaches that depend on each other. They should exchange information, and the nature of the questions asked and the available *a priori* knowledge will define which approach may be more appropriate.

10. Interactions between the Japanese and US neuroscientists must further evolve. Although there has been an improvement in the contacts and collaboration between the scientists of the two countries over the last several years, there is much left to accomplish. This must be a true, equal commitment partnership, and the BRCP program is the ideal framework to ensure the further evolvement of joint collaborative projects, information exchange and scientist exchange between the two neuroscience communities. The size of the meeting was ideal to facilitate intensive discussion and initiation of the future collaborations.