NEWS

Translational Biomarkers in Alzheimer Disease Research, Part 1

VOTE/SUBMIT A COMMENT

20 February 2006. Ask any Alzheimer disease researcher about the most pressing need in the field today and many will cite biomarkers. A consensus surrogate marker that predicts disease and responds to treatment tops investigators' priority lists. Indeed, such a marker would be an invaluable tool for obtaining the ultimate prize, the mechanism-based therapeutic that is more effective than current treatments. But scratch the surface a bit more, and you'll quickly discover that scientists' notions of that seemingly uniform thing, an AD biomarker, can diverge quite widely. While academic and industry researchers easily agree on what they ultimately want—a practicable, robust, noninvasive, inexpensive readout that enables early diagnosis and indicates if a therapy is working—it is also true that the word "biomarker" can mean different things to different people. In particular, there is a great need for mutual exchange on how best to employ translational research and animal models toward the development of the shared goal.

Last November, a satellite workshop to the 35th Annual Conference of the Society for Neuroscience in Washington, DC, drew scientists from academia, the biotechnology, and the pharmaceutical industry for a day of talks and discussion. Entitled "Translational Biomarkers in AD Drug Discovery: From Animal Models to the Clinic," the workshop was free of charge and sponsored by the Alzheimer Research Consortium (see ARF related news story). This public-private initiative supports the development of new research models that mimic features of AD and requires that the models be made freely available to investigators in academia and industry.

Patrick May of Eli Lilly and Company in Indianapolis and **Lennart Mucke** of the Gladstone Institute of Neurological Disease in San Francisco assembled a program of six speakers, each from academia and industry. They defined the terms, the problems, the opportunities in taking biomarkers from bench to bedside, and introduced the latest data in AD biomarker studies. The day began with a lesson on successes and failures from previous efforts in AD and other fields by **Ivan Lieberburg** of Elan Pharmaceuticals in South San Francisco. May described how his company used soluble Aβ as a biomarker all the way through a development program of γ-secretase inhibitors, and **Peter Seubert**, of Elan, did the same for the use of amyloid plaques as a biomarker in Elan/Wyeth's joint immunotherapy program. **Karen Ashe** at the University of Minnesota, Minneapolis, Mucke, at University of California, San Francisco, and **Greg Cole** of the University of California, Los Angeles, each introduced new biomarker candidates related to cognitive function and to the biology of synapses that new academic research has uncovered. **Henry VanBrocklin**, also of UCSF, outlined the path to developing radiopharmaceutical probes so researchers can one day image changes in these new markers. **Michael Greicius**, of Stanford University in Palo Alto, California, described early work on a surprising new functional imaging opportunity based on measuring changes in a "stream of consciousness" network of brain activity. **Bill Klunk** related a cautionary tale about reliance on mouse models by showing that the Pittsburgh Compound B (PIB) imaging agent that beautifully displays amyloid deposits in human brains fails utterly in mice. **Floyd Bloom** of Neurome Inc. reinforced the utility of those models in other aspects of biomarker research. He illustrated how new techniques for high-throughput morphometrics can analyze biomarkers in mouse models in a more powerful way than did earlier approaches. **Peter Davies**, of Albert

RELATED NEWS
Sorrento: They've Just Got to Have It—Where's That Biomarker? Part 1
Sorrento: They've Just Got to Have It—Where's That Biomarker? Part 2
SfN: Early Steps toward Parkinson Disease Blood Test
Sorrento: ADNI Imagines the Future of AD Imaging
Two New Developments in the Search for Alzheimer's Biomarkers
Translational Biomarkers in Alzheimer Disease Research, Part 3
Translational Biomarkers in Alzheimer Disease Research, Part 4
Translational Biomarkers in Alzheimer Disease Research, Part 5
Translational Biomarkers in Alzheimer Disease Research, Part 2

OTHER RELATED
Antecedent Biomarkers in Alzheimer's Disease: Uses, Limitations, and Future Directions for Research
Antecedent Biomarkers for Alzheimer's Disease Workshop II
Live Discussion: Making a BioMark on Alzheimer Disease
Antecedent Biomarkers for Alzheimer's Disease Workshop II
5th Leonard Berg Symposium

Einstein College of Medicine, The Bronx, New York, updated the audience on where diagnosis based on measuring CSF phospho-tau concentration stands to date after having been tested in thousands of samples, and **Tony Wyss-Coray** of Stanford University, Palo Alto, California, introduced a new blood test built on a proteomic fingerprint of AD-specific inflammation.

By day's end, the one complaint heard was that the program deprived the audience of those refreshing catnaps that are part and parcel to surviving a day packed with 12 lectures and discussion; sleep was not an option because every talk was excellent.

This report first pulls together the main themes that emerged during the day, and then expands on them in summaries of each presentation. The workshop drove home the message that it is critical to distinguish clearly among types of biomarkers, and to work those distinctions into one's study design. For example, a target-related biomarker differs from a disease-related, and again from a surrogate marker. A target-related biomarker represents a readout that changes close to the action of the drug at hand, that is, Aβ concentration in a test of a secretase inhibitor, or cytokine levels in tests of an anti-inflammatory drug. Disease-related biomarkers tend to change further downstream from the drug intervention; examples for this are cognitive readouts, or the measurement of tau changes in an Aβ immunotherapy trial.

On this issue, researchers agreed that it is important that drug trials measure defined target-related biomarkers. This assures researchers that the drug hits the intended target in the relevant tissue, and allows them to test hypothesis of whether that target truly plays an important role in the disease process. This is not always done. For example, trials of NSAIDs failed on disease-related endpoints that were far removed from the action of the drug without also determining whether the drug actually tamped down brain inflammation. For this reason, the trials left the field unable to learn whether inflammation remains a valid approach. In the area of target-related biomarkers, mouse models that are often criticized for mimicking only aspects of AD, such as models of amyloidosis, can be extremely useful, speakers agreed. In general, preclinical studies in animal models need to employ target-related biomarkers separately from disease-related endpoints so researchers can draw conclusions about the value of the target and the intervention.

A second take-home message was that candidate diagnostic biomarkers tend to bump up against a ceiling set by an imperfect clinical diagnosis. Their accuracy is judged against a clinical diagnosis that itself is prone to error. This raises the question of whether any biochemical or imaging diagnostic can ever reach 100 percent accuracy short of postmortem confirmation, and suggests that the field consider finding a consensus on what is good enough.

Thirdly, scientists agreed that it is important to use a diverse panel of animal models, not just one, in a given translational research program. Frequently in the history of drug development, translational studies in one model did not fully predict the human response, but integrated data from several models did. For example, Elan's preclinical work on mice did not predict the inflammation seen in the phase 2 trial of their first-line active vaccine. The PIB amyloid marker does not work in APP-transgenic mice, and had the program hinged on mouse data, it might have ended at this stage. Using an array of different animal models today is a less onerous standard than even 5 years ago, in part because a variety of brain imaging techniques for mice, rats, and monkeys have come online in recent years.

Researchers also took away a distinction among the different uses of a given biomarker. A marker that is useful for predicting AD may fail at responding to treatment, and different markers may work for different stages of this decade-long disease process. Finally, researchers shared a sense that to get a better grip on the pathophysiology of AD, the field needs to move past classical markers of plaques/tangles and soluble Aβ/tau. For one, it must develop ways of quantifying a range of different incarnations of Aβ and tau; for another, it should begin to exploit their interacting proteins from across the emerging area of synaptic biology. Such an effort might finally yield functional readouts that are likely to be clinically relevant. —Gabrielle Strobel.

NEWS

RELATED NEWS
Translational Biomarkers in Alzheimer Disease Research, Part 3

Translational Biomarkers in Alzheimer Disease Research, Part 4

Translational Biomarkers in Alzheimer Disease Research, Part 1

Translational Biomarkers in Alzheimer Disease Research, Part 5

Translational Biomarkers in Alzheimer Disease Research, Part 2

21 February 2006. This is Part 2 of a 5-part series. See part 1 for introduction.

From Industry: A Historical Perspective
Interest in biomarkers is growing rapidly, and not only among scientific investigators and medical providers, **Ivan Lieberburg**, Elan Pharmaceuticals, noted in his lecture. The FDA and its European counterpart, EMEA, are paying increasingly close attention and have raised the bar for how rigorously putative biomarkers must be vetted before a new product using that marker is approved. Insurance and government payers, too, have become more selective in what biomarker procedures they will reimburse.

The typical definition of the term "biomarker" in a clinical setting involves a physical sign or lab measurement that occurs in association with a disease and has diagnostic value. Once that biomarker is vetted to substitute for a clinically meaningful result, it then achieves the status of a surrogate marker. Few biomarkers have achieved this desirable status, and a critical variable in getting there lies in the choice of a meaningful clinical endpoint, Lieberburg said.

Cancer research offers lessons for AD, which is at a much earlier stage of developing clinically relevant biomarkers. A decade ago, showing that a tumor responded to a drug by shrinking in a CT or PET scan for a certain period of time passed as a satisfactory clinical endpoint. Now, only survival meets that standard in most cancers, because it has since turned out that initial tumor response to a treatment does not capture its benefit. People responded but still died early. It is important to distinguish clearly between various intermediate endpoints and the ultimate outcome, and to know to which endpoint a given biomarker relates. The key concept involves a distinction between target-based versus disease-based biomarkers.

For example, when a person takes a β-blocker, his or her heart rate slows within hours. This primary effect of the drug can be measured, and it constitutes the target-based biomarker. Subsequent physiological effects then follow this slowed heart rate and ultimately result in the clinical outcome of reduced death from congestive heart failure. Of this entire series of in-between effects, some will be of intermediate clinical benefit and some can be measured. In theory, they can serve as disease-based biomarkers so that studies do not have to wait years for quantification of the final clinical endpoint of survival, but there is a caveat. “The further removed the target-based biomarker is from the clinical outcome, the greater the chance it is wrong,” Lieberburg said.

First, the good news. There are two well-documented examples—hypertension and hypercholesterolemia—where the correlation has worked so well that the target-based biomarker has become the disease itself. Strictly speaking, hypertension and hypercholesterolemia are only biomarkers, but their link to the ultimate clinical outcome of death from cardiovascular disease has proved so tight that physicians, the FDA, payers, and the public at large view them as diseases. “That is where we eventually want to go with surrogate AD markers, as well,” Lieberburg said.

Sadly, failures of surrogate markers are more numerous. Medical history

contains many examples where investigators took a biomarker, converted it in their own minds into a surrogate marker, and then went astray, Lieberburg said. Either the putative biomarker did not correlate with a desired clinical outcome, or it did not capture the risk to which an intervention exposes a patient. Consider the example of atrial fibrillation. In this condition, the atrium flutters inefficiently instead of contracting normally; as a consequence, clots form and the risk of embolism and stroke rises dramatically. The classic procedure of cardioversion was widely used to jolt the heart into resuming a normal sinus rhythm (NSR), and NSR then assumed the status of surrogate marker for a good clinical outcome for patients with atrial fibrillation. Because electrical cardioversion is unpleasant, doctors switched over to chemical cardioversion, using digoxin and quinidine to reestablish NSR. It worked initially but then the patients' mortality increased because the drugs touched off a process that leads to a form of rapid heartbeat called ventricular tachycardia. "Chemical conversion to the surrogate marker NSR turned out to be a Pyrrhic victory," Lieberburg said.

Another example is bone density, considered a surrogate marker for fractures in osteoporosis. Most drugs, such as estrogen or SERMs, increase bone density and reduce fracture rates. But one of the most potent boosters of bone density, sodium fluoride, turned out to generate brittle bones that were as prone to fracture as non-treated osteoporotic bone. Bone density, then, is a flawed surrogate marker for that drug, Lieberburg said.

Then how does the AD field vet a biomarker? This is where science still needs to do some heavy lifting. A roadmap exists from other fields, and it includes these criteria:

- epidemiology to associate the biomarker with a clinical endpoint
- clinical relevance
- sensitivity, specificity
- reliability, robustness
- practicability; it must be noninvasive
- simplicity to adapt into clinical practice

For the gold standards of hypertension and hypercholesterolemia, decades of efforts went into gathering data on all these criteria. Indeed, the relationship between cholesterol and coronary artery disease appeared in the literature as early as 1938 and grew stronger with epidemiology data from the Framingham study in the 1970s and subsequent studies on many thousands of patients. Additional time went into building a consensus with FDA and payers. A more recent example of this gradual consensus building is prostate-specific antigen (PSA) testing for prostate cancer. Where this biomarker is the most useful depends on the stage of disease in a given patient and on the specific clinical question asked. Even though PSA has been widely used for years, the field is still defining exactly what it does well (i.e., predict mortality when measured as rate of change) and what it does poorly (i.e., guide chemotherapy adjustments). In AD, too, the value of a given biomarker will likely depend on the stage of the disease and exactly what it is asked to do, Lieberburg said.

Where are biomarkers used? A prominent area is drug discovery, and an established example to consider is viral load in HIV-AIDS. It is accepted as a marker for drug screening and predicts well whether a compound will eventually lead to a drug on the market. More frequently, however, industry researchers do not know whether the biomarker they use correlates with the clinical endpoint. That remains true for Aβ as a biomarker in AD drug discovery, Lieberburg said.

Biomarkers are used routinely in pharmacodynamic/pharmacokinetic (PK/PD) measures to determine whether the experimental drug at hand reaches the site of action, and in designing human trials. There, too, surrogate markers can lead investigators astray. Lieberburg cited an example where a presumed antidepressant that had worked well in carefully designed animal studies and in human trials using an imaging surrogate marker later turned out not to affect depression at all. "Despite the best efforts, you may still be going down the wrong path with unvetted surrogate markers," he said. AD researchers face a similar problem, where animal models capture aspects of what is ultimately a uniquely human disease.

Having told these cautionary tales, Lieberburg added that the standard the FDA imposes is surmountable. Subpart H of FDA rules stipulates that the agency can approve a drug based on a surrogate endpoint. The sponsor needs to submit clinical trial data that show the drug has a pathophysiologically-based effect on a surrogate marker and that this change in the marker reasonably predicts a clinical benefit. The FDA has applied Subpart H to expedite approval for serious and life-threatening diseases such as HIV infection, where drugs go through solely based on their ability to reduce viral load, not mortality, and for cancer (for recent drugs approved in this way, see FDA page).

AD also falls into this disease category. This means that accelerated approval for AD drugs is within reach but the quality of the surrogate marker will be crucial, Lieberburg said. AD research has produced a number of interesting candidates. The key challenge for researchers to keep in mind as they explore them is that any candidate needs to come armed with solid data on the bullet point list above in order to pass muster with the agency. To date, no AD surrogate marker that is vetted on all criteria exists, but the language of Subpart H suggests that it may be possible to use instead a portfolio of individually less-vetted markers so long as they each respond to the drug in a pathophysiologically relevant way. No drugs have as yet been approved in that way, but it would be worth making the case, Lieberburg added.

One problem holding back AD research is that it is still unclear which aspects of the pathophysiology one must treat to get an improvement of the clinical endpoint. Pharmaceutical companies can help with that by showing in their clinical trials that the drug at hand actually controls the immediate pathophysiological endpoint of the drug target in addition to measuring more distal endpoints. That would inform scientists about the role of that part of the pathophysiology in the course of disease. For example, trials of COX inhibitors measured the drugs' effects on cognitive and overall clinical endpoints, but did not show that the drugs actually controlled inflammation by tracking CSF cytokine levels. This left a large gap between the treatment and the endpoint and made it impossible to learn much from a failed trial. If the trials had shown that inflammatory markers were indeed down yet disease progressed unchanged, researchers could have ruled out inflammation as a target (at least for the stage of disease tested in the trial).

Previously, few instruments to test intermediate biomarkers related to a candidate drug were available. Company scientists tended to have a candidate drug but no tools to show it enters the brain, hits its target, and has an immediate effect on it. In effect, their trials tested only the molecule, not the hypothesis about its role in disease. That is slowly changing, Lieberburg and May noted. For example, Elan's ongoing phase 2 trial 9) of passive immunization with an Aβ antibody is using Pittsburgh Compound B (PIB) and glucose PET to test pathophysiologic markers broadly, and Lilly's γ-secretase inhibitor trials assess Aβ levels in CSF and plasma along with cognitive measures.

Industry Experience with Amyloid Biomarkers

Patrick May described the story of how Eli Lilly and Co. has used biomarkers as tools in its preclinical and early clinic program of γ-secretase inhibitors. In short, the lessons there are that animal models are indispensable for target validation and for assessing clinically relevant biomarkers. Preclinical biomarkers can help the researcher prepare clinical trials and get a sense of what to expect when the drug enters people; however, one should not rely on one animal model alone but integrate data from several different species.

APP proteolysis offers several potential biomarkers, he noted. The secreted fragment sAPPβ could report on β-secretase activity, while the Aβ peptide is the most immediate biomarker for γ-secretase activity. (The other cleavage product, AICD, occurs in amounts too small to be easily traceable.)

A bona fide biomarker that one tracks in the clinic in accessible tissues (i.e., blood, urine, saliva) differs from a preclinical biomarker one can track more invasively to ensure the candidate drug really acts on the intended target. Lilly uses biomarkers to validate the target in vivo and to

assess the pharmacodynamic effects of γ-secretase inhibitors in a variety of animal models, first in line being the PDAPP717 mouse originally characterized by scientists at Athena Neurosciences, now a subsidiary of Elan. As it ages, this animal model of Aβ amyloidosis mimics some of the pathologic hallmarks of AD.

The first use of animal models in Lilly's program lies in target validation. For that purpose, the scientist needs to show that the candidate drug affects the pharmacodynamic biomarker in a dose-dependent way. For example, rising doses of experimental γ-secretase inhibitors increasingly lower levels of hippocampal Aβ in young PDAPP mice. A range of compounds does this, but Lilly scientists have selected one named LY450139 to advance into the clinic. Another aspect of target validation is to ensure that the effect one measures relates in its size and timing to drug exposure. This requires measuring at what time the inhibitor achieves peak concentration before it decays, and relating that to the size and time course of change in its biomarker, that is, hippocampal Aβ.

Lilly's second use of animal models lies in defining clinically relevant biomarkers. This has been a challenge with CNS drugs because such clinically relevant biomarkers require easy access to tissue, and one cannot assess biochemically whether an experimental compound changes Aβ in human brain. Researchers make inferences about brain Aβ from sampling accessible tissues, but for that to work, they must first understand the peptide's trafficking from brain through CSF to plasma and its subsequent degradation in the liver. Research on antibodies and chaperones is beginning to do that (see, e.g., Deane et al., 2005; Cirrito et al., 2005). Animal models are indispensable for correlating changes in a clinically relevant tissue with a desired pharmacological response in clinically intractable brain tissue, May said. Not all animals are available for this kind of pharmacodynamic work. It is routinely done with mice, but in dogs and primates, the necessary brain biopsies are usually reserved for the end toxicology studies, May said.

Pharmacodynamic studies in PDAPP with LY450139 showed that a transient drop in plasma Aβ 24 hours after injection correlated with a drop of Aβ in hippocampus, cortex, and CSF, May said. The mouse models established that plasma Aβ can indeed report on changes of Aβ in the central nervous system. At the same time, one cannot simply extrapolate from a transgenic model, May cautioned. This is because non-transgenic mice showed a more complex plasma Aβ response to LY450139, probably because their Aβ contribution to plasma is not driven entirely by a transgene expressed in the brain.

For this reason, the Lilly scientists moved their translational biomarker studies of LY450139 efficacy into the beagle dog, a non-transgenic model large enough to allow repeated drawing of fluid sample big enough for a detailed analysis. Pharmacokinetics and pharmacodynamic studies in this species also showed that plasma Aβ acutely dropped at the same time that the secretase inhibitor reached its maximal concentration in plasma, but then Aβ rebounded, much like it had done in wild-type mice. This held true for a single dose or for a six-month treatment with daily inhibitor oral doses. The decrease in plasma Aβ correlated with significant reductions in CSF Aβ in the dogs. Like in wild-type mice, Aβ levels in plasma showed complex changes over time, but in CSF they did not.

These translational biomarker data prepared the ground for first forays into the clinic. In single-dose and 2-week safety trials in healthy volunteers, plasma Aβ dropped robustly for 6 to 8 hours after LY450139 injection and then rebounded to baseline. The dose-dependence and the pattern of the response mimicked exactly that seen in dog, May said, showing that careful biomarker studies in animal models allow the scientist to predict what to expect in humans. The correlation appeared to break down, however, where it mattered most: Despite the clear drop in plasma, human CSF Aβ levels did not budge significantly. A phase 1b study of 60 people with mild to moderate AD who received placebo or LY450139 once a day for 6 weeks showed, again, that plasma Aβ went down as predicted, but CSF Aβ did not change robustly. In trying to understand this disappointing finding, the scientists discovered that CSF Aβ values varied greatly between subjects, and even across time in a given person. Aβ concentrations swung wildly between 4,000 and 12,000 picograms per milliliter, making it difficult to ascertain a definitive drug effect in CSF. It

is unclear if this variability is part of Aβ's biology—for example, because its levels vary with excitatory activity—or if it reflects the biophysics of Aβ, that is, its stickiness and tendency to aggregate. "Fifteen years ago when we started this program, we called Aβ the peptide from hell. Now after all this intense research, we still think it is the peptide from hell," quipped May.

Taken together, the translational and clinical studies have shown that this γ-secretase inhibitor appears to be safe and able to reduce plasma Aβ, but an obvious decrease in CSF Aβ remains elusive and the company has not launched a phase 2 trial yet. The biomarker studies in mice and dogs helped set the drug dose and helped predict how Aβ would change in humans, but they have not answered the question of whether LY450139 can be a drug one day, May said. (For a study published last month on how Lilly's competitor Merck, Sharp, and Dohm tested a new γ-secretase inhibitor of their own in rat brain versus CSF, see Best et al., 2006.)

Lilly's competitor Elan has had its own share of trouble from preclinical mouse studies, most famously when they failed to predict the side effect that hobbled the phase 2 trial of its first-generation, active Aβ immunotherapy AN-1792. Like May, **Peter Seubert** also conceded that researchers have been humbled by the difficulty they encountered in developing a biomarker for AD, a disease whose pathology is so glaring that its major features—plaques and tangles—have been known for a hundred years.

Seubert reviewed preclinical research leading up to the company's clinical trials of AN-1792, which ended dosing prematurely when 18 (or 6 percent of) patients developed meningoencephalitis, though their follow-up continues. The Alzheimer Research Forum has covered this effort extensively (see, e.g., ARF conference story; ARF related news story; ARF news story; ARF Live Discussion), and this report therefore presents only points specific to biomarker use in this research program.

For one, the meningoencephalitis prompted a follow-up study relevant to biomarker development. Margot O'Toole and colleagues at Elan's partner Wyeth Research in Cambridge, Massachusetts, compared participants' blood samples in search of an immunological gene expression fingerprint that could potentially serve to screen prospective patients in a future trial. Ideally, one would want to exclude people prone to developing T cell-driven inflammation and include those who are likely to mount a desirable B cell-mediated immune response. Results suggested that combinations of gene expression patterns could potentially identify such subjects; however, this is a tentative conclusion because the authors had only five encephalitis cases available for the analysis (O'Toole et al., 2005).

Seubert then turned to cognitive outcomes. Fifty-nine people produced antibody titers above a predetermined threshold in response to the one to three shots of AN-1792 they received. Comparing them to the roughly 300 non-responders, the researchers found no difference in most clinical measures, but they did see a small but significant, titer-related effect in the Wechsler verbal-delayed memory test. They also saw an effect in the composite Z-scores of the memory-related elements of the Neuropsychological Test Battery (NTB).

The AN-1792 trial used CSF markers in a subset of patients. CSF Aβ42 showed no clear change, but total tau levels, which are typically elevated in AD cases, went down significantly in the responders (Gilman et al., 2005). "I take that as a very encouraging sign, that a biomarker (tau) of presumed neurodegenerative origin and distinct from the amyloid target was reduced," Seubert said. The researchers did not measure phospho-tau in the AN-1792 trial but are considering it in current ones, Seubert said.

The trial also used serial brain volumetric MRI imaging. (It did not include PIB imaging, but a second-generation trial does, Seubert noted.) The MRI biomarker study accompanying the AN-1792 trial lobbed a surprise at the field when it turned out that responders saw their brain volume shrink more than the non-responders (Fox et al., 2005). This was counterintuitive because numerous studies had established that the brain and hippocampus shrink with progressing AD. Beyond speculation about the reasons for this result—loss of amyloid and gliosis, fluid shifts—follow-up data and

further analysis are not yet available. It is unclear at present what this finding means for the future of MRI volumetry as a biomarker in AD diagnosis as opposed to one for treatment monitoring. Data keep coming in to suggest it may be useful for the former (e.g., den Heijer et al., 2006) perhaps more than the latter. Meanwhile, the finding has raised questions over regulatory demands that this biomarker be included in pivotal AD trials. Seubert would not say whether Elan still uses volumetric MRI in ongoing trials, or whether it is predicting more or less brain shrinkage.

Postmortem studies of brains of trial participants who have since died confirmed that the vaccine removed amyloid deposition in swaths of brain parenchyma. Activated microglia appeared to engulf this form of amyloid, but its cousin deposited around blood vessels stayed in place, as did neurofibrillary tangles.

In summary, the trial responders showed changes in these biomarkers: Their parenchymal amyloid burden and their CSF tau decreased, as did their brain volume. Their Wechsler verbal-delayed memory improved, as did NTB memory component Z-scores. The trial implies, then, that immunotherapy could treat processes related to amyloidosis as well as to tau, Seubert said. Consequently, biomarkers assessing both classic pathologies will be useful in the development of newer forms of this therapeutic approach (see Drugs in Clinical Trials).—Gabrielle Strobel.

For introduction, see part 1 of this series.

NEWS

Translational Biomarkers in Alzheimer Disease Research, Part 3

22 February 2006. This is part 3 of a 5-part series. See also part 1 and part 2.

From Academia: New Leads for Next Generation
Not surprisingly, biomarker research on amyloid plaques and soluble Aβ has a long track record given how central these players have long been to AD research. But are they really the best markers, or merely the oldest, most entrenched ones? And what else is coming down the pike?

One academic perspective on new biomarkers based on soluble Aβ and tau came from **Karen Ashe** at the University of Minnesota. Ashe, who is one of this year's recipients of a MetLife award, suggested that novel biomarkers could be found by focusing squarely on the particular forms of Aβ and tau that cause the memory impairment in AD. Ashe laid out recent studies in her lab that attempt to identify those, and her studies on a 12mer called Aβ* is in press at Nature. The Alzforum has recently covered this work in detail; see SfN conference story on Aβ; SfN conference story on tau; and ARF recent news story).

Likewise, **Lennart Mucke**, of University of California, San Francisco's Gladstone Institute, invited the field to reach beyond traditional readouts toward new, functional markers, especially of synaptic biology. He began by saying that academia fundamentally shares industry's view of what makes an ideal biomarker. Above all, it must be clinically meaningful, and this particularly has been a knotty problem in AD. New marker candidates are being found in animal studies, and in the future, the field will have to transfer knowledge about them into new radiological imaging agents in humans. In addition, scientists can look for equivalents of mouse markers in human brain, CSF or, ideally, blood and urine.

In AD research, the major pathogenic players are also the biomarkers. Aβ accumulates and forms different types of larger assemblies that then impair neural transmission and reduce expression of activity-dependent markers. For its part, tau accumulates into neurofibrillary tangles but increasingly is also thought to be pathogenic in small, oligomeric forms. For a recent review on these proteins as diagnostic CSF biomarkers, see Andreasen and Blennow, 2005. Studies on thousands of patients, largely in Europe, have established that AD patients have increased CSF tau, particularly forms phosphorylated at specific residues, whereas their CSF Aβ42 tends to be decreased. Phospho-tau cleanly distinguishes AD from normal aging, but when compared with related illnesses such as vascular or frontotemporal dementia, it separates the groups less well. Hence, the current trend in the field has become to combine phospho-tau with MRI hippocampal volumetry and Aβ measurements. (See also Davies presentation.)

Such combined measures still don't track satisfactorily with cognitive measurements such as MMSE performance. Why is this, Mucke asked? Part of the answer may have to do with the great variety of Aβ and tau assemblies present in the brain. It is unclear exactly which forms a given biomarker assay captures relative to all forms that are present, particularly relative to the forms that do the most damage to the brain at the time of measurement. For example, a total Aβ assay may not reflect oligomers, which may be as pernicious for cognition as plaques. Here, too, new methods of measuring Aβ oligomers are coming on line (Georganopoulou

RELATED NEWS
Translational Biomarkers in Alzheimer Disease Research, Part 1
Translational Biomarkers in Alzheimer Disease Research, Part 2
Translational Biomarkers in Alzheimer Disease Research, Part 4
Translational Biomarkers in Alzheimer Disease Research, Part 5

et al., 2005).

Mucke outlined proposed pathways for how AD develops subsequent to APP expression. The question of whether Aβ oligomers, or plaques and dystrophic neurites do the most damage to cognition in AD remains unsettled. Regardless of the answer, therapies reducing Aβ production (i.e., secretase inhibitor drugs) should change biomarkers and neurologic outcome measures in parallel. In practice, many types of therapy targeting one species of Aβ may indirectly draw down other species, as well. But this may not be true for other therapies that specifically target a more downstream segment of the pathway such as amyloid deposition, Mucke noted. Examples would be plaque-busters, or drugs that selectively inhibit fibril formation but not oligomerization. To the extent that they increase pathogenic oligomer levels, such drugs might actually worsen cognitive deficits, and in this case, using amyloid plaques as the biomarker of choice could create the problem Lieberburg outlined in part 2 of this series, where the biomarker responds to the drug but is clinically irrelevant. On the flip side, a drug that moves oligomers into deposition might improve neurologic deficits even while increasing amyloid load. Therefore, drug studies could benefit from monitoring cognition-related biomarkers in addition to standard measures of plaque and amyloid load, Mucke suggested. There was widespread agreement that for the field to move past these questions, it will be critical to develop more sophisticated ways of measuring Aβ that distinguish its different forms during the aggregation process.

This is more than speculation, as animal model data have begun to separate amyloid deposition from cognition—at least what passes for cognition in mouse strains. For example, consider new mouse models of mutant APP made by **Irene Cheng** in Mucke's lab. A related ARF conference story already summarizes this work. In brief, the study allows comparison of a particularly fibrillogenic mutant form of Aβ to wild-type Aβ in otherwise similar transgenic mouse strains. The new lines develop amyloid plaques and associated neuritic dystrophy earlier than do the J20 comparison line, but this rapid deposition does not track with neural function. On the contrary, the mice with the fewest plaques performed more poorly on the Morris water maze than did lines whose brains were laden with plaques. The scientists interpret this to mean that the fibrillogenic mutation diminishes the pool of bioactive oligomers even as it speeds up fibrillization, suggesting that plaques are relatively protective compared to oligomers.

Mucke then moved on to describe new studies of other biomarkers that correlate more closely with cognitive impairment than does Aβ. His lab explores markers whose levels depend on excitatory synaptic transmission, and found that these tend to be depleted in relevant brain areas in the presence of elevated Aβ. Researchers including Bloom and Cole have shown before that dendrites of hippocampal granule cells are susceptible to the effects of Aβ, and Mucke's group reported a drop in the synaptic calcium-binding protein calbindin in dentate gyrus of APP transgenic mice and in people with AD (see ARF related news story). This is not due to neuronal loss; rather, the neurons were still there, but their synapses were altered. Calbindin levels correlated well with water maze performance in mice, and with cognitive scores in humans.

It's not just calbindin, either. Other activity-dependent markers change in APP-transgenic mice, especially in the dentate gyrus subregion of their hippocampus. An example is the immediate early gene arc, which acts locally at activated synapses and helps maintain LTP and consolidate memories. Experiments using an enriched environment showed that APP-transgenic mice are dramatically less able to induce this marker than are wild-type mice (Palop et al., 2005). Prior mRNA measurements in hippocampus done in David Morgan's laboratory have suggested that already (Dickey et al., 2003). Arc interacts with many other synaptic proteins. The pathways of synaptic molecular biology, and the effect of Aβ and tau on them, deserve more study in the search for biomarkers linked to cognition. One of arc's upstream regulators of interest is the extracellular matrix protein reelin, which functions in LTP and dendritic reorganization. A separate mediator of Aβ's synaptic toxicity is the kinase Fyn (Chin et al., 2005). In this context, tau only heightens its notoriety, as it is necessary for APP-transgenic mice to develop behavioral deficits (see ARF

conference story).

In summary, recent work on animal models suggest that plaques remain an important outcome measure for secretase inhibitors and for Aβ removal, that is, immunotherapy. Yet, where the goal is to assess cognition, Mucke recommends that researchers also develop measures of synaptic activity. For that, new markers established in mouse studies should be moved into human imaging, where they could complement existing measures of regional volume loss, cerebral glucose utilization, and amyloid imaging.

Greg Cole of UCLA continued the theme of searching for new biomarkers in the withering synapses of AD. Cole's prior work on testing different therapeutic approaches in transgenic mice has identified new candidates, some of which are known to play a role in mental retardation. It has also identified dietary ways of influencing the underlying pathways. Cole's recent work points toward the interplay of Aβ oligomers and proteins involved in the cytoskeletal rearrangements of synapse and spine formation.

Broadly speaking, the conventional view that neuron loss drives synapse loss in AD is gradually yielding to one where neuron loss is not the initial problem. The real question has become what is driving cognitive deficits, and the answer to it may yield not only new markers of cognition, but also more treatable targets, Cole noted.

What biomarkers of cognitive function already exist? Synaptophysin is the best-known one, and its levels closely track tangle formation. Paul Coleman and others have shown a tight and compelling correlation between tangle formation and synaptophysin loss, where human tangle-bearing neurons have a large reduction in synaptophysin message. This marker is less useful in AD transgenic mice Cole uses, because they do not show marked synaptophysin loss. The mice do, however, have cognitive deficits, and they also have in common with AD that both show loss of spines, shrinking dendritic arbors, and loss of dendritic area (for more on this, see section on Bloom, part 5). One way of approaching the molecular biology of cognition is to look at mental retardation genes, said Cole, because different inherited forms of mental retardation share spine defects that are similar to those seen in AD and Down syndrome. Cole is testing the hypothesis that Aβ aggregates cause these dendritic spine defects, and that solving the cognitive problem would require repair of the spine defects.

Aβ aggregates likely affect these processes. They are known to induce rapid LTP deficits, probably by mechanisms that include microglial activation and down-regulation of components NMDA and AMPA receptors (see ARF related news story). All this is consistent with the broader idea that AD entails a postsynaptic attack on excitatory spines, Cole said (see Bloom; Moolman et al., 2004; Spires et al., 2005; Dickey et al., 2004). Cole is particularly interested in the synaptic protein drebrin, an actin-binding protein in spines that occurs primarily in cells containing PSD95. Both proteins are lost in AD brain and in APP transgenic brains.

Cole believes that diet and oxidative stress modulate the proposed attack on spines. A diet that depletes the omega-3 fatty acid docosahexaenoic acid (DHA) exacerbates it (Calon et al., 2004; Calon et al., 2005), and dietary DHA deficiency in transgenic mice accentuates oxidative damage and correlates with the loss of a panel of synaptic markers including the NMDA receptor subunits NR1 and NR2, as well as CamKII. The idea that people can develop cognitive impairment from deficits in an NMDA receptor component has support in mouse models. In particular, Cole suggests that an interaction between the transgene (or in AD, Aβ oligomers) and diet leads to the oxidization of DHA's double bonds in neurons and synapses. Part of the mechanism could be that DHA availability influences the activation of AKT, the nexus of a prominent survival pathway in neurons (Akbar et al., 2005). In theory, AKT activators could be helpful, though in practice they would have to be weighed carefully against the established role of AKT in fueling some cancers.

Incidentally, DHA is being added to some brands of infant formula to support brain development. The brains of infants who have not yet developed elaborate dendritic arbors look strikingly similar in PET scans

of glucose utilization to those of late-stage AD patients, who have lost them, Cole noted.

To search for pathways that mediate Aβ's role in synapse formation, Cole's group, in collaboration with **Sally Frautschy**'s, performed a microarray analysis of their transgenic mice. The researchers measured expression differences in a panel of previously identified mental retardation genes between animals on a DHA-rich and a DHA-depleting diet. They focused on PAK kinases because these enzymes control actin dynamics in dendritic spines and appear to protect against loss of drebrin, Cole said. Other groups have shown that PAK inhibition by itself can cause cognitive deficits. APP-transgenic mice on the unhealthy diet lost PAK message, and their remaining PAK protein clustered around amyloid plaques instead of being distributed near synapses, where it normally occurs. AD brains showed severe depletion of soluble PAK and similar changes in PAK staining, Cole reported. Furthermore, in a culture model of Aβ oligomers/ADDL colocalization with NMDA receptor and PSD95-containing sites, PAK also changes its normal diffuse staining around synapses to a clustered pattern two hours after oligomers are added and taken up into the neurons. This suggests a PAK translocation has occurred in response to the Aβ oligomers. Drebrin staining went down in tandem with PAK, but adding intact PAK protected the cultures against drebrin loss. Back in vivo, infusing anti-Aβ antibody into APP-transgenic mice increased PAK just as levels of a 12mer form of Aβ decreased, Cole said. (This is the same 12mer Ashe studies, article in press at Nature.) In short, Cole is implicating Aβ aggregates in a PAK pathway defect.

It is vexing that the first step of any proposed pathway leading from Aβ via synaptogenesis proteins to spine loss is still a mystery, Cole noted. This is, of course, the identity of the Aβ receptor; integrins are sometimes mentioned as a candidate and some labs have hinted of having found a receptor but are staying mum until formal publication. Even so, this work suggests common pathways between developmental and age-related cognitive impairment. In the former, mutations in mental retardation genes are often the cause. In the latter, Aβ would induce loss of PAK and drebrin, and that, in turn, would impair formation of the protein complexes that are needed for remodeling the actin cytoskeleton in dendrites and spines as the brain tries to form new synapses. Part of Cole's work was recently published (see ARF PAK/p21 news story). Cole suggested PAK as a new biomarker that might be amenable to treatment. Experimental treatment with curcumin lowered Aβ levels and reversed PAK and drebrin deficits in APP transgenic mice. Insulin also affects PAK levels. In this context, a growing literature describing overlaps between insulin dysregulation and dementia is of interest, most recently an MRI study documenting that people with type 1 diabetes have subtle reductions in gray matter in brain areas responsible for memory, language processing, and attention (Musen et al., 2006). —Gabrielle Strobel.

See also part 1 and part 2.

NEWS

RELATED NEWS
Translational Biomarkers in Alzheimer Disease Research, Part 1

Translational Biomarkers in Alzheimer Disease Research, Part 2

Translational Biomarkers in Alzheimer Disease Research, Part 3

Translational Biomarkers in Alzheimer Disease Research, Part 5

Translational Biomarkers in Alzheimer Disease Research, Part 4

23 February 2006. This is part 4 of a 5-part series. See also part 1, part 2, and part 3.

Light Them Up: Imaging Biomarkers with Radiotracers
New markers candidates such as PAK, drebrin, arc, or calbindin would be more powerful if there was a way to see them in human brain. **Henry VanBrocklin** of University of California, San Francisco, described how the AD field could begin working toward that goal by using a combination of chemistry and radiological imaging, and he quoted ongoing work in Parkinson disease (PD) as an example. To stimulate early efforts in the AD field, VanBrocklin gave an overview of the development of radiopharmaceutical probes and their application in drug discovery. It is none too soon to start, as developing radiotracers for clinical use can be almost as lengthy and expensive a process as developing a drug.

Two recent changes at the FDA may expedite things, VanBrocklin said. For one, its Critical Path Initiative is part of an effort to get more drugs into the pipeline. Besides identification of biomarkers, its key component is imaging biomarkers or surrogates to assess the value of experimental drugs earlier in the process. In principle, radiotracers for clinical use are subject to the same development process as are new drugs, but recently, the FDA has introduced the "Exploratory IND." It lowers the threshold for getting new pharmaceuticals, and also new radiotracers, into humans for study, and represents an effort to cut the time this usually takes from a decade to a year or less, VanBrocklin noted.

Researchers generally use two ways to incorporate radiopharmaceutical probes into the drug development process. They either label the drug itself, with tritium or 14C, to see where drug is and how it gets metabolized. Or they label a receptor system or enzyme that the drug perturbs. In this regard, too, it is important to distinguish between biochemical targets and disease-related targets (see Lieberburg lecture).

Much like development of any other biomarker, radiotracer development involves three steps. The scientist must assure that the tracer hits the target of interest, that it is sensitive enough to track the expected change in target density (see Klunk presentation below for an example where this failed), and to validate it in various animal models of the disease.

For radiotracers to work, the density of the intended target must be adequate so one can see the tracer. If a tracer is to label a low-density receptor, its specificity must be high, and its on-off rate must be within a certain range. It must be selective to avoid nonspecific binding reducing the contrast between background and signal. Blood-brain barrier penetration is vital, and its metabolism must be understood because metabolites can cause background activity.

Neural receptor proteins have been a cornerstone of neuroimaging, and neuropeptides, pre- and postsynaptic neurotransmitter systems, and transporters have seen a variety of tracers made against them, as well. Few of those were ever fully developed and validated, however, VanBrocklin noted. One example relevant to AD is the α7 nicotinic acetylcholine receptor, which a growing number of studies are reporting to be lost in AD-relevant brain areas early in the disease (see Oddo et al., 2005 for a recent demonstration in a triple-transgenic mouse model of AD). Researchers at Johns Hopkins University, working in a joint project with

Novartis, recently published the first imaging agent for it (Pomper et al., 2005). Using 11Carbon, its affinity and selectivity for the receptor were good but did not penetrate the brain well. “This is a start, and it is an example for where partnerships between industry and academic PET labs can make great advances,” VanBrocklin said.

Another area that has benefited from neuroimaging with radiolabeled tracers is the study of enzyme activity. The key example in PET is glucose metabolism with 18-fluorodeoxyglucose (FDG), which clears the blood-brain barrier by means of the glucose transporter. This tracer differentiates AD from frontotemporal dementia.

Radiotracer imaging can complement gene therapy, where a therapeutic gene together with a reporter gene gets packaged into a viral vector, introduced into cells, and a radioligand probe picks up expression of the transgenes. For example, the dopamine D2 receptor, a target of antipsychotic drugs, has been used for such an approach. Ideally, the therapeutic and the reporter gene should be the same, and this is the goal of an ongoing study for Parkinson’s conducted jointly by VanBrocklin and collaborators at UCSF and UC Berkeley. The researchers use the enzyme amino acid decarboxylase, which converts L-dopa to dopamine, and detect its expression with 6-fluorometatyrosine (FMT) or fluoro-dopa PET imaging. Such an approach would make it possible to reduce the dose of current L-dopa therapy. This works in a primate model of MPTP lesion on one side of the brain, VanBrocklin said. Both the FMT PET signal and some function returned after injection of the construct into the striatum, and the approach has moved into a phase 1 trial, VanBrocklin added.

Beyond this particular approach, a host of small-animal imaging devices mimicking the human imaging armamentarium has sprung up for use in translational studies in recent years. Micro-PET, microMR, microCT, microSPECT, as well as ultrasound, fluorescence, and bioluminescence are all in place for testing preclinical radioligands in animal models, VanBrocklin said. PET, for example, now works well in mouse, rat, monkey, and human, and thus can support drug discovery all along the way from the bench to the bedside. VanBrocklin finished by saying that imaging agents should be plugged into the development program of a drug as early as possible.

Stream of Consciousness Network for Diagnosis?

Even while researchers have been whipping into shape techniques for animal brain imaging, human brain imaging has made its own advances toward new diagnostic methods. Perhaps the most novel approach comes partly out of **Michael Greicius**’s group at Stanford University in Palo Alto, California. Greicius is working on a way of exploiting a network of brain activity during mental rest, or free association, for a functional imaging method to predict and monitor AD.

First, some background on fMRI of active and deactivated brain areas. Unlike PET, functional magnetic resonance imaging (fMRI) requires no injection of a radioactive tracer or contrast agent but instead relies on an intrinsic signal contrast known as the blood oxygen level dependent (or BOLD) signal. fMRI exploits different magnetic resonance properties of oxygenated versus deoxygenated hemoglobin, and it can measure brain function because active brain regions contain more of the former.

In a typical fMRI experiment, a person performs a cognitive task in an alternating 30-second task-rest pattern for a few minutes. This generates a “task wave form,” and a statistical program then searches the brain for regions whose BOLD signal time series correlates with this wave. fMRI has relatively low spatial resolution but when overlaid on a high-resolution structural MR scan, the wave signal yields an activation map. This standard scan looks for regions of activation whose BOLD signal increases during the task period and decreases during rest periods.

Deactivation is the opposite. It occurs in regions where the BOLD signal increases during the rest periods and decreases while the person performs the task. In essence, these brain areas get shut off while a person focuses on the task at hand. Deactivation drew interest because it is a consistent phenomenon, whereby the same set of brain areas emerges during rest periods across different tasks and different subjects. Marcus Raichle at Washington University in Saint Louis, Missouri, first put deactivation on

the field's collective radar screen with a study showing that the posterior cingulate cortex, the inferior parietal lobes, and the medial prefrontal cortex were all deactivated across a variety of cognitive tasks (Raichle et al., 2001). Raichle suggested that these regions constitute a network whose activity is suppressed when one has to perform a cognitively demanding task, and he called it the default mode of brain function.

Of course, the resting brain is not truly resting. Brain activity continues even in the absence of a task cued from the outside. How do scientists measure that? In resting-state functional connectivity MRI, scientists do not search the brain with the activation task wave form; instead they pick spontaneous BOLD signal oscillations emanating in a particular region of interest (called the seed region) as the wave form and search the wider brain for regions that correlate with it. Such regions are considered functionally connected. The scientist first isolates that brain area in an activation scan, then the subject is scanned during 5 minutes of rest. To generate the connectivity map, the investigator takes the spontaneous resting-state activity from the seed region (for example, the motor cortex as defined by a preceding finger-tapping task) and searches the brain for regions whose BOLD signal is tightly correlated with it. This provides a resting state network for regions connected to the motor cortex. Subsequent work has convinced the field that these networks truly represent functional connections between brain areas and not mere blood flow artifacts, Greicius noted.

Greicius's group used this kind of functional connectivity imaging to test Raichle's hypothesis of the resting-state default mode. They had healthy young adults perform a working memory task to isolate the posterior cingulate as a region of deactivation. Then they used that as a seed region during a second, resting-state scan to look for connected areas. This analysis confirmed that the posterior cingulate indeed formed resting-state connections with all the brain regions Raichle had proposed to be part of the default-mode network.

Different resting-state networks linked to particular functions, such as motor, language, etc., exist. Among them, the default-mode network for memory is the only one that is normally active but needs to be suppressed when a person performs a cognitively demanding task, Greicius said. It has since become clear that most cognitive tasks—working memory, calculation, matching—suppress this network. An intriguing exception is tasks that involve the retrieval of memories; such tasks actually activate this network. Taken together, these findings suggest that the network mediates "stream of consciousness" processing, a silent mode of reminiscing and mulling over recent events in which we spend quite a bit of time every day, but which we suppress when focusing our attention on a specific task.

The network became interesting for AD research when it turned out that the regions involved in this default-mode network overlap remarkably well with regions of decreased metabolism in AD. Along with hippocampal atrophy, the most widely accepted imaging finding in AD is that, when scanned at rest, people with AD have decreased metabolism in the posterior cingulate and inferior parietal lobes on both sides (e.g., Alexander et al., 2002). This raised the question of whether examining resting-state, default-mode activity in these areas might become a diagnostic marker.

Task-activation fMRI is a highly specialized form of imaging that is not used in standard clinical settings. Resting-state fMRI is easier because one need not feed stimuli into the scanner and analyze the person's responses. As a step toward its use in clinical diagnostics, Greicius developed a simpler, automated means of detecting the network using a statistical method called independent component analysis (ICA). Applied to a publicly available fMRI data set that includes AD patients, the method visualized a robust default-mode network in young and elderly healthy volunteers but a greatly degraded network in the AD group (see Greicius et al., 2004 and, for more detail, ARF related news story). For further free fMRI data sets, Greicius referred researchers to the fMRI Data Center maintained at Dartmouth College.

Group data as in this study are different from what ultimately matters most, that is, a method the neurologist can apply for a patient who comes

to the clinic. To this end, the researchers tested whether a measure called the goodness-of-fit score could be useful in distinguishing single AD patients from healthy controls. This score compares the strength of the network measured in a given person to a standard template of the network averaged from a set of healthy controls. Based on it, this first-pass study achieved 85 percent sensitivity and 77 percent specificity, Greicius noted.

Since then, scans with a separate set of 18 patients and 13 age-matched controls have confirmed this first finding. Current work with other patients aims to probe the method's ability to distinguish AD from other forms of dementia. Early indications from this ongoing, preliminary work are that changes in the posterior cingulate cortex might not only distinguish AD from frontotemporal dementia but also correlate with MMSE scores, Greicius added.

Greicius closed by pointing to a recent collaborative study by Randy Buckner's group and Klunk and Mathis, which he said convinced him of pursuing this network's disruption in AD (Buckner et al., 2005). In it, the scientists showed that the default-mode network defined by deactivation, that is, the brain areas that are jointly suppressed while a person focuses on a cognitive task, showed remarkable damage in AD across three separate methods of imaging. Their glucose metabolism was down, as was their volume, and they also comprised the brain areas (posterior cingulate and inferior parietal cortex) that retained the most PIB, that is, had a high amyloid load. For some reason, then, the stream of consciousness network appears to bear the brunt of AD pathology.

In summary, Greicius suggested that resting-state fMRI could be repeated at short intervals in the same person because it is relatively easy to perform and requires neither radiation exposure nor a nearby cyclotron. Ultimately, it could complement the standard structural MRI that is becoming part of the routine workup of patients in memory clinics.

No Bed of Roses: Making PIB Work in Mice

Bill Klunk, of the University of Pittsburgh, Pennsylvania, is best known these days for the success, so far, of the amyloid imaging agent PIB in human studies (Klunk et al., 2004). At the workshop, however, Klunk talked about one of his failures. It is a cautionary tale for any scientist trying to decide how much to stake on mouse studies in a translational biomarker research program.

Klunk and his colleague, radiochemist Chet Mathis, started their search for an amyloid tracer in the late 1980s, first with derivatives of Congo red and then, in the late 1990s, with derivatives of the amyloid-binding dye thioflavin T called BTA compounds. Tinkering with the compounds' side groups, they gradually improved their affinity to amyloid in the test tube, as well as their ability to enter the brain and leave it again within 10 minutes. They finally settled on a 6-hydroxy derivative called Pittsburgh Compound B.

Klunk and Mathis's first mouse study worked just fine. They teamed up with Brian Bacskai and Brad Hyman at Massachusetts General Hospital, Charlestown, to validate this compound with multiphoton imaging of live, plaque-ridden Tg2576 mice. This allowed them to watch over the course of 30 minutes how the fluorescent compound labeled the blood vessels, diffused out of the vessel into the brain parenchyma, lit up plaques, and then disappeared from the brain (Bacskai et al., 2003; see image at ARF related news story). Binding studies using nanomolar concentrations of Pittsburgh Compound B with brain homogenates confirmed that the compound selectively recognizes amyloid plaques, not tau.

PIB has since moved into human studies to a stage where centers around the world are beginning to report similar results on its ability to distinguish early AD and MCI from normal aging and related conditions (see ARF related conference story) and where the first reports are appearing on the interplay between PIB retention and CSF biomarkers (Fagan et al., 2005). Amersham and G.E. Healthcare have licensed commercial development, and some natural history studies (Coats and Morris, 2005) have begun using PIB.

But even as the human studies are progressing well, the mouse work with PIB has been a vexing defeat, Klunk said. His group wanted to exploit the

powerful technique of micro-PET in transgenic mice to fine-tune PIB and find even better tracers. But the project ran into trouble when PIB sailed through the brain of amyloid-laden PS1/APP transgenic mice as fast as through control brains. This lack of binding also proved true in extensive subsequent ex-vivo studies of various forms of Aβ from mouse brain, including highly sensitive, classic "grind-and-bind" assays with brain homogenate. In all of these tests, PIB stuck poorly to mouse Aβ. The double-transgenic mouse brain had less than one PIB binding site per 1,000 molecules of Aβ. In stark contrast, human AD brain has a PIB binding site for every two molecules of Aβ. Intriguingly, synthetic Aβ aggregated in the test tube is just as invisible to PIB as is Aβ that aggregates in a transgenic mouse brain (Klunk et al., 2005), although all three sources of aggregated Aβ start with the same human Aβ peptide sequence.

The difference lies not in the affinity of binding sites, but in their frequency, Klunk said. The multiphoton experiments worked well because they used 10,000-fold higher PIB concentrations than one uses for micro-PET. At 2 microns, multiphoton microscopy also has almost 1,000-fold better resolution than micro-PET, and this allowed the investigators to focus in on spots where the concentration is the highest, that is, selected plaques, Klunk noted.

Why does PIB "see" only such a tiny subset of synthetic and mouse Aβ? Klunk does not know but suggests that something about the aggregation of Aβ in human and mouse brain is fundamentally different, at least with respect to generating PIB binding sites. There could be cofactors that determine the tertiary structure human Aβ assumes during aggregation. There could be post-translational modifications in humans that don't occur during faster aggregation in mice, or there could be environmental differences in the brains' pH and ion composition that account for the difference.

It is unclear at this point what, if anything, this difference between human and mouse implies for AD pathogenesis. Understanding the reasons for the difference might enable researchers to design better animal models of AD and yield new insight into the fundamental process of amyloid aggregation and toxicity itself, Klunk noted. Perhaps differences in Aβ aggregation might even explain why the behavioral deficits in PS1/APP mice are relatively subtle even though their brains are loaded with the human peptide.

For now, transgenic APP/PS mice (and several other transgenic mouse strains that Klunk tested) appear to be of no use for in-vivo micro-PET studies of amyloid deposition, Klunk said. This applies to tracers other than PIB, as well. This experience raises questions about the value of mice for this aspect of AD research, Klunk said. The development of promising PET tracers may stop in its tracks if investigators make human experiments contingent on prior success in mice. "You have to be careful how important you make animal studies along the way," Klunk said. "I am really glad it worked in humans and not in mice, not the other way round."—Gabrielle Strobel. See also part 1, part 2, and part 3.

NEWS

Translational Biomarkers in Alzheimer Disease Research, Part 5

RELATED NEWS
Translational Biomarkers in Alzheimer Disease Research, Part 1

Translational Biomarkers in Alzheimer Disease Research, Part 3

Translational Biomarkers in Alzheimer Disease Research, Part 4

Translational Biomarkers in Alzheimer Disease Research, Part 2

24 February 2006.This last section of our 5-part news report on a recent workshop on translational biomarkers summarizes contributions by biotech companies, and academic-biotech joint efforts, to the search for new biomarkers and diagnostic tools. See also part 1, part 2, part 3, and part 4.

What's Cooking in Biotech: Fledgling Products
Discouraging as Bill Klunk's experience with APP/PS transgenic mice appears, it stands opposite other new data that show, on the contrary, just how useful even the existing models can be when examined with new technologies. **Floyd Bloom** is a former editor in chief of Science magazine, who in 2000 co-founded the biotech company Neurome, Inc. after leaving the journal. His premise was that the mouse and its genome will be the premier model for molecular neuroscience for some time to come, and transgenic mice for CNS diseases. He believed that knowing which neurons and circuits express a particular gene is a necessary first step toward a functional characterization of the genes one wants to target therapeutically. Because classical chemical neuroanatomy was largely based on rats and cats, the field needed new tools to map and compare gene expression in mouse strains, Bloom reasoned, and he set up Neurome to create high-throughput morphology and histology protocols that can assess ingredients of circuits quickly and comprehensively.

An early project at the company focused on estimating volume changes in brain regions of the PDAPP mouse. Using a 9.4 tesla magnet at CalTech, the scientists obtained digital volumes of specific areas of mouse brain rendered in three dimensions. They observed that between 40 and 90 days of age, the hippocampus in wild-type mice grows significantly, but in PDAPP mice it does not. Something about the expression of the transgene leads to a 14 percent volume reduction already by day 90, and it may set the ground rules for the pathology developing later (see Redwine et al., 2003).

From this initial experience with image analysis, the researchers learned how to align brain sections done in one plane (i.e., coronal) into smooth, virtual atlases in another (i.e., sagittal or horizontal). This enabled them to analyze abnormalities in gene function and view their effects on brain structure in those 3D atlases. The scientists can also perform neurochemical or anatomical experiments on sections, including stereological assessments of the volumes of small components of a given brain region, for example, hippocampus. This showed that the dentate gyrus was the subarea that failed to grow in the PDAPP mice; at 90 days it was 28 percent smaller than in wild-type mice.

To understand why the dentate gyrus languished in this way, the Neurome scientists adapted a DiOlistic technique developed by Jeff Lichtman, then at Washington University in St. Louis, in which a "gene gun" shoots a lipophilic dye on microscopic gold particles into a fixed slice of the desired brain region. This yielded 3D images of fluorescent dentate granule cells in a high-throughput fashion. To make this assay practicable and reproducible, the Neurome scientists developed software for analyzing the dendritic complexity of a granule cell in a short time period and with limited computer memory. This study found that the dendritic trees of dentate granule cells were stunted, and that not all cells were equally affected. Of the six granule cell layers in the dentate gyrus, dendritic

length was more reduced in superficial cells than in deep cells, and more reduced in the dorsal blade of the hippocampus than in the ventral blade (Wu et al., 2004). The reason for this difference is unknown, but Bloom noted that work by Fred Gage's lab has shown the superficial layer of the dentate gyrus to incorporate fewer new neurons from neurogenesis than do the deeper layers.

Morphometric analysis with DiOlistic labeling and a second labeling method based on Golgi impregnations are ongoing at Neurome; the goal is to compare the fate of granule cell dendrites in aging mice to that seen in young PDAPP mice. Early data suggest that PDAPP mice start losing dendritic spines in the most vulnerable layers early in life and that wild-type mice begin showing a similar loss at around 15 months. These could be two independent processes, or they could represent premature aging of the dendritic spines of a particularly vulnerable set of neurons driven by the transgene, Bloom speculated.

These morphometric tools also have made it possible to quantify the distribution of diffuse and compact amyloid in different brain areas of the aging animal in a newly comprehensive way. For example, this study found that diffuse amyloid increases greatly between 12 and 15 months of age in PDAPP mice, suggesting that it could serve as a biomarker for drugs targeting this form of Aβ. "If we had a medication that prevents diffuse amyloid formation, we could test it in that age bracket and get an answer in 3 months," said Bloom.

A further finding was that the amyloid load in subareas of the dentate gyrus correlates with their incoming nerve circuitry: PDAPP mice carry a much higher amyloid load in the outer molecular layer and the lateral entorhinal cortex (which projects to the outer molecular layer) than in the middle molecular layer and the medial entorhinal cortex (which projects to the middle molecular layer). Something about that latter circuit allows it to resist the process by which amyloid is laid down, and finding out what it is would hold clues to the underlying disease process, Bloom suggested. There is no explanation for what molecules distinguish these circuits. One candidate for a spatially defined modifier is SUMO2/3, which Barbara Cordell's group at Scios Inc. in Sunnyvale, California, reported to restrict APP expression (see Li et al., 2003). Bloom's group noticed that the hard-hit lateral entorhinal cortex at 100 days of age begins to express less SUMO3 than does the medial entorhinal cortex. Incidentally, the lateral entorhinal cortex is an element of the default network discussed by Greicius.

Furthermore, this new system of morphometric analysis reinforced the field's realization in recent years that the genetic background of APP transgenic mice can greatly influence their phenotype and that data from one strain should not be interpreted in isolation. For example, the MRI volumetric findings in the PDAPP mice were not reproducible in Tg2576 mice, Bloom noted. This strain did, however, lose spines in areas of the hippocampus and cortex that later laid down amyloid, as also found by Greg Cole.

Peter Davies, of Albert Einstein College of Medicine, consults for Applied Neurosolutions on the development of a diagnostic test for AD. This Illinois-based biotechnology company develops an assay based on measuring CSF concentrations of the protein tau phosphorylated on the amino acid threonine 231. At the workshop, Davies evaluated the quality of existing tests for tau; he argued, in essence, that the p-tau231 assay is as good as it can be within the confines of an imperfect clinical diagnosis.

Davies began by noting that tau phosphorylation clearly is increased in AD, and that antibodies visualizing this process light up much more than neurofibrillary tangles. They stain what looks at first glance like intense background but in truth represents evidence of widespread threonine 231 phosphorylation of tau beyond the actual tangles themselves. This biochemical abnormality in tau labels whole cell groups and their processes in large areas of the hippocampus. Threonine 231 and serine 202 are sites on tau that become hyperphosphorylated early in hippocampal pyramidal neurons of very mild AD, well before tangles form there. These markers are useful for detecting early disease, whereas total tau has proved unreliable, Davies said.

Led by Davies's collaborator Harald Hampel at Ludwig Maximilian University in Munich, Germany, as well as other European investigators, researchers to date have tested the p-tau231 assay in more than 3,000 CSF samples. Most are from patients who came to the clinic for a diagnostic workup (Hampel et al., 2004). AD cases come up positive while controls do not, but the assay also picks up a fraction of patients diagnosed with other dementias. This weakness is often cited. The assay's sensitivity lies at 90 percent; its specificity ranges from 80 to 100 percent but is lower for vascular dementia and diffuse Lewy body disease. Davies argued that where results diverge from the clinical diagnosis, he suspects that the assay is correct and the neurologist may have made the wrong diagnosis. "Even the world's best clinicians are not always accurate. I think the patients diagnosed with vascular dementia who were positive in our assay actually have AD," Davies said.

Overlap in underlying pathologies plays into this issue, as well. A substantial number of vascular dementia cases prove upon autopsy to have had amyloid and tau pathology, as do people clinically diagnosed as having diffuse Lewy body disease. Ironically, people with frontotemporal dementia—the quintessential tauopathy—do not. Their predominant pathology is cell death, and tangles do not accumulate massively as they do in AD brains, Davies said. Accordingly, they come up negative in this assay.

The p-tau231 assay shows no obvious relationship with the MMSE, a crude but widely used cognitive assessment. Why not? Davies pointed to a study that, to date, has measured CSF p-tau231 of 103 mild cognitive impairment (MCI) cases, 163 AD cases, healthy controls, and samples from other neurological diseases. It shows that the signal in MCI already is nearly as high as that of full-blown AD (Buerger et al., 2002), suggesting that by the time a person begins to fall off on the MMSE test, p-tau231 has long accumulated. Davies's collaborator Mony De Leon is trying to determine exactly when p-tau231 first begins to rise in a cohort of healthy people and MCI patients he is following longitudinally (de Leon et al., 2002). This ongoing study indicates that p-tau231 is fourfold above normal even at the MCI stage when MMSE performance is still fairly high; however, the higher a person's p-tau231 concentration is at that point, the faster he or she declines on the MMSE in the next few years. "This means we may be able to use this marker to identify patients at a very early stage of disease and predict their progression," Davies said.

In summary, Davies argued that the clinical diagnosis of AD cannot be an ideal yardstick by which to measure the accuracy of a biochemical assay above the ninetieth percentile, and other investigators agreed. "All biological markers are going to run up against a ceiling effect of the clinical diagnosis. At the present time, our assay has reached this ceiling," Davies said.

Besides creating circular arguments, this situation makes validating the assay a challenge. Validation on autopsy is difficult in practice because the needed several thousand samples are difficult to obtain in the U.S., where autopsy rates are low and falling. Moreover, people on average live another decade after receiving a diagnosis, and during this time some who were negative at testing would likely develop tau pathology, again muddying the waters. It's questionable whether such an expensive, long-range study must be done, especially for a combination test of Aβ42 and p-tau231. "We may already have excellent biomarkers; we just do not know it yet," Davies concluded. (See also ARF related news story on CSF Aβ/p-tau; for new reviews on CSF-based biomarkers, see Formichi et al., 2006; also Wallin et al., 2006).

The well-tested p-tau231 assay illustrates some of the hurdles that a separate, fledgling test introduced at the workshop has yet to clear. It is a proteomic blood test developed in a joint effort between **Tony Wyss-Coray** at Stanford University and Sandip Ray, who co-founded the startup biotech company Satori with Wyss-Coray.

The process of AD features a vigorous inflammatory response. Astrocytes and microglia become activated and cause the secretion of a large number of proteins, including cytokines, chemokines, growth factors, proteases, and protease inhibitors, which together mediate communication between

these cells in the brain. Lymphocytes chime in, as well, especially in vascular forms of the disease. An increasing number of studies indicate that this CNS reaction communicates with the periphery, particularly with peripheral macrophages, lymphocytes, and myeloid cells, Wyss-Coray said. Some of these can travel into the brain, assess its state, and either leave or induce production of factors or even initiate immune responses.

The larger point is that every disease, in every organ, leads to changes in plasma, Wyss-Coray said. The blood is the body's most complex organ in terms of protein moieties, and Wyss-Coray started his study from the question of whether one can understand a disease process by studying plasma. Scientists have measured individual markers by ELISA, but the low power of this approach has left many studies that report initial discrimination with individual factors without replication by other labs. Mass spectrometry has matured as a tool for mining the proteome, but problems persist there, too, as the method is asked to keep apart many tens to 100,000 different proteins, fragments, and post-translational modifications. Abundant proteins such as albumin tend to overload the system, and efforts to deplete them take down other proteins that may be of interest. This approach is not reproducible enough yet to be clinically useful, Wyss-Coray said. (See also an independent recent attempt to identify a new candidate biomarker set focused around neuroprotective and complement proteins, see Selle et al., 2005.)

The scientists decided on a middle-of-the-road approach between these two extremes. The scientists first picked a set of proteins that might be important in the disease process. "We call this a candidate-based approach by tuning in to the language of cells. How do they communicate when healthy, how when stressed and diseased? Hopefully, we get a disease-specific picture," Wyss-Coray said.

The scientists gradually whittled down an initial group of 300 proteins from among cytokines, chemokines, growth factors, neurotrophins, hormone-like proteins, acute-phase proteins, complement factors, soluble receptors, proteases, and inhibitors to a set of 12 predictors. They developed an ELISA array of a membrane with monoclonal antibodies specific against these proteins, incubated it with patient plasma samples, and read the signal with chemiluminescence. This yields a picture of relative levels of expression of these 12 factors, which can be quantified.

A first, small study used 48 cases in various stages of AD as well as 50 age-matched controls from seven centers around the world. Of the 17 who have since died, the test had predicted their condition with 100 percent accuracy, Wyss-Coray said. Of the cases that have not yet come to autopsy, the difference in the relative probability of having AD between control and AD groups was large, Wyss-Coray said.

Software developed at Stanford, called significance analysis of microarray (SAM), pulled up 44 proteins whose blood levels differed between AD and controls. Individually, none of these factors predict AD, but together they do, Wyss-Coray said, and a separate procedure of unbiased clustering of these 44 markers based merely on their expression levels reproduced the AD and control groups.

To analyze further whether these plasma differences could predict AD, the scientists turned to another form of analysis called predictive analysis of microarrays (PAM). This algorithm tries to identify a minimal set of markers that can discriminate and predict the proper sample groups without having seen the primary data. In multiple iterations, it adds proteins from within a training data set and calculates their predictive power until it has reached maximal accuracy. This minimal set included 12 proteins, which predicted whether a sample came from AD or control with 97 percent accuracy (i.e., a composite score of 100 percent sensitivity and 94 percent specificity). When the algorithm then applied this information to a different test set it had not seen before, it classified 32 of 33 samples correctly.

The top 12 factors are involved in immune function, energy metabolism, and vascular function. Wyss-Coray proposed that the most abundant changes are consistent with immune and macrophage impairment. There is scarce data on this topic, but a growing trickle of studies is suggesting that mononuclear cells or macrophages isolated from AD patients are impaired

in a number of ways and respond poorly to stimulation. (For a current review on serum-based proteomics of neurodegenerative diseases, see Sheta et al., 2006.)

Next on Wyss-Coray's list is to study related dementias. Initial work on ALS, Parkinson disease, multiple sclerosis, and peripheral neuropathy indicates the AD fingerprint is specific to this disease and does not merely reflect a generic inflammation. The hope is that related dementias will prove to have their own unique pattern of plasma predictors, suggesting that a top 12 set may be found for them, as well.

One important caveat with blood tests is that infections or flu could mask AD in plasma samples. While the scientists have not ruled this out, Wyss-Coray said individual markers clearly change in response to a flu, but a defined set of 12 may not. Confounders such as this imply, however, that an ultimate test for AD may need more than 12 predictors. The present data aim to prove the concept; it is not a commercial test just yet, Wyss-Coray said. —Gabrielle Strobel.

See also part 1, part 2, part 3, and part 4 of this series.