Discovery and development of biomarkers of neurological disease

Travis Dunckley, Keith D. Coon and Dietrich A. Stephan

The identification of clinically relevant biomarkers for neurological diseases poses unique challenges. These include an historical lack of availability of relevant tissues from the site of pathology, relatively poorly matured techniques for disease diagnosis, the complexity and cellular heterogeneity of the brain, and a clear deficiency of models for functional validation of candidate biomarkers. Here, the unique challenges that neurological disorders introduce to biomarker discovery are described and how modern technological advances in genomics, proteomics and metabolomics are overcoming these obstacles and are driving the discovery of novel biomarkers to improve early diagnosis and therapeutic treatment is discussed.

Ultimately, the main goals of research on human diseases are to cure the disorders or to increase the length and quality of life of those affected. Novel biomarker identification in neurological disorders will facilitate the achievement of these goals, first, by providing sensitive and selective clinical correlates for the evaluation and diagnosis of those affected and, second, by providing insights into disease mechanisms that can be used to identify therapeutic targets and to develop efficacious compounds to target them. Novel biomarker identification for neurological disorders will address the current shortcomings in diagnosis and therapeutics for this broad class of disorders, including devastating neurodegenerative and neurobehavioral diseases that affect millions of people annually.

Biomarkers are pharmacological and physiological measurements, or specific biochemicals in the body, that have a particular molecular feature that makes them useful for measuring the progress of disease or the effects of treatment. Biomarkers that will be useful for either disease prediction or treatment should have one or more of several properties, including: (i) specific and selective association with illness in a population; (ii) heritability; (iii) state independence and presence, whether or not the clinical phenotype of the disease is present; (iv) co-segregation with disease within families; and (v) presence in relatives of affected individuals at a higher rate than in the general population [1,2]. Importantly, multiple relevant disease biomarkers that can be examined concurrently will increase diagnostic specificity. Because of the inherent difficulties in characterizing and accessing neurological disorders, biomarkers that satisfy these criteria have been difficult to identify.

This review will discuss some of the challenges in attempting to identify biomarkers in neurological disease and how recent advancements in genomics, proteomics and metabolomics will help to overcome these challenges and lead to improved diagnostics and therapeutics.

Challenges to biomarker discovery in neurological disorders

Four basic challenges to biomarker identification in neurological disease are: (i) the availability of tissue at the site of pathology; (ii) poor clinical diagnostics and extent of disease progression at the time of

Travis Dunckley*
Keith D. Coon
Dietrich A. Stephan
Neurogenomics Division,
The Translational Genomics
Research Institute,
Phoenix,
AZ 85004, USA
*e-mail: tdunckley@tgen.org

1359-6446/04/$ – see front matter PII: S1359-6446(04)033534-9

diagnosis; (iii) the complexity of the brain and tissue heterogeneity; and (iv) the lack of functional endpoints and models for validation.

Availability of tissue at the site of pathology

Many of the difficulties in biomarker identification in neurological disorders are related to the acquisition and quality of the necessary tissues, especially those at the actual site of pathology. The rarity and dangers associated with brain biopsies necessitate the use of post-mortem tissue samples from affected individuals [3]. In these cases, the disease is most often at end-stage and the affected tissues have been ravaged by the disease process, leaving little experimental material of good enough quality to investigate early etiologies [4,5].

For the purposes of developing disease diagnostics, peripheral tissues such as blood, urine and saliva are easily attainable ante-mortem. However, for discovering etiologically relevant genes, proteins or small molecules, the preferred biological source is often those pathologically affected tissues that are more difficult to attain. Progress in overcoming the problem of tissue availability and acquisition has been achieved with advances in brain banking. New freezing techniques and shorter post-mortem intervals (PMIs) are making higher-quality tissue available more rapidly. For example, a brain bank at Duke University has developed a rapid brain autopsy protocol, which processes brains within one hour after death [6] and Sun Health Research Institute in Arizona maintains an average PMI of 2.6 h [T. Beach, personal communication]. Institutes all over the country are attempting to achieve similar PMIs to standardize protocols as well as to provide the highest quality samples to researchers. The expedited processing of the brain samples minimizes the loss of tissue and, thus, increases the availability of tissue at the site of pathology.

Many attempts at bypassing the problem of tissue availability have used *in vitro* and animal models of neurological disease. However, given the complexities of human neurological disorders, which often contain significant behavioral components, these models are often imperfect. Thus, where possible, well-characterized human tissues are the preferred substrates for neurological studies, placing significant emphasis on the need to further advance and standardize brain-banking programs.

In addition to the speed at which the specimens are being processed, new computerized databasing methods are cataloging and organizing the donor submissions in ways that maximize the amount of information available to the researcher. Detailed knowledge of the neuropsychiatric, neurologic, neuropathologic, familial/social histories and medical characteristics (i.e. comorbid disease, substance abuse, renal/hepatic complications, pharmacological treatments) of the tissue donors, which can influence brain biochemistry and impact evaluation of inclusion or exclusion criteria of appropriate samples for a clinical investigation [7], are now easily and rapidly available to the researcher. This information will enable 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 and the extent of disease progression at the time of diagnosis

Clinical diagnostics and (sub)stratification of patient populations are poorly developed for most neurodegenerative diseases, most notably multiple sclerosis [8,9], but also to a lesser degree with atypical forms of Parkinson's disease (PD) [10] and Alzheimer's disease (AD) [11]. Even in the better case scenario of AD, clinicopathological diagnosis has been demonstrated to have a specificity ranging between 76% and 88% and sensitivity between 53% and 65% [12–14] with a confirmation rate of 'probable' AD as low as 65% [15]. In neurodegenerative disorders as a whole, clinical diagnosis has been reported to be accurate in only ~70–80% of cases [16]. Although beyond the scope of this review, advances in modern brain imaging techniques show promise in providing more definitive ante-mortem diagnosis of neurological disorders.

Currently, definitive clinical diagnoses of AD, PD, Lewy body disease (LBD), and many other neurodegenerative diseases, can only be achieved through evaluation of their respective pathological traits within the brain upon autopsy. Many of these neurodegenerative diseases are differentiated by a complex set of neuropathological features, which share a significant number of common characteristics. Cases of mixed pathology are common. For example, AD pathology is present in 66% and 77% of LBD and vascular dementia patients, respectively [17]. In addition, the deposition of amyloid, a hallmark of AD, has also been demonstrated in many cases of PD [18]. Another hallmark of AD, tau pathology, is common in frontotemporal dementia with Parkinsonism [19], dementia with Lewy bodies [20] and progressive supranuclear palsy [21]. Thus, diagnosis has become extremely stringent. In the case of AD, both cerebral β-amyloid deposition and neocortical neurofibrillary tangles are necessary for diagnosis [22]. However, neuropathological diagnosis can only be performed post-mortem, obviating any opportunities for early therapeutic interventions. In this regard, biomarker identification for early diagnostics will be crucial for improving treatment of affected individuals.

Complexity of the brain and tissue heterogeneity

In many cases, the complexity of the brain itself presents a severe roadblock to identification of useful biomarkers. In most organs (e.g. liver, muscle), cells are more homogenous in their phenotypes, transcriptomes, proteomes and cellular interactions. However, in the brain, transcriptomes, proteomes, morphological phenotypes and interactive

connections vary widely within the neurons and glia. Diverse cellular experiences can be interpreted as differences that manifest on the biochemical and epigenetic level. Additionally, the complex experiences and interactions of each individual must be considered. Thus, human behavior, the phenotypic output of the brain, is much more than the sum of its parts.

Heterogeneity of the representative neuropathologies further confounds biomarker identification. If the neuropathological or neuropsychiatric characterization of the sample is incorrect or unavailable to the researcher, the conclusions drawn from molecular biological and/or neurochemical investigations will be invalid.

Lack of functional endpoints and models for validation

The paucity of model systems for functional validation in neurological diseases makes confirmation of candidate biomarkers extremely difficult. Neurological diseases are partially, or wholly, behavioral in nature, therefore it is difficult to ascertain many of the phenotypic characteristics as they occur *in vitro* or *in vivo*. Thus, surrogate endpoints (i.e. biomarkers) need to be designated, which can help identify these characteristic pathologies without necessarily being able to observe the underlying behavioral attributes and these endpoints must then be functionally validated. It is important to evaluate whether or not changes in surrogate endpoints (observed in the RNA, DNA, protein, metabolites, and so on) 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 and increase expression of specific target genes or proteins. These approaches can be extremely informative when combined with powerful functional validation assays that measure a specific cellular, neurologically relevant phenotype.

RNA interference (RNAi) is a strategy to suppress gene expression and to subsequently validate surrogate endpoints in cellular models. RNAi is an evolutionarily conserved mechanism whereby 20–25 nucleotides of double stranded RNA molecules [small interfering RNAs (siRNAs)] direct the degradation of RNA molecules of complementary sequence [23]. 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 [24] recently used siRNA to elucidate the precise functions of several cofactors involved in presenilin (PS1)/γ-secretase-mediated cleavage of β-amyloid precursor protein (βAPP) in AD. These 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 [24]. RNAi has more recently been extended to animal models, which are the most likely functional validation endpoints in neurological diseases. To this end, efficient gene silencing has been achieved *in vivo* in mouse brain using picomolar quantities of siRNA [25]. However, the foreseeable reality is that functional validation efforts in animal models will be of relatively low-throughput and should be undertaken following sufficient evidence for a candidate biomarker gained from experiments in cell culture validation models, which are amenable to high-throughput approaches.

Gene overexpression studies, which increase expression and/or activity of specific genes of interest, provide a complementary approach to RNAi. Application of overexpression and knockdown 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, production of β-amyloid peptides and β-amyloid toxicity [26]. Measurement of these fundamental molecular pathways provides insight into the formation and effects of neurofibrillary tangles and amyloid plaques, two pathologies that are intimately involved in the development of the disease. 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 *in vitro* methods described previously are likely to become increasingly important in the study of neurological diseases because they can be adapted to high-throughput platforms and applied to the study of any disease for which one has a relevant cell line and phenotypic assay.

The proper *in vitro* validation assays will provide useful information about the role of biomarkers in disease. However, animal models provide the best *in vivo* measure of functional validation. Model systems are common in other diseases, such as obesity [27,28], diabetes [29,30], autoimmune disease [31,32] or cancer [33–37], where there is no overt behavioral phenotype. However, the involvement of complex human behavior in neurological disorders has hindered the development of animal models that are truly representative of the disease process.

Nonetheless, animal models have been invaluable for validating new biomarkers for a variety of neurological diseases. These models rely more on the molecular and pathological findings of the disease rather than the psychological or behavioral aspects. Several animal models of AD have been generated for biomarker validation. For example, *Drosophila melanogaster* and *Caenorhabditis elegans* models have been used to identify genes involved in neuronal cell death of AD [38,39]. In murine models, transgenic mice harboring overexpressing mutations in either partial [40–44] or whole [45,46] human βAPP or PS1/PS2 [47–49] alone are not sufficient to develop the

TABLE 1

Summary of –omics technologies used for biomarker discovery

	Preferred tissue source	Purpose
Genomics		
Positional cloning	Any nucleated cell	Mapping of disease loci
SNP genotyping	Any nucleated cell	Identification of disease gene
Microsatellites	Any nucleated cell	Mapping of disease loci
Expression arrays	Pathologically affected cells and blood	Identification of dysregulated genes and signaling pathways and/or diagnostics
Exon arrays	Pathologically affected cells	Detect gene amplification, LOH and alternative splicing
CGH arrays	Pathologically affected cells	Detect gene amplification and LOH
Proteomics		
2DE MS	Urine, blood, saliva, CSF and affected tissues	Identification of hydrophilic, nonbasic proteins (pI <10)
LC–MS	Urine, blood, saliva, CSF and affected tissues	Identification of hydrophobic, low abundance proteins
ICAT–MS	Urine, blood, saliva, CSF and affected tissues	Identification of large, hydrophobic, low abundance proteins
DIGE–MS	Urine, blood, saliva, CSF and affected tissues	Identification of hydrophilic, nonbasic proteins (pI <10) that are differentially expressed
Metabolomics		
NMR	Urine, blood, saliva and CSF	Small molecule identification
MS	Urine, blood, saliva and CSF	Small molecule identification and characterization

The numerous tools currently available will speed the discovery of diagnostic and therapeutic biomarkers at all levels: genes, mRNA, proteins and small molecules. There have been numerous successful applications of combinations of these technologies, for example, LC–MS followed by ICAT, tandem LC for the fine purification of protein mixtures, and MS–MS for quantitation and protein identification [82,83]. Abbreviations: CGH, comparative genomic hybridization; DIGE, 2D fluorescence difference gel electrophoresis; LOH, loss of heterozygosity; pI, isoelectric point.

level of necessary AD-like pathology. However, double-transgenic animals with two homozygous mutations (APP/PS) have robust Aβ/amyloid deposition [50,51].

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 [52]. Despite the advancements in the APP/PS double transgenic AD mice, the development of the tau pathology concurrent with the Aβ/amyloid deposition has been an elusive phenotype. A recent investigation by Oddo *et al.* [53] has developed the first triple-transgenic mouse model by directly introducing two additional transgenes into the germline of an already genetically modified mouse. This is the first known mouse model to display both amyloid plaque and neurofibrillary 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 probably precedes plaque and tangle pathology, offering new insight into the mechanism of degeneration in this debilitating disease.

The efforts to date illustrate clearly that validation of human biomarkers of disease is not sufficient when using an *in vitro* or rodent model, as exemplified by the apparent difficulty in mimicking the dramatic histopathology of humans. This highlights the need to ultimately perform retrospective and prospective diagnostic clinical studies to determine the accuracy, sensitivity and specificity of any biomarker for a clinical phenotype.

Systematic approach to biomarker identification

Ideally, a systematic approach to biomarker identification will involve multiple technologies to investigate a disease process at all levels, including whole genome association studies to identify etiologic mutations or polymorphisms, as well as expression profiling, proteomics and metabolomics to identify expression signatures and protein and small-molecule profiles that are either specific to the disease process or provide mechanistic insights into disease pathology. For biomarker identification in neurological disorders, unique challenges will necessitate the simultaneous use of each of these technologies; common uses for genomics, proteomics and metabolomics in biomarker discovery are summarized in Table 1. Genomics is used to identify relevant disease genes, aberrant cellular signaling pathways and expression signatures correlated with disease. Proteomics is used to identify aberrant protein expression, post-translational modification, protein interactions and protein profiles that are specific to a particular disorder. Finally, metabolomics is implemented to identify the presence of abnormal levels of small-molecule metabolites that are specific to and indicative of an underlying disease process. The focus of this review is the application of these technologies to biomarker discovery in neurological diseases.

Genomics

Identification of pathogenic mutations

New genetic and genomic tools have revolutionized the way in which neurologic diseases are investigated. Genotyping of vast numbers of genetic polymorphisms

in large populations is increasingly important for the identification of etiologically relevant mutations. Many of these mutations have been discovered as a direct result of recent monumental advances in high-throughput genome screening techniques. These techniques use the recently completed human genome sequence to construct array-based platforms, including expression arrays, exon arrays, sequencing arrays and single nucleotide polymorphism (SNP) arrays. Whole-genome SNP-genotyping is most effectively applied to well-defined pedigrees and has become a robust tool for the rapid identification of pathogenic mutations. Current technology (Affymetrix) enables the simultaneous assessment of up to 100,000 SNPs, which are spaced, on average, 30 kb apart in the genome. Because this is an array-based platform, a single researcher can easily genotype these 100,000 SNPs from over 200 individuals in a single week, vastly exceeding the throughput of more conventional microsatellite approaches for disease gene identification. Furthermore, this technology has already been expanded to enable the assaying of ~1.5 million SNPs using Perlegen Sciences wafer-based technology. SNP genotyping is a rapidly advancing technology that will lead to the identification of pathogenic mutations and biomarkers with ever increasing speed. The challenge will then be to develop appropriate functional validation models for the neurological disorders to establish the role of the genes in the disease process.

One analysis that has revolutionized the identification of biomarkers is positional cloning, which is a strategy for identifying disease-causing genes 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 [54]. 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, SNP arrays and sequencing arrays, have recently begun to have a role as adjuncts to physical mapping and identification of gene targets associated with disease [55]. Once disease-associated genes have been identified, functional validation approaches are crucial to confirm the role of the candidate genes in a disease process.

Identification of pathogenic signaling pathways and cascades

Expression profiling can be an effective approach for developing disease diagnostics [56–58], predicting the prognosis of specific forms of disease [56,59], and for the identification of dysregulated genes and signaling pathways that contribute to cellular dysfunction and disease [60]. Thus far, whole-genome expression analysis has been applied primarily to numerous forms of cancer, where the tissue of interest is clonally derived and relatively homogeneous with respect to the cellular phenotype [56–58,61–63]. Neurological disorders provide a striking contrast. For neurological diseases, the elucidation of pathogenic signaling pathways that contribute to the disease phenotype often requires a specific analysis of neuronal or glial cell subpopulations. The development of laser capture microdissection (LCM) technology has revolutionized our ability to perform expression analyses on populations of single cell types. LCM is now being applied extensively to the study of multiple neurological disorders, including AD, PD and schizophrenia [64]. In the case of AD, where cortical neurofibrillary tangles are thought to have a significant role in the disease pathology, LCM has been used to isolate neurofibrillary tangle (NFT) containing neurons as well as adjacent NFT-free neurons from the same Alzheimer's affected individuals, eliminating the significant expression variability between different individuals in the population. Recent advances in RNA amplification procedures enable these two homogeneous cell populations to be expression profiled and compared to identify dysregulated genes that contribute to the development of this dementia-inducing pathology. This approach can be extended to a variety of neurological diseases that show known histopathology (e.g. PD and LBD), as well as neuropsychiatric diseases with no known pathology, based on biomarkers defined by neuroimaging studies. For expression profiling studies, reliable and reproducible expression correlates of disease can be identified with as few as 1000 cells. Current sensitivity limits of proteomic technologies and the lack of any protein amplification techniques necessitates the use of 10^5 to 10^6 cells for protein analyses, a number that requires a much more extensive investment of labor and time, significantly decreasing throughput until the sensitivity of proteomic technologies increases.

Expression profiling strategies identify genes and/or signaling pathways for which there is a correlation between dysregulation and disease, therefore, they will result in the identification of strong candidate targets for therapeutic intervention. Current treatments for most neurological disorders are either ineffective or minimally effective (acetylcholinesterase inhibitors in AD) or produce extremely deleterious side-effects (typical or atypical antipsychotic medications in schizophrenia). Thus, there is a generalized clear requirement to elucidate disease-causing pathways and to develop novel compounds for more effective treatments based on new molecular targets.

Expression profiling techniques have also been used to develop potential diagnostics in neurological diseases through the identification of expression signatures that correlate with the disease phenotype. For multiple sclerosis, expression analysis of peripheral mononuclear blood cells clearly distinguished affected individuals from unaffected individuals [65]. A similar, yet significantly larger scale approach is currently underway to identify a reliable early diagnostic for children with the neurobehavioral disorder autism.

The development of suitable functional validation models to study candidate genes emerging from expression profiling studies will be a crucially important area in the future. For neurodegenerative diseases, such as AD or PD, strong models for validation based on the known pathologies provide immediate outlets for confirming the importance of candidate genes. However, developing neurobehavioral models, where no known pathologies exist, will be more problematic.

Proteomics

Proteomics refers to the systematic study of every protein and protein modification produced by the cell. Proteomic technologies, including MS platforms, liquid chromatography (LC), and 2D gel electrophoresis (2DE), can be used on the affected tissues for diagnostic purposes and to identify etiologic contributors to disease, or can be used on biofluids such as serum, saliva and cerebrospinal fluid (CSF) to identify protein profiles that are diagnostic or prognostic indicators of disease course. The development of proteomic technologies is still in its infancy and their effectiveness is limited because of the complexity of sequences present and the vast number of possible post-translational modifications. On average, each mammalian transcript encodes ten different protein isoforms, which might be differentially regulated at the protein rather than mRNA level, greatly increasing the complexity of the genome and the potential information available. However, in the near future MS approaches coupled with high-throughput methods for protein separation, such as LC, will be applied with increasing effectiveness in neurological disorders and should surpass the lower throughput gel-based methods that require 2DE for protein separation.

Thus far, cancer has served as the model disease for the application of proteomics to human disease. Proteomic technologies have resulted in the identification of novel protein biomarkers that will be useful for early diagnosis, prediction of outcome and monitoring of disease course for multiple forms of cancer [66,67]. More recently, proteomics technologies are being applied to the study of neurodegenerative and neurobehavioral diseases [68]. For example, analysis of CSF from Alzheimer's patients and unaffected controls has demonstrated elevated levels of over half a dozen proteins in AD affected individuals, including apolipoprotein E [69].

The more recently developed technique of isotope-coded affinity tag (ICAT) is being increasingly applied to neuroproteomics [70]. This technique involves tagging a protein of interest for the purposes of affinity purification, thereby providing a gel-free system for reducing the complexity of the protein mixture present in cell lysates and for isolating proteins associated with the tagged protein of interest: the isotope labeling enables quantification by MS. ICAT–MS has been used to identify p53 dependent changes in protein expression that occur during neuronal cell death, which have important implications for a host of neurodegenerative diseases [71]. Interestingly, in this study, many changes in protein expression were not reflected at the mRNA level as assessed by expression array analyses, clearly illustrating the benefits of a combinatorial approach to biomarker identification, using all available '-omics' technologies.

LC–MS, a promising high-throughput, gel-free proteomics technique, has been applied to neuroproteomic studies of animal models of neurological disease. These studies have had some success in characterizing transmembrane proteins from mouse fore- and hind-brain [72] and in analyzing differences in neuropeptides from pituitary extracts from wild-type and mutant mice [73]. This approach provides the noted benefit of resolving membrane proteins, which are difficult to isolate using more traditional 2DE approaches and which are likely to be important in neurological diseases.

Finally, the ultimate goal of high-throughput, whole proteome analysis is drawing closer with the development and maturation of 2D liquid-phase separation of proteins involving various methods of HPLC. Electrospray ionization (ESI)-time-of-flight (TOF) MS can then be used to quantitate the proteins, which are identified by mass fingerprinting using matrix-assisted laser desorption/ionization (MALDI)-TOF MS and MALDI-quadrupole TOF (QTOF)-tandem MS (MS–MS) [74]. Further advancements in the sensitivity of these techniques will facilitate the identification of cell-type specific biomarkers obtained from the analysis of microdissected homogenous cell types, as is the trend for genomics approaches to biomarker identification.

The proteomics approach for diagnosis is advantageous as saliva, serum, urine or CSF are more easily available than diseased tissue and can be screened rapidly to identify a protein profile that could be predictive of currently developing or ongoing disease processes. Because this approach uses easily procured peripheral tissues, it is readily amenable to use as a diagnostic adjunct to the clinical assessment of disease, leading to earlier and more reliable diagnoses. More targeted, disease-specific therapies based on predicted disease course and outcome can then be initiated.

Metabolomics

Metabolomics refers to the analysis of the entire complement of small-molecule metabolites produced by a biological system. Metabolomics has recently begun to have a more prominent role in efforts at biomarker identification and will increasingly be applied to neurological disorders. Metabolomics is typically performed on biofluids, such as serum, urine, saliva and CSF, and can be useful for physiological evaluation, drug safety assessment, diagnosis of human disease, drug therapy monitoring and characterization of genetically modified animal models of disease [75]. Given the novelty of this technology, applications to neurological diseases are just beginning.

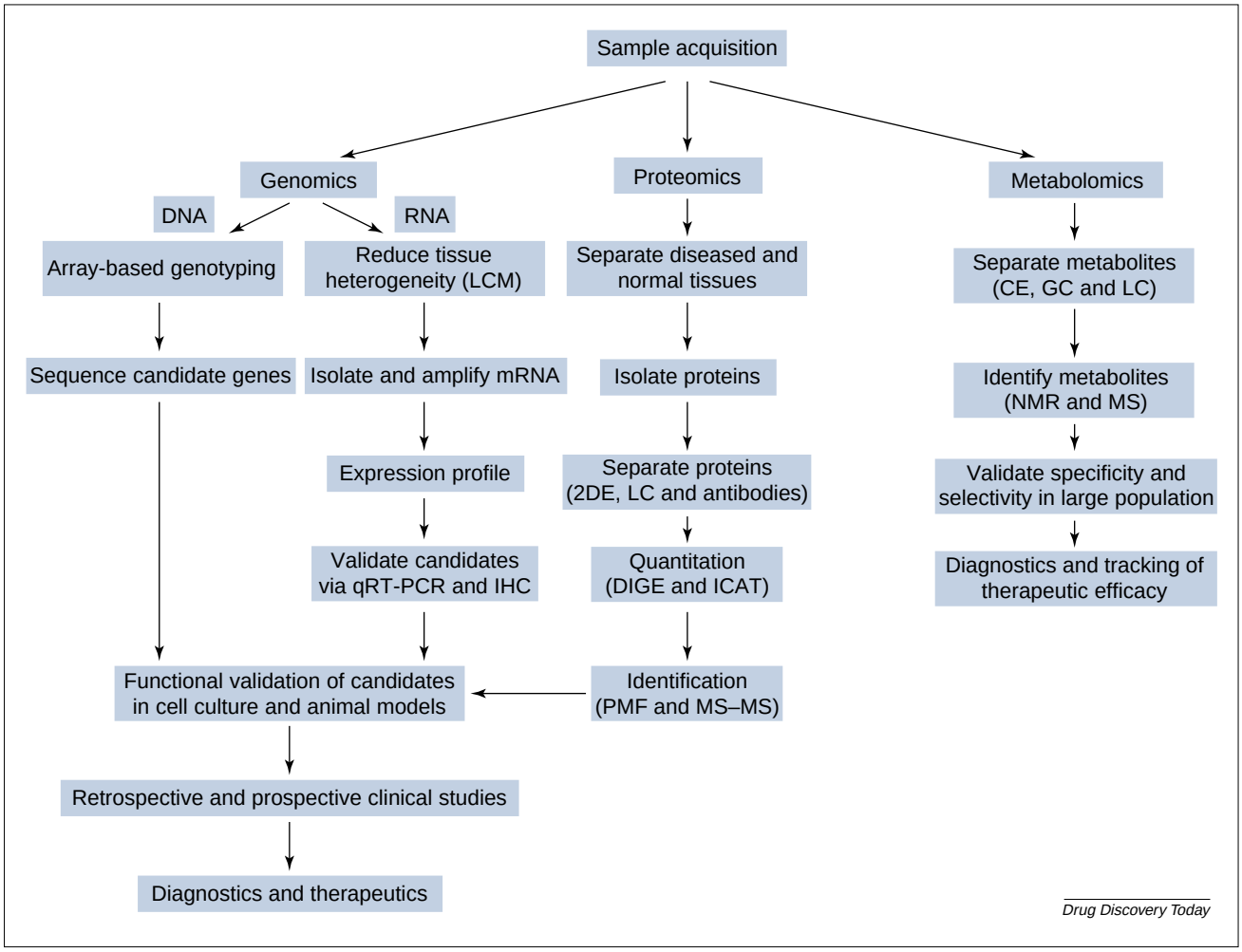

FIGURE 1

Work flow for systematic biomarker discovery in neurological diseases. An integrative systems biology approach to biomarker discovery is outlined wherein all aspects of -omics technologies are used to arrive at selective and specific biomarkers of neurological diseases. Abbreviations: CE, capillary electrophoresis; IHC, immunohistochemistry; PMF, peptide mass fingerprinting; qRT-PCR, quantitative real-time reverse transcription PCR.

High-resolution ^{1}H-NMR spectroscopy has been used to characterize a transgenic rodent model of spinal cerebellar ataxia 3 (SCA3), the most common form of dominant inherited ataxia [76]. Increased levels of glutamine and decreased levels of γ-aminobutyric acid, choline, phosphocholine and lactate were observed in the cerebellum and cerebrum, an area not previously implicated in SCA3.

Metabolomics has also been used successfully to develop potential diagnostic methods in neurological disease. For example, elevated levels of the metabolites hypoxanthine, xanthine, urate and guanine have been identified in patients with Lesch-Nyhan syndrome (LNS) using a combination of gas chromatography and MS [77]. The authors further demonstrate that patients with LNS currently undergoing treatment can still be identified using this approach. Thus, metabolomics strategies hold significant potential for disease diagnosis. This study also highlights the point that the identity of the metabolites in a disease process can lead to hypotheses about the underlying disease mechanisms, which can later be tested in either *in vitro* or animal models of disease.

Summary

Biomarker identification in neurological disorders has been hindered by the unique cellular and phenotypic complexity of the brain. However, a combination of approaches to biomarker identification using modern genomics, proteomics and metabolomic technologies promises to enable significant advances in overcoming these obstacles; a generalized workflow for conducting a systematic approach to biomarker identification in neurological disorders is presented in Figure 1. Current advances in high-throughput methods to identify etiologic mutations, in single cell expression profiling, in high-throughput MS-based proteomics and in metabolomic approaches combined with improved clinical diagnosis of affected individuals will accelerate biomarker identification. These advances will result in earlier and more specific diagnoses, identification and validation of therapeutic targets, monitoring of treatment effects, and tailoring of treatment to specific individuals based on the prediction of disease course and outcome. With this strategy, biomarkers are identified at all levels of biology: DNA, RNA, protein and small molecules. A

significant challenge lies in compiling, warehousing and synergizing the vast amounts of data that result from this type of combinatorial, systems biology approach to biomarker discovery. To this end, bioinformatics and biostatistics will have a significant role in biomarker discovery. Current statistical methods for identifying disease relevant correlates, including principal components analysis, partial least squares, permutational statistics and numerous clustering algorithms will be essential [78–81]. Through the successful integration of these datasets to make meaningful advances in biomarker discovery, neurological disorders will not be the devastating diseases that they are currently.

References

1 Gershon, E.S. and Goldin, L.R. (1986) Clinical methods in psychiatric genetics. I. Robustness of genetic marker investigative strategies. *Acta Psychiatr. Scand.* 74, 113–118

2 Leboyer, M. *et al.* (1998) Psychiatric genetics: search for phenotypes. *Trends Neurosci.* 21, 102–105

3 Hulette, C.M. *et al.* (1992) Evaluation of cerebral biopsies for the diagnosis of dementia. *Arch. Neurol.* 49, 28–31

4 Perry, E.K. *et al.* (1982) The influence of agonal status on some neurochemical activities of postmortem human brain tissue. *Neurosci. Lett.* 29, 303–307

5 Ferrer, I. and Blanco, R. (2000) N-myc and c-myc expression in Alzheimer disease, Huntington disease and Parkinson disease. *Brain Res. Mol. Brain Res.* 77, 270–276

6 Hulette, C.M. *et al.* (1997) Rapid brain autopsy. The Joseph and Kathleen Bryan Alzheimer's Disease Research Center experience. *Arch. Pathol. Lab. Med.* 121, 615–618

7 Haroutunian, V. and Davis, K.L. (2002) Issues and perspectives on brain tissue banking. *Curr. Psychiatry Rep.* 4, 233–234

8 Riaz, S. and Nowack, W.J. (1998) Diagnostic problems in multiple sclerosis: overreliance on neuroimaging. *South. Med. J.* 91, 270–272

9 Levin, N. *et al.* (2003) Patterns of misdiagnosis of multiple sclerosis. *Isr. Med. Assoc. J.* 5, 489–490

10 Poewe, W. and Wenning, G. (2002) The differential diagnosis of Parkinson's disease. *Eur. J. Neurol.* 9 (Suppl. 3), 23–30

11 Mangino, M. and Middlemiss, C. (1997) Alzheimer's disease: preventing and recognizing a misdiagnosis. *Nurse Pract.* 22, 58–59, 63–68, 70 passim

12 McKhann, G. *et al.* (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34, 939–944

13 American Psychiatric Association (1994) American Psychiatric Association. Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders: DSM-IV* (American Psychiatric Association, Washington, DC).

14 Chui, H. and Young Lee, A. (2001) *Clinical criteria for various dementia subtypes: Alzheimer's disease (AD), vascular dementia (VaD), dementia with Lewy bodies (DLB), and frontotemporal dementia (FTD).* (Qizilbash, N., *et al.* eds), Blackwell Science

15 Petrovitch, H. *et al.* (2001) Accuracy of clinical criteria for AD in the Honolulu-Asia Aging Study, a population-based study. *Neurology* 57, 226–234

16 Gearing, M. *et al.* (1995) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part X. Neuropathology confirmation of the clinical diagnosis of Alzheimer's disease. *Neurology* 45, 461–466

17 Barker, W.W. *et al.* (2002) Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Dis. Assoc. Disord.* 16, 203–212

18 Mastaglia, F.L. *et al.* (2003) Prevalence of amyloid-beta deposition in the cerebral cortex in Parkinson's disease. *Mov. Disord.* 18, 81–86

19 Hutton, M. *et al.* (1998) Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* 393, 702–705

20 Iseki, E. *et al.* (2003) Dementia with Lewy bodies from the perspective of tauopathy. *Acta Neuropathol. (Berl.)* 105, 265–270

21 Sanchez, M.P. *et al.* (2002) Progressive supranuclear palsy and tau hyperphosphorylation in a patient with a C212Y parkin mutation. *J. Alzheimers Dis.* 4, 399–404

22 Braak, H. and Braak, E. (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol. (Berl.)* 82, 239–259

23 Weiner, A. (2003) Soaking up RNAi. *Mol. Cell* 12, 535–536

24 Luo, W.J. *et al.* (2003) PEN-2 and APH-1 coordinately regulate proteolytic processing of presenilin 1. *J. Biol. Chem.* 278, 7850–7854

25 Hassani, Z. *et al.* Lipid-mediated siRNA delivery down-regulates exogenous gene expression in the mouse brain at picomolar levels. *J. Gene Med.* (in press)

26 Lue, L.F. and Walker, D.G. (2002) Modeling Alzheimer's disease immune therapy mechanisms: interactions of human postmortem microglia with antibody-opsonized amyloid beta peptide. *J. Neurosci. Res.* 70, 599–610

27 Takahashi, K. *et al.* (2002) Targeted disruption of H3 receptors results in changes in brain histamine tone leading to an obese phenotype. *J. Clin. Invest.* 110, 1791–1799

28 Williams, T. *et al.* (2003) Diet-induced obesity and cardiovascular regulation in C57BL/6J mice. *Clin. Exp. Pharmacol. Physiol.* 30, 769–778

29 Davies, E.L. *et al.* (2001) Functional enhancement of electrofusion-derived BRIN-BD11 insulin-secreting cells after implantation into diabetic mice. *Int. J. Exp. Diabetes Res.* 2, 29–36

30 Lesniewski, L.A. *et al.* (2003) Mechanisms of force loss in diabetic mouse skeletal muscle. *Muscle Nerve* 28, 493–500

31 Taki, S. (2002) Type I interferons and autoimmunity: lessons from the clinic and from IRF-2-deficient mice. *Cytokine Growth Factor Rev.* 13, 379–391

32 Ishimaru, N. *et al.* (2003) Development of autoimmune exocrinopathy resembling Sjogren's syndrome in estrogen-deficient mice of healthy background. *Am. J. Pathol.* 163, 1481–1490

33 Jackson, E.L. *et al.* (2001) Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* 15, 3243–3248

34 Li, X. *et al.* (2002) Optically imageable metastatic model of human breast cancer. *Clin. Exp. Metastasis* 19, 347–350

35 Cox, C. *et al.* (2003) Inhibition of the growth of squamous cell carcinoma by tetrathiomolybdate-induced copper suppression in a murine model. *Arch. Otolaryngol. Head Neck Surg.* 129, 781–785

36 Garcia-Castro, J. *et al.* (2003) Purging of leukemia-contaminated bone marrow grafts using suicide adenoviral vectors: an *in vivo* murine experimental model. *Gene Ther.* 10, 1328–1335

37 Sfacteria, A. *et al.* (2003) Cyclin D1 expression in pre-cancerous and cancerous lesions of the canine mammary gland. *J. Comp. Pathol.* 128, 245–251

38 Link, C.D. *et al.* (2003) Gene expression analysis in a transgenic *Caenorhabditis elegans* Alzheimer's disease model. *Neurobiol. Aging* 24, 397–413

39 Finelli, A. *et al.* (2004) A model for studying Alzheimer's Abeta42-induced toxicity in *Drosophila melanogaster*. *Mol. Cell. Neurosci.* 26, 365–375

40 Yamaguchi, F. *et al.* (1991) Transgenic mice for the amyloid precursor protein 695 isoform have impaired spatial memory. *Neuroreport* 2, 781–784

41 Wirak, D.O. *et al.* (1991) Deposits of amyloid beta protein in the central nervous system of transgenic mice. *Science* 253, 323–325

42 Kammesheidt, A. *et al.* (1992) Deposition of beta/A4 immunoreactivity and neuronal pathology in transgenic mice expressing the carboxyl-terminal fragment of the Alzheimer amyloid precursor in the brain. *Proc. Natl. Acad. Sci. U. S. A.* 89, 10857–10861

43 Higgins, L.S. *et al.* (1994) Transgenic mouse brain histopathology resembles early Alzheimer's disease. *Ann. Neurol.* 35, 598–607

44 Moran, P.M. *et al.* (1995) Age-related learning deficits in transgenic mice expressing the 751-amino acid isoform of human beta-amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A.* 92, 5341–5345

45 Lamb, B.T. *et al.* (1993) Introduction and expression of the 400 kilobase amyloid precursor protein gene in transgenic mice. *Nat. Genet.* 5, 22–30

46 Pearson, B.E. and Choi, T.K. (1993) Expression of the human beta-amyloid precursor protein gene from a yeast artificial chromosome in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* 90, 10578–10582

47 Borchelt, D.R. *et al.* (1996) Familial Alzheimer's disease-linked presenilin 1 variants elevate Abeta1-42/1-40 ratio *in vitro* and *in vivo*. *Neuron* 17, 1005–1013

48 Citron, M. *et al.* (1997) Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. *Nat. Med.* 3, 67–72

49 Chui, D.H. *et al.* (1999) Transgenic mice with Alzheimer presenilin 1 mutations show accelerated neurodegeneration without amyloid plaque formation. *Nat. Med.* 5, 560–564

50 Borchelt, D.R. *et al.* (1997) Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins. *Neuron* 19, 939–945

51 Holcomb, L. *et al.* (1998) Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes. *Nat. Med.* 4, 97–100

52 Strazielle, C. *et al.* (2003) Regional brain cytochrome oxidase activity in beta-amyloid precursor protein transgenic mice with the Swedish mutation. *Neuroscience* 118, 1151–1163

53 Oddo, S. *et al.* (2003) Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction. *Neuron* 39, 409–421

54 Bertelson, C.J. *et al.* (1986) Localisation of Xp21 meiotic exchange points in Duchenne muscular dystrophy families. *J. Med. Genet.* 23, 531–537

55 Dobrin, S.E. and Stephan, D.A. (2003) Integrating microarrays into disease-gene identification strategies. *Expert Rev. Mol. Diagn.* 3, 375–385

56 Dybkaer, K. *et al.* (2004) Molecular diagnosis and outcome prediction in diffuse large B-cell lymphoma and other subtypes of lymphoma. *Clin. Lymphoma* 5, 19–28

57 Nambiar, S. *et al.* (2004) Applications of array technology: melanoma research and diagnosis. *Expert Rev. Mol. Diagn.* 4, 549–557

58 Petty, R.D. *et al.* (2004) Gene expression profiling in non-small cell lung cancer: from molecular mechanisms to clinical application. *Clin. Cancer Res.* 10, 3237–3248

59 Yamada, S.I. *et al.* (2004) Expression profiling and differential screening between hepatoblastomas and the corresponding normal livers: identification of high expression of the PLK1 oncogene as a poor-prognostic indicator of hepatoblastomas. *Oncogene* 23, 5901–5911

60 Hoek, K. *et al.* (2004) Expression profiling reveals novel pathways in the transformation of Melanocytes to Melanomas. *Cancer Res.* 64, 5270–5282

61 Li, S.R. *et al.* (2004) Differential expression patterns of the insulin-like growth factor 2 gene in human colorectal cancer. *Tumour Biol.* 25, 62–68

62 Woo, I.S. *et al.* (2004) Expression of placental growth factor gene in lung cancer. *Tumour Biol.* 25, 1–6

63 Mazzanti, C. *et al.* (2004) Using gene expression profiling to differentiate benign versus malignant thyroid tumors. *Cancer Res.* 64, 2898–2903

64 Ginsberg, S.D. *et al.* (2004) Single-cell gene expression analysis: implications for neurodegenerative and neuropsychiatric disorders. *Neurochem. Res.* 29, 1053–1064

65 Bomprezzi, R. *et al.* (2003) Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. *Hum. Mol. Genet.* 12, 2191–2199

66 Rai, A.J. and Chan, D.W. (2004) Cancer Proteomics: Serum Diagnostics for Tumor Marker Discovery. *Ann. N. Y. Acad. Sci.* 1022, 286–294

67 Rodland, K.D. (2004) Proteomics and cancer diagnosis: the potential of mass spectrometry. *Clin. Biochem.* 37, 579–583

68 Davidsson, P. *et al.* (2003) Clinical mass spectrometry in neuroscience. Proteomics and peptidomics. *Cell. Mol. Biol. (Noisy-le-grand)* 49, 681–688

69 Puchades, M. *et al.* (2003) Proteomic studies of potential cerebrospinal fluid protein markers for Alzheimer's disease. *Brain Res. Mol. Brain Res.* 118, 140–146

70 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999

71 Johnson, M.D. *et al.* (2004) Proteome analysis of DNA damage-induced neuronal death using high throughput mass spectrometry. *J. Biol. Chem.* 279, 26685–26697

72 Olsen, J.V. *et al.* (2003) HysTag – a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol. Cell. Proteomics* 3, 82–92

73 Che, F.Y. and Fricker, L.D. (2002) Quantitation of neuropeptides in Cpe(fat)/Cpe(fat) mice using differential isotopic tags and mass spectrometry. *Anal. Chem.* 74, 3190–3198

74 Hamler, R.L. *et al.* (2004) A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification. *Proteomics* 4, 562–577

75 Lindon, J.C. *et al.* (2004) Metabolomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* 9, 1–31

76 Griffin, J.L. *et al.* (2004) Defining a metabolic phenotype in the brain of a transgenic mouse model of spinocerebellar ataxia 3. *Physiol. Genomics* 16, 334–340

77 Ohdoi, C. *et al.* (2003) Chemical diagnosis of Lesch-Nyhan syndrome using gas chromatography-mass spectrometry detection. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 792, 123–130

78 Rosenfeld, S. (2004) New developments in cancer-related computational statistics. *Ann. N. Y. Acad. Sci.* 1020, 22–31

79 Looney, S.W. (2002) Statistical methods for assessing biomarkers. *Methods Mol. Biol.* 184, 81–109

80 Gieser, P. *et al.* (2002) Introduction to microarray experimentation and analysis. *Methods Mol. Biol.* 184, 29–49

81 Lazaridis, E.N. and Bloom, G.C. (2002) Statistical contributions to molecular biology. *Methods Mol. Biol.* 184, 1–14

82 Kim, S.I. *et al.* (2004) Neuroproteomics: expression profiling of the brain's proteomes in health and disease. *Neurochem. Res.* 29, 1317–1331

83 Freeman, W.M. and Hemby, S.E. (2004) Proteomics for protein expression profiling in neuroscience. *Neurochem. Res.* 29, 1065–1081