Original Article

Neuroinformatics: A Spotlight on Various Databases and the Importance of Their Integration

Erica Sequeira and Saraswathy Nagendran*

SVKM`s NMIMS Shobhaben Pratapbhai Patel School of Pharmacy and Technology Management, Vile Parle (W), Mumbai, India

Address for Correspondence

SVKM`s NMIMS Shobhaben Pratapbhai Patel School of Pharmacy and Technology Management, Vile Parle (W), Mumbai, India.

E-mail: saraswathy76@gmail.com

ABSTRACT

Brain is one of the most complex organs of the body and thus the challenge in neuroscience would seem the effort of understanding the complex structure, function, and development of the nervous system in healthy as well as diseased condition. Such understanding requires the integration of huge amounts of heterogeneous and complex data collected at various levels of investigation. *Neuroinformatics* combines neuroscience with information science/technology and deals with the creation and maintenance of web accessible databases that will be required to achieve such integration. There is significant interest amongst neuroscientists in sharing neuroscience data and analytical tools which provides the opportunity to differently re-analyze previously collected data and encourage new neuroscience interpretations that facilitates further development. However, information is usually stored in various databases, managed by heterogeneous database management systems or files, spreadsheets etc. which results into inaccuracy and inconsistencies in data acquisition and processing, lack of coordination resulting in duplication of efforts and of resources. Thus an integration of databases is a vital solution to these problems. This article, analyses various databases in the field of neuroscience along with the importance of integration of databases.

Keywords: Neuroscience, Databases, Nervous system, Brain, Information technology.

INTRODUCTION

Neuroinformatics as a field for understanding the nervous system in healthy as well as the diseased condition was initiated in the early 1900[1]. The subsequent decade witnessed a development which involved dealing with the complex types of data that is required during the research. In 2000, a paper had suggested that databases were needed to organize the numerous databases and tools[2]. Around 2004, a database (SfN (Society for Neuroscience) Neuroscience Database Gateway 2007)[3] that emerged to satisfy this need. Many investigators realized that the computer was a vital tool to keep a track of the series of letters that represented the base sequence that makes up the nucleic acids (e.g. GCAT) or the sequence of amino acids that makes up the protein molecule[4-7]. European Molecular Biology laboratory's EMBL-Bank and the National Center for Biotechnology Information GenBank, were launched in the early 1980s for databases of genes[8,9]. Neuroinformatics is a field that capitalizes on the potential synergies between neuroscience and information technology (IT) such as: deals with the application of advanced IT methods to deal with the flood of neuroscientific data by developing and applying data analysis methods for the study of the brain; by providing both analytical and numerical tools for theoretically modeling brain function; and by exploiting our insights into the principles underlying brain function to develop new IT technologies. With the development of many methods of studying the brain, researchers have studied the brain using various technologies such as magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and computerized tomography (CT) to study functions, connectivity, and structures of the brain; microarray, in situ hybridization (ISH) and next generation sequencing (NGS) to study the molecular state of the brain; electroencephalography (EEG) and magnetoencephalography (MEG) to study the electrophysiology of the brain[10]. Thus, neuroscience has become essential to the neuroscience investigators and clinicians for conducting scientific inquiry, understanding the cause and pathogenesis of a brain disorder and practicing medicine.

Neuroinformatics databases

Brain imaging database

Current techniques in brain imaging are divided into mainly two different areas, firstly, structural imaging technique for studying the anatomy of the brain, for example, computerized tomography (CT) and magnetic resonance imaging (MRI) and secondly, functional imaging techniques for studying its functions or the connectivity of the brain, for example, functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and diffusion tensor imaging (DTI).

Brain-development.org

It is a part of the research project in the Imperial college London which contains datasets of over 600 MRI images of healthy subjects which utilizes T1 T2 and PD weighted images, MRA images, DTI images, and Diffusion weighted images of normal and healthy human brains. It also gives the demographic information of each subject[11].

Brainmuseum.org

It is a public web site that provides the images and the information of well preserved, sectioned and stained brains of over 100 species of mammals including humans. Stained sections of several species such as humans, chimpanzees, monkeys, carnivores, rodents, California sea lion, Big brown bat, American badger, American Racoon, yellow mongoose, zebra, cow and Atlantic Bottle nose Dolphin is available and it explains how the evolution of brain occurred[12].

Harvard whole brain atlas

It is a comprehensive and a well designed site that contains a huge compilation of modern cross sectional signaling including CT, MRI and SPECT in health and diseased condition of brains such as cerebrovascular diseases, neoplastic diseases, degenerative diseases, and infections and inflammatory diseases[13].

Brainmap.org

It is a database of published functional and structural neuroimaging experiments with coordinate-based results (x, y, z) in Talairach or Montreal Neurological Institute (MNI) space. It involves the development of software and tools to share neuroimaging results and enable meta-analysis of studies of human brain function and structure in healthy and diseased subjects[14].

Brainmaps.org

It is an interactive multiresolution next generation brain atlas that is based on over 20 million megapixels of submicron resolution, annotated, scanned images of serial sections of both primates, e.g. Homo sapiens and non-primate brain, e.g. *Carassius auratus*. In addition, it also contains connectivity maps, Connectomes.org of some species such as *Caenorhabditis elegans*[15,16].

Neuromorpho.org

It is the largest public repository of 3-D digital neuronal reconstruction and associated metadata. It is an inventory that contains 7,986 digitally reconstructed neurons and is associated with peer-reviewed publications, it can be used for analysis, visualization and modelling of neuronal data[17].

Electrophysiological databases

Brains achieve efficient functioning due to different neurons that serve distinct computational role. One striking way in which the neuron types differ is their electrophysiological properties. Electro-physiology of the brain is an important aspect since the biological information is converted into electrical signals for the brain to interpret. However there are no firm standards how electrophysiological data is to be described and integrated due to the complexity and diversity of it. In spite of all these difficulties some databases and tools are developed and available now.

NeuroElectro

It is a database describing the electrophysiology properties of different neuron types so as to understand the role of diversity across neuron types. Using a combination of manual and automated methods it describes a methodology to curate the neuron electrophysiology information into a centralized database and facilitates the discovery of neuron-to-neuro relationships[18].

Code Analysis, Modeling and Repository for E-Neuroscience (CARMEN) Project

It provides a web based portal platform through which users can share and collaboratively exploit data such as the neural activity recording including signals and series of images; analysis code and expertise in neuroscience[19].

Collaborative Research in Computational Neuroscience (CRNCS)

It provides a forum of discussion over electrophysiological datasets as well as involved in sharing data. Data sets of this forum include the visual cortex, auditory cortex, hippocampus, eye movements, lateral geniculate nucleus in the mouse. It gives information about data collection, experimental conditions, identifying the species, surgical procedures and the special recording techniques[20].

Brain connectivity databases

To transfer information between anatomical structures of a brain, the brain is wired with a dense network of electrical signals such as the axonal connections and synapses. But there are many connections between the brain components of various scales to an entire brain regional scale, and different technologies are required for the analysis of each scale of connection. Connectomes are divided to micro connectome and macro connectome on the basis of data scale of connections.

Micro connectome is a matrix of all axonal connection between all individual neurons. Open connectome project is open to anybody to view, manipulate, analyze and contribute. It is interested in all levels, from nano (using electron microscopy) to macro (using magnetic resonance imaging). It provides the neuroanatomical images and the ultimate aim is to reconstruct the 3D model of brain. Meso connectomes is a matrix of all axonal connections between all neuron types. Macro connectomes (connectivity matrices at brain regions level), focuses on the strength of connections and the number of neurons between two brain regions[21]. Macro connectome data is usually obtained from fMRI (functional Magnetic Resonance Imaging), Diffusion MRI, and DTI (Diffusion Tensor Imaging)[22].

CoCoMac

It is the main database and contains data from tracing studies on anatomical connectivity in the macaque cerebral cortex. The combined results are used for constructing the macaque macro connectome. It provides connectivity matrix data of primate brains. Since connectivity matrix is well-quantized such as each cell of matrix has a number of neurons composing each connection, its data can be easily interpreted and its use is facilitated by data mining technique. CoCoMac also provides large amount of data and 3D graphics[23].

BAMS (Brain architecture management system)

It is an online database having the information about the neural circuitry, brain connectivity database and an ontology of brain regions. One of the special features of BAMS is that it has atlas and parsing engine. The parsing engine makes connectivity map containing brain regions and connection strength. Since it has 5 atlases so it can easily be related to the other neuroscience databases[24].

Databases for genetic information

Gene expressions in various brain regions and various cell types are important information for understanding a relationship between regulation of molecular level and physiological level of brains. The volume of genetic data has grown exponentially and researches have studied gene expression of brains.

SigCS base

It is an integrated genetic information resource for human cerebral stroke. SigCS base is an effective tool that can assist researchers in the identification of the genetic factors associated with stroke by utilizing existing literature information, selecting candidate genes and variants for experimental studies, and examining the pathways that contribute to the pathophysiological mechanisms of stroke. The current version of SigCS base documents 1943 non-redundant genes with 11472 genetic variants and 165 non-redundant pathways[25].

Gene expression omnibus (GEO)

It was a project initiated in response to the growing demand for public repository for high-throughput gene expression data. It provides a flexible and an open design that facilitate submission, storage and retrieval of heterogeneous data sets from gene expression and genomic hybridization experiments[26].

ArrayExpress

It is a new public database of microarray gene expression data which is a generic gene expression database designed to hold data from all microarray platforms.

ArrayExpress uses the annotation standard minimum information about a Microarray Experiment (MIAME) and the associated XML data exchange format Microarray Gene Expression Markup Language (MAGE-ML) and it is designed to store well annotated data in a well organized way. Both, the GEO and ArrayExpress are representative repositories of publicly available gene expression data from National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EBI). They stored vast amount of gene expression profiles from brains of various species, various brain regions, including the brains with some disorders[27].

Allen brain atlas

A growing collection of public resources integrating gene expression and neuroanatomical data. Interestingly, in addition, it also has launched Allen Developing Mouse Brain Atlas, Allen Human Brain Atlas. Allen Developing Mouse Brain Atlas is gene expression data in the mouse brain beginning with mid-gestation through to juvenile / young adult. The Allen Developing Mouse Brain Atlas shows temporal and spatial regulation of gene expression of mouse brain. Allen Human Brain Atlas provides gene expression from human whole brain regions which is the first and unique multi-modal gene expression atlas of the human brain. The interesting features include both simple and sophisticated methods for gene searches, colorimetric and fluorescent ISH image viewers, graphical displays of ISH, microarray and RNA sequencing data, Brain Explorer software for 3D navigation of anatomy and gene expression, and an interactive reference atlas viewer. In addition, cross data set searches enable users to query multiple Allen Brain Atlas data sets simultaneously[28].

Databases related to brain disorders

The various brain disorders such as Alzheimer’s disease, Parkinson’s disease, Huntington’s disease, depression, bipolar disorder, schizophrenia, and so on are treated by various therapies and drugs. Due to the complexity involved in the mechanism and treatment strategies, it is of interest to collate the different brain disorders in the form of a database.

Alzheimer’s disease neuroimaging initiative (ADNI)

Unites the researchers with study data as they work to prevent and treat the progression of Alzheimer’s disease. ANDI researchers collect, validate and utilize data such as MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers as predictors for the disease. Using this database, researchers can validate their biomarkers obtained from MRI / PET imaging, blood tests and tests of cerebrospinal fluid for Alzheimer's disease (AD) clinical trials and diagnosis[29].

Major depressive disorder neuroimaging database (MaND)

Provides an Excel spreadsheet file containing database and meta-analysis. The database contains information of 225 studies which have investigated brain structure (using MRI and CT scans) in patients with major depressive disorder compared to a control group. 143 studies and 63 brain structures are included in the meta-analysis[30].

The open access series of imaging studies

A series of magnetic resonance imaging data sets that is publicly available for study and analysis. The cross sectional MRI data contains images of 416 subjects aged 18 to 96 and longitudinal MRI data in non demented and demented older adults. Longitudinal MRI data contains longitudinal collection of 150 subjects aged 60 to 96. In

Longitudinal MRI datasets, 72 subjects were not demented throughout the study, 64 subjects were characterized as demented at initial image, and 14 subjects were not demented at initial image and subsequently characterized as demented through a longitudinal study[31].

PDGene, SZGene, ALSGene, MSGene and Alzgene are examples of databases for providing Parkinson's disease, schizophrenia, amyotrophic lateral sclerosis, multiple sclerosis, and Alzheimer's disease genetic association studies with sufficient data[32-36].

MDPD (Mutation database for parkinson's disease)

Provides not only an integrated genetic information resource for Parkinson's disease but also each genetic substitution and the resulting impact with its reference[37].

DND (Database of Neurodegenerative Disorders)

An on-line web based database that contains more than 100 neuro related disease concepts and provides with a covering of all related genes, proteins, pathways and drug information. The interface offers different options to learn about the disease causing molecular factors[38].

Environment Wide Association Study (EWAS)

Using environmental factors with brain disorders can be conducted for brain disorders[39].

National Health and Nutrition Examination Survey (NHANES)

Provides studies designed to assess the health and nutritional status of adults and children in the United States, carried out by interviews and physical examinations. It contains information of samples about not only environmental factors but also clinical status of mental illness (e.g. depression) and that information can be used for surveying the brain disorders. Thus it gives the information about the prevalence of the major diseases and the risk factors for the diseases[40].

Integration of databases

With the increasing amount of data of brain generated from technologies there has been an increasing need of integration of databases as well as codifying them in standard formats.

Neuroscience Information Framework (NIF) is one such example which provides the information network and has integrated over 174 datasets and databases. It is meant for researchers to discover and access to a large amount of public neuroscience data and tools easily, thus serves as a single hub for neuroscience information[41].

Neuronames is a comprehensive hierarchical nomenclature for structures of the primate (human and macaque) brain used as an indexing structure for computer systems. It defines the brain using 550 primary structures and includes other related structures, names, and synonyms. Neuronames contains 15,000 neuroanatomical terms, and shows hierarchical relationship between each structures and neuroanatomical terms[42].

Combination of databases have their own advantages, for example, different gene expressions from postmortem brain studies of mental illness can be caused by not only effects of mental illness but also by treatments of the brain disorder. Leon French *et al*., 2011, studied relationship between gene expression and neuroanatomical connectivity in the adult rodent brain[43] by utilizing a large data set of the rat brain "connectome" from the Brain Architecture Management System and used statistical approaches to relate the data to the gene expression signatures of 17,530 genes in 142 anatomical regions from the Allen Brain Atlas. Their results showed that gene expression signature of mouse brain regions are significantly related to connectivity of mouse brain regions and also found a set of genes that are closely correlated with neuroanatomical connectivity. Lior Wolf *et al*., 2011, also studied the link between gene expressions of mouse brain regions and neural connectivity patterns of mouse brain regions[44] by using Allen Brain Atlas and Brain Architecture Management System. They found that through the gene expression levels of mouse brain regions it is possible to predict the connectivity of mouse brain regions. Byungkyu Part *et al.,* combined genome-wide microarray- based gene expression profiles and diffusion tesnsor images of human brain to study Alzheimer's disease[45]. They modelled the interactions of Alzheimer related genes from the fiber pathway by using microarray data of Allen Human Brain Atlas and diffusion tensor images of Allen Brain Atlas.

BRAINnet Database

It is another example of the integrative neuroscience and is the largest available library of human brain health information, where multiple sources of data are available on the same individual. As of June 2009, data from the Brain Resource International Database has been made available to BRAINnet from 5,000 subjects, who were healthy and 1,000 subjects with clinical disorders such as Major Depressive Disorder, First Onset Schizophrenia, Post Traumatic Stress Disorder, Alzheimer's disease, Mild Cognitive Impairment, Panic disorder, Sleep apnea, Anorexia nervosa, Traumatic Brain injury and Attention deficit hyperactivity disorder. Thus it has the capacity to provide members with different types of data in the same individual and these different kinds of data can be integrated straight away because they are available on the same study subjects. The amount of data is enormous and still continuous to grow[46].

CONCLUSION

Through this survey, various databases of different information types has been represented. Various databases are already available for the researchers so as to understand the brain in every aspect and more information would be available in future to the researchers. Thus, the researchers have increased opportunities to study the brain from new perspectives using these databases. Integrating the databases is a novel approach towards understanding the brain which decreases the chances of errors and helps in better understanding of the brain. Despite the research being very active in the past 20 years, integration of databases has to be achieved at all levels and the effort to solve the integration issues should be continued so that the proposed methodologies are evaluated through experiments.

REFERENCES

1. Huerta MF, Koslow SH, Leshner AI. The Human Brain Project: an international resource. *Trends Neurosci* 1993; 16(11):436–8.
2. Smaglik P. Internet gateway planned for neuroinformatics data. *Nature* 2000; 405(6787): 603.
3. SfN. Neuroscience Database Gateway. 2007; NDG [online] URL:http://ndg.sfn.org/ and Databases having Atlas as Categories. URL:http://ndg.sfn.org/eavObList.aspx?cl=81&at=262&vid=28603.
4. Ouzounis CA, Valencia A. Early bioinformatics: the birth of a discipline-a personal view. *Bioinformatics* 2003; 19:2176–90.
5. Smith TF. The history of the genetic sequence databases. *Genomics* 1990; 6(4):701–7.
6. Dayhoff MO: Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC, USA 1978; 7(3).
7. Bernstein FC, Koetzle TF, Williams GJB. The protein data bank: a computer-based archival

file for macromolecular structures. *J. Mol. Biol* 1977; 112:535–42.
8. *National Library of Medicine, National Institute of Health*. Growth of GenBank 1982; 2005 [online] 2005b, URL:http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html.
9. *National Library of Medicine, National Institute of Health*. Medical Subject Headings (MeSH) [online] 2007; URL:http://www.nlm.nih.gov/mesh/meshhome.html.
10. Ha Sun Yu, Joon Bang, Yousang Jo. Combining Neuroinformatics Databases for Multi-level Analysis of Brain Disorders. *Interdisciplinary Bio Central* 2012; 4(7); 1-8.
11. http://www.brain-development.org/.
12. http://www.brainmuseum.org/index.html.
13. http://www.med.harvard.edu/AANLIB/home.html.
14. Laird AR, Lancaster JL, Fox PT. BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* 2005; 3(1):65-78.
15. Mikula S, Trotts I, Stone JM, Jones EG. Internet-enabled high-resolution brain mapping and virtual microscopy. *Neuroimage* 2007; 35(1):9-15.
16. http://connectomes.org/.
17. Ascoli GA, Donohue DE, Halavi M. NeuroMorpho.Org: a central resource for neuronal morphologies. *J Neurosci* 2007; 27(35):9247-9251.
18. http://www.neuroelectro.org
19. http://www.carmen.org.uk/.
20. http://crcns.org/.
21. http://openconnectomeproject.org/.
22. Swanson LW, Bota M. Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proc Natl Acad Sci* 2010; 107(48):20610-7.
23. Stephan KE, Kamper L, Bozkurt A, Burns GA, Young MP, Kotter R. Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). *Philos Trans R Soc Lond B Biol Sci* 2001; 356(1412):1159-1186.
24. Bota M, Dong HW, Swanson LW.Brain architecture management system. *Neuro-informatics* 2005; 3(1) :15-48.
25. sysbio.kribb.re.kr/sigcs/.
26. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; 30(1):207-210.
27. Rocca-Serra P, Brazma A, Parkinson H. ArrayExpress: a public database of gene expression data at EBI. *C R Biol* 2003; 326(10-11):1075-1078.
28. Lein ES, Hawrylycz MJ, Ao N. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 2007; 445(7124):168-176.
29. Jack CR, Bernstein MA, Fox NC: The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* 2008; 27(4):685-691.
30. Kempton MJ, Salvador Z, Munafo MR. Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. *Arch Gen Psychiatry* 2011; 68(7): 675-690.
31. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, non demented, and demented older adults. *J Cogn Neurosci* 2007; 19(9): 1498-1507.
32. Lill CM, Roehr JT, McQueen MB. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet* 2012; 8(3):e1002548.
33. Allen NC, Bagade S, McQueen MB. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008; 40(7):827-834.
34. Lill CM, Abel O, Bertram L, Al-Chalabi A. Keeping up with genetic discoveries in amyotrophic lateral sclerosis: the ALSoD and ALSGene databases. *Amyotroph Lateral Scler*. Jul 2011; 12(4):238-249.
35. Lill CM RJ, McQueen MB, Bagade S, Schjeide BM, Zipp F, Bertram L. The MSGene Database. Alzheimer Research Forum.2012b; http://www.msgene.org/.
36. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet.* 2007; 39(1):17-23.

37. Tang S, Zhang Z, Kavitha G, Tan EK, NG SK. MDPD: an integrated genetic information resource for Parkinson's disease. *Nucleic Acids Res.*2009; 37: D858-862.
38. bioschool.iitd.ac.in/NeuroDNet/
39. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 2010; 5(5):10746.
40. http://www.cdc.gov/nchs/NHANES.htm.
41. Gardner D, Akil H, Ascoli GA. The neuroscience information framework: a data and knowledge environment for neuroscience. *Neuroinformatics* 2008; 6(3):149-160.
42. Bowden DM, Song E, Kosheleva J, Dubach MF. NeuroNames. An ontology for the Brain Info portal to neuroscience on the web. *Neuroinformatics* 2012; 10(1):97-114.
43. French L, Pavlidis P: Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comput Biol* 2011; 7(1):1001049.
44. Wolf L, Goldberg C, Manor N, Sharan R, Ruppin E. Gene expression in the rodent brain is associated with its regional connectivity. *PLoS Comput Biol* 2011; 7(5):1002040.
45. Park B, Lee W, Han K. Modeling the interactions of Alzheimer-related genes from the whole brain microarray data and diffusion tensor images of human brain. *BMC Bioinformatics* 2012; 13(7):S10.
46. www.brainnet.net/.