

## **Enabling Technologies 2001 Workshop Summary**

By Gabrielle Strobel and June Kinoshita

In August 2001, a diverse group of academic and industry investigators from within and outside of Alzheimer's disease research participated in this workshop in Bar Harbor, Maine. The goal was to identify critical knowledge gaps that slow down the search for diagnostics and treatments, and to develop a strategy for directing new approaches and technologies towards bridging these gaps.

The first day was spent summarizing the current state of knowledge. Following a primer on AD biology by Brad Hyman, scientists gave presentations on topics including AD genetics and epidemiology and the role of the cell cycle, inflammation, and cell death in the nervous system. Two scientists introduced "maverick" ideas: that AD is an autoimmune rather than a degenerative disease, and that epigenetic changes to DNA underlie some sporadic cases of AD. This day also included presentations of novel technologies that have matured sufficiently to become useful to AD research, including RNA arrays, high-throughput transfection in cultured brain slices, MR imaging of gene expression, multiphoton imaging in live animals, mass spectrometry, as well as high-throughput screening for drug discovery and medium-throughput functional assays in mammalian neurons. Rounding off the first day were a talk on the management of mouse genetic information and two perspectives from representatives of the National Institute on Aging and the National Institute of Neurological Disorders and Stroke.

The second day saw discussions of biological research priorities and of bioinformatics requirements to enable high-level analysis of information gathered in the course of collective research. In the afternoon, Dennis Selkoe led a discussion in which he suggested that a new Alzheimer Research Institute be launched, in which a permanent staff of scientists would work with rotating groups of visiting scientists to pursue research projects identified in the workshop. This proposal sparked lively debate on Friday and Saturday, without consensus.

On the third morning, the group distilled their discussions into an action plan, which they presented to foundation representatives.

### **The Participants**

- Robert Balaban, Chief, Laboratory of Cardiac Energetics, National Heart, Lung and Blood Institute
- James Bacion, Assistant Professor, Center for Molecular Imaging Research, Massachusetts General Hospital
- Robert Baughman, Associate Director for Technology Development, National Institute for Neurologic Disorders and Stroke
- Judith A. Blake, Mouse Genome Informatics, Jackson Laboratory
- Harvey Cantor, Chair, Dept. of Cancer Immunology, Dana-Farber Cancer Institute
- Timothy Clark, Director of Bioinformatics, Millennium Pharmaceuticals, Inc.

- Paul D. Coleman, Professor, Department of Neurobiology and Anatomy, Director, Alzheimer's Disease Center, University of Rochester Medical Center
- Peter Davies, Department of Pathology & Neuroscience, Albert Einstein College of Medicine
- C. Forbes Dewey, Professor of Mechanical Engineering/Bioengineering, Massachusetts Institute of Technology
- Peter DiStefano, Senior Director of Neurobiology, Millennium Pharmaceuticals, Inc.
- Alison Goate, Department of Psychiatry, Washington University School of Medicine
- James Heywood, Executive Director, ALS-Therapy Development Foundation
- Bradley T. Hyman, Professor of Neurology, Mass. General Hospital
- Rudolph Jaenisch, Professor of Biology, Whitehead Institute
- June Kinoshita, Executive Editor, Alzheimer Research Forum
- Peter T. Lansbury, Associate Professor of Neurology, Brigham and Women's Hospital
- Donald C. Lo, Chief Scientific Officer, Cogent Neuroscience
- Richard Mayeux, Director, Behavioral Neurology, Sergievsky Center, Columbia-Presbyterian Medical Center
- Marcelle Morrison-Bogorad, Associate Director, National Institute on Aging
- Dennis J. Selkoe, Professor of Neurologic Diseases, Neurology-BWH, Harvard Institutes of Medicine
- Gabrielle Strobel, Managing Editor, Alzheimer Research Forum
- Rong Wang, Department of Human Genetics, Mt. Sinai School of Medicine

### **Summary of Recommendations**

The working groups need a tight management structure, a clear work plan, and mechanisms for accountability and follow-up on initiated projects and collaboration. Their composition is key to success. They need administrative resources and decision-making power. The working groups must have regularly scheduled phone conferences. The Dana-Farber Harvard Cancer Center and Harvard Center for Neurodegeneration and Repair can model the structure of technology cores interacting with disease-based groups.

The working groups will:

1. Define specific AD research questions in their area and match them with technologies that can address these questions.
2. Draw up an inventory of available resources as well as those that need to be developed or purchased.
3. Identify institutions and people best suited to addressing the defined problems.
4. Initiate projects to generate pilot data that can attract subsequent industry investment.

5. Serve as a mediator between resources (e.g. well-characterized samples from epidemiological studies) and scientists who can analyze these resources with novel assays to assure that available samples are more widely used. In general, the working groups should set up mechanisms to improve communication between scientists that do not usually talk to one another.

The Biomarkers Working Group needs to develop strict standards for what constitutes a quality biomarker. This requires novel research information from the Pathways Working Group. In cancer, effort was wasted developing ultimately worthless biomarkers early on when there was not yet good understanding of underlying mechanisms. First, research opportunities in basic neurobiology and development must be identified. Exploit Patterson's in-vitro systems for neurons to avoid chasing poor biomarkers. Then focus on imaging as a next step. Imaging techniques are ready but awaiting good AD biomarkers.

### **Some additional short-term steps**

- Explore modus operandi of Hereditary Disease Foundation in identifying knowledge gaps and funding postdoc salaries for them.
- On ARF website: in cooperation with the technology core, list available technologies, explain what they can do for AD research, who has them and is willing to use them for AD questions brought up by others. List equipment review, asking people who have expensive equipment to share info on what works and what does not.
- Create an on-line directory of Alzheimer specimen resources available from laboratories and centers around the world.

### **SUMMARY OF SCIENTIFIC DISCUSSIONS:**

Robert Balaban opened the discussion by observing that heart disease and AD research share certain characteristics, including a weak familial linkage, poor penetrance of some of the known risk factors, and a poor understanding of the interplay between environmental and behavioral factors with gene-gene interactions. He challenged the room by stating that the basic cellular pathology mechanism in AD remains unknown. He predicted that this undiscovered mechanism will be specific to certain cells and that it will be cataclysmic because neurons are disappearing, two characteristics that may make therapy development easier. A human pathology mechanism urgently needs clarification in order to validate AD animal models for screening and to develop biomarkers, readouts for cell-based assays, and markers for imaging techniques. Overall, Balaban said, a rational scheme for therapy development does not exist. He recommended a concerted effort to unravel mechanistic pathways.

This introduction provoked debate about how much of the etiology can be explained by elevated A $\beta$ , and whether the underlying signaling cascades must be known before therapies can be developed. Selkoe argued that a complete mechanism of elevated cholesterol had not been worked out prior to the development of statin drugs, and that it was not necessary because LDL production was clearly connected to atherosclerosis.

Likewise, elevated A $\beta$  is strongly linked to the AD endpoint, therefore a step-by-step understanding of how accumulated A $\beta$  damages neurons need not be the priority at this point.

Lansbury argued that the pathologic pathways may be manifold and complex. Redundant pathways are at work in the neuron's slow death, and interfering with one would not be effective. Sorting them out will take many years and is not required for testing the hypothesis that blocking the inducing event-A $\beta$  accumulation and fibrillization-will be therapeutic. Lansbury favored embarking on high-throughput screening approaches now to save time.

Others disagreed, saying that understanding the underlying pathways, especially those regulating the cell cycle, inflammatory cascades, disrupted cell signaling, mitochondrial involvement in apoptosis, metabolic failure, and potential autoimmunity was a prerequisite to identifying novel biomarkers and new therapy strategies.

Everyone agreed these pathways ought to be unraveled, but participants disagreed on what should be top priority now. Some argued for hypothesis-driven research into mechanisms to generate biomarkers and better mouse models, others prefer investing in high-throughput screening to generate lead compounds, targets, and novel hypotheses.

On animal models, Jaenisch said that the genetics employed for current AD mice are outdated and could be much improved. Mayeux defended current models, saying they simulate aspects of the disease and enable testing of the basic premise that removing amyloid from the brain improves symptoms. Balaban countered that the pathology seen in mice may not be the real pathology occurring in AD.

Selkoe said synaptic loss deserves more study, rather than loss of cell bodies, because the synapses are key functionally and their loss may precede that of cell bodies. Synaptophysin changes, electrophysiological changes including LTP maintenance and ESP alterations, all occur in mouse models of disease even in the absence of plaques or massive neuronal loss. To this extent, mouse models of elevated A $\beta$  correlate with toxicity.

He said data on A $\beta$  suggest it is not directly toxic to neurons but precedes neuronal injury. Presenilin and A $\beta$ PP mutations lead to elevated A $\beta$ , as seen in plasma and Down syndrome. No one knows exactly how A $\beta$  damages neurons. Myriad in-vitro studies suggesting its toxicity are inadequate. Animal models show that elevated A $\beta$  causes synaptophysin loss and changes in electrophysiological properties, but that does not support the leap that A $\beta$  is toxic to all neurons lost in AD.

Goate and others said genetics clearly points to A $\beta$ PP processing abnormalities as key to disease in those cases. Wang said that while genetics clearly points to importance of amyloid, A $\beta$  levels are not elevated in serum early. Therefore other molecular changes must be occurring prior to onset of symptoms, and identifying these other proteins is a priority.

Lo said the key knowledge gap that is making AD difficult for his company to approach is the lack of proteins on which to base assays. Most medium to high-throughput systems are based on a cellular readout. A readout in whole animals is ideal but throughput suffers, so his company developed a brain slice assay to keep the context of cells temporarily intact. In AD he does not know what to look for, Lo said. He needs a proxy that represents some state of progression of the pathology at the cellular level. David Sabatini agreed that reliable markers at any point of the long pathogenic process would accelerate the identification of targets, even before a comprehensive mechanism is worked out. Some of these test points are going to occur years before there is anything to image.

How do we get those markers? Goate said that genetics is trying to uncover them, but a total of six genetics labs competing to pin down the same few candidate genes is too small an effort.

Balaban said FAD genetics are fine but sporadic disease has no mutations. Normal A $\beta$ PP processing contributes to disease if a different genetic defect or environmental condition leads to downstream sensitization in vulnerable neurons. The lack of complete penetrance of overexpressors and the fact that people without A $\beta$  overproduction get AD, point to other gene-gene interactions and environmental interactions in downstream pathology. All agreed that an explanation for the differential vulnerability of certain neuronal populations is a priority.

Selkoe said it is clear that amyloid accumulation is toxic in other diseases, as well, not how it is but that it is. He said the present discussion follows behind what has been found in 30 years of research on these other disorders, as some of them are treated successfully by inhibiting amyloid production. Rather than focus on holes in the amyloid hypothesis, he urges discussion of better ways to block A $\beta$ ?

Jaenisch asked whether neurons in AD die because of intrinsic problems or influences from its environment, a question that neuronal transplantation studies as developed in embryology could address. Others reply that this was an early question in the AD field, subsequent studies have pointed to extrinsic defects. This issue sparked a discussion of inadequate data in AD on the effects of tissue-specific expression of transgenes. Heywood said that in ALS, the SOD mouse develops disease only if the transgene is expressed in spleen and liver; its expression in neurons and/or astrocytes alone is not sufficient to cause disease.

Coleman, Hyman, and Davies pointed to the hierarchical layers of vulnerability of brain areas as a knowledge gap that can be addressed. Neuronal loss begins in layer 2 of entorhinal cortex and then progresses through the brain in a fairly predictable anatomical sequence. Why? A key priority is to describe what distinguishes the affected cells from the unaffected cells. Nobody has really exploited this opportunity.

Jaenisch said Rett syndrome may be relevant to this question because the protein and molecular mechanism at play acts as a general suppressor of transcription in every tissue,

every cell, without any specificity whatsoever. Yet the Rett phenotype is extremely specific. So the question is: is there something especially sensitive in those neurons affected in AD that defines their response to a less specific insult?

The question of apoptosis in AD was discussed. In in-vitro, in-vivo, development, and disease models, cell death always occurs over short periods of time. In AD, cell death seems to occur over decades in an individual neuron. What is the reason for this? Coleman said array studies show that the neuron-postmitotic and designed to last a lifetime-mounts defensive mechanisms. Synapse loss and neurite shrinkage is one such mechanism, "moving troops back from outposts." Many forms of defense prolong the neuron's path to death, and understanding those and devising ways to boost them can lead to therapy.

Inflammation was mentioned as a protective mechanism. DiStefano said that generally, NF $\kappa$ b activation in the nervous system is considered protective, whereas stimulation of caspases is detrimental. He sees in many experimental systems a balance between these two currents, and immune-type or inflammatory functions appear to dictate the balance. They act as life/death checkpoints and may explain why it takes so long for neurons to die. Crudely withdrawing trophic support in vitro causes neurons to die within 24 hours, but in vivo, the assaults are more subtle and protective mechanisms are in place.

DiStefano said activated microglia and astrocytes are the source of some of these immune-like functions and deserve more attention. Are they protective or damaging? Others object to using the term inflammation, because AD does not feature a classic immune response involving peripheral lymphocytes and macrophages.

On the question of normal aging versus AD, all agree that AD is not just accelerated aging. This has become a fringe view. Hyman said his data clearly prove that no massive neuronal loss occurs in normal aging but does in AD. Coleman elaborates that the pattern of cell loss distinguishes AD from normal aging. In CA1 of hippocampus, Coleman found massive neuronal loss in AD but no loss in normal aging, whereas in subiculum, neuron loss is similar in AD and normal aging.

Balaban closed by saying that the current trials of secretase inhibitors and A $\beta$  immunotherapies may fail and that the meeting's charge is to define a research strategy for identifying other Achilles' heels to go after while these trials take place. On that, there is fuzziness beyond amyloid.

All agree A $\beta$  should be pursued therapeutically. All agree a better understanding of underlying pathways would make it possible to identify other sensitive, non-redundant points in the interplay of signaling cascades that may provide new targets. All agree that uncovering the molecular pathology and unbiased screening ought to be pursued in parallel.

**These Scientific Priorities Drew Some Consensus:**

- Identify additional genetic and environmental risk factors. None known for 50 percent of AD cases.
- Develop readouts for use in medium-throughput cell-based and slice assays.
- Focus effort on understanding differential vulnerability of neuronal cell types, even single cells within a given area. Begin by characterizing more fully the difference in surface markers expressed on most versus least susceptible neurons. Wang says Mass Spec is able to filter out noise. Do a systematic study on both mRNA and protein.
- Test decisively whether an autoimmune response is at play in AD.
- Define protective versus damaging aspects of the inflammatory mechanisms at play; define the pathways of microglia and astrocyte activation and their role in neuronal death/survival.
- Refine animal models to more faithfully mimic AD. Limitations are not inherent to species; rather, current models miss a genetic aspect. For example, make mice transgenic for (micro)glial receptors to mimic induction of inflammation.
- Bring high-throughput proteomic screening (e.g. in cell-based assays analyzed by mass spectrometry) to bear on AD research.
- Develop better psychometric instruments.

## **SUMMARY OF TECHNOLOGY DISCUSSIONS**

Tim Clark began by laying out recommendations about the information infrastructure required if many groups want to be able to do collective experimentation, to share data, and to exploit automated pattern recognition in that shared data. One example where this is indispensable is data mining to elucidate complex pathways, Clark said.

Clark stated that researchers' frequent inclination to "just do screens and then mine the data" is wrong because, without a world model about the data one measures, that data does not later yield meaning. By world model he means ample annotation about the system under study. Clark added that even rudimentary data needs a world model as support. Until a few years ago, people used to think that having the proper technical architecture to support database management was sufficient. People thought using relational databases would enable data mining across all biology. This proved to be wrong.

Blake concurred, saying one needs to work out an information architecture prior to running array experiments. Databases set up to allow data sharing and integration are necessary but not sufficient for effective data mining. She urged development early on of a controlled vocabulary, a hierarchy of terms. This can be a simple classification system or become very sophisticated, as in artificial intelligence methods that enable full-fledged world knowledge representation and allow automated inferencing and other functions.

Moreover, in any complex bioinformatics project one needs to separate two fundamentally different kinds of information: the technical architecture and the information architecture, Clark said. The former is code (e.g. a data management system), hard to lay down and requires a lot of programming. The latter is concepts guided by

world knowledge about the biology under study; it must be lightweight and easily modified as concepts of biology change. For example, Millennium has terabyte systems that do sequence analysis and others that do expression analysis. If tissue type on one says brain and on the other says hypothalamus, the system integrates the data only if it knows the anatomical relationship between the two. That information should go into the information architecture. In AD, a defined "ontology" is necessary if one wants a pathways database to talk with an expression database, for example.

Clark and Blake urged researchers to ask: Would it make sense to construct a controlled vocabulary specifically geared to their area? If one wants to make interoperable the data from different labs running all kinds of high-throughput experiments, the answer should be yes. If the effort involves individual labs doing their own experiments and storing their results in their own databases, then the answer can be no. Certain types of analysis, however, cannot be done without data sharing and constructing an information architecture.

This triggered disagreement from many who said this approach would create too much prior information, restricting the system to what is already known. Clark and Blake said too little prior information is equally problematic. They clarified that annotation does not mean building assumptions, hypotheses, and bias from the literature into the system. It means developing a controlled nomenclature plus taking accepted knowledge from the literature. Blake said that in mouse genetics, simple things pose big problems. For example, researchers do not publish in what strains they did their experiments or are not using standard terms to describe strains. In brain research, naming cell types could cause similar problems.

BLAST searches were cited as an example of a successful common information architecture. A database to share microarray data is being built by industry. Baughman said he would like to see specific recommendations for microarray database sharing coming out of future AD informatics discussions to be incorporated into data-sharing standards that need to be developed for NIH microarray centers currently being built.

DiStefano said that when his group created their own chips, they tried to ascribe function to the microarray spots based on BLAST analysis. Without annotation, the first data analysis iteration took months. With Clark, they then entered everything they knew empirically about what they put on the microarray. Since then they have doubled the arrays to 15,000 points each and still conduct much faster analysis. The resulting clustering and organizing maps are better and have yielded interesting clusters that they would not have been able to conceive from the literature and their primary array data. Forbes Dewey concurred, saying that without a world model in mind one cannot design the right experiments (i.e. put the right things on a chip.)

Clark went on to say that the information architecture must be transferable to other systems, such mouse databases, mass-spectrometry expression databases, and that the annotation must be continually updated to mature over time. Dewey agreed, saying biologists are too afraid of complex systems. He quoted as a successful example the

cardiac myocyte, saying his group can predict how it will react to extracellular changes in e.g. calcium and confirm the computer model in the lab. This was worked out with 250 coupled equations but the system originated with 10 equations and grew over time. Another example is computational protein folding at the cell surface, which has matured to now predict what proteins get folded from first-principle physics. He says that the capability to deal with very complex, maturing systems is highly underutilized in biology.

Heywood said that annotation is a management process that is difficult to put in place in academia. Biology analysis at the level discussed here has outgrown the classic academic lab and requires technical staff and a tight management structure. Lo said the same applies to data reading in his system.

Coleman brought the discussion back from information handling to measurement, asking "what should we measure?" The basic cellular pathology of AD is unknown. There is heterogeneity of cell types and of individual cell reactions in the brain. To understand this heterogeneity, Coleman urged that two existing technologies be brought together to look at brain in AD and in normal aging. One is imaging that gives information at the level of individual cells or even within a cell. This must be developed for live humans with AD. The other is laser capture microdissection that lets one pick single cells to determine the molecular fingerprint of single cells. Coleman said that he wants this kind of information to be part of the data going into the informatics system. Others agreed.

Lansbury suggested an experiment involving cluster analysis of serum samples to identify gene clusters that can later be correlated with postmortem diagnosis of AD versus Lewy body disease versus other dementias or even Parkinson's, in which clinical diagnosis is only 75 percent correct. What annotation would be required? The expected diagnosis? Blake answered that no expectations should be included, just other factual information available on study participants:, such as symptoms, etc.

Wang suggested a systematic study of AD versus normal aging, planning what data to collect and what information infrastructure to put it into now, and starting to collect this data even though it is not fully annotated in the beginning. Lansbury asked how best to organize such a study. Should one start small with a group of well-known data that one can heavily annotate and add to, or should one start big because of the huge timesaving from putting in all that annotation?

DiStefano suggested all participants define a "laundry list" of needed databases. Then Clark and Blake can provide guidance on what empirical annotation is needed to compare experiments. Offhand, for AD one needs: transcriptional profiling, genetic sequence of affected patients, proteomic information, MRI databases, tissue block, mouse database sharing data across transgenic lines and new strains that come in, future cell-based databases.

Basilion cautioned that transcriptional profiling misses aspects of AD and PD, where patients accumulate certain proteins while the message is disappearing. True mechanistic understanding is impossible without proteins. Coleman concurred with the example of a

class-1 assembly protein, whose message stays stable in AD but protein levels drop dramatically. His group found that normally, this protein is protected from degradation by glycosylation but in AD, glycosylation fails and the protein is degraded.

Balaban agreed, saying a proteomics evaluation of AD samples will lead to therapies faster than mere clustering of transcriptional profiles. Proteomics incorporates gene-gene interactions and environmental influences but requires more starting material than transcriptional profiling, making analysis of individual cells difficult. Proteomics only takes the problem one step further because activity and phosphorylation status is not captured. One needs a whole range of databases that can talk to each other.

Wang reminded the audience that mass spectrometry can help here. Examples are PS1 function, where mass spectrometry can identify all components of the complex and their interacting partners. On the question of ApoE4 function, mass spectrometry can identify its interacting partners faster and with fewer false-positives than yeast-two-hybrid screens. Then one pulls up annotation, literature, and compares with other databases.

## **These Technological Opportunities Drew Some Consensus**

### **Tissue Blocks**

To make better use of human samples, Balaban suggested establishing tissue blocks of clinical trial material and of blood/serum collected in epidemiological studies. Blood or brain samples can be embedded in tissue blocks, frozen, and then sliced when a new hypothesis or biomarker has come up that needs testing. This can be done in an array-based setup to support high-throughput tests or quantitative proteomics (measuring concentrations of many proteins simultaneously). Blocks also can support laser capture microdissection.

### **Imaging**

Balaban suggested looking into imaging early physiological consequences of the pathologic mechanism in AD with high-field strength PET and fMRI methods currently in use for heart disease in some ERs. These methods have sufficient sensitivity and resolution to perhaps detect an altered metabolic response to memory tasks, or even reduced background activity, as an indicator of early neuronal dysfunction and possible early diagnostic. Mayeux objected that everyone in the field agrees neurons change prior to clinical symptoms, but how early is unclear. Epidemiology from Framingham cohort suggests 20 years, a Scottish study suggests 40. One needs to scan people longitudinally. Balaban agreed, saying one would first try to find a neuronal change that imaging can pick up in people with diagnosed early AD (this is doable now in 8-tesla magnets), then image that in presymptomatic people at high risk, then do prospective longitudinal scans of groups.

Balaban said inflammation is an early physiological consequence of the pathologic mechanism that could be useful for imaging, because in inflammation one cell affects a large area. Macrophage recruitment with concomitant change in epithelial cell

fenestration (easily imaged with contrast agents in heart disease) does not occur in AD, but other inflammatory-type characteristics in AD must be amenable to imaging.

Furthermore, Balaban suggested diffusion imaging as a promising technology, for example to image how tau inclusions disturb normal diffusion patterns of water inside neurons. This has been demonstrated in stroke and offers resolution of about 4 microns. All agreed that imaging methods are available but early markers are needed.