Neuromethods 97

Springer Protocols

Kewal K. Jain *Editor*

Applied Neurogenomics

Humana Press

NEUROMETHODS

Series Editor
Wolfgang Walz
University of Saskatchewan
Saskatoon, SK, Canada

For further volumes:
http://www.springer.com/series/7657

Applied Neurogenomics

Edited by

Kewal K. Jain

Jain PharmaBiotech, Basel, Switzerland

Humana Press

Editor
Kewal K. Jain
Jain PharmaBiotech
Basel, Switzerland

ISSN 0893-2336 ISSN 1940-6045 (electronic)
ISBN 978-1-4939-2246-8 ISBN 978-1-4939-2247-5 (eBook)
DOI 10.1007/978-1-4939-2247-5
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014958451

Printed on acid-free paper

Humana Press is a brand of Springer
Springer is part of Springer Science+Business Media (www.springer.com)

Series Preface

Experimental life sciences have two basic foundations: concepts and tools. The *Neuromethods* series focuses on the tools and techniques unique to the investigation of the nervous system and excitable cells. It will not, however, shortchange the concept side of things as care has been taken to integrate these tools within the context of the concepts and questions under investigation. In this way, the series is unique in that it not only collects protocols but also includes theoretical background information and critiques which led to the methods and their development. Thus it gives the reader a better understanding of the origin of the techniques and their potential future development. The *Neuromethods* publishing program strikes a balance between recent and exciting developments like those concerning new animal models of disease, imaging, in vivo methods, and more established techniques, including, for example, immunocytochemistry and electrophysiological technologies. New trainees in neurosciences still need a sound footing in these older methods in order to apply a critical approach to their results.

Under the guidance of its founders, Alan Boulton and Glen Baker, the *Neuromethods* series has been a success since its first volume published through Humana Press in 1985. The series continues to flourish through many changes over the years. It is now published under the umbrella of Springer Protocols. While methods involving brain research have changed a lot since the series started, the publishing environment and technology have changed even more radically. Neuromethods has the distinct layout and style of the Springer Protocols program, designed specifically for readability and ease of reference in a laboratory setting.

The careful application of methods is potentially the most important step in the process of scientific inquiry. In the past, new methodologies led the way in developing new disciplines in the biological and medical sciences. For example, Physiology emerged out of Anatomy in the nineteenth century by harnessing new methods based on the newly discovered phenomenon of electricity. Nowadays, the relationships between disciplines and methods are more complex. Methods are now widely shared between disciplines and research areas. New developments in electronic publishing make it possible for scientists that encounter new methods to quickly find sources of information electronically. The design of individual volumes and chapters in this series takes this new access technology into account. Springer Protocols makes it possible to download single protocols separately. In addition, Springer makes its print-on-demand technology available globally. A print copy can therefore be acquired quickly and for a competitive price anywhere in the world.

Wolfgang Walz

Preface

Following the decade of the brain at the end of the twentieth century and completion of the sequencing of the human genome, the postgenomic era started in the twenty-first century with a marked increase in research on genomic technologies along with applications relevant to clinical medicine. Since the publication of the first review on applied neurogenomics in 2001, considerable new developments have taken place in this area, particularly with the rapid progress in next-generation sequencing. It would be impossible to describe all technologies in one book. This volume contains detailed methods relevant to the genomics of neurodegenerative diseases, particularly Alzheimer's disease and Huntington's disease. There is a chapter on targeted sequencing of psychiatric disorders as well. An introductory chapter reviews all technologies relevant to neurogenomics with an emphasis on next-generation sequencing. Finally, one chapter deals with the role of genomics in the development of personalized neurology.

Basel, Switzerland *Kewal K. Jain*

Contents

Contributors

Andre Altmann • *Functional Imaging in Neurodegenerative Disorders Lab, Department of Neurology & Neurological Sciences, Stanford University, Stanford, CA, USA*

Carol A. Barnes • *Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ, USA; The Arizona Alzheimer's Consortium, Phoenix, AZ, USA; The Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA; ARL Division of Neural Systems, Memory and Aging, University of Arizona, Tucson, AZ, USA; Department of Psychology, University of Arizona, Tucson, AZ, USA; Department of Neurology, University of Arizona, Tucson, AZ, USA; Department of Neuroscience, University of Arizona, Tucson, AZ, USA*

Fran Borovečki • *Department for Functional Genomics, Center for Translational and Clinical Research, School of Medicine, University Hospital Center Zagreb, University of Zagreb, Zagreb, Croatia; Department for Neurology, Center for Translational and Clinical Research, School of Medicine, University Hospital Center Zagreb, University of Zagreb, Zagreb, Croatia*

Alberto Capurro • *Department of Cell Physiology and Pharmacology, University of Leicester, Leicester, UK*

Monica K. Chawla • *Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ, USA; The Arizona Alzheimer's Consortium, Phoenix, AZ, USA; The Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA; ARL Division of Neural Systems, Memory and Aging, University of Arizona, Tucson, AZ, USA; Department of Psychology, University of Arizona, Tucson, AZ, USA; Department of Neurology, University of Arizona, Tucson, AZ, USA; Department of Neuroscience, University of Arizona, Tucson, AZ, USA*

Jason J. Corneveaux • *Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ, USA; The Arizona Alzheimer's Consortium, Phoenix, AZ, USA*

Matt De Both • *Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ, USA; The Arizona Alzheimer's Consortium, Phoenix, AZ, USA*

Kristina Gotovac • *Department for Functional Genomics, Center for Translational and Clinical Research, School of Medicine, University Hospital Center Zagreb, University of Zagreb, Zagreb, Croatia*

Mario Habek • *Department of Neurology, Center for Translational and Clinical Research, School of Medicine, University Hospital Center Zagreb, University of Zagreb, Zagreb, Croatia*

Matthew J. Huentelman • *Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ, USA; The Arizona Alzheimer's Consortium, Phoenix, AZ, USA; The Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA*

Kewal K. Jain • *Jain PharmaBiotech, Basel, Switzerland*

Alexandre Kuhn • *Microfluidics Systems Biology Lab, Institute of Molecular and Cell Biology, Agency for Science, Technology, and Research, Singapore, Singapore*

Ruth Luthi-Carter • *Department of Cell Physiology and Pharmacology, University of Leicester, Leicester, UK*
Ashley L. Siniard • *Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, AZ, USA; The Arizona Alzheimer's Consortium, Phoenix, AZ, USA*
Peter Weber • *Molecular Genetics of Affective Disorder, Max Planck Institute of Psychiatry, Munich, Germany*

Chapter 1

An Overview of Methods Used in Neurogenomics and Their Applications

Kewal K. Jain

Abstract

This chapter is an introduction to and an overview of neurogenomics—an analysis of genes in the nervous system and their application for diagnosis as well as potential therapeutics of disorders of the nervous system. The most important technologies are those for sequencing. Traditional as well as new techniques are described briefly including next-generation sequencing. Important applications include discovery of genomic biomarkers, brain mapping as well as connectomics, molecular diagnostics, drug discovery, and potential new therapeutics for neurologic disorders. Knowledge of the genes relevant to the nervous system will improve gene therapies and RNA interference approaches for neurologic disorders. Overall it will contribute to development of personalized neurology.

Key words Brain mapping, Gene therapy, Genes, Genomics, Molecular diagnostics, Neurogenetics, Neurogenomics, Neuroproteomics, Personalized neurology, Sequencing

Abbreviations

NGS	Next-generation sequencing
WGS	Whole-genome sequencing
WES	Whole-exome sequencing
CNV	Copy number variation
SNP	Single-nucleotide polymorphism
SMS	Single-molecule sequencing

1 Introduction

Genomics is the study of all of the genes in an organism, their sequences, structure, regulation, interaction, and products. As a scientific discipline, genomics involves mapping, sequencing, and analysis of the genomes and can be described as structural or functional. Structural genomics deals with construction of high-resolution genetic, physical, and transcript maps of an organism.

Kewal K. Jain (ed.), *Applied Neurogenomics*, Neuromethods, vol. 97,
DOI 10.1007/978-1-4939-2247-5_1,

The ultimate physical map of an organism is its complete DNA sequence. However, DNA sequence information provides only a static snapshot of the various ways in which the cell might use its proteins, whereas the life of the cell is a dynamic process. The human genome is extremely complex, and the estimated number of genes has varied considerably during the past years. GENCODE 19 contained 20,719 protein-coding genes.

A study has mapped peptides detected in seven large-scale proteomics studies to ~60 % of the protein-coding genes in the GENCODE annotation of the human genome (Ezkurdia et al. 2014). The investigators described a set of 2,001 potential noncoding genes based on features such as weak conservation, a lack of protein features, or ambiguous annotations from major databases. Peptides were identified for only 3 % of these genes. Most of these genes behave like noncoding genes rather than protein-coding genes and are unlikely to code for proteins under normal circumstances. If one excludes them from the human protein-coding gene catalog, the total number of genes in the human genome is reduced to ~19,000.

Approximately 80 % of the genes are expressed in the brain, and 5,000 of these exclusively in the brain and not in other organs. Neurogenomics is the study of genes in the nervous system. Of particular interest in neurology are the genes involved in neurologic disorders. In a broad sense, neurogenomics is the study of how the genome as a whole contributes to the evolution, development, structure, and function of the nervous system. The closely related term "neurogenetics" deals with the role of genetics in development and function of the nervous system as well as investigation and management of genetic disorders of the nervous system. Neurogenomics has applications in basic research, pharmaceutical industry, and the management of neurologic disorders (Jain 2001). Many of the methods used in neurogenomics are the same as those used for genomics in general. Sequencing is the most important activity in this area and a considerable space in this chapter is devoted to various methods of sequencing with exclusion of those that are specifically suited for study of microorganisms and cancer. Most of these methods are either in development or available in the commercial sector and the relevant companies are identified.

1.1 Historical Aspects

Historical landmarks in the development of genomics are listed in Table 1.

1.2 Variations in the Human Genome

Because of the diversity of the human species, there is no such thing as a normal human genome sequence. Variations are specific locations in the human genome where differences between

Table 1
Historical landmarks in the development of genomics

Year	Discovery/landmark/reference
Pregenomic era	
1871	Discovery of nucleic acids
1889	Hugo de Vries postulated "pangene" to be a living, self-replicating unit of heredity. His postulation was adapted from Darwin's "pangenesis" (the process by which cells might produce offspring)
1909	Introduction of the word "gene" (second half of pangene) into the German language as "Gen" by Wilhelm Ludvig Johannsen
1940	Beadle and Tatum linked genes to unique protein products and formulated the "one gene, one protein" concept
1951	Discovery of the first protein sequence
1953	Identification of the double-stranded structure of DNA (Watson and Crick 1953)
1960s	Modern concept of gene expression developed following discovery of messenger RNA, deciphering of genetic code, and description of the theory of genetic regulation of protein synthesis Establishment of the complete genetic code
Dawn of the genomic age	
1972	Production of the first recombinant DNA organism (Cohen et al. 1972)
1975	DNA hybridization analysis (Southern 1975)
1975	Introduction of 2-dimensional electrophoresis of proteins (O'Farrell 1975)
1977	Advent of DNA sequencing
1978	Discovery of restriction fragment length polymorphism (Maat and Smith 1978)
1981	Gene mapping by in situ hybridization becomes a standard method
1982	GenBank is established
1983	Demonstration of Huntington's disease gene (Gusella et al. 1983)
1985	Discovery of polymerase chain reaction (Mullis et al. 1986)
1986	Dr. Roderick coined the word "genomics" as the title of the journal that started publication in 1987 (Kuska 1998)
1987	Identification of dystrophin, the protein product of Duchenne muscular dystrophy gene, which now forms basis of gene therapy for this disorder (Hoffman et al. 1987)
Genomic age	
1990	Launch of the Human Genome Project, National Institutes of Health, United States (a $3 billion/15-year project)
1990	First human gene therapy experiment. Correction of adenosine deaminase deficiency in T lymphocytes using retroviral-mediated gene transfer (Blaese et al. 1990)
1991	Venter found that expressed sequence tags can provide a cheap, rapid way to skim the genome for practical information. Starting point of commercialization of genomics
1995	Definition of the proteome (Wilkins et al. 1995)
1996	Completion of the first whole-genome sequence of an organism: the budding yeast *Saccharomyces cerevisiae*
1999	First human chromosome sequenced: chromosome 22
2000	Completion of the sequencing of the human genome ahead of the anticipated date
Postgenomic era	
2000–2010	Increase in amount of sequence data; integration of information from genomics with that from other omics, such as proteomics and metabolomics; and applications for the development of personalized medicine

Table 2
Genetic variations in the human genome

Variation	Features
Complex chromosomal rearrangements (CCRs)	CCRs account for a large fraction of nonrecurrent rearrangements at a given locus
Copy number variation (CNV)	DNA segments >1 kb in length, whose copy number varies with respect to a reference genome. ~12 % of human genes vary in DNA sequences they contain
Insertions and deletions in the human genome (indel)	Indels are an alternative form of natural genetic variation that differs from SNPs
Interspersed repeated elements	Long and short interspersed nuclear elements are a significant portion of human genome
Large-scale variation in human genome	Large portions of DNA can be repeated or missing for no known reason in healthy persons
Segmental duplication	Duplicons have >90 % sequence homology to another region in the genome
Single-nucleotide polymorphisms (SNPs)	SNPs are sequence variations at single-base-pair level with a population frequency of >1 %
Structural variations (SVs)	SVs involve kilobase- to megabase-sized deletions, duplications, insertions, inversions, and complex combinations of rearrangements
Tandem repeats	Tandem sequences repetitions represent ~10 % of the genome

individuals are found, and the term "normal" or "wild type" refers to the most common variant at a location in a given population group. Variants are referred to as "alleles," but if the frequency of an allele is greater than 1 %, such variants are called polymorphisms. The term "mutation" is generally used for changes in DNA that are associated with disease.

Events contributing to genomic variation fall into three categories: (1) single-base-pair changes or point mutations that disturb the "normal" DNA nucleotide sequence, (2) insertions and deletions of nucleotides from the DNA, and (3) structural rearrangements that reshuffle the DNA sequence, thus changing the order of nucleotides (Feero et al. 2010). Replication-based mechanisms can result in complex genomic rearrangements. Genetic variations in the human genome are listed in Table 2.

1.3 Neurogenomics in Relation to Other Omics

There are numerous "omics" and relation of some of these is shown in Fig. 1. More are listed at the Website: http://www.genomicglossaries.com/content/omes.asp.

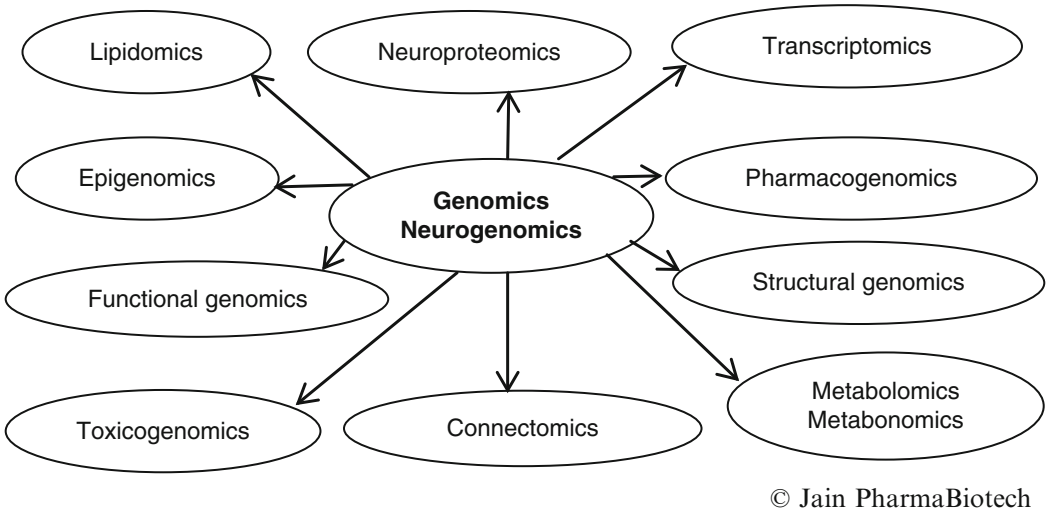

Fig. 1 Relationships of neurogenomics with other omics

Proteomics is the systematic analysis of protein profiles of tissues and parallels the related field of genomics. The term "proteomics" combines the words "protein" and "genome"; the spelling indicates PROTEins expressed by a genOME. Neuroproteomics refers to the protein profile of the nervous system. The massive amount of information generated by genomics and other omics has led to the development of bioinformatics and various tools that are required to analyze this data.

2 Methods of Study of Neurogenomics

2.1 Gene Expression

The activity of a gene, so called gene "expression," means that its DNA is used as a blueprint to produce a specific protein. Only a limited number of the genes are expressed in a typical human cell, and the expressed genes vary from one cell to another. Gene expression can be detected by various techniques. The discovery that eukaryotic genes are not contiguous sequences of DNA but consist of coding sequences (exons) interrupted by intervening sequences (introns) led to a more complex view of gene expression. The temporal, developmental, typographical, histological, and physiological patterns in which a gene is expressed provide clues to its biological role. Malfunctioning of genes is involved in most diseases, not only inherited ones.

All functions of cells, tissues, and organs are controlled by differential gene expression. Gene expression is used for studying gene function. Genes are now routinely expressed in cultured cell lines by using viral vectors carrying cDNA, the transcription of

which yields the gene's mRNA. RNA–RNA interaction can induce gene expression and RNA can regulate its activities without necessarily requiring a protein. The protein produced from mRNA may confer specific and detectable function on the cells used to express the gene. It is also possible to manipulate cDNA so that proteins are expressed in a soluble form fused to polypeptide tags. This allows purification of large amounts of proteins that can be used to raise antibodies or to probe protein function in vivo in animals. Knowledge of which genes are expressed in healthy and diseased tissues would allow us to identify both the protein required for normal function and the abnormalities causing disease. This information will help in the development of new diagnostic tests for various illnesses as well as new drugs to alter the activity of the affected genes or proteins.

Current techniques for analysis of gene expression either monitor one gene at a time, e.g., RT-PCR methods, or can do simultaneous analysis of thousands of genes, e.g., microarray hybridization or serial analysis of gene expression. A flexible, alternative PCR-based method, RAGE (rapid analysis of gene expression) has been developed which enables expression changes to be determined in either a directed search of known genes or an undirected survey of unknown genes. A single set of reagents and reaction conditions allows analyses of most genes in any eukaryote. The method is useful for assaying on the order of tens to hundreds of genes in multiple samples. Control experiments indicate reliable detection of changes in gene expression twofold and greater and sensitivity of detection better than 1 in 10,000.

2.1.1 Methods for the Study of Gene Expression in the Brain

The human brain has a more complex pattern of gene expression than any other region of the body. The molecular events in neurologic disorders are caused or paralleled by specific gene expression changes. Analysis of these changes provides an understanding of the disease at the molecular level. Gene expression profiling also provides some information about mitochondrial disorders because of a bidirectional information flow between the mitochondrion and the cell nucleus (Mende et al. 2007).

Several technological advances enable the analysis of thousands of expressed genes in a small brain sample. These techniques include expressed sequence tags, sequencing of cDNA libraries, differential display, subtractive hybridization, serial analysis of gene expression, and the high-density DNA microarrays. Gene expression measurements may be used to identify genes that are abnormally regulated as a secondary consequence of a disease state or to identify the response of brain cells to pharmacological treatments.

The usual method for the study of gene expression in the brain is by obtaining tissue sections and examining them for the expression of a particular gene using a fluorescent probe. When these genes are

illuminated under a fluorescence microscope, the regions where the gene is most highly activated within the nervous system are clearly shown. The nervous system provides abundant opportunities to study gene expression because of the presence of numerous genes that carry out a wide range of functions. However, the development of a probe for each gene that could potentially be expressed in the brain, and then the utilization of these probes to test for the presence or absence of gene expression, is a challenging task.

2.1.2 Study of Gene Expression by Brain Imaging

Molecular imaging is an emerging field of study that deals with imaging of disease on a cellular and molecular level. It can be considered as an extension of molecular diagnostics. In contradistinction to "classical" diagnostic imaging, it sets forth to probe the molecular abnormalities that are the basis of disease rather than to image the end effects of these molecular alterations. Radionuclide imaging, magnetic resonance imaging (MRI), and positron emission tomography (PET) can be used to visualize gene expression.

Three-dimensional gene expression patterns in the brain can be mapped by analysis of spatially registered voxels (cubes) by a process analogous to the images reconstructed in functional brain imaging systems. Consistent gene expression differences between normal and Alzheimer's disease brains can be demonstrated by this approach.

2.1.3 Study of Genetic Variation by Brain Imaging

Large-scale neuroimaging studies can be used to discover genetic variants that affect the brain. Screening of brain circuits for testing genetic associations in connectome-wide and genome-wide scans is feasible (Medland et al. 2014). Analysis of massive data, however, will be challenging.

2.2 Genotyping

Single-nucleotide polymorphisms (SNPs) serve to distinguish one individual's genetic material from that of another. There are no exact figures on the frequency of occurrence of single SNPs in the human genome, but they occur about once every 1,250 bases along the six billion base pairs, i.e., the "letters" that make up the genetic code. Studies suggest ~5 SNPs per gene, but not every gene has an SNP. Approximately nine million SNPs have been identified already in various databases but only a small fraction of these are well characterized and validated. SNPs comprise ~80 % of all known polymorphisms. Several technologies are used for their identification, of which the most important are based on DNA microarrays or biochip technology. SNPs have the following relation to an individual's disease and drug response:

- SNPs are linked to disease susceptibility.
- SNPs are linked to drug response, e.g., insertions or deletions of ACE gene determine the response to beta-blockers.