Biomarker identification in neurologic diseases: improving diagnostics and therapeutics

Keith D Coon, Travis Dunckley and Dietrich A Stephan†

Identification of biomarkers in neurological disease remains impeded by many obstacles. Among them are the availability of tissue at the site of pathology, poor clinical diagnostics, the complexity of the brain and a general dearth of functional end points and models for validation. However, advances in technology have helped to overcome these challenges. Some of these advances include standardization and increased efficiency in brain banking, novel techniques for brain imaging, improved methods for reducing tissue heterogeneity including laser capture microdissection, high-throughput genomics, new functional validation techniques such as RNA interference, and the development of new animal models of neurologic disease. In order to efficiently handle the wealth of information that will be gleaned from these new technologies, new integrated databasing protocols will be necessary. Access to these databases by researchers and clinicians is critical to the continued progress being made in biomarker identification in neurological disease. These challenges and ways to overcome them are presented here in the context of a disease known to be a robust model for biomarker identification, Alzheimer's disease.


Although effective biomarkers should display most, if not all, of these factors, a full discussion of each of the points in relation to an individual biomarker is beyond the scope of this review. Traditionally, these biomarkers have been especially difficult to discern in neurological disease due to several factors inherent in the study of human behavior and brain pathology.

This review seeks to investigate the challenges that can be expected when attempting to identify biomarkers in neurological disease and how advancements in science have helped to overcome these challenges. The four basic challenges to biomarker identification in neurological disease that will be discussed are:

• Availability of tissue at the site of pathology
• Poor clinical diagnostics and extent of disease progression at the time of diagnosis
• Complexity of brain and tissue heterogeneity
• Lack of functional end points and models for validation
The continued identification of these biomarkers will permit further investigations, including genetic or chemical manipulation of animal models, which can help characterize disease pathology and lead to the development of novel methods of diagnostics and/or therapeutics. As Alzheimer’s disease (AD) has become a robust model for biomarker identification in neurological disease in recent years, this review will focus mainly on AD as an example. However, other examples of neurodegenerative and neurobehavioral disease will be included when appropriate.

**Availability of tissue at site of pathology**

Identification of biomarkers in neurological diseases is considerably more challenging than in other diseases due to a variety of factors. Many of these difficulties are related to the acquisition and quality of the necessary tissues, especially those at the actual site of pathology. Obviously, tissue availability in the affected area (i.e., brain), although potentially obtainable via the rarely performed brain biopsy, is realistically only available post-mortem [4]. By this stage, the affected tissues have likely been ravaged by the disease leaving little experimental material of good quality to investigate early etiologies [5,6].

Traditional methods for preserving post-mortem brain tissue consisted of formalin fixation, however, it has since been demonstrated that formalin fixation is a barrier to performing many modern molecular biological techniques. Thus, at present most brain banks use a variety of freezing techniques [7,8], in addition to fixation, to preserve the valuable specimens. As some biomarkers have extremely short half-lives [9], post-mortem interval (PMI), the amount of time that elapses between death and preservation, as well as the type of preservation, has become increasingly important. PMI is known to vary dramatically in brain banks across the nation from 1.6 to 32.5 h and even longer [10,11]. In addition to the dilemmas facing post-mortem neuropathological characterization of those suffering from neurologic disease, ante-mortem clinical diagnostic procedures are also difficult.

The problem of tissue availability and acquisition has been largely overcome with advances in brain banking. As has been mentioned, new freezing techniques and shorter PMIs are making higher quality tissue available more rapidly. A brain bank at Duke University has developed a rapid brain autopsy protocol, which processes brains within 1 h after death [12], and the Sun Health Research Institute in Arizona maintains an average PMI of 2.6 h [BEACH T, PERS. COMMUN.]. Institutes all over the country are attempting to achieve PMIs like this in order [30] to standardize protocols as well as provide the highest quality samples to researchers. The expedited processing of the brain samples also minimizes the loss of tissue, thus increasing the availability of tissue at the site of pathology.

In addition to the speed at which the specimens are being processed, new computerized databasing technologies are cataloging and organizing the donor submissions in ways that maximize the amount of information available to the researcher. Detailed knowledge regarding the neuropsychiatric, neurologic, neuropathologic, and other clinical parameters of the tissue donors are now easily and rapidly available to the researcher [13]. Collectively, these general physiologic variables can also influence brain biochemistry and impact evaluation of inclusion or exclusion criteria of appropriate samples for a clinical investigation. With the new databases, it is not unusual for up to 1000 different ante- and post-mortem data factors to be generated and cataloged for a given individual. This availability of information will allow the researcher to make informed, appropriate choices about sample inclusion that will result in more informative investigations. Of course, the accuracy of this database information, particularly in relation to neurological and neuropsychiatric evaluation, is only as good as the initial clinical diagnosis.

**Poor clinical diagnostics & extent of disease progression at time of diagnosis**

Clinical diagnostics and (sub)stratification of patient populations are poorly developed for most neurodegenerative diseases, most notably multiple sclerosis (MS) [14,15] but also to a lesser degree with atypical forms of Parkinson’s disease (PD) [16] and AD [17]. Even in the better case scenario of AD, clinicopathological diagnosis has been demonstrated to have a specificity between 76 and 88% and sensitivity between 53 and 65% [18–20], with a confirmation rate of probable AD as low as 65% [21]. Using the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) criteria, AD could only be reliably distinguished from frontotemporal dementia 23% of the time [22]. In neurodegenerative disorders as a whole, clinical diagnosis is only accurate in approximately 70–80% of cases [23].

With neurobehavioral diseases, for example schizophrenia, the problem becomes even more difficult as many symptoms, such as the presence of hallucinations and delusions, are disclosed through verbal self-report of affected individuals and often questionable second-hand patient histories from family members and/or friends [24,25]. Additionally, many of the psychotic symptoms that are the hallmarks of schizophrenia can result from other behavioral [26] and nonbehavioral insults, such as mental retardation [27] and/or brain tumors [28]. Further compounding the dilemma is the fact that the diagnoses of neurobehavioral disorders are frequently based on negative symptoms, flattened affect, refusal to speak, loss of motivation in schizophrenia [29] and lack of manic episodes in bipolar disorder [30]. These can be difficult to confirm by the clinician in outpatient scenarios and may require more intensive observation, for example, hospitalization or institutionalization.

New developments in brain imaging techniques have helped overcome some of the problems associated with clinical diagnosis of neurological disorders. Used in the past primarily as a technique to exclude structural lesions as etiological in neurodegenerative disease [31], neuroimaging has become increasingly utilized as a tool for diagnosis of the conditions themselves. The American Academy of Neurology suggests that neuroimaging should be used to facilitate the diagnosis of all cases of dementia [32]. Imaging of progressive cerebral atrophy by computerized tomography (CT) has contributed to the diagnosis of dementia.
since the early 1980s [33], however, the emergence of targeted therapeutics has made early diagnosis and scrutiny of disease evolution increasingly important [34,35]. Magnetic resonance imaging (MRI) extended the ability of researchers to assess the microscopic changes that accompany progressive brain atrophy. However, several drawbacks inherently associated with the cross-sectional nature of this analysis (inter individual variability, laborious manual outlining and subjectivity in defining structural boundaries) has limited the precision, and thus efficacy, of the procedure [36].

In the detection and tracking of AD, it has been helpful to characterize and measure the volume of a hippocampal region of interest [37,38]. Whereas this and other regional or whole-brain measurements can be generated separately from each image, researchers have begun to obtain some of these changes from the direct comparison of sequential, coregistered MRIs. Techniques such as the brain boundary shift integral [39,40] or an iterative principal component analysis [41] are used to compute rates of whole-brain atrophy from sequential MRIs, whereby the individual serves as his or her own control. Accordingly, researchers have demonstrated the ability to distinguish abnormally high rates of whole-brain atrophy in patients with probable AD from normal aging, with no overlap between groups, to estimate the power of this approach to assess the effects of a putative disease-modifying treatment in randomized, placebo-controlled trials. Although the findings are more preliminary, researchers have begun to use automated voxel-based image-analysis strategies to detect and track changes in regional brain size and shape [39,41] and gray matter density [43,44]. Some of these methods capitalize on the transformation of an individual’s brain image, before or after it has been segmented into gray matter, white matter and cerebrospinal fluid (CSF) or different regions of interest. In the case of voxel-based morphometry (VBM), for instance, spatially standardized gray matter images can be used to detect or track alterations in gray matter in patients with probable AD.

Using a variety of visual and linear measures (e.g., qualitative visual rating scale, height of left hippocampus, radial width of temporal horn of lateral ventricle) to assess medial temporal lobe atrophy in neuroimages obtained using MRI from a number of investigations, Scheltens and coworkers calculated a corrected sensitivity and specificity for detection of AD, compared with controls, of 85 and 88%, respectively [45]. These imaging techniques are even more important in other neurodegenerative diseases (e.g., frontotemporal dementia and semantic aphasia) where asymmetric atrophy is maintained throughout the progression of the disease and helps differentiate these conditions from AD [46,47]. This has been determined as essential for the diagnosis of vascular dementia, according to the National Institute for Neurological Disorders and Stroke-Association pour la Recherche et L’Enseignement en Neurosciences, an international consortium on vascular dementia [48].

Positron emission tomography (PET) scanning, which has been used extensively to assess metabolic activity within the brains of those suffering from neurologic disease, and single-photon emission computed tomography (SPECT), which measures blood flow, can also help illustrate neuronal degeneration and/or cognitive deterioration. Imaging data from PET scans have been shown to discriminate (with 93% accuracy) mild-to-moderate AD cases from normal controls [49]. PET evaluation of relative cerebral glucose metabolic rate demonstrated that 36% of mild cognitive impairment (MCI) patients converted to AD within 1 year and thus concluded that a reduction of cerebral glucose metabolic rate in prefrontal cortical areas is associated with the transition from MCI to AD [50]. Another common pattern that has been observed in AD patients, this time using SPECT, is the progressive decrease of temporoparietal blood flow that increases with disease severity [51,52].

Neuroimaging has also become important for diagnosis [53-55], as well as testing the effectiveness of treatments [55-57], for behavioral disorders such as schizophrenia and bipolar illness. Early studies using CT repeatedly demonstrated a lateral ventricular enlargement in the brains of schizophrenic patients that could be used to forecast subsequent cognitive impairment [58]. This phenomenon was further characterized as the ventricle-brain ratio (VBR) and demonstrated to be heritable within families of schizophrenic patients [59-61], as well as those affected with bipolar disorder [62,63].

Profiles of accelerated gray matter loss, assessed by MRI, have been demonstrated repeatedly in schizophrenia [54,55,64,65] and lack of functional connectivity of brain regions, also assessed by MRI, has been implicated in the etiology of both schizophrenia [66] and bipolar disorder [67]. PET and SPECT have been used extensively to elucidate the mechanism of action of antipsychotic drugs in schizophrenia [68-71], as well as mood stabilizers in affective disorders [72-75]. For example, the principal pharmacological treatment for bipolar disorder, lithium, has been reported to be neuroprotective and to prevent gray matter volume loss in chronically treated neuropsychiatric patients [76,77].

Reduced neuronal activity in the dorsolateral prefrontal cortex (DLPFC), as assessed by PET/SPECT, is common in schizophrenia [78,79] and has been linked with such symptoms as lack of motivation and impaired abstract thinking [80]. On the contrary, increased neuronal activity in the DLPFC, measured via functional MRI (fMRI), has also been reported in association with certain cognitive tasks such as item recognition [81-83]. Some investigators have used the morphometric alterations in the DLPFC to distinguish between bipolar disorder and schizophrenia [84], which some believe to be different manifestations of psychosis along the same continuum [85-88].

Complexity of the brain & tissue heterogeneity

In many cases, the complexity of the brain itself presents a severe road block in the identification of utile biomarkers. In most organs, for example liver and muscle, cells are generally homogenous in their phenotypes, transcriptomes, proteomes and cellular interactions. This is not the case in the brain,
been demonstrated in many cases of PD [100]. Another hallmark
addition, the deposition of amyloid, a hallmark of AD, has also
progressive supranuclear palsy [90] as well as AD [91]. These new
neuropathological features, which share a significant number of common characteristics. For example, immunohistochemical staining for the genetic PD biomarker, α-synuclein, identified numerous, varied, and as yet unidentified subtypes of Lewy body pathology along an LBD continuum [99], some showing LBD pathology alongside that of progressive supranuclear palsy [90] as well as AD [91]. These new findings have instigated a review of the Lewy body pathology in both PD [92] and dementia with Lewy bodies [93].

Cases of mixed pathology are more common than expected. Barker and coworkers reported that AD was present in 66 and 77% of LBD and vascular dementia patients, respectively [99]. In addition, the deposition of amyloid, a hallmark of AD, has also been demonstrated in many cases of PD [100]. Another hallmark of AD, tau pathology, is also common in frontotemporal dementia with parkinsonism [94], PD [95,96], dementia with Lewy bodies [97] and progressive supranuclear palsy [98]. Thus, diagnosis has become extremely stringent. In the case of AD, both cerebral β-amyloid deposition and neocortical neurofibrillar tangles are necessary for diagnosis [101]. Cases with either of the two lesions or cases with both lesions but in the incorrect brain location, are excluded from the diagnosis of AD and classified otherwise [7].

As there are few pathological markers discovered as yet for the neurobehavioral diseases, these observations apply mainly to the neurodegenerative diseases. Therefore, the presence of histopathological lesions in neurodegenerative diseases provides a distinct advantage over neurobehavioral diseases, which generally lack pathological entities that can help confirm or deny the previous clinical diagnosis.

Several approaches are available to help unravel the complexities of neuropathological tissue heterogeneity in the brain. One example is to purify individual cells of interest and one technique for doing so is laser capture microdissection (LCM). In the past, utilization of microarray strategies has required the use of large amounts of RNA (5–10 µg) and thus large tissue dissections from the brain. LCM, combined with new RNA amplification protocols, has permitted microarray analyses to be performed with as little as 10 ng of starting material and has revolutionized the study of transcriptomes from single cell populations [102]. Transparent, thermoplastic film on a standard glass histopathology slide to create strong focal adhesion allows selective procurement of the targeted cells by employing a carbon dioxide laser pulse that specifically activates the film above the cells of interest. LCM thus permits a rapid, one-step acquisition of selected human cell populations from a section of complex, heterogeneous tissue [103]. This breakthrough has not only helped reduce the level of tissue heterogeneity inherent to these types of investigations, it has enabled researchers to access tissue precisely at and around the site of pathology. In AD for example, RNA from individual neurons containing neurofibrillary tangles versus that from nearby nontangle-containing neurons have been isolated and examined via expression arrays. This results in expression profiles of tangled versus nontangled neurons in AD.

This approach can be extended to a variety of neurologic diseases that show known histopathology (e.g., PD and LBD), as well as neuropsychiatric diseases with no known pathology, based on the biomarkers defined by neuroimaging studies.

Several groups have undertaken large-scale efforts to create publicly accessible atlas of the human brain in an effort to help deal with the issue of tissue heterogeneity. One of these endeavors is GEN SAT, the Gene Expression Nervous System Atlas [201]. This ambitious project seeks to create a database of CNS gene expression during development at the cellular level to facilitate investigation of cell-specific genetic and physiological phenomena in the brain. Utilizing a combination of in situ hybridization and bacterial artificial chromosome (BAC) vectors containing green fluorescent protein-tagged genes of interest in transgenic mice, Gong and coworkers have been able to identify several genes and gene pathways involved in neural specification (Gsd), axon target interactions (Sema3b), neuronal migration (Lhx6 and Pde1c) and hundreds of other genes [104]. These results underscore the power of this approach for identifying and characterizing anatomical, genetic and physiological factors in specific cell populations.

The Allen Institute for Brain Science has initiated the Allen Brain Atlas [202], which proposes to capitalize on recent advances in computer science, bioinformatics, imaging analysis and the sequencing of the human genome. The aim is to create the most comprehensive map possible of the brain at the cellular level, illustrating the functional anatomy of the brain through a collection of gene expression maps, brain circuits and cell locations. In the future, the Allen Brain Atlas plans to extend the database to include multiple strains of mice and to compile data from nonmurine species, such as human and nonhuman primates, as well as adding pathway, connectivity, proteomics and behavioral data to augment the anatomic and genetic data.

A commercial group has also begun to produce, collect and integrate accurate, 3D volumetric data on gene expression within the mouse brain and to correlate that data with the
developing wealth of learning on the architecture and functions of brain structures, circuits and cells. Neurome, Inc., has developed several new technologies for teasing out the complexities of tissue heterogeneity in the brain, including MiceSlice™, NeuroZoom™, BrainArchive™ and BrainPrints™ [203]. These tools incorporate standardized preparation of brain sections followed by precise, computer-aided extraction, analyses and display of quantitative data from the resulting microscope images. These data will be archived in a publicly accessed web database (BrainArchive) that will permit integration and comparison of brain structure and circuitry data as well as automated comparison of quantitative, spatial and volumetric data from the experimental mice. These novel mechanisms will undoubtedly contribute to an increased understanding of brain structure and function at the cellular level and will likely become a model for continuing investigation into the etiology of neurological diseases.

Lack of functional end points & models for validation

Finally, in the list of roadblocks to obtaining neurological biomarkers is the dearth of functional and model systems for validation. Since neurological diseases are partially, or wholly, behavioral in nature, it is difficult to ascertain many of the phenotypic characteristics as they can occur in vitro or in vivo. Thus, surrogate end points (i.e., biomarkers) need to be designated, which can help to identify these characteristic pathologies without necessarily being able to observe the underlying behavioral attributes. These end points must then be functionally validated. In other words, it must be evaluated whether or not changes in surrogate end points (observed in the RNA, DNA or protein) have any measurable effect on the actual phenotype of a cell or animal model.

In general, there are two methods for in vitro functional validation of gene expression studies in neurological diseases. These include methods to both decrease expression of specific target genes or proteins and methods to increase expression of specific genes or proteins. These approaches can be extremely informative when combined with powerful functional validation assays that measure a specific cellular neurological phenotype.

RNA interference (RNAi) is a novel strategy to suppress gene expression and subsequently validate surrogate end points in cellular models. RNAi is an evolutionarily conserved mechanism whereby long (typically >200 nucleotides [nt]), double-stranded RNA (dsRNA) molecules specifically suppress the expression of genes bearing their complementary sequences [105]. Introduction of long dsRNA (>30 nt) in mammalian cells, however, initiates a potent antiviral response, characterized by nonspecific inhibition of protein synthesis and RNA degradation [106]. This antiviral response can be circumvented by the introduction and/or expression of small interfering RNAs (siRNAs). siRNAs are dsRNAs that have been processed into 20–25-nt RNAs by an RNase III-like enzyme, for example Dicer. The siRNAs assemble into endoribonuclease-containing complexes known as RNA-induced silencing complexes (RISCs), which become active upon linearization of the siRNA.

The unwound siRNAs serve to direct the RISCs to RNA molecules of complementary sequence, where they cleave near the middle of the region bound by the siRNA and destroy the associated RNA [107].

RNAi has been used successfully to silence gene expression in a variety of systems. Of particular relevance to the study of neurological diseases, Luo and colleagues recently used siRNA to elucidate the precise functions of several cofactors involved in presenilin (PS) 1/γ-secretase-mediated cleavage of β-amyloid precursor protein (βAPP) in AD [108]. The cofactors APH-1 and PEN-2 have been shown to physically interact with PS1 and are necessary for γ-secretase activity. RNAi of PEN-2 abolished endoproteolytic cleavage of PS1, whereas overexpression of PEN-2 elevated the level of processed PS1 by-products, suggesting a primary role for PEN-2 in PS1 endoproteolysis. RNAi of APH-1 diminished the accumulation of PS1 resulting from PEN-2 RNAi and overexpression of APH-1 facilitated PEN-2-mediated PS1 proteolysis, indicating a more facilitative role for this cofactor [108].

A complementary approach to RNAi is to induce expression of specific genes in transgenic cell lines, thereby allowing one to assess the effects of increased expression and/or activity of individual genes. Application of overexpression and knock-down studies, when coupled to relevant in vitro assays, can be a powerful tool for functional validation. For example, several in vitro assays have been developed to study phenotypic markers of AD, such as hyperphosphorylation of tau protein and production of β-amyloid peptides [109]. Measurement of these fundamental molecular pathways provides insights into the formation of neurofibrillary tangles and amyloid plaques, two pathologies that are intimately involved in the development of AD. Importantly, with the availability of the entire human genome sequence and rapid proliferation of genomics technologies, high-throughput functional validation assays will be required to interpret the results of genomics experiments. The above in vitro methods are likely to become increasingly important in the study of neurological diseases since they can be applied to the study of any disease for which one has a relevant cell line and phenotypic assay.

While RNAi and gene overexpression studies are valuable tools for functional validation in cellular models (in vitro), animal models provide the best in vivo measure of functional validation. Although model systems are common in other diseases, when no behavioral phenotype manifests, such as in obesity [110,111], diabetes [112,113], autoimmune disease [114,115] and cancer [116–120], models for neurological disorders are much more difficult to come by.

Animal models have been invaluable for validating new biomarkers for a variety of neurological diseases. Several animal models of AD, mostly mouse, have been generated to test biomarker validation. Although transgenic mice harboring overexpressing mutations in either partial [121–125] or whole [126,127] human βAPP or PS1/PS2 [128–130] alone are not sufficient to develop the level of necessary AD-like phenotype, double-transgenic animals with two homozymous mutations (APP/PS) have robust Aβ amyloid deposition [131,132].
Investigators using these double-transgenic AD mice have demonstrated that mitochondrial dysfunction (hypo- and/or hypermetabolism of cytochrome oxidase activity in different regions of the brain) could be a potential biomarker for AD [133]. Despite advancements in the development of APP/PS double-transgenic AD mice, the development of the tau pathology concurrent with the Aβ/amloid deposition has been an elusive phenotype. A recent investigation by Oddo has developed the first triple-transgenic mouse model by directly introducing two additional transgenes into the germline of an already genetically modified mouse [134]. This is the first known mouse model to display both the plaque and tangle pathology. This model has enabled researchers to investigate the series of events leading to AD in vivo. These investigations suggest that synaptic dysfunction, including long-term potentiation deficits, could be a biomarker for AD and that this phenomenon likely precedes plaque and tangle pathology, offering new insight into the mechanism of degeneration in this debilitating disease.

Of course, these models will only be utilized to their full potential when synergistically combined with other types of biomarker identification. For example, knockout mice have been used to test the efficacy of novel PET radioligands that are being tested as potential pharmacotherapeutic targets. For instance, an antipsychotic, the putative dopamine D2 and 5-HT2A receptor antagonist M100907, was tested for binding efficiency to its receptors and shown to have a novel mechanism that occurred in the absence of D2 receptor binding [135]. Another group used this technology to measure the neurotransmitter concentration in response to treatment with another pharmaceutical D2 receptor antagonist, 11C raclopride, to study dopamine release in schizophrenia in response to amphetamines [136]. Eckelman suggests that this symbiotic relationship could facilitate the process of drug discovery by several methods, including authenticating the drug localization method, establishing the transport efficiency of drugs to their intended targets, determining the saturaibility of receptor sites and/or measuring pharmaceutical half-life [137].

In an alternative application, Dickey and coworkers used the APP/PS mice in combination with microarray analyses to identify genes that have altered expression in Aβ/amyloid deposition [138]. Using this approach they identified several genes associated with long-term potentiation and memory formation that showed decreased expression in the absence of altered expression of known presynaptic markers. These analyses would suggest that, in contrast to a previous study [134], synaptic dysfunction is not necessarily important for development of an AD-like pathology. However, this example highlights the factors that need to be addressed in future investigations, including standardization of protocols to ensure proper interpretation of the results. In the above example with apparently contradictory results, it is important to note that the mouse strains used are significantly different, one containing both amyloid and tau pathology [134] and the other containing only amyloid pathology [138]. This difference could suggest that tau pathology, present in the triple-transgenic, is responsible for the synaptic dysfunction not detected in the APP/PS mice.

Advances in biomarker identification

In addition to the rapid expansion of neuroimaging in biomarker identification, new genetic and genomic tools have revolutionized the way in which neurologic diseases are investigated. Genotyping vast numbers of genetic polymorphisms in large populations is increasingly important for the identification of etiologically relevant mutations. Many of these etiologically relevant mutations have been discovered in response to monumental advances in high-throughput genome screening techniques that have been made in recent years. High-throughput screening started concurrent to the Human Genome Project. With this multi-institutional undertaking, the DNA sequences for humans and other common model organisms have and are being generated at an unprecedented rate. The release of the complete human genome sequence has precipitated vast numbers of investigations incorporating human genome sequence data into diverse applications including the previously mentioned models of structural and functional brain mapping via GEN SAT, the Allen Brain Atlas and Neurome.

Another analysis that has revolutionized the identification of biomarkers is positional cloning. Positional cloning is a strategy for identifying genes that are etiological in nature, based upon their location within the genome. This is generally accomplished through genetic mapping but can also be achieved by cytogenetic visualization of chromosomal abnormalities, as was the case with Duchenne muscular dystrophy [148]. After physical mapping of the region, which necessitates the identification of gene content in that interval of the genome, functional candidate genes are sequenced to identify etiological sequence variants. Microarray technologies, including expression arrays, exon arrays, single nucleotide polymorphism (SNP) arrays and sequencing arrays, have recently come to play a role as adjuncts to physical mapping and identification of gene targets associated with disease [149].

Concurrent with the benefits of neuroimaging in diagnosis of neurologic disorders, there is the added benefit of being
able to look at early, predisease states of the brain and assess longitudinally the effects of the disease over time. It should be noted that although diagnostic utility of neuroimaging is increasingly important, the ultimate goal of neuroimaging in neurologic disorders is to identify physical changes in the brain that can be used as surrogate markers (e.g., quantification of blood glucose in diabetes) for the pathologic changes that bring about the disease (i.e., biomarkers).

**Existing biomarkers in neurological disease**

DeKosky and Marek have divided biomarkers into four categories: clinical, neuroimaging, genetic and biochemical [150]. In addition to established and evolving clinical diagnostic measures, such as psychometric and/or neuropsychological testing, the identification of specific genetic and protein biomarkers is rapidly accelerating. Although a small number of general imaging studies have demonstrated gross metabolic/structural changes, most of these have not been shown to be diagnostic or prognostic.

In contrast, genetic strategies have already yielded a number of AD-specific biomarkers. To date, most of the genetic research on AD has focused on four confirmed factors contributing to the inherited form of the disease, and all of these factors have been shown to increase the production and/or deposition of amyloid β-protein in the brain [151]. Mutations in the presenilin genes (PS1 and PS2) lead to the most aggressive form compared with sporadic AD patients [154,155]. Recent research suggests that presenilin itself may actually carry out the β-secretase activity that cleaves APP [156].

The missense mutations are substantially more rare within the βAPP gene itself. These mutations, located at or near sites of secretase (α, β and/or γ) cleavage, lead to AD by altering proteolytic processing at these secretase sites that enhances cleavages that generate Aβ and promote amyloidogenesis. Nine βAPP missense mutations have been identified so far. Last but not least, of the four confirmed AD-associated gene factors is the ε4 allele of the apolipoprotein E (ApoE4) gene. ApoE4 is, at present, the most significant predictor of outcome for sporadic AD. Inheritance of one or two copies of the ε4 allele increases the probability of developing AD and lessens mean age of onset in a dose-dependent manner compared with those carrying the ε2 or ε3 allele [157,158]. ApoE4 appears to augment the steady-state levels of Aβ (particularly Aβ40) by diminishing its elimination from the brain [159]. It should be noted that the presence of an ε4 allele is a contributing, not causal, factor in the etiology of AD, as some humans homozygous for the ε4 allele show no AD symptoms and some with no ε4 alleles still develop the disease.

It is certain that other genetic factors mediating the etiology of AD exist and several new candidate genes have been identified. Alterations in or near the α2-macroglobulin (α2M) gene have been shown to segregate with the AD phenotype in some late-onset subjects [160]. In recent years, mutations in the homocysteine gene [161], ATP-binding cassette transporter A1 gene [162], Down’s syndrome critical region 1 gene [163], the serotonin transporter gene [164] and several others have all been implicated in the etiology of AD. A comprehensive review of existing biomarkers for AD has been published by the Biological Markers Working Group and many more of these potential candidate genes are certain to emerge in the months and years to come [165].

Advances in genomics have made the investigation of genes associated with AD much more robust. Specifically, microarray technologies have allowed the assessment of thousands of genes at a time under different AD-related conditions. Recent research using microarrays of various types has identified hundreds of genes that are selectively up- or downregulated in patients displaying a range of AD symptomologies [166–168].

Although not as far advanced as AD, research into genetic markers for schizophrenia is also on the rise. Several biomarkers, many of them already independently reproduced, have been identified in just the last 5 years. SNPs in the neuregulin 1 [169,170], dysbindin [171,172], regulator of G-protein signaling-4 [173–175] and especially catechol-O-methyltransferase [176–179] genes have been replicated and are all good candidates for schizophrenia biomarkers. In addition, two of these potential loci already have relevant animal models that are currently being tested [169,176]. Though there are most certainly other genetic markers that have been associated with schizophrenia in the 100 years of investigation that this disease has undergone, few of the candidates have been successfully reproduced. It is hoped that these new potential markers will be further confirmed and lead to the types of therapeutic advancements that have been made in neurodegenerative disorders such as AD and PD.

**Expert opinion**

The identification of new biomarkers for neurological disease is essential for the continued advancement of clinical diagnostics and therapeutics. However, due to the inherent complexities involved in brain-related research, there are also intrinsic challenges. Four basic challenges in the identification of biomarkers in neurological disease were discussed in this review. These include:

- Availability of tissue at the site of pathology
- Poor clinical diagnostics and the extent of disease progression at time of diagnosis
- Complexity of the brain/tissue heterogeneity
- Lack of functional end points and models for validation studies

Several ways to overcome these basic challenges were also presented in the framework of one neurological disorder, AD. In short, advanced brain banking protocols will enable researchers to have access to high-quality tissues at the site(s)
of pathology. Highly developed imaging technologies can be used to improve diagnostics and therefore, permit clinicians to recognize and circumvent the rapid disease progression that is characteristic of these types of disorders. Sophisticated microscopical methodologies can be utilized to tease apart the innate heterogeneity that exists in an organ as complex as the brain by isolating individual neurons for examination. Lastly, several novel techniques for functional end-point identification and validation were presented.

Combining these new state-of-the-art strategies with the rapidly evolving strides in clinical diagnostic, genetic and therapeutic paradigms will enable a more rigorous diagnostic classification based on the identification of new and more efficacious biomarkers. These techniques will continue to reshape the research and development strategies for drugs, vaccines and novel gene therapies. The use of high-throughput screening technologies (such as cell- and tissue-based microarrays and mass spectrometry-based proteomic scanning) has rapidly accelerated the quantity and quality of molecular factors entering the drug development conduit. This has resulted in an increasing number of therapeutic targets, which necessitates new validation paradigms to ensure that these new therapeutics make it from the laboratory bench to the patient bedside.

In an effort to combine these various technologies to create a better, more sensitive assay, the Translational Genomics Research Institute (TGen) has undertaken a large, multi-institutional investigation into the etiology of AD. This National Institute on Aging-funded project will utilize the most up-to-date brain imaging techniques to identify regions of interest based on metabolic changes associated with the disease state. Once identified, individual neurons from these regions will be isolated from well-characterized patient samples by LCM and expression profiling of isolated RNA will be performed. Finally, RNAi will be utilized for functional validation of the resulting gene lists.

Another project being undertaken by TGen, which illustrates the scope of what the new high-throughput technologies can offer to biomarker identification in neurological disease, is a large whole-genome association study. This project is designed to look at almost every variable within the human genome and correlate it to a phenotype. It is to be performed in thousands of cases and compared with controls. This is a striking example of a revolutionary new approach to the identification of common pathogenic variants or combinations of variants that are associated with neurological disease.

Five-year view

In order for biomarkers to be utilized to their full potential, the key components of the different available technologies should be combined. Thus, standardization of diagnostic protocols to include the addition of neuroimaging and individual genotyping for the identification of genetic/genomic biomarkers to the customary clinical protocols of today is essential to achieving better precision in disease diagnostics. In addition, the constant advances in neuroimaging will eventually lead to their use as a more reliable diagnostic tool. This increased diagnostic efficacy should allow therapeutic intervention before more significant pathologies develop and, thus, result in more effective treatments for these patients who have, as of yet, had little relief from these devastating diseases.

In addition to its role in clinical diagnostics, neuroimaging data will likely be used to identify specific biomarkers in the future. The development of specific ligands that target defined pathologies is of the utmost importance and will revolutionize biomarker identification. For instance, research is underway to develop specific ligands that target microglial activation, a pathological change believed to occur in the early stages of AD and PD [180]. In addition, a novel amyloid-imaging PET tracer, Pittsburgh compound B (PIB), has been shown to provide quantitative information on amyloid deposits in living subjects [181]. This technique appears to accurately and consistently detect amyloid deposition longitudinally in patients from the early stages of AD through conversion to full Alzheimer’s dementia. Although the sample group was small, the results of this preliminary study suggest that this compound may constitute the first true biomarker derived solely from imaging data.

Additionally, 5 years will bring to the forefront mass spectrometry protocols, which are still in their infancy at present. Matrix-assisted laser desorption ionization with time of flight detectors (MALDI-TOF) and surface-enhanced laser desorption/ionization (SELDI-TOF) methods of mass spectrometry from serum and CSF are going to revolutionize the way biomarkers can be identified. These new technologies should generate rapid protein expression profiles of the more than half a million proteins or protein variants isolated from any of a variety of tissue samples. It can and will in the future be used for biomarker identification and diagnostics, as well as the study of protein–protein and protein–DNA interactions. In addition, its versatility will likely permit its use as a functional validation mechanism for potential biomarkers identified through other high-throughput techniques. For instance, the products of individual genes derived from gene lists generated via microarrays can be functionally validated on a MALDI- or SELDI-TOF protein chip to elucidate whether the altered messenger RNA expression pattern is translated into altered protein expression or is masked by some post-transcriptional, post-translational or RNA-stabilization mechanism.

Clinical diagnoses will be increasingly refined and sub-classifications of common diseases will most certainly occur as we increase our ability to discover genetic and histologic variants. In addition, some diseases that are thought to be distinct may actually contain overlapping pathological features. Thus, it is likely that new classifications, such as synucleinopathies and/or tauopathies, will be used in the future to describe what could conceivably become a continuum of neurodegenerative disease, analogous to that being debated in neuropsychiatric disease in relation to bipolar disorder and schizophrenic psychoses [182–184].
T he future of functional validation lies in the hands of novel high-throughput assays that can precisely and reproducibly analyze the huge amounts of data that are being generated by current and future experimental methodologies. Existing array-based chip technologies are advancing at an astounding rate. For example, the 10K SNP chip (Affymetrix) that can analyze more than 10,000 known human SNPs simultaneously was debuted at the European Society for Human Genetics Conference in Birmingham, U.K. in 2003. It is now possible to assay an estimated 1.5 million SNPs at a time using a new wafer-based technology from Perlegen, Inc. (Mountain View). New developments in SNP technology include a 250K SNP array anchored in the Hapmap [204], an internationally developed haplotype map of the human genome that will describe the common patterns of human DNA sequence variation and will be available in the next few years. Another recent development in SNP typing is under development by Illumina. It attempts to combine the use of fiber optic bundles and specially coated beads that self assemble into an array. There can be up to 1500 different types in each individual bundle, with each bead containing oligonucleotides for a specific sequence. DNA from individual samples bind their complementary oligonucleotide on the coated bead permitting the tens of thousands of SNPs to be assayed at the same time [185]. These rapidly advancing technologies will continue to expand our knowledge of the human genome and help direct our search for efficacious biomarkers.

With these kinds of assays generating enormous amounts of data, functional validation becomes the bottleneck in the process of biomarker identification. Thus, it is likely that the future of functional validation rests in the capable hands of the siRNA technologies. With the completion of the Human Genome Project, the development of siRNAs that will silence every gene in the genome is currently underway. Once generated, the applications for this type of assay are boundless and will undoubtedly revolutionize the speed at which the vast amounts of data generated by the new, high-throughput chip technologies can be validated. In vitro analyses are currently fairly laborious and time consuming. Thus, the key will be to develop systems that will allow the in vitro siRNA assays to be employed in a high-throughput manner in order to keep up with the amount of data that will be generated by the data-generating infrastructure.

Lastly, there will be a great need for integrated databasing solutions to catalog and analyze the vast amounts of data that will be generated for these complex systems. The creation of a standardized protocol for brain banking guidelines used for collection, dissection and preservation of tissue samples is the key to achieving the level of tissue availability that is needed to fulfill the research demands of the future. In order to accommodate these research needs, the banking protocols and procedures must be designed for a broad set of experimental approaches. This may include significantly decreasing the PMI of many brain collections, implementing multiple methods of preservation (e.g., freezing and various fixation protocols) and thorough characterization and databasing of donors and specimens. Ideally, all data collected from primary research of banked materials should be routed back to the brain banking organization for assimilation into an ever-expanding database. Although data sharing can be a controversial topic with respect to publication, intellectual property rights and data access, one example of this, the National Alzheimer’s Coordinating Center, is already utilizing this type of data-sharing strategy to facilitate collaborative research among its investigators. The creation of a centralized repository for archived tissues concurrent with a national database network of ante and post-mortem characterization of each donor and specimen that can be securely accessed by researchers via the internet would enable researchers across the globe to have access to the best possible samples for their respective experiments.

Although an extremely valuable resource, systems such as the National Center for Biotechnology Information/Genbank RNA, DNA and protein sequence database [205], will need to be associated with newer databasing paradigms such as the NINDS/National Institute of Mental Health array consortium [206], housed at the TGen. This network facilitates the acquisition and dissemination of high-quality expression profiling array data to many researchers, by establishing, cataloging and perhaps eventually standardizing the way in which this data is accessed. These types of databases will likely be established for the various types of data output (e.g., expression profiles, SNP genotypes and protein profiles) as well as for individual diseases (e.g., AD and schizophrenia). In addition, these databases, new and existing, will have to be expanded to include other types of data that will be garnered from the new methodologies that will be employed in the next 5 years.

Thus, a functional paradigm that might reflect the general course of neurological disease biomarker identification in the next 5 years can be summarized as follows:

- Analysis and dissection of highly complex, heterogeneous tissue samples using high-performance imaging systems (e.g., LCM)
- Examination of cellular macromolecules (e.g., RNA, DNA and protein) via some high-throughput chip-based mechanism (SNP typing, microarray, MALDI- and/or SELDI-TOF mass spectrometry)
- Automated data generation and archiving
- Data extraction
- Integration of all data into internet-based information management systems

High-throughput strategies such as this will successfully ensure that the identification of biomarkers in neurological diseases can continue to advance. It will also facilitate further development of novel diagnostics and therapeutics to aid the patients and families that are suffering from these tragic neurological diseases.
Key issues

- Identification of biomarkers is particularly challenging in neurologic diseases due to a variety of factors, including availability of tissue at the site of pathology, poor clinical diagnostics (clinical heterogeneity), complexity of the brain (tissue heterogeneity) and lack of functional end points and models for validation.
- Advances in research technology, including brain banking, brain imaging, laser capture microdissection and online databases (e.g., GENSAT and Allen Brain Atlas), high-throughput genomics and new functional validation techniques such as RNA interference can help to overcome these challenges.
- New databasing protocols are necessary to accommodate the vast amounts of data generated by the new high-throughput genomics and brain-imaging technologies.
- Access to generated data is critical to continued progress in biomarker identification so that this information can be used to devise new diagnostics and therapeutics that will help people suffering from these insidious diseases.

References

Papers of special note have been highlighted as:

- of interest
- of considerable interest

Schlosser R, Schlegel S, Hiemke C

Goodwin GM, Cavanagh JT, Glabus MF


Colangelo V, Schurr J, Ball MJ et al. Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor


183 Angst J. Historical aspects of the dichotomy between manic-depressive disorders and schizophrenia. Schizophr. Res. 57(1), 5-13 (2002).
