The Genomic Biomarkers for Neurological Diseases

Essay
Submitted for partial fulfillment of master degree in Neuropsychiatry

By
Ahmad Ibraheem Abdel-Bary
M.B., B.Ch.

Under supervision of

Prof. Dr. Mohammad Osama Abdulghani
Professor of Neuropsychiatry
Faculty of Medicine – Ain Shams University

Prof. Dr. Nevine Medhat El-Nahas
Professor of Neuropsychiatry
Faculty of Medicine – Ain Shams University

Dr. Salma Hamed Khalil
Assistant Professor of Neuropsychiatry
Faculty of Medicine – Ain Shams University

Ain Shams University
2010

Contents

Tables & Figures

Abbreviations

AAA	*a*denosine triphosphatase [ATPase] *a*ssociated with various cellular *a*ctivities
AD	autosomal dominant / Alzheimer's disease
ADNFLE	autosomal dominant nocturnal frontal lobe epilepsy
AIVED	ataxia associated with isolated vitamin E deficiency
ALS	amyotrophic lateral sclerosis
AMN	Adrenomyeloneuropathy
APP	amyloid precursor protein
AR	autosomal recessive
AT	ataxia telangiectasia
aCGH	array comparative genomic hybrid
BHC	benign hereditary chorea
BMD	Becker's muscular dystrophy
CADASIL	cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
CAG	cytosine-adenine-guanine
CCA	cerebral amyloid angiopathy
CJD	Creutzfeldt-Jakob disease
CMAP	compound muscle action potential
CMD	congenital muscular dystrophies
CTX	cerebrotendinous xanthomatosis
CNS	central nervous system
DLBD	diffuse Lewy body dementia
DMD	Duchenne's muscular dystrophy
DNA	deoxyribonucleic acid
DRPLA	dentatorubral-pallidoluysian atrophy
DSD	Dejerine-Sottas disease
EIEE	early infantile epileptic encephalopathy
EME	early myoclonic encephalopathy
EM	essential myoclonus
ET	essential tremors
FA	Friedreich's ataxia
FALS	familial amyotrophic lateral sclerosis
FAP	familial amyloid polyneuropathy
FD	familial dysautonomia
FFI	fatal familial insomnia
FISH	flurescent in situ hybridization
FSHD	facioscapulohumeral dystrophy
FTD	frontotemporal dementia
GAN	giant axonal neuropathy
HD	Huntington's disease
Hex-A	hexosaminidase-A
HHT	hereditary hemorrhagic telangiectasia
HI	hypomelanosis of Ito

HNPP	hereditary neuropathy with liability to pressure palsies
HSAN	hereditary sensory and autonomic neuropathy
HSP	hereditary spastic paraplegia
ICH	intracerebral hemorrhage
KSS	Kearns-Sayre syndrome
LDH	lactate dehydrogenase
LMN	lower motor neuron
LOAD	late onset Alzheimer's disease
MD	myotonic dystrophy
MERRF	myoclonic epilepsy with ragged-red fibers myopathy
MELAS	mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes
MJD	Machado-Joseph disease
MLD	metachromatic leukodystrophy
MND	motor neuron disease
mRNA	messenger ribonucleic acid
NAIP	neuronal apoptosis inhibitory protein
NARP	Neuropathy, ataxia, retinitis pigmentosa syndrome
NCV	nerve conductive velocity
NF	Neurofibromatosis
NKH	nonketotic hyperglycinemia
PCR	Polymerase chain reaction
PADC	parkinsonism-ALS-dementia complex
PD	Parkinson's disease
PDC	Parkinsonism-dementia complex
PEO	Progressive external ophthalmoplegia
PLP1	prteolipid protein 1
PMD	Pelizaeus-Merzbacher disease
RLS	restless leg syndrome
RRF	ragged red fibers
SALS	sporadic amyotrophic lateral sclerosis
SCA	spinocerebellar ataxia
SMA	spinal muscular atrophy
SMN	survival motor neuron
SNAP	sensory nerve action potential
SNP	single nucleotide polymorphism
TSC	tuberous sclerosis complex
TTR	Transthyretin
UMN	upper motor neuron
VHL	von Hippel-Lindau syndrome
VLCFAs	very-long-chain fatty acids
XL	X-chromosome linked
XP	xeroderma pigmentosum
XR	X-chromosome linked recessive

Preface

All the thanks are to Allah, the lord of all creatures, for his innumerable and uncountable blessings and gifts, which we will never ever be able to thank him for. And I ask him to make his peace and grace to be perpetually upon the master of and the best ever of all the creatures, our prophet 'Muhammad' and upon his family, companions, and followers up to the last day.

This modest essay is a necessary step in achieving the master degree in neuropsychiatry. It is a considerable hurdle in the way of most of us, i.e. the master degree students, where we are in an internal conflict between completing this work and preparing for the final examination --the time is crucial. No doubt all of our works were to be more better if we were given an extra time. We all know that striving for perfection is an instinct of the human being, nevertheless, defects do always exist. Perfection is only and only for Allah. Once upon a time, I read a dictum belonging to a Chinese wise man, he said " if I published my book only when I feel satisfied with, it will never see the light as long as I am alive". So, we will never be satisfied with our works, whatever we did, and whatever it was laborious.

But --to be honest-- it is not mere my work. There are others who were my guides and mentors in this journey. They were the light which illuminated to me throughout the task of fulfilling this mission. I owe them much more. I have impressed with their kindness, decency and gentleness. They are Prof. Dr. M. Osama Abdulghani, Prof. Dr. Nevine Medhat El-Nahas, and Ass. Prof. Dr. Salma Hamed Khalil. Really, they were not only decent guides, but, at first, an intimate friends. As the words fail to express about my deep gratitude, all what I can do is to bow to them...

Introduction

In a trial for estimating the prevalence of genetic disorders in the spectrum of neurological diseases, it has been found that as much as 10% of neurological patients have single gene mutation disorder **(Manji, 2006)**. Additionally, it is thought that the majority of complex neurological diseases have a genetic component that has historically been difficult to uncover using traditional genetic methods **(Robeson, 2008)**. For example, about 20% of the risk of ischemic stroke remains unexplained by conventional risk factors and genetic predisposition has been widely speculated to account for some of this unexplained risk **(Markus, 2004)**.

In many neurological diseases, such as Alzheimer's disease and Parkinson's disease, definitive diagnosis occurs only at autopsy with sometimes as many as 20% of cases being found to have been clinically misdiagnosed. That is due to ante-mortem lack of availability of tissue at the site of pathology for most neurological diseases. Moreover, there is no currently available laboratory test for early detection and/or accurate diagnosis of such a disease. Furthermore, the familial genetic approaches have been successful in identifying the genetic causes of only some very prominent familial forms of the disease which account for small fraction of the individuals afflicted with the disease, but not the sporadic forms which are much more prevalent **(Robeson, 2008)**.

The study of genomes, referred to as 'genomics,' has made substantial progress toward understanding of human disease and promises to direct the future of medicine toward more personalized diagnostics and therapeutics based on an individual's specific genetic variations and predispositions to disease. A significant challenge lay both in developing biomarkers for and understanding the cause of the much more prevalent sporadic forms of complex neurological diseases **(Scherzer, 2007)**.

There are numerous examples of the use of genomic technologies to identify novel biomarkers for multiple neurological diseases. For examples, Single nucleotide polymorphisms (SNP)-based screening is useful when trying to identify biomarkers that can be used in presymptomatic risk assessment. In contrast, Gene expression profiling are being useful when some component of the disease has already manifested itself and can be used to identify the underlying molecular changes which may serve as targets for therapeutic development **(Sharp, 2006)**.

In a study aimed at linking copy-number alterations to complex neurological diseases using the array comparative genomic hybridization (aCGH), it was found that there are a reciprocal duplications and deletions on chromosome 16p13.1 that are associated with autism and/or mental retardation. With further validation of this genetic association, aCGH may provide a more accurate diagnostic tool for a subset of autistic individuals. This is particularly important in autism where the difficulty in making a definitive clinical diagnosis often delays administration of appropriate behavioral therapy, which is substantially more effective when initiated early in life **(Ulmann, 2007)**.

In a recent study which looked for potential diagnostic biomarker for Parkinson's

disease (PD) in blood using Gene expression profiling, it was strikingly found that dysregulated genes identified in peripheral blood samples reflect many of the known pathogenic mechanisms occurring in the brain of patients with PD. This suggests that in some cases insights into biological pathogenesis of neurological diseases may be gained from identifying aberrations in blood samples, despite the presence of intact blood brain barrier. Thus, the obstacles such as lack of availability of the affected tissue, and lack of diagnostic laboratory test can be overcome **(Scherzer, 2007)**.

In another study which examined gene expression profiles in a group of patients suffering from Huntington's disease (HD), it identified subset of mRNAs that successfully distinguished controls from presymptomatic individuals with the HD mutation and from symptomatic HD patients. This allowed discrimination of the different stages of the disease, suggesting that this marker set is able to monitor disease progression. In addition, numerous genes from this biomarker set were also dysregulated in HD post-mortem frozen brain samples **(Borovecki, 2005)**.

Aim of the Work

- ✓ To review the current genomic technologies used for novel biomarkers discovery.
- ✓ To review the genomic biomarkers for the different neurological diseases, as a diagnostic tools and therapeutic targets.

Chapter I

THE GENOMIC BIOMARKERS AND TECHNOLOGIES

The genomic biomarkers and technologies

Biomarker (biological marker) is a general term coined to any character that occurs in association with a pathological process and has diagnostic and/or prognostic utility. So, biomarker could be a physical sign, laboratory measurement, radiological finding, or electrical signal. The different biomarkers in association might be helpful to improve diagnosis, get insight in pathophysiology, and facilitate treatment choices. Genomic biomarkers are defined as genetic variations (mutations and polymorphisms) that can help in diagnosis, predict disease susceptibility, disease outcome, or treatment response and toxicity **(Robeson, 2008).**

The ultimate goals for research focused on complex neurological diseases are to either prevent or to cure the diseases. These are ambitious goals which will be greatly facilitated by the identification of new biomarkers that can serve as novel diagnostic or prognostic indicators of disease course, and can be used as surrogate disease markers to track the efficacy of novel treatment strategies, or that may provide new targets for the treatment of the diseases **(Dunkley, 2005).**

Neurological disorders present multiple unique challenges to biomarker discovery. Briefly, the essential and vital role that the brain and central nervous system play in all aspects of life ensures the lack of availability of tissue at the site of pathology for most neurological disorders. Additionally, the ante-mortem clinical diagnostics, although continually improving in most neurological diseases, remains problematic. In many diseases, such as Alzheimer's disease and Parkinson's disease, definitive diagnosis occurs only at autopsy with sometimes as many as 20% of cases being found to have been clinically misdiagnosed. Reliable genetic based biomarkers specific for these diseases would clearly be useful for earlier and more accurate diagnoses and treatment interventions. Lastly, there are few animal models fully representative of the diseases that can be used for validation of candidate biomarkers. This is likely due in part to the increased complexity of the human brain, of human behavior, and the possibility that neurological diseases, many of which develop later in life and require an aging component, are difficult to recapitulate in shorter lived animals **(Robeson, 2008)**.

High-throughput technologies for studies of the human genome are advancing at a rapid pace and are increasingly being applied toward the study of neurological diseases to help overcome some of these hurdles. There are numerous examples of the use of genomic technologies to identify novel biomarkers for multiple neurological diseases, including Alzheimer's disease, Parkinson's disease, autism and Huntington's disease among others **(Borovecki, 2005)**.

The genetic underpinnings of Mendelian forms of disease have been pieced together using familial genetic approaches, such as linkage studies, where whole families are studied to find commonly inherited mutations associated with disease. This approach has been successful in identifying the genetic causes of some very prominent familial forms of

neurodegenerative diseases, such as Alzheimer's disease (AD) and Parkinson's disease (PD). However, even with these common disorders, familial cases account for a small fraction of individuals afflicted with the disease. A significant challenge lay both in developing biomarkers for and understanding the causes of the much more prevalent sporadic forms of complex neurological diseases, such as late onset AD (LOAD). The main genomic technologies currently being applied to this problem include array CGH (array comparative genomic hybridization), gene expression profiling, whole genome association studies using single nucleotide polymorphisms (SNPs), and, more toward the future, whole genome resequencing to identify causal disease-associated DNA sequence variants **(Scherzer, 2007)**.

Array Comparative Genomic Hybridization

On a functional level, array comparative genomic hybridization (aCGH) measures copy-number variations at multiple locations in the genome simultaneously by comparing DNA content from two differentially labeled genomes: one being the patient (or test) genome and the second being the control (or reference) genome. Using this technology one can show changes such as deletions, duplications, or amplifications at any locus that is represented on the array. Unlike FISH (fluorescence in situ hybridization), which is limited in the number of loci that can be simultaneously interrogated, aCGH is able to detect DNA copy number changes at hundreds of thousands of loci in a genome in a single experiment and thus provides a far more rapid and facile method for detecting genomic copy number variations associated with disease **(Bejjani, 2006)**.

Array CGH has been suggested as a potential diagnostic tool to personalize treatment strategies. Recently, aCGH has been used to detect copy number alterations in the PLP1 region for prenatal diagnosis of Pelizaeus-Merzbacher disease (PMD), a rare X-linked dysmyelinating disorder of the central nervous system. Genomic duplications of the PLP1 (proteolipid protein 1) gene is one of the main causes of PMD, and has been used to molecularly diagnose the disease by interphase FISH and quantitative multiplex PCR methods using blood samples in children and adults. It is also well established in prenatal diagnosis by using either amniotic fluid or chorionic villus sampling. Array CGH has been used to successfully detect PLP1 copy number in the developing fetus and provides the noted benefit of being more efficient than one FISH experiment. The specific size of genomic duplication can also be determined using aCGH provided that the probe density is sufficient, something that is more labor intensive using FISH. One limitation of aCGH however is that it cannot detect other rearrangements, such as inversions or balanced translocations. Nevertheless, aCGH provides a rapid diagnostic for PMD and establishes the utility of this technology for assisting in the diagnosis of neurological diseases **(Lee, 2005)**.

Gene expression profiling

Gene expression microarrays are used to rapidly assess the expression of thousands of genes in a single experiment, generating specific "expression profiles" of normal and disease states. A typical experiment will compare tissue from a healthy control to tissue from an individual affected by a specific disease. Comparison of the expression profile for the healthy tissue to that from the disease tissue will identify specific gene dysregulation correlated with the disease of interest. This disease related expression signature (Figure 1) provides information about the underlying cellular dysfunction involved in disease pathogenesis, or could be used as

a diagnostic gene expression signature to help guide treatment options **(Scherzer, 2007).**

Gene expression profiling has been very useful for identifying the underlying cellular changes involved in idiopathic and multifactorial diseases, including neurodegenerative disorders such as Alzheimer's disease and Parkinson's disease. One of the advantages of gene expression profiling is its applicability to any tissue sample that contains intact mRNA, such as brain tissue or blood. Importantly, identification of disease-specific differences in gene expression in blood would provide a facile diagnostic biomarker for neurological disease **(Papapetropoulos, 2007).**

In recent study looked at potential diagnostic gene expression markers for Parkinson's disease (PD) in blood. Using a two-stage experimental design with a training and validation population, a set of dysregulated genes was significantly associated with risk of PD. The set of genes identified predict PD significantly more effectively than current prediction methods, which include the risk factors of age and sex. The genes within this set do not correlate to a single pathway, but all are known to be expressed in the human brain. Also, many of the genes seem to be previously identified or associated with PD or with a major process that could be contributing to PD. With no current laboratory test available and early detection of PD being clinically difficult, these findings may advance the development of diagnostic biomarkers for PD. Furthermore, it is striking that dysregulated genes identified in peripheral blood samples in this study reflect many of the known pathogenic mechanisms occurring in the brain of patients with PD. This suggests that in some cases insights into biological pathogenesis of neurological diseases may be gained from identifying aberrations in blood samples, despite the presence of an intact blood-brain barrier **(Scherzer, 2007)**.

Another interesting study examined gene expression profiles in a group of patients suffering from Huntington's disease (HD). In that study it has been able to identify a subset of mRNAs that successfully distinguished controls from pre-symptomatic individuals with the HD mutation and from symptomatic HD patients. Thus, using these dysregulated genes we are able to discriminate the different stages of the disease, suggesting that this marker set is able to monitor disease progression. In addition, numerous genes from this biomarker set were also dysregulated in HD post-mortem frozen brain samples. Thus, this example and the one above for PD suggest that dysregulation in blood may reflect dysregulation occurring in the brain during the course of neurodegenerative diseases. For HD, the correlation between brain and blood may mechanistically result from mutated huntingtin protein similarly affecting brain and blood targets. These provide useful examples of the study of peripheral tissues to gain insights into the pathology occurring in a less accessible, but more functionally relevant tissue, the brain **(Lee, 2007)**.

The findings from these studies and many others illustrate the potential and evolving capability of gene expression profiling to determine distinct profiles attributed to different phenotypes in a disease, to identify individual subtypes of complex diseases and others that possibly have central environmental and genetic factors contributing to the disease, and also to develop a way to monitor disease progression, early recognition and correlate activities in the brain to markers in blood **(Papapetropoulos, 2007).**

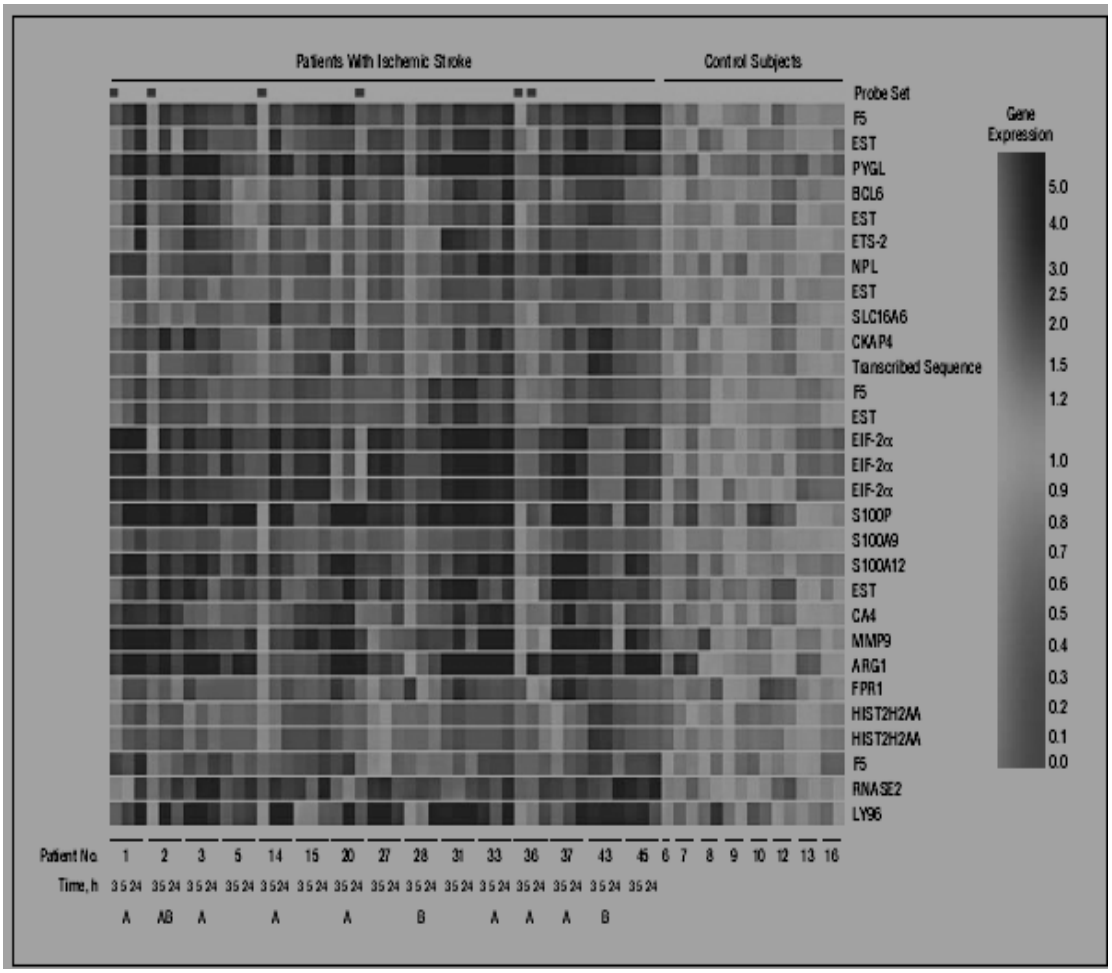

Figure 1. an example of gene expression profile microarray comparing gene profile between patients with stroke and control group (Sharp, 2006).

Single nucleotide polymorphisms

Any two humans are 99.9% identical at the DNA sequence level. However, it is the 0.1% or ~ 1 nucleotide in 1000 variations that defines our uniqueness as well as our differential susceptibility to diseases. Single Nucleotide Polymorphisms (SNPs) represent single base pair differences between individuals at a specific location in the genome. There are an estimated 10 million SNPs that differ between individuals in the population **(Carlson, 2003).**

SNPs are grouped into three functional classes. The first group contains the classically defined Mendelian inherited single base-pair mutations. These are single SNPs occurring in the coding region of genes that alter the protein sequence of the gene product or that result in premature truncation of the encoded protein. These are predicted to have strong functional effects on their own. The second class of

SNPs, referred to as functional SNPs, consist of those that have subtle effects on gene function or expression, contributing to disease only when occurring in the context of additional genetic variants or environmental influences. The third category consists of nonfunctional SNPs, those that are functionally completely silent, but may nevertheless be of interest due to genetic linkage to a nearby functional DNA sequence variant. The majority of SNPs fall into this category. Functional SNPs are of particular interest in the study of common and complex neurological disorders since they are thought to occur at high frequencies in the general population and to result in disease when occurring in specific combinations. Recent advances in high-throughput SNP genotyping technologies provide an opportunity to identify these SNPs in large case-control association studies of complex neurological diseases. SNP analysis also enables more rapid identification of Mendelian inherited diseases through linkage analyses **(Robeson, 2008)**.

SNP genotyping technology holds great potential for discovering multigenic contributions to complex neurological disorders. These disorders are usually not inherited in a Mendelian fashion and, in many cases, may be referred to as "sporadic" cases of disease, which are ideally studied using case-control whole genome association studies of outbred populations. The development of SNP scanning technology now allows the simultaneous testing of more than a million genetic variants, enabling initial studies into the genetics of complex neurological disorders **(Dunkley, 2005)**.

The utility of identifying genetic variants for complex diseases lies in their use as a predictor of disease risk, as a diagnostic for disease, as a predictor of response to therapy, or for the identification of therapeutic targets. For example, currently genetic testing is done one gene at a time using a candidate gene approach. That is, one has a family history of a particular disease for which a common genetic variant is known, such as cystic fibrosis, and can be tested for the presence of that variant within their genome. This information can then be used to guide life decisions, such as reproductive choices or exercise and eating habits in instances of other disorders. However, most human disease is sporadic and multigenic. Risk for these diseases cannot be diagnosed using traditional approaches. SNP analysis can be performed on a genome-wide scale in large case-control association studies of outbred populations to identify all of the genetic variants that contribute simultaneously to a specific disease. These variants can then be packaged into a prognostic test to predict an individual's overall genetic risk for developing a given disease. These prognostic tests would be most effective when coupled to an effective therapeutic or prevention strategy **(Carlson, 2008)**.

However, such a test to assess disease risk could have a significant impact on human health, even in the absence of a specific therapy because environmental influences, which are modifiable, also affect the development and course of disease.