Federating and Integrating What We Know About the Brain at All Scales: Computer Science Meets the Clinical Neurosciences

Richard Frackowiak, Anastasia Ailamaki, and Ferath Kherif

Abstract Our everyday professional and personal lives are irrevocably affected by technologies that search and understand the meaning of data, that store and preserve important information, and that automate complex computations through algorithmic abstraction. People increasingly rely on products from computer companies such as Google, Apple, Microsoft and IBM, not to mention their spinoffs, apps, WiFi, iCloud, HTML, smartphones and the like. Countless daily tasks and habits, from shopping to reading, entertainment, learning and the visual arts, have been profoundly altered by this technological revolution. Science has also benefited from this rapid progress in the field of information and computer science and associated technologies (ICT). For example, the tentative confirmation of the existence of the Higgs boson (CMS Collaboration et al. Phys Lett B 716:30–61, 2012), made through a combination of heavy industrial development, internet-based scientific communication and collaboration, with data federation, integration, mining and analysis (Rajasekar et al. iRODS primer: integrated rule-oriented data system. Synthesis lectures on information concepts, retrieval, and services. Morgan & Claypool, San Rafael, 2010; Chiang et al. BMC Bioinformatics 12:361, 2011; Marks. New Sci 196:28–29, 2007), has taken our understanding of the structure of inorganic matter to a new level (Hay et al. The fourth paradigm: data-intensive scientific discovery. Microsoft, Redmond, WA, 2009). But within this vision of universal progress, there is one anomaly: the relatively poor exploitation and application of new ICT techniques in the context of the clinical neurosciences. A pertinent example is the genetic study of brain diseases and associated bioinformatics methods. Despite a decade of work on clinically well-defined cohorts, disappointment remains among some that genome-wide association studies (GWAS) have not solved many questions of disease causation, especially in psychiatry (Goldstein. N Engl J Med 360:1696–1698, 2009). One question is

R. Frackowiak (✉) • F. Kherif
Department of Clinical Neurosciences, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland
e-mail: richard.frackowiak@gmail.com

A. Ailamaki
Department of Computer Science, Ecole Polytechnique Fédérale de Lausanne, 1021 Lausanne, Switzerland

G. Buzsáki, Y. Christen (eds.), *Micro-, Meso- and Macro-Dynamics of the Brain*,
Research and Perspectives in Neurosciences, DOI 10.1007/978-3-319-28802-4_10

whether we have the appropriate disease categories. Another factor is that gene expression is affected by environmental and endogenous factors, as is protein function in different circumstances (think of the effects of age, developmental stage and nutrition). It is clear that any genetic associations with disease expression are likely to be highly complex. Why then are the world's most powerful supercomputers not being deployed with novel algorithms grounded in complexity mathematics to identify biologically homogeneous disease types, or to understand the many interactions that lead to the integrated functions that arise from DNA metabolism, such as cognition? Is it from a lack of appropriate data and methods or are the reasons related to our current clinical scientific culture?

Introduction

Reductionist methods of hypothesis falsification have dominated science in the last two centuries, and rightly so, given the major advances in knowledge about the living and non-living worlds they have afforded. However, there is much evidence to suggest that a uniquely reductionist approach may be blinkered. Indeed it may always have been so—think of Linnaean categorization of the plant kingdom as a scoping exercise prior to a more modern hypothesis-led, genetically based description of plant biology. Darwin had no idea about the physical nature of the hereditary process he so cleverly deduced (he received one of the few original copies of Gregor Mendel's manuscript, but did not read it, judging by the fact that it was found uncut at his death). Yet he catalogued the animal kingdom, discovering hidden patterns that gave rise to his theory about adaptive mechanisms and successful procreation underlying the evolution of species. But do biomedical granting agencies fund work that does not express a firm and clear hypothesis? What modern biomedical grants agency will fund scoping studies involving observation and classification (though again, perhaps that's what GWAS studies are)? Outside epidemiology, such a scenario is difficult to entertain. And in epidemiology, how many studies emerge from the correlative world of univariate statistics, and how many founder on inadequate power?

The most powerful means of examining the spread of influenza epidemics is now achieved by analyzing the geographical spread of incidence posted on Google (Brownstein et al. 2008; http://www.google.org/flutrends/). This is a real-time example of the interconnected power of global computing; crowd sourcing is another (Brabham 2008). How organic matter self-organizes across spatial and temporal scales to produce the diversity of living, reproducing, adaptive creatures and their nervous systems is a question that is slowly being addressed with a new methodological agenda. The complexity of the human brain demands modern methods that address, describe and quantify interactions in large integrated systems.

Clinicians need to take note of this trend, both in terms of the science and art of medicine and also in any effort to rapidly identify and develop effective treatments.

Syndromic Diagnosis

What is the challenge? Firstly, the clinical-pathological paradigm of the last century and a half, attributed to Broca in the clinical neurosciences, has reached the limits of its usefulness. Syndromes, composed of groups of symptoms narrated by patients with varying degrees of cognitive impairment, or by their relatives, to individual practitioners, overlap too much to remain useful as a basis for the precise diagnosis of brain diseases. This is not a new insight, as demonstrated by the variability in presentation of diseases such as syphilis and diabetes mellitus, but it is an increasingly pertinent one. Recently it has been reported that the five major classes of psychiatric illness share a similar set of associated genes that predispose not to one or other class but to mental illness in general (Cross-Disorder Group of the Psychiatric Genomics Consortium 2004). The spinocerebellar ataxias are associated with well over 20 dominant, often partially penetrant, mutations, each of which generates a similar pattern of clinical features, at times causing diagnostic confusion (Schöls et al. 2012). The dementia syndrome is caused by a range of pathological mechanisms, a few of which are genetically determined, the vast majority of which are of unknown aetiology, to the extent that the diagnosis of Alzheimer's disease (AD) is wrong in the best centers about 30 % of the time, if post mortem features are used to define disease (Beach et al. 2012). Longitudinal syndromic studies demonstrate that even diagnoses of "pure" syndromes fail to remain applicable through life, and correlation with post mortem features is poor if not random (Kertesz et al. 2005). Finally, the same single genetic mutation can present with a variety of syndromes. A simple example is that of Huntington's disease, where a behavioral or psychiatric presentation is recognized, as are presentations with movement disorders or gait abnormalities. Though the phenomenon of generational anticipation in male presentation of the disease is associated with the length of CAG repeats in the huntingtin gene, it is not understood how this happens. In short, there is a pressing need to move from an observational and simple correlational approach to clinical neuroscience to one that is mechanistic and multifactorial.

A Theory of the Brain

That is easier said than done, for a simple reason. Unlike the materials sciences, where there is a clear if still often approximate (except at the quantum level) understanding of the organization of inorganic matter across spatial and temporal scales, no such theory of living matter is available. However, this is not an intractable problem with infinite degrees of freedom, as some have suggested.

The building block of organic matter, DNA, is composed of a limited set of highly specific base pairs. We have a good understanding of how transcription to RNA and translation to proteins occur, and what mechanisms control these processes. The human genome is known and much if not all of the variation in it has been catalogued. Much of it consists of (mysterious) non-coding sequences. That takes care of a lot of degrees of freedom and sets parameters on how life itself emerges, as well as cognition, emotion, perception and action. The rules that determine mechanistic interactions at these basic levels are constantly being discovered but remain unconnected without a global theory of brain organization from the lowest to the highest levels: from base pairs, to genes, to functional and structural proteins, to neurons and glia, to cortical columns and subcortical nuclei, to redundant networks and functioning, learning adapting systems, and eventually to cognition and more. Each level with its rules constrains the structure and function of the next, more complex ones. There are many examples of such rules. The Hodgkin-Huxley equations are the best known and among the oldest (Hodgkin and Huxley 1952). In principle, then, all the levels of brain organization should eventually become expressible in terms of mathematical relationships, and that would constitute a brain theory, or model.

Computers

A decade or two ago, the idea that an inestimable number of simultaneous non-linear equations could represent a theory of brain organization, if dreamt of by a few, seemed such an unlikely proposition that it merited no more than a passing *frisson*. There were two fundamental problems: how to make the calculations, and how to amass the data on which to make them. The first problem is largely solved, at least in principle and partly in practice. The most powerful super-computers currently available are at the peta-flop level (http://www.top500.org/list/2013/06/).

The IBM roadmap predicts the production of an exascale computer around 2018 (1×10^{18} flops/s). Extrapolating today's Blue Brain Project numbers, exascale is probably the minimum required to simulate the entire brain. This level of performance is just sufficient for the simultaneous computation of the present estimate of the number of equations needed to provide a first holistic version of a brain model, one that instantiates the nonlinear interactions that give rise to the emergent properties of living brains. As to data storage, this is a practical problem that has effectively been solved by cloud computing and distributed storage with appropriate addressing; it is data analysis and aggregation with efficient database queries that are challenges at this scale.

Data and Data Mining

Clinical scientists are used to dealing with highly controlled, "clean" data sets, despite the messy nature of their observational constructs. Hence their data sets are often small, precious and closely guarded, being a critical part of the discovery process. This mind set is invalidated by advances in data mining algorithms that have become commonplace in industry (banking, nuclear power, air transportation, space and meteorology, to name but a few) (http://en.wikipedia.org/wiki/Data_mining).

Such algorithms identify patterns in big data that are characterized by invariable clusters of (mathematical) rules. In other words, they are rule-based classifiers. They offer a potential escape from the world of correlations into the world of causes. However, strictly, rule-based classification generates correlations, not causality (although it depends on how narrowly causality is defined). It shows what occurs together but not what causes what. Homogeneous clusters are useful for disease signatures, but for treatments causality will have to be understood by integration of knowledge and simulation results from genetics, biochemistry, physiology and medical description into randomized experiments (Fig. 1).

These powerful and computer-sensitive, data-hungry algorithms often use novel mathematics. They have been developed because the new generations of computers can verify and validate them. They deal with multivariate and "dirty" data, missing data, textual or semantic data and data from different sources or with different ranges. They can work in non-linear, non-Euclidean, non-stochastic, high-dimensional spaces (Loucoubar et al. 2013). Others are more statistically based, such as machine learning techniques. Some attempt to exhaustively test all possible models describing the data to discover the most parsimonious set that explains them. Which will be the best tools and methods for use in the clinical neurosciences is not yet clear, but one can be sure that data mining will generate many hypotheses for testing! And so the perspective emerges that the comprehension of brain organization and the causes of brain disease are not to be found by a reductionist approach alone but by a combination of hypothesis falsification that follows a constructivist, simulation-based approach using novel classifiers working on large amounts of real biological data.

Simulating the Brain

An initial proof-of-concept program has recently communicated very encouraging results. The Blue Brain Project (http://bluebrain.epfl.ch) at the Brain Mind Institute of the Ecole Polytechnique Fédérale de Lausanne (EPFL) took as its starting point data on the functionality of ion channels and their distributions along axons and dendrites of different neural types (Peck and Markram 2008; Khazen et al. 2012). Proceeding with a simulation-based approach, using biological data about matters

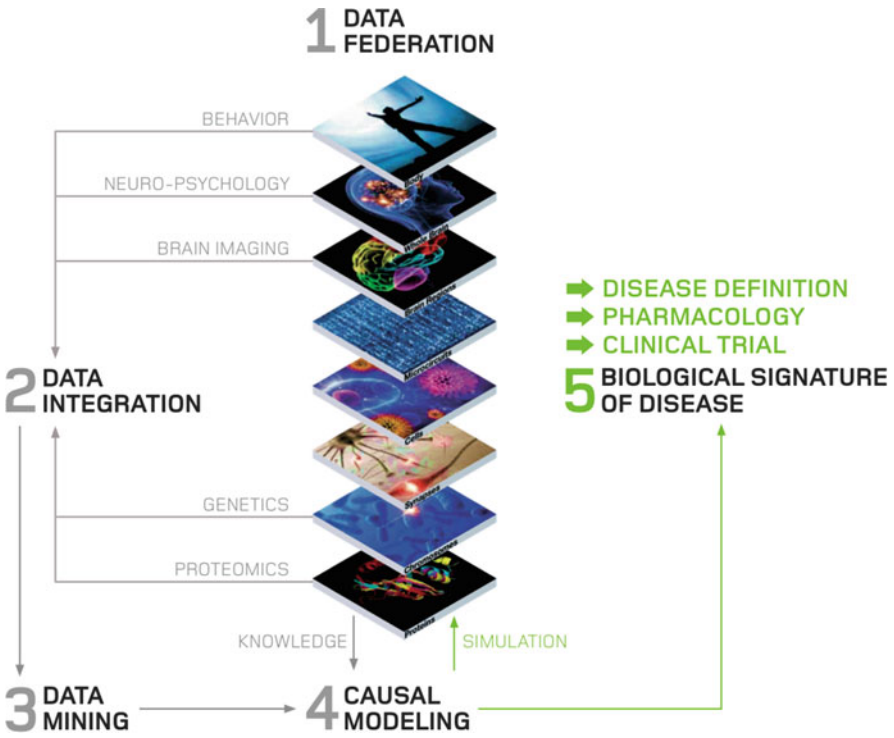

Fig. 1 Schematic representation of steps in applying modern informatics to clinical neuroscience. The Human Brain Project (HBP) aims to collect, explain and simulate the functions of the human brain at different levels of hierarchical complexity. Within the HBP framework, a strategically feasible approach to understand brain diseases is depicted in this figure. The idea is to federate (1) and integrate (2) the data, thus making use of an abundance of biological information from the different levels of brain organisation. Data mining (3) will be used to extract sets of rules that constitute definitions of homogeneous groupings of patients or subjects. Causal modelling with new data (real or simulated) will be performed for external validation (4), which will complete the process of defining (5) the biological signatures of diseases. The signatures of diseases will constitute the basis for a new, biologically determined nosology that should facilitate drug target identification and selection of homogeneous groups of patients for clinical trials as well as simulation of the effects of pharmacological treatment and secondary event profiles

such as cortical volume, the distribution of cortical layers, the distribution of capillaries, the variation in numbers and distributions of morphological and functional types of neurons in the various cortical layers, and statistical data on the probability of connections between different cell types, it built a preliminary model of a rodent cortical column using an IBM Blue Gene/Q computer. A correspondence of functionality and morphology between the model and ex vivo slices of brain tissue has been demonstrated (Ramaswamy et al. 2012). Predictions about the distribution of synaptic connections (Hill et al. 2012) and the occurrence of spontaneous activity (in the gamma band) have also been made that in themselves constitute strong hypotheses for further empirical verification.

Data Provenance

The computing power needed for the extension of such a project to whole brains is now within our reach. Data provenance remains a problem. In the research domain, there are 30 years of data described in millions of scientific papers lodged in repositories such as the National Library of Science in Washington DC. There are many basic science laboratory databases, often publically funded, held in universities and research laboratories around the world. These data have often been used once and exist for archival reasons alone. In the clinical field, there are databases in each hospital that contain clinical and diagnostic information on innumerable patients. Again, the data are used for the benefit of an individual and are normally kept for medico-legal reasons or as a baseline for returning patients. In countries with socialized medicine, these data are paid for by taxes and so, at least in part, belong to the public. This mass of legacy data represents an enormous, untapped research resource. How can such heterogeneous data be usefully exploited?

Real-time data addressing is a fact of life for anyone who uses the Internet and a search engine today. Therefore, in principle, the infrastructure and software are available. It remains to be seen whether specialized integrated hardware and software infrastructures will become acceptable to hospitals and researchers for scientific activity. Issues such as privacy protection in the context of anonymization are technically solvable and already acceptable on the grounds of proportionality (the potential benefit to members of society as a whole, compared to the potential risk to an individual) in worlds such as those of Internet banking and crime prevention (http://www.scienceeurope.org/uploads/Public%20documents%20and%20speeches/ScienceEuropeMedicalPaper.pdf; but see Gymrek et al. 2012). Following the CERN model, asking for scientists' data in return for giving them access to many other databases should be a huge incentive, especially since it will accelerate the process of scientific discovery by increasing the efficiency of data usage. The acceptability of such systems will depend on their ability to avoid displacement and corruption of source data, which is already a practical possibility (Alagiannis et al. 2012; Fig. 2). The advantage to society is that taxpayers will contribute to medical research at no extra cost while benefiting from its fruits. In other words, every datum collected in the course of standard medical care will also serve to promote medical and health research based on big data (Marx 2013).

Disease Signatures

Initially, the strategy is to federate data through a dedicated, connected infrastructure and then to integrate them appropriately so that they can be mined for answers to specific questions. In all cases the results will relate to groups and not to individuals, so guaranteeing an appropriate and proportionate degree of privacy.

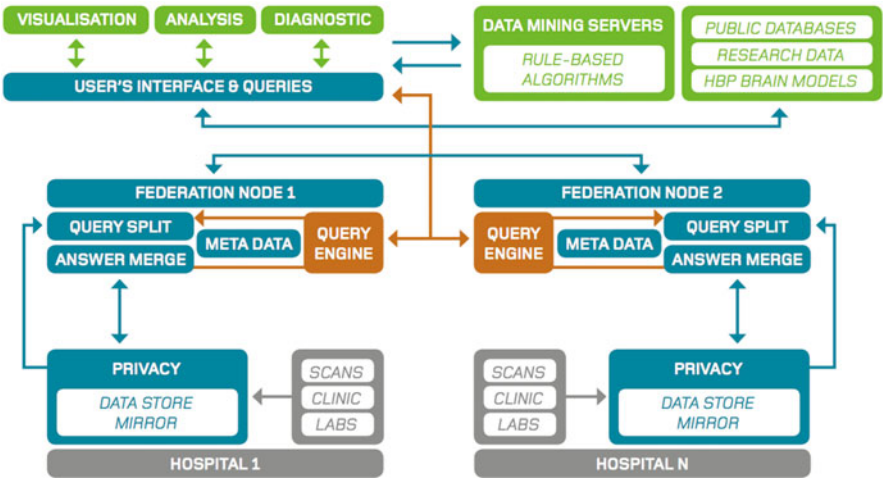

Fig. 2 Schematic describing the clinical neurosciences big data infra-structure. In the context of the Human Brain Project, research will be undertaken based on distributed processing of medical informatics infrastructures. The Medical Informatics Platform will provide a software framework that allows researchers to query clinical data stored on hospital and laboratory servers, without moving the data from the servers where they reside and without disproportionately compromising patient privacy (in situ querying). Tools and data queries will be made available to a participating community via a web-based technology platform adapted for neuroscientific, clinical, genetic, epidemiological and pharmacological users. The information made available will include brain scans of various types, data from electrophysiology, electroencephalography and genotyping, metabolic, biochemical and hematological profiles and also data from validated clinical instruments. Tools will be designed to aggregate data for analysis by state of the art high-performance computing that automatically provides a basic descriptive statistical overview as well as advanced machine learning and discovery tools

"The Human Brain Project," awarded one billion euros in a European Commission Flagship of Enterprise and Technology competition in 2013, seeks to use this strategy in its medical informatics division (http://www.humanbrainproject.eu/#). One type of question will involve identifying groups of patients who show identical patterns of biological abnormality based on the results of clinical investigation. These patterns, called "disease signatures," will comprise sets of causative features including clinical findings, results of validated questionnaires of mood and emotion, brain images of various types, electrophysiological recordings, blood tests, genotypic characteristics, and protein composition of cerebrospinal fluid or blood. To obtain maximal differentiation and sensitive discrimination between different diseases, the strategy will be to use data from as wide and inclusive a range of brain diseases (both neurological and psychiatric) as possible. This approach runs directly counter to standard techniques of epidemiology based on tightly defined syndromes or single characteristics, such as a unique pattern of single nucleotide polymorphisms or protein expression, by seeking to understand and resolve the one syndrome—multiple mutations and one mutation—multiple phenotypes problems. The disease space, sampled in multiple dimensions, each of which is described by a

specific vector of biological variables, will provide a new diagnostic nosology that is in principle quantitative and expressed by a complete, exclusive set of characteristic clinical features and results.

In the context of a medical consultation, a doctor might take a set of measurements and order a set of tests to provide a results vector, which can be presented to a database for matching to disease type, a clear step towards personalized medicine. Biologically characterized diagnostic criteria will facilitate drug trials in that diagnostic ambiguities in control and disease cohorts will be drastically attenuated, leading to small groups with reduced error variances and adequate power for drug discovery in humans. In dementia, as mentioned earlier, the error in AD diagnosis approaches 30 %. Certain aged normal people have a degree of AD-related pathological change, which is compensated for at the behavioral or cognitive level. It is claimed that 39 % of elderly subjects supposed to be normal show AD pathology post mortem (Schöls et al. 2012). Twenty percent of 80-year-old adults have some form of recognizable cognitive decline, so the error variance in currently constituted normal control groups may also be substantial. Clinical trials with groups that are as inhomogeneous as these are likely to fail, even with specifically targeted drugs. A search for preclinical abnormality in populations may lead to a definition of types of "normality" in large enough data sets, and the dementias may become more usefully defined by shared clinical and biological characteristics.

Data Mining and Medical Data

One data mining tool—Hypercube©—has already been used in medical research (Loucoubar et al. 2011). We have preliminary data with this algorithm on a set of 200 patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.three-city-study.com/les-publications-scientifiques-de-l-etude-des-trois-cites.php) and also from a subset of 500 elderly subjects from the 3 Cities study (http://adni.loni.ucla.edu) and associated image-genetics-clinical-psychology cohorts followed in France for 10 years. Our analyses are somewhat flawed, in that the entire disease space is not sampled and the numbers of patients are pitifully small (though we now have over 6000 donated datasets from the same sources and from the pharmaceutical company Sanofi-Aventis), but encouraging patterns have emerged. Of the subjects in the first dataset, 199 of 200 fell into six distinct subgroups on the basis of "disease signature." In the second, where substantial genotyping data were also available, separate, normal-aged groups can be distinguished from a number of groups associated with cognitive decline. The largest of the latter includes APP and ApoE4 in its "disease signature." Of great interest will be secondary phenotyping, returning to groups of patients with the same "disease signature" to identify specific clinical characteristics or variability in them with factors such as age, which will give further insight into how brain diseases manifest (Fig. 3).

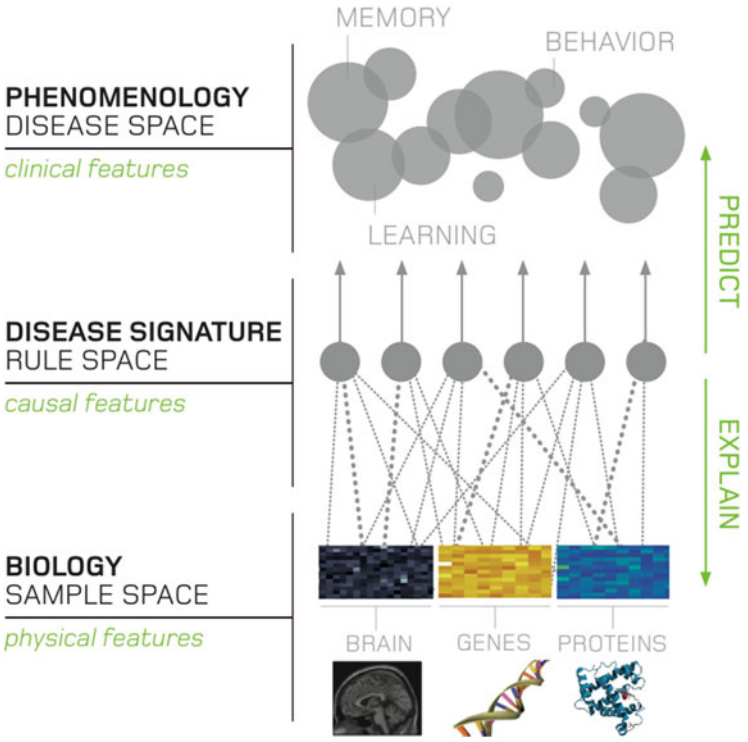

Fig. 3 Theoretical schema describing the relationship between different levels of description and the role of the disease signature in relating biology to phenomenology. The biological signatures of diseases are deterministic mathematical constructs that aim to describe both variability at the phenomenological level (clinical features with symptoms and syndromes) and at the biological level (genetic, proteomic, etc.). The key property of a biological signature of disease is that it accounts for the fact that a symptom of brain dysfunction can be due to many biological causes (one-to-many symptom mapping) and that a biological cause can present with many symptoms (many-to-one symptom mapping). In reality, the situation is often one of many-to-many mappings between symptoms and biological causes. With advanced computing power, nearly exhaustive searches of a data space can be performed to identify sets of rules that describe homogeneous populations, to explain their biological data and to predict the pattern of symptoms

Human Brain Project

The Human Brain Project has, in addition to a medical informatics division, a basic neuroscience component that is charged with creating an in silico blueprint (model) of the normal human brain. Replacement of normal biological characteristics in such a model by disease-associated values should, if correct, give an idea after propagation through the model of what associated functional or structural changes to expect. Likewise, modifications of parameters induced by a neuromodulator or other factor should provide ideas about the spectrum of both desired and undesired effects of any such medication (Harpaz et al. 2013). It may be worth enlarging this

perspective to system-based approaches, too (Zhang et al. 2013). In a real sense the normal brain simulation program and the medical informatics effort will serve to test each other in a cycle of repeated virtuous iteration until adequate accuracy can be achieved for medical practice.

Europe has provided funds for a major coordinated effort in this field, supported by leading edge computer science and technology, which has its own agenda of using knowledge about human brain organization to inspire novel chip and computer architectures. The aim is to move on from von Neumann digital binary machines to neuromorphic probabilistic architectures that are much more energy-efficient (Pfeil et al. 2013; Indiveri et al. 2011). The vision described here is broad but practical. Its implementation will demand new competencies in medical researchers and doctors, greater cross-disciplinary collaboration (along the lines pioneered by physicists in CERN) and major changes in culture and practice.

Clinical Neuroscience-Related Big Data Initiatives

The scientific world is taking on the challenge faced by clinical neuroscience to create a culture and foster competences that will be needed for the effective use of big data research (see Box 1 for more details). Examples include BioMedBridges, a joint effort by ten European biomedical sciences research infrastructures in which the project partners will develop a shared e-infrastructure—the technical bridges—to allow interoperability between data and services in the biological, medical, translational and clinical domains; One Mind, which has the vision of a technology-enabled data-sharing community focused on psychiatric disease and brain injury, brought together through a federated data platform; the Allen Brain Atlas, a growing collection of online public resources integrating extensive gene expression and neuroanatomical data, including that of humans, complete with a novel suite of search and viewing tools; ELIXIR, which unites Europe's leading life science organizations in managing and safeguarding the massive amounts of data being generated every day by publicly funded research; and ENIGMA (Enhancing Neuro-Imaging Genetics through Meta-Analysis), which brings together researchers in imaging genomics to understand brain structure and function based on MRI, DTI, fMRI and GWAS data.

Box 1. Big data initiatives: a selection

Allen Brain Atlas—http://www.brain-map.org
BioMed Bridges—http://www.biomedbridges.eu
One Mind—http://1mind4research.org/about-one-mind
Elixir—http://www.elixir-europe.org
Enigma—http://enigma.loni.ucla.edu/about/
European Bioinformatics Institute—http://www.ebi.ac.uk
International Neuroinformatics Coordinating Facility—http://www.incf.org/about
Machine Learning—http://en.wikipedia.org/wiki/Machine_learning
Data mining—http://en.wikipedia.org/wiki/Data_mining

Changing the Culture

Far-seeing higher educational establishments such as the EPFL have been developing strategies of recruitment and faculty development that bring engineering and ICT together with life and clinical sciences in preparation for such a revolution.

The public will need to be convinced of the privacy issues, and researchers will need to acknowledge that it is ideas and not just data that generate Nobel Prize-winning work. Finally, politicians and industrialists will need to be convinced that there are substantial efficiency savings to be made by preventing the endless repetition of underpowered studies with unrepeatable results that characterize much of present-day life science. They will presumably be open to exploiting the added value that federating data offers at no extra cost and to the business opportunities that arise from developing, installing and maintaining local infrastructures to feed big data-based medical and health sciences research on a global scale (Hood and Friend 2011).

Acknowledgments This work benefited from funding by the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 604102 (Human Brain Project).

References

Alagiannis I, Borovica R, Branco M, Idreos S, Ailamaki A (2012) NoDB: efficient query execution on raw data files. In: ACM SIGMOD international conference on management of data, ACM, 978-1-4503-1247-9/12/05

Beach TG, Monsell SE, Phillips LE, Kukull W (2012) Accuracy of the clinical diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. J Neuropathol Exp Neurol 71:266–273

Brabham DC (2008) Crowdsourcing as a model for problem solving: an introduction and cases. Convergence Int J Res New Media Technol 14:75–90

Brownstein JS, Freifeld CC, Reis BY, Mandl KD (2008) Surveillance sans frontieres: internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med 5:e151. doi:10.1371/journal.pmed.0050151

Chiang G-T, Clapham P, Qi G, Sale K, Coates G (2011) Implementing a genomic data management system using iRODS. BMC Bioinformatics 12:361

CMS Collaboration, Chatrchyan S, Khachatryan V, Sirunyan AM, Tumasyan A, Adam W, Aguilo E, Bergauer T, Dragicevic M, Erö J, Fabjan C, Friedl M, Frühwirth R, Ghete VM, Hammer J, Hoch M, Hörmann N, Hrubec J, Jeitler M, Kiesenhofer W, Knünz V, Kramme M, Krätschmer I, Liko D, Majerotto W, Mikulec I, Pernicka M, Rahbaran B, Rohringer C, Rohringer H, Schöfbeck R, Strauss J (2012) Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys Lett B 716:30–61

Cross-Disorder Group of the Psychiatric Genomics Consortium (2004) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 381:1371–1379

Goldstein DB (2009) Common genetic variation and human traits. N Engl J Med 360:1696–1698

Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2012) Identifying personal genomes by surname inference. Science 339:321–324

Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C (2013) Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. doi:10.1038/clpt.2013.125

Hay A, Tansley S, Tolle K (2009) The fourth paradigm: data-intensive scientific discovery. Microsoft, Redmond, WA. ISBN 978-0-9825442-0-4

Hill SL, Wang Y, Riachi I, Schurmann F, Markram H (2012) Statistical connectivity provides a sufficient foundation for specific functional connectivity in neocortical neural microcircuits. Proc Natl Acad Sci USA 109:E2885–E2894

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. Physiology 117:500–544

Hood L, Friend SH (2011) Relevance of network hierarchy in cancer drug-target selection. Nat Rev Clin Oncol 8:184–187

Indiveri G, Linares-Barranco B, Hamilton TJ, van Schaik A, Etienne-Cummings R, Delbruck T, Liu S-C, Dudek P, Häfliger P, Renaud S, Schemmel J, Cauwenberghss G, Arthur J, Hynna K, Folowosele F, Saighi S, Serrano-Gotarredona T, Wijekoon J, Wang Y, Boahen K (2011) Neuromorphic silicon neuron circuits. Front Neurosci 5:73. doi:10.3389/fnins.2011.00073

Kertesz A, McMonagle P, Blair M, Davidson W, Munoz DG (2005) The evolution and pathology of frontotemporal dementia. Brain 128:1996–2005

Khazen G, Hill SL, Schuermann F, Markram H (2012) Combinatorial expression rules of ion channel genes in juvenile rat (Rattus norvegicus) neocortical neurons. PLoS One 7:e34786. doi:10.1371/journal.pone.0034786

Loucoubar C, Paul R, Huret A, Tall A, Sokhna C, Trape J-F, Ly AB, Faye J, Badiane A, Diakhaby G, Sarr FD, Diop A, Sakuntabhai A, Bureau J-F (2011) An exhaustive, non-Euclidean, non-parametric data mining tool for unraveling the complexity of biological systems—novel insights into malaria. PLoS One 6:e24085. doi:10.1371/journal.pone.0024085

Loucoubar C, Grange L, Paul R, Huret A, Talll A, Telle O, Roussilhon C, Faye J, Diene-Sarr F, Trape JF, Mercereau-Puijalon O, Sakuntabhai A, Bureau JF (2013) High number of previous Plasmodium falciparum clinical episodes increases risk of future episodes in a sub-group of individuals. PLoS One 8:e55666. doi:10.1371/journal.pone.0055666

Marks P (2007) Massive science experiments pose data storage problems. New Sci 196:28–29

Marx V (2013) The big challenges of big data. Nature 498:255–260

Peck C, Markram H (2008) Identifying, tabulating, and analyzing contacts between branched neuron morphologies. IBM J Res Dev 52:43–55

Pfeil T, Grubl A, Jeltsch S, Muller E, Muller P, Petrovici MA, Schmuker M, Bruderle D, Schemmel J, Meier K (2013) Six networks on a universal neuromorphic computing substrate. Front Neurosci 7:11. doi:10.3389/fnins.2013.00011

Rajasekar A, Moore R, Hou CY, Lee CA, de Torcy A, Wan M, Schroeder W, Chen SY, Gilbert L, Tooby P, Zhu B (2010) iRODS primer: integrated rule-oriented data system. Synthesis lectures on information concepts, retrieval, and services. Morgan & Claypool, San Rafael, 143p

Ramaswamy S, Hill SL, King JG, Schurmann F, Wang Y, Markram H (2012) Intrinsic morphological diversity of thick-tufted layer 5 pyramidal neurons ensures robust and invariant properties of in silico synaptic connections. J Physiol (Lond) 590:737–752. doi:10.1113/jphysiol.2011.219576

Schöls L, Bauer P, Schmidt T, Schulte T, Riess O (2012) Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. Lancet Neurol 3:291–304

Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, Zhang C, Xie T, Tran L, Dobrin R, Fluder E, Clurman B, Melquist S, Narayanan M, Suver C, Shah H, Mahajan M, Gillis T, Mysore J, MacDonald ME, Lamb JR, Bennett DAB, Molony C, Stone DJ, Gudnason V, Myers AJ, Schadt EA, Neumann H, Zhu J, Emilsson V (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153:707–720