



REVIEW

How Big Data and High-performance Computing Drive Brain Science



Shanyu Chen^{1,2,a}, Zhipeng He^{1,2,b}, Xinyin Han^{1,2,c}, Xiaoyu He^{1,2,d}, Ruilin Li^{1,2,e},
Haidong Zhu^{1,2,f}, Dan Zhao^{1,2,g}, Chuangchuang Dai^{1,2,h}, Yu Zhang^{1,2,i},
Zhonghua Lu^{1,j}, Xuebin Chi^{1,2,4,k}, Beifang Niu^{1,2,3,*,l}

¹ Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China

² University of Chinese Academy of Sciences, Beijing 100190, China

³ Guizhou University School of Medicine, Guiyang 550025, China

⁴ Center of Scientific Computing Applications & Research, Chinese Academy of Sciences, Beijing 100190, China

Received 14 November 2018; revised 12 September 2019; accepted 29 September 2019

Available online 2 December 2019

Handled by Xiangdong Fang

KEYWORDS

Brain science;
Big data;
High-performance computing;
Brain connectomes;
Deep learning

Abstract Brain science accelerates the study of intelligence and behavior, contributes fundamental insights into human cognition, and offers prospective treatments for brain disease. Faced with the challenges posed by imaging technologies and **deep learning** computational models, **big data** and **high-performance computing** (HPC) play essential roles in studying brain function, brain diseases, and large-scale brain models or connectomes. We review the driving forces behind big data and HPC methods applied to brain science, including deep learning, powerful data analysis capabilities, and computational performance solutions, each of which can be used to improve diagnostic accuracy and research output. This work reinforces predictions that big data and HPC will continue to

* Corresponding author.

E-mail: niubf@cnic.cn (Niu B).

^a ORCID: 0000-0002-1163-7635.

^b ORCID: 0000-0003-0740-4959.

^c ORCID: 0000-0001-8871-4689.

^d ORCID: 0000-0002-5481-2763.

^e ORCID: 0000-0001-9593-1235.

^f ORCID: 0000-0001-7038-1070.

^g ORCID: 0000-0002-5446-0307.

^h ORCID: 0000-0002-1461-4826.

ⁱ ORCID: 0000-0001-6261-8568.

^j ORCID: 0000-0002-1554-8429.

^k ORCID: 0000-0002-0150-5815.

^l ORCID: 0000-0002-7448-7793.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2019.09.003>

1672-0229 © 2019 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

improve brain science by making ultrahigh-performance analysis possible, by improving data standardization and sharing, and by providing new neuromorphic insights.

Introduction

A human brain has about 100 billion neurons [1]. Each neuron must pass through approximately 10^{15} neurons to communicate with another neuron [2]. The brain is responsible for human intelligence. Compromised functioning of the brain resulting from brain disease causes more than 6.3% of the global disease burden in terms of disability-adjusted life years. Moreover, the World Health Organization (WHO) noted that the proportion of brain disease in the global disease burden is projected to increase by 12% from 2005 to 2030 [3]. In 2010, around \$2.5 trillion spent on research on brain disease, an amount that is estimated to increase to \$6 trillion by 2030 [4]. According to statistics from the WHO, brain diseases account for 12% of the world's total deaths [5]. Also, in low-

and middle-income countries and high-income countries, encephalopathy accounts for 16.8% and 13.2% of the total death toll, respectively [5].

Driven by the rising incidence of brain diseases, brain research is important for understanding brain function mechanisms, promoting the diagnosis and treatment of brain diseases, and improving the development of brain-like intelligence. In 2013 [6], the European Union announced the Human Brain Project (HBP) to strengthen neuroscience research and gain a comprehensive understanding of brain function through worldwide research [6]. Moreover, the USA, China, and many other countries and organizations had also focused on and invested in brain-research projects (Table 1). The governments of these countries had attempted to build platforms for studying the human brain using neu-

Table 1 National brain projects

Project	Platform	Sub-project	Database	Supercomputer center	URL
Human Brain Project	Neuroinformatics; brain simulation; high-performance analytics and supercomputing; medical informatics; neuromorphic computing; neurorobotics; ethics and society	Organization of human and rodent brains; systems and cognitive neuroscience; theoretical neuroscience	Data support and curation laboratories; neuroscience simulation and data laboratories	Barcelona Supercomputing Centre; Swiss National Supercomputing Centre; Jülich Supercomputing Centre; Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca	www.humanbrainproject.eu/en/
Allen Brain Institute	/	Brain science (mouse brain atlas, spinal cord atlas, developing mouse brain atlas, human brain atlas, and mouse brain connectivity atlas); cell science; frontier of bioscience	Allen cell-type database	/	www.alleninstitute.org
US Brain Initiative	/	Advanced neurotechnology; facilitation of dynamic imaging to explore brain functionality; linking function and behavior; advanced consumer applications	/	/	www.braininitiative.nih.gov
Brain Canada Foundation	MIRI; PSG; training awards	Knowledge translation	/	/	https://braincanada.ca/
Japan Brain/MINDS project	Brain/MINDS research platform database; Brain/MINDS publication platform database	Structural and functional mapping of the marmoset brain; development of innovative neurotechnology for brain mapping; human-brain mapping and clinical research; advanced technology and application development	Marmoset Gene Atlas (<i>in situ</i> hybridization database); Marmoset genome database; portal for marmoset neural circuit data for open research sources	/	brainminds.jp/en/

Note: Brain/MINDS, Brain Mapping by Integrated Neuro technologies for Disease Studies; MIRI, Multi-Investigator Research Initiative; PSG, platform support grants.

roinformatics, brain simulation, and brain-tissue science to create a validation model that runs on supercomputers [7].

The development of big data technology and HPC has contributed to insights gained from these brain-research projects. Big data methods have improved the details of brain scans in several ways, thereby laying the foundation for the generation of new knowledge that can drive understanding of the human brain even further. HPC methods have advanced data storage, computational accuracy, and computational speed, and thereby assist in the processing of vast and very complex data sets. The development and popularization of brain science, big data, and HPC methods are shown in **Figure 1**. Data, methods, and computing power are being continuously added to brain science research. Particularly after 2006, brain science combined with big data and deep learning has become a research hotspot. Subsequently, the support of neural networks and HPC for brain science research has also been enhanced considerably. An overall increase in interdependence has been observed from 2000 to 2018.

Combination and revolution

Brain science is based primarily on biological insights and data-driven, bottom-up experimental research. **Figure 2** shows the current research fields of brain science: brain function/mechanisms, diagnosis of brain disease, brain-like intelligence, and refinement of sub-areas. These seemingly different research areas are interrelated. In biology, “classical reductionism” suggests that each entity comprises smaller parts. That is, the aging process, decision-making principles, pathogenesis, or brain cognition are the “macroscopic” consequences of “microscopic” behavior and reaction in the brain. In this sense, “microscopic” denotes less vision and more knowledge and, most importantly, big data. In addition, these connections between the microscopic parts are usually linear [8], and there are approximately 10^{15} such linear connections in the brain [9]. Thus, exploring microscopic characteristics in such a large number of structures and complex connections results in extensive use of computational resources. Moreover, as brain science advances, the resolution becomes infinite.

The cycle of acquisition of functional magnetic resonance imaging (fMRI) data has accelerated from 4 s to 1 s in speed, and the resolution has increased from 5 mm^3 to 1 mm^3 [10]. Faster speeds and more detailed visions bring challenges while bringing amazing scientific discoveries. For example, genome-wide association studies combined with functional and structural brain-imaging phenotypes to verify genes for iron transport and storage are associated with the magnetic susceptibility of subcortical brain tissue [11]. However, a single genome contains around 180 GB of uncompressed data that is equivalent to 30 copies of 3 billion bases, and 100 GB of compressed information needs to be retained over time [12]. Therefore, the ability to process data analysis and workflows, meet data storage/simulation requirements, and even break the limits of computational speed are critical factors in brain science. Here, big data and HPC bear the brunt of these challenges. Advances in data storage and mining technologies allow users to retain increasing amounts of data directly or indirectly and to analyze such data for valuable discoveries [13]. HPC provides a high-speed computing environment that meets high-throughput and multitasking computing features, including

the use of clusters of multiple processors or computers as part of a single machine [14].

In recent decades, research funding agencies and scientists have placed great importance on the use of big data and HPC techniques in brain science. With regard to big data, the HBP's fifth sub-project is to build a Neuroinformatics Computing Platform (NCP) for integrating multidimensional brain data (*e.g.*, molecules, genes, cells, networks) and provide powerful analysis of brain data to simplify models of actual neuronal circuits [15]. The Allen Human Brain Atlas is a unique multi-modal atlas that maps gene expression across adult brains [16], including magnetic resonance imaging (MRI), diffusion tensor imaging, histology, and gene-expression data [17]. The BrainSpan Atlas of the developing human brain is a unique resource for exploring the development of the human brain. It comprises in situ hybridization data, RNA sequencing, microarray data, as well as broad and detailed anatomical analysis of gene expression during human brain development [17]. The seventh sub-project of the HBP has been to establish a platform for high-performance analytics and computing designed to provide the HBP and the neuroscience community with HPC systems to meet their particular [18]. Currently, the HBP has four tier-0 supercomputer centers: Barcelona Supercomputing Centre, Cineca, Centro Svizzero di Calcolo Scientifico and Jülich Supercomputing Centre (JSC) [6]. The supercomputer JUR-OPA at the JSC has developed an ultrahigh-resolution three-dimensional (3D) human brain model called BigBrain [19]. An account of research in brain science combined with other fields is shown in **Figure 3**. The amount of literature on brain science that encompasses the other four fields detailed in **Figure 3** increased from 2000 to 2018. The correlation coefficients in **Figure 3** demonstrate that these combinations have stimulated productivity in brain science.

Research advances enabled by big data and HPC

Function and mechanisms within the brain

Brain activity consists of dynamic sets of input sensations and spontaneous responses [20]. However, the brain is very active even in the absence of explicit inputs or outputs [21]. Researchers often create brain-simulation models (*e.g.*, simulations of brain molecules and brain cells) or use fMRI data of living human brains to gain insight into brain structure. For example, one MRI time-course of 512 echo-planar images in a resting human brain obtained every 250 milliseconds showed fluctuations of physiologic origin in signal intensity in each pixel. Such data can reveal functional connections in the motor cortex of the resting human brain [22]. Furthermore, recent functional imaging studies have revealed co-activation in a distributed network of cortical regions that characterize the quiescent state or “default mode” of the human brain [23]. Subtle imaging provides a large data set that requires a large storage space and capability to undertake high-resolution analysis [24]. Applications that mimic brain structures could simulate about 100 billion neurons at the molecular level, and each requires 0.1 megabyte (MB) to 10 terabytes (TB) of memory [25]. The fMRI divides the brain into tens of thousands of voxels and then images the entire brain continuously with high time resolution. One scan is used as a processing

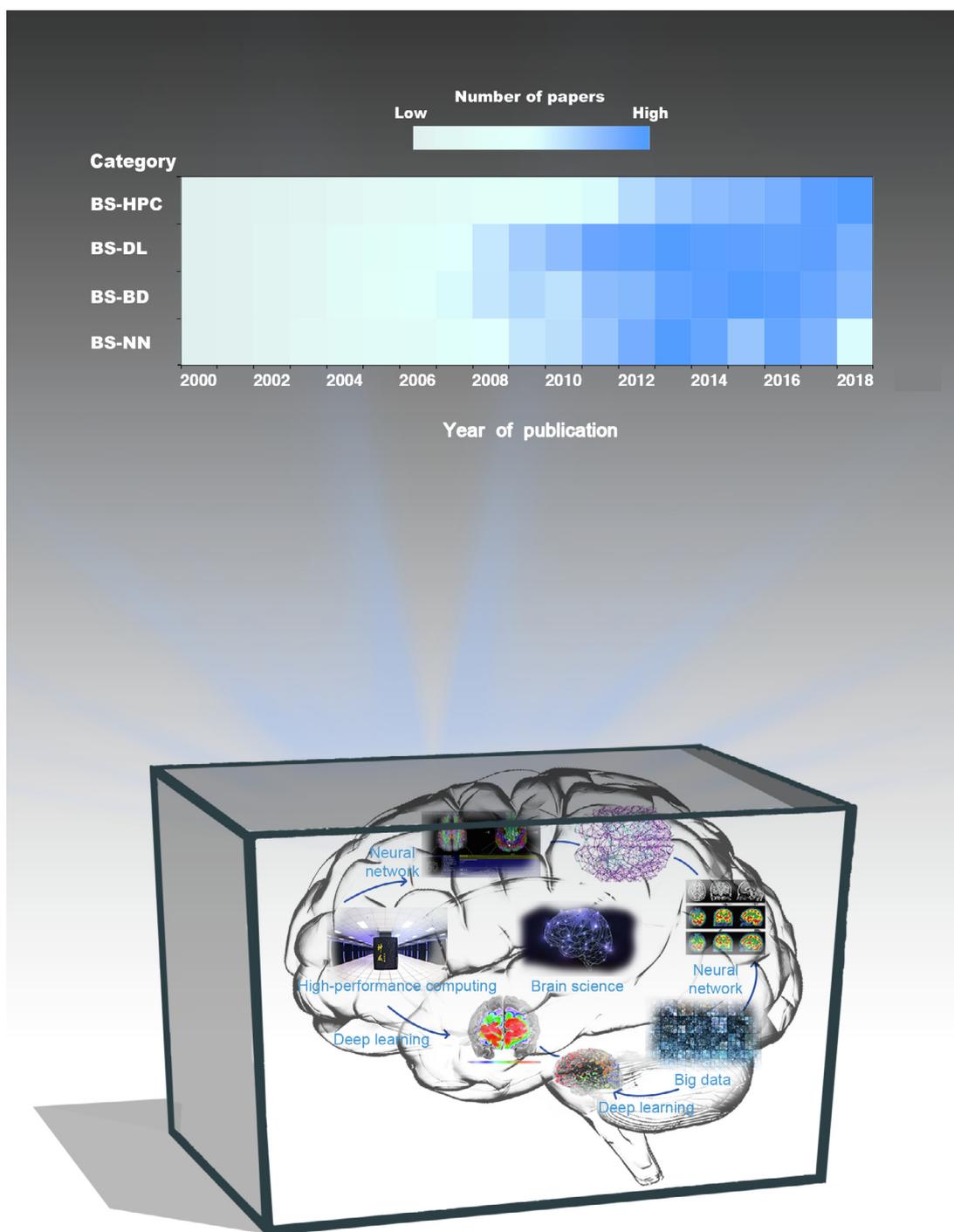


Figure 1 Research overview of big data and HPC methods in brain scienceA.

The heatmap shows the changes in the number of articles published annually from 2000 to 2018 in four research directions of brain science: brain science with HPC; brain science with deep learning; brain science with big data; and brain science with neural networks. Articles in brain science with deep learning, brain science with big data, and brain science with neural networks reached their highest numbers in 2013, whereas articles in brain science with HPC reached its highest number in 2018. All articles were retrieved by searching using keywords “brain science, HPC” (BS-HPC), “brain science, deep learning” (BS-DL), “brain science, big data” (BS-BD), or “brain science, neural network” (BS-NN) in Google Scholar in September 2019. **B.** Combinations between brain science and big data or HPC methods. Big data provide a wealth of knowledge and data, from which neural networks and deep learning methods can extract features that represent brain functions, mechanisms, or diseases. Big data can also be used to build computational models. HPC provides storage space and formidable computing power for the study of brain science. HPC, high-performance computing.

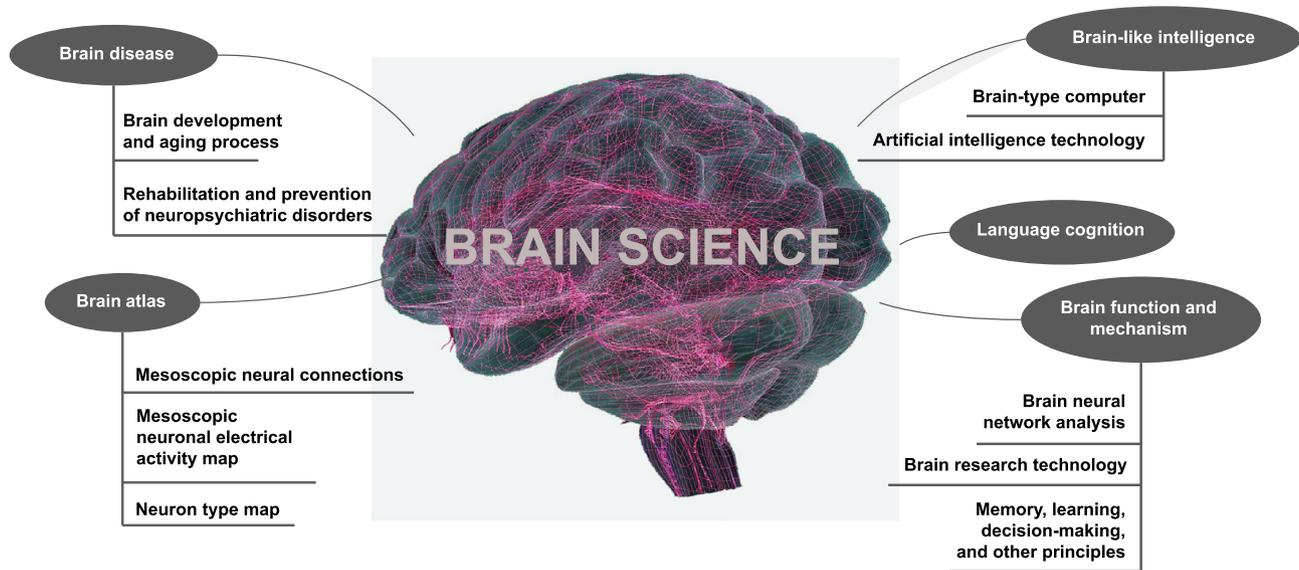


Figure 2 General classification of research activities in brain science
This figure shows the main research directions in contemporary brain science.

unit, so the entire brain generates a very large amount of high-dimensional data [26]. Although extant computational models often carry out feature combinations (*e.g.*, Suk et al. [27]) in studies on brain function and brain mechanisms and even multimodal fusions of features (Liu et al. [28]) to reduce “dimensional disasters”, their capacity to leverage computational power to analyze highly multidimensional data is limited.

Deep learning computational models use eigenvector sets to represent biological information. This strategy enables computational models that consist of multiple processing layers to learn data representations with multiple levels of abstraction [29]. Li et al. [30] used image-analysis technology combined with a Statistical Parametric Mapping (SPM) and Voxel-Based Morphometry (VBM) toolbox to visualize cerebrospinal fluid and other brain components. Besides, Vidaurre et al. [31] proposed a framework that used a hidden Markov model to infer a consistent and interpretable dynamic brain network in different data sets. Spitzer et al. [32] proposed a model based on a convolution neural network (CNN) to predict 13 regions of the human visual system. Even though deep learning provides important insights, the deep neural network is a complex hierarchical structure similar to a biological neural network. Each layer consists of several “artificial neurons”; the more layers in the neural network, the better the insights. However, mapping relationships from input layers to output layers often reveals non-linear relationships, and some deep learning models may even extend to brain mechanisms at the 3D level (*e.g.*, Payan et al. [33], Hosseini-Asl [34]).

Whether using data or computational models, it is clear that the exploration of brain function and brain mechanisms demands a great deal of storage capacity and powerful computing capabilities. Big data and HPC approaches are, therefore, necessary in all brain science studies that generate such data. This is especially true if supercomputers are used to model brain function. For instance, Apache Spark with autoencoder for fMRI [35], and JUROPA in JSC have both implemented ultrahigh-resolution 3D models of the human brain [36], thereby greatly improving computational perfor-

mance and the speed of data processing. In addition, researchers in China have used the Tianhe-1A supercomputer for NEST (NEural Simulation Tool), which showed that a single compute node implements a neural-network simulation of 7.3×10^4 neurons and implementation of 5.6×10^6 neurons on 128 compute nodes [12]. Use of an NVIDIA graphic processing unit (GPU) accelerates the Groningen Machine for Chemical Simulation (GROMACS) [37] by a factor of 3–4, thereby reducing the time spent on simulations of molecular dynamics in the brain from days to hours. HPC helps virtual epilepsy (VEP) brain model to explore system parameter space, fit and validate brain model, thus promoting large-scale brain models to encourage the development of personalized treatment and intervention strategies [38]. HPC methods not only allow new studies on brain science to overcome traditional hardware and software constraints, they also foster a new research dynamic whereby studies of intrinsically biological problems can be studied by relying on HPC methods only. The advantage of this development is that HPC methods enable in-depth quantification that was impossible previously [19]. For instance, the Salk Institute has established a 3D molecular computational model to understand the human brain through neural communication processes in the ciliary ganglion of the chicken, which model adjusts data parameters with the aid of a central HPC system in the San Diego Supercomputer Center.

Research on brain disease

The diagnosis and treatment of brain diseases—especially Alzheimer’s disease (AD) and Parkinson’s disease (PD)—is an important focus of clinical research into brain disease in many countries. AD is a progressive neurologic disease that presents as memory and cognitive dysfunctions [39]. Major clinical manifestations include memory impairment, aphasia, agnosia, altered personality, and behavior changes [40]. The failure rate of treatment in the human brain in clinical trials for AD reached 99.6% from 2002 to 2013 [41]. In 2010, the total num-

autoencoder to understand the features of 2D patches through MRI. Yang and co-workers [49] used MRI data from the Open Access Series of Imaging Studies and Alzheimer's Disease Neuroimaging Initiative (ADNI) to develop a classification of AD. Machine learning and deep learning methods for disease diagnosis (*e.g.*, support vector machine (SVM) [50], Gaussian kernel SVM [51], enhanced logistic regression model [52], deep belief networks [53]) have also been found to be indispensable in improving efficiency. Big data combined with deep learning models not only provide increased information generation and analytical efficiency but also improve accuracy. Experimental accuracy with respect to AD, mild cognitive impairment (MCI), and MCI-converter diagnoses have reached 95.90% in the ADNI data set [25]. Moreover, Sarraf et al. [54] used fMRI data with CNN and LeNet-5 to diagnose AD, and achieved an accuracy of 96.86% using test data. Payan [28] used 3D-CNN to achieve accuracy $\leq 95.39\%$ when classifying patients as AD or healthy controls.

While big data and deep learning methods have yielded notable benefits, they have also introduced new computational challenges that can be resolved only by HPC methods. The machine from Hewlett-Packard Development Company contains 160 TB of single memory (equivalent to 160 million books of active memory). This helps the German Center for Neurodegenerative Diseases speed up the genomics pipeline by nine times, meaning that the process would have taken only approximately 36 seconds instead of the original 25 min [22]. Moreover, a research team at the Friedrich-Alexander-Universität (Germany) used a supercomputer from the Regionales Rechenzentrum Erlangen to carry out all-atom molecular dynamics simulations in an explicit solvent of 0.7 μs in total on five A β -42 oligomers (monomers through to pentamers) to reveal A β peptides, an important factor in AD [55]. All in all, these examples show the indispensable nature of HPC approaches.

Brain models and connectomes

The human brain is a complex multi-scale structure in space and time, and produces fine molecular, cellular and neuronal phenomena [56,57]. Neuroimaging can provide brain images with high temporal and spatial resolution, but dynamic information for the brain is lacking. Therefore, in brain science research, simulation tools, brain models, and connectomes have been developed gradually and built to provide simulation information of neurons, brain structures, and networks. Simulation tools focus on individual neurons and the corresponding models of ion channels [58]. For example, the GENeral Network SIMulation System (GENESIS) was designed to simulate neural networks using standard and flexible methods to obtain detailed and realistic models [59]. A model of the human brain is a "reference brain" that provides important biological details. It outlines the spatial framework and brain composition from a macroscopic perspective and helps researchers extract and analyze microscopic data from molecular processes to various behaviors for modeling and simulation. The brain connectome is the "Google™ Maps" for brain models. It provides precise human brain coordinates and helps researchers transform detailed neural connections with human brain cognition and behavior. Connectomes map elements to human brain networks dynamically [60], where circuit abnormalities

presumably reflect a complex interplay between genes and the environment [61]. Large-scale models of the human brain and connectomes not only provide basic insights of brain structure [62] but also serve as biomarkers of brain diseases to help researchers explain diseases such as AD and PD [63–67] and even help researchers understand the sex-based differences in human behavior [68].

Vast human brain structures and high-resolution imaging technology determine the essence of brain models and connectomes to be a big data set. The brain model of the HBP consists of 100 neocortical columns [69]. Defense Advanced Research Projects Agency (DARPA)'s Synapse project 500 billion neurons [70–72]. Also, Izhikevich et al. published a detailed large-scale thalamocortical model that simulates one million multicompartmental spiking neurons [73]. On a microscopic scale, the number of neurons and synapses contained in brain connectomes is approximately 10^{10} – 10^{11} and 10^{14} – 10^{15} [74]. At the macroscopic scale, the cortical hypothalamic junction contains hundreds of brain regions and thousands of comprehensive pathways data sets [75]. Currently, the Human Connectome Project (HCP) already has 7-T diffusion magnetic resonance imaging (dMRI) and 3-T resting-state fMRI (R-fMRI) data [76,77,78]. Not only structure but also high-resolution imaging. When building a brain model, an optical microscope sufficient to track a single neuron has a resolution of 0.25–0.5 microns, and an electron microscope capable of displaying synaptic or chemical signals has a resolution of nanometers [79]. Diffusion tensor imaging [80] and four main MRI modes (structural MRI, task fMRI, dMRI, R-fMRI) can be used to measure connectivity in the brain [61] with resolution of 1–3 mm or even smaller [81]. Brain models and connectomes as big data sets provide abundant information and knowledge to drive the development of brain-research programs. The Izhikevich neuron model has influenced more than 3000 academic studies by 2019. Based on the 500-subject release of the HCP, the Budapest Reference Connectome Server v2.0 and v3.0 has generated the common sides of connectomes in 96 and 477 different cortical layers [82,83]. Also, disease research by the HCP applies HCP-style data to people at risk or suffering from brain disease (*e.g.*, anxiety, depression, epilepsy) [84].

The human brain is not only a simple big data set, but also a complex mathematical object. Hence, building models of the human brain and connectomes requires powerful platforms for data storage and processing. To meet the HPC requirements for brain models, the Izhikevich neuron model [85] used the Beowolf Cluster [86] with 60 processors of 3-GHz each, the HBP and Synapse project used the IBM Blue Gene supercomputer, and Spaun used eight Core Xeon processors (2.53 GHz) [71]. The HCP infrastructure uses IBM HPCS from the WU Center for High Performance Computing to execute pipelines and user-submitted jobs to meet the high-throughput data-processing requirements of approximately 200,000 inputs and outputs per second [71,87]. In addition, the HCP has established a set of informatics tools, including ConnectomeDB [88] and Connectome Workbench [89], to collect, process, share and visualize high-throughput data [90]. Not only brain models and connectomes, workflows and simulation tools are moving toward high-performance computing and distribution. Simulation tools such as GENESIS [59], NEURON [91], NEST [84], Network and Cache Simulator (NCS) [92], Neosim [93], and SpikeNET [94] have also been extended to support

Table 2 Databases related to brain science

Name	Sample population	Description	Data type	URL
Allen Brain Atlas	Healthy	Brain sections stained to represent development and gene expression	Images	www.brain-map.org
BigBrain	Healthy	3D reconstruction of the complete brain		bigbrain.loris.ca/main.php?test_name=bigbrain
BCP database	Healthy	Dataset from 500 children with typical development between birth and 5 years of age	tfMRI, fMRI, and MRI datasets	https://db.humanconnectome.org
HCP Young Adult Data Releases	Healthy	Including high-quality neuroimaging data in >1100 healthy young adults		www.humanconnectome.org/study/hcp-young-adult
AD Neuroimaging Initiative	Healthy /AD	Structural MRI	MRI datasets	http://adni.loni.usc.edu
OASIS Brains Project	Healthy /AD	Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and AD		www.oasis-brains.org/
BIRN fMRI and MRI data	Healthy /AD	fMRI, MRI, and atlases for human and mouse brains	fMRI and MRI datasets	www.openeuroscience.com/data-collaboration-projects/neuroimaging-databases/
BP neuroimaging database	BP	Meta-analysis and MRI study databases	Descriptive or numerical data	https://sites.google.com/site/bipolarDATABASE/data
HCP database	Healthy /various	Various data from various people	Multimodal imaging data	https://db.humanconnectome.org
ABCD Data Repository	Healthy /various	Data from >10,000 children aged 9–10 years	Images	https://data-archive.nih.gov/abcd
OpenNEURO /OpenfMRI	Healthy /various	Database for analyzing and sharing neuroimaging data	fMRI and MRI datasets	https://openneuro.org/ https://openfmri.org

Note: AD, Alzheimer's disease; BIRN, Biomedical Informatics Research Network; BP, bipolar disorder; MRI, magnetic resonance imaging; fMRI, functional MRI; tfMRI, task-based fMRI; HCP, Human Connectome Project; BCP, Baby Connectome Project; ABCD, Adolescent Brain Cognitive Development; OASIS, Open Access Series of Imaging Studies.

parallel processing systems to improve performance. After parallelization, NEURON achieved almost linear acceleration. It requires an integration time of 9.8 seconds and communication time of 1.3 seconds if running 40,000 realistic cells on 2000 processors on the IBM Blue Gene supercomputer [95]. Also, the BigBrain-based 3D-PLI workflow uses the JUDGE and JUR-OPA supercomputers from the JSC to meet the requirements for data reading, analysis, and calculation [96]. The JUROPA supercomputer owns 26,304 cores, and its Rpeak reaches 308.3 TP/s. With HICANN from Heidelberg University, brain activity can be simulated at 10,000 times the normal speed to compress one day into ten seconds [72]. The HBP estimates

that supercomputers with exaflop computing speed and exabyte computing memory can simulate a complete human brain (1000 times that of the rodent brain) [69]. With the development of ultrahigh-performance computers and computing environments, a model of the whole brain and dynamic brain connectomes will be completed eventually.

Prospects for brain science

Using big data and HPC methods to address problems in high-dimensional brain science research is important. Hence, research programs on brain science (and especially large

Table 3 Supercomputers related to brain science

Name	System model	Cores	Memory (GB)	Processor	Interconnect	Rmax (TFlop/s)	Power (kW)	URL
Sequoia	BlueGene/Q	1,572,864	1,572,864	Power BQC 16C 1.6 GHz	Custom Interconnect	17,173.2	7890.00	www.top500.org/system/177556
K	K computer	705,024	1,410,048	SPARC64 VIIIfx 8C 2 GHz	Custom Interconnect	10,510	12,659.89	www.top500.org/system/177232
JUWELS Module 1	Bull Sequana X1000	114,480	264,096	Xeon Platinum 8168 24C 2.7 GHz	Mellanox EDR InfiniBand/ParTec ParaStation ClusterSuite	6177.73	1361.00	www.top500.org/system/179424
JURECA	T-Platforms V-Class/Dell C6320P	155,150	287,744	Xeon E5-2680v3 12C 2.5 GHz	Infiniband EDR	3782.57	1345.28	www.top500.org/system/178718
Tianhe-1A	NUDT YH MPP	186,368	229,376	Xeon X5670 6C 2.93 GHz	Proprietary	2566	4040.00	www.top500.org/system/10587
JUQUEEN	BlueGene/Q	458,752	458,752	Power BQC 16C 1.6 GHz	Custom Interconnect	5008.86	2301.00	www.top500.org/system/177722
JUROPA	Sun Constellation, NovaScale R422-E2	26,304	80,806	Xeon X5570 4C 2.93 GHz	Infiniband QDR Sun M9/Mellanox/ParTec	274.8	1549.00	www.top500.org/system/176488

government-led programs) are becoming more reliant on supercomputers and large databases. Table 1 introduces the basic concepts of some national brain programs, including commonly used platforms and supercomputers. Table 2 lists several large databases on brain science research. At this stage, governments and funding agencies have resolved economic problems, and several mature HPC environments (and some customized for brain science research) are available. However, three main limitations exist. First, computing resources are scattered. As shown in Table 3, Sequoia, K, JURECA, and Tianhe-1A are among the top-500 supercomputers in the world, and are ranked 13th, 20th, 52th, and 87th, respectively (June 2019). These supercomputers are in the USA, Japan, Germany, and China, respectively. In the near future, we hope that a cooperative high-performance platform, shared high-performance resources, and universal standards for analysis of brain science data can be created. The second limitation is based on storage capacity and computational performance. As shown in Figures 1 and 3, the correlation coefficient between brain science and the HPC is smaller than that of the others, but the research output in these fields has been increasing from 2000 to 2018. In contrast, although the correlation coefficient is larger, the number of published articles in brain science and other fields decreased slightly after 2013. These results could be because brain science using HPC is constantly evolving, but the support provided by HPC is insufficient and limits the further development of brain science with other fields. As one of the most important elements of the brain model, the amount of imaging data of a neuron can reach the petabyte (PB) level readily at a resolution of microns or even nanometers, which is far beyond that in the human genome. The Sequoia supercomputer can simulate 530 billion neurons and 137 trillion synapses, but this is less than 1% of the information-processing capacity of humans [10]. The K supercomputer analyzed approximately 1% of neuron connections in the simulated human brain; this implies that simulation of the entire brain neuron requires 100 times the power of the best supercomputer performance. The JSC

has developed an algorithm that can simulate 100% of neuron connections but, unfortunately, no supercomputer can run this algorithm [97]. Alternatively, a neuromorphic computer that breaks the conventional Von Neumann architecture could be developed [98] or even a combination of traditional supercomputers with neuromorphic computing could be initiated. The third limitation is the lack of randomness and dynamics. To fully grasp the relationship between the human brain and consciousness, emotions, and even thinking, a dynamic and complete brain model is, ultimately, what is needed. The “perfect” brain should be able to simulate all dynamic neural activity (not just static images), support the ability to switch visual imaging at any resolution, and even provide brain coordinates at any scale. This perfection requires not only sufficient storage space and flexible and variable computing environments, but also meticulous process design and impeccable visualization tools. To further meet random and dynamic requirements, these tools should be developed to be distributed and parallel, and they should even be highly portable.

In general, the development of brain science is destined to be inseparable from the support of big data and HPC. Big data, HPC and brain science promote each other and will break all the technical bottlenecks. A detailed and accurate atlas of the brain, and a fully simulated dynamic computer model will be produced through the use of big data and HPC methods, which could be used as an adjunct to the mouse model for developing new drugs associated with brain disease.

Conclusion

Big data and HPC have become indispensable for: (i) exploration of brain function; (ii) determining the mechanisms of brain disease; (iii) building a whole-brain model and dynamic connectomes. Big data provides large databases on knowledge of the brain, such as the Allen Human Atlas, and efficient frameworks for big data analysis, such as Apache Spark. HPC methods use platforms such as the JSC and supercomputers such as Sequoia, and can solve the challenges of compu-

tational performance caused by large data sets and complex models. HPC methods require more storage space but provide increasingly powerful simulation capabilities to reduce run-times for complex simulations from days to hours. Big data combined with deep learning models can increase the diagnostic accuracy of AD to more than 90%. HPC has also transformed biology into a science of deep quantitative analysis, and has made breakthroughs in characterizing neural communications.

Brain science will continue to develop in a more comprehensive, precise, and detailed direction with the support of governments and the scientific community. Over time, big data analysis for brain science will be standardized, HPC will be shared and coordinated, and new, ultrahigh-performance forms of computer will become a universal reality. However, as big data methods generate ever-larger pools of data, ever-faster and more powerful computational methods will be required to analyze them. Therefore, the “one-way ratchet” of this relationship suggests that brain science will become more reliant on big data and HPC methods in the future.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31771466), the National Key R&D Program of China (Grant Nos. 2018YFB0203903, 2016YFC0503607, and 2016YFB0200300), the Transformation Project in Scientific and Technological Achievements of Qinghai, China (Grant No. 2016-SF-127), the Special Project of Informatization of Chinese Academy of Sciences, China (Grant No. XXH13504-08), the Strategic Pilot Science and Technology Project of Chinese Academy of Sciences, China (Grant No. XDA12010000), and the 100-Talents Program of Chinese Academy of Sciences, China (awarded to BN).

References

- [1] Bartheld VSC. Myths and truths about the cellular composition of the human brain: a review of influential concepts. *J Chem Neuroanat* 2017;93:2–15.
- [2] Murre JMJ, Sturdy DPF. The connectivity of the brain: multi-level quantitative analysis. *Biol Cybern* 1995;73:529.
- [3] Organization WH. Annexes and index. In: Campanini B, editor. *Neurological Disorders: Public Health Challenges*. Switzerland: WHO Press; 2006, p. 183–218.
- [4] Marquez PV, Saxena S. Making mental health a global priority. *Cerebrum* 2016;2016:cer-10-6.
- [5] Dua T, Cumbreira M, Mathers C, Saxena S. Global burden of neurological disorders: estimates and projections. In: B C. (ed) *Neurological Disorders: Public Health Challenges*. Switzerland: WHO Press; 2006, p. 27–39.
- [6] Shepherd GM, Mirsky JS, Healy MD, Singer MS, Skoufos E, Hines MS, et al. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. *Trends Neurosci* 1998;21:460–8.
- [7] Markram H, Meier K, Lippert T, Grillner S, Frackowiak R, Dehaene S, et al. Introducing the Human Brain Project. *Procedia Comput Sci* 2011;7:39–42.
- [8] Geerts H, Dacks PA, Devanarayan V, Haas M, Khatchaturian Z, Gordon MF, et al. From big data to smart data in Alzheimer’s disease: the brain health modeling initiative to foster actionable knowledge. *Alzheimers Dement* 2016;12:1014–21.
- [9] Poo MM. Whereto the mega brain projects? *Natl Sci Rev* 2014;1:12–4.
- [10] Liu YD, Hu DW. Computing in the viewpoint of brain research. *Chin J Comput* 2017;40:2148–66 (in Chinese with an English abstract).
- [11] Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 2018;562:210.
- [12] Becker M, Schultze H, Schultze JL. Personalized medicine: the need for exascale data handling. *Schriften des Forschungszentrums Jülich IAS Series* 2018;40:18–9.
- [13] Michael K, Miller KW. Big data: new opportunities and new challenges [guest editors’ introduction]. *Computer* 2013;46:22–4.
- [14] Buyya R. High performance cluster computing: architectures and systems (volume 1). 1st ed. Upper Saddle River: Prentice Hall; 1999.
- [15] Calimera A, Macii E, Poncino M. The Human Brain Project and neuromorphic computing. *Funct Neurol* 2013;28:191–6.
- [16] Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012;489:391–9.
- [17] Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* 2013;41:D996–1008.
- [18] Amunts K. The EU’s Human Brain Project (HBP) Flagship—accelerating brain science discovery and collaboration. *CEUR-WS* 2017;2022:187–8.
- [19] Casanova H, Berman F, Bartol T, Gokcay E, Sejnowski T, Birnbaum A, et al. The virtual instrument: support for grid-enabled mcell simulations. *Int J HighPerform C* 2004;18:3–17.
- [20] Wu X, Xu L, Yao L. Big data analysis of the human brain’s functional interactions based on fMRI. *Chin Sci Bull* 2014;59:5059–65.
- [21] Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci* 2007;8:700–11.
- [22] Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 1995;34:537–41.
- [23] Greicius MD, Srivastava G, Reiss AL, Menon V, Raichle ME. Default-mode network activity distinguishes Alzheimer’s disease from healthy aging: evidence from functional MRI. *Proc Natl Acad Sci U S A* 2004;101:4637–42.
- [24] Luo SQ. Study on digitized atlas of the human brain. *Zhongguo Yi Liao Qi Xie Za Zhi* 2001;25:91 (in Chinese with an English abstract).
- [25] Lippert T, Orth B. Supercomputing infrastructure for simulations of the human brain. *BrainComp* 2013;8603:198–212.
- [26] Zhao XW, Junzhong JI, Liang PP. The human brain functional parcellation based on fMRI data. *Chin Sci Bull* 2016;61:2035–52.
- [27] Suk H-I, Shen D. Deep learning-based feature representation for AD/MCI classification. *Miccai* 2013;8150:583–90.
- [28] Liu S, Liu S, Cai W, Che H, Pujol S, Kikinis R, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer’s disease. *IEEE T Bio-Med Eng* 2015;62:1132–40.
- [29] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436.
- [30] Li KC, Yang XP. Imaging diagnosis of Parkinson’s disease. *Diagn Theory Pract* 2005;4:273–4 (in Chinese with an English abstract).

- [31] Vidaurre D, Abeysuriya R, Becker R, Quinn AJ, Alfaro-Almagro F, Smith SM, et al. Discovering dynamic brain networks from big data in rest and task. *Neuroimage* 2017;180:646–56.
- [32] Spitzer H, Amunts K, Harmeling S, Dickscheid T. Parcellation of visual cortex on high-resolution histological brain sections using convolutional neural networks. *Proc IEEE Int Symp Biomed Imaging* 2017:920–3.
- [33] Payan A, Montana G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. *Comput Sci arXiv* 2015;1502.02506.
- [34] Hosseini-Asl E, Ghazal M, Mahmoud A, Aslantas A, Shalaby AM, Casanova MF, et al. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci* 2016;23:584–96.
- [35] Makkie M, Huang H, Zhao Y, Vasilakos AV, Liu T. Fast and scalable distributed deep convolutional autoencoder for fMRI big data analytics. *Neurocomputing* 2017;325:20–30.
- [36] Amunts K, Lepage C, Borgeat L, Mohlberg H, Dickscheid T, Rousseau MÉ, et al. BigBrain: an ultrahigh-resolution 3D human brain model. *Science* 2013;340:1472–5.
- [37] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Software* 2015;1–2:19–25.
- [38] Jirsa VK, Proix T, Perdikis D, Woodman MM, Wang H, Gonzalez-Martinez J, et al. The virtual epileptic patient: individualized whole-brain models of epilepsy spread. *Neuroimage* 2016;145:377–88.
- [39] Forman MS, Trojanowski JQ, Lee VM. Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. *Nat Med* 2004;10:1055–63.
- [40] Budson AE, Solomon PR. Chapter 4 – Alzheimer's disease dementia and mild cognitive impairment due to Alzheimer's disease. *Mem Loss Alzheimers Dis Dement* 2016;2016:47–69.
- [41] Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther* 2014;6:37.
- [42] Murray CJ, Lopez AD, World Health Organization. World Bank, Harvard School of Public Health. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary. In: Murray CJL, Lopez AD, World Health Organization. editors. Cambridge: Harvard University Press; 1996, p. 41.
- [43] Xiong JZ, Xia SR, Li JS. The Study on the auto-classification of Parkinson's disease based on MR imaging. *Chin Digit Med* 2016;11:8–10.
- [44] Xu J, Gong D, Man C, Fan Y. Parkinson's disease and risk of mortality: meta-analysis and systematic review. *Acta Neurol Scand* 2014;129:71–9.
- [45] Okano H, Sasaki E, Yamamori T, Iriki A, Shimogori T, Yamaguchi Y, et al. Brain/MINDS: a Japanese national brain project for marmoset neuroscience. *Neuron* 2016;92:582.
- [46] Poo MM, Du JL, Ip NY, Xiong ZQ, Xu B, Tan T. China brain project: basic neuroscience, brain diseases, and brain-inspired computing. *Neuron* 2016;92:591–6.
- [47] Shamir RR, Dolber T, Noecker AM, Walter BL, McIntyre CC. Machine learning approach to optimizing combined stimulation and medication therapies for Parkinson's disease. *Brain Stimul* 2015;8:1025–32.
- [48] Gupta A, Maida AS, Ayhan M. Natural image bases to represent neuroimaging data. *Proc Int Conf Mach Learn* 2014:987–94.
- [49] Yang W, Lui RLM, Gao JH, Chan TF, Yau ST, Sperling RA, et al. Independent component analysis-based classification of Alzheimer's MRI data. *J Alzheimers Dis* 2011;24:775–83.
- [50] Klöppel S, Stonnington CM, Barnes J, Chen F, Chu C, Good CD, et al. Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method. *Brain* 2008;131:2969.
- [51] Janušová E, Vounou M, Wolz R, Gray KR, Rueckert D, Montana G. Biomarker discovery for sparse classification of brain images in Alzheimer's disease. *Ann Bmva* 2012;2012:1–11.
- [52] Batmanghelich N, Taskar B, Davatzikos C. A general and unifying framework for feature construction, in image-based pattern classification. *Inf Process Med Imaging* 2009;21:423–34.
- [53] Al-Fatlawi AH, Jabardi MH, Ling SH. Efficient diagnosis system for Parkinson's disease using deep belief network. *IEEE CEC* 2016:1324–30.
- [54] Sarraf S, Tofghi G. Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. *Ftc* 2016;2016:816–20.
- [55] Horn AHC, Sticht H. Amyloid- β 42 oligomer structures from fibrils: a systematic molecular dynamics study. *J Phys Chem B* 2010;114:2219–26.
- [56] Bassett DS, Gazzaniga MS. Understanding complexity in the human brain. *Trends Cogn Sci* 2011;15:200–9.
- [57] Ed B, Anna B, Bassett DS, Alex F, Manfred K, David M, et al. Generic aspects of complexity in brain imaging data and other biological systems. *Neuroimage* 2009;47:1125–34.
- [58] Schutter ED. A consumer guide to neuronal modeling software. *Trends Neurosci* 1992;15:462–4.
- [59] Wilson MA, Bhalla US, Uhley JD, Bower JM. GENESIS: a system for simulating neural networks. *Adv Neural Inf Process Syst* 1989;1989:485–92.
- [60] Sporns O. The human connectome: a complex network. *Ann N Y Acad Sci* 2012;136:109–25.
- [61] Van Essen DC, Barch DM. The human connectome in health and psychopathology. *World Psychiatry* 2015;14:154–7.
- [62] Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, et al. Mapping the structural core of human cerebral cortex. *PLoS Biol* 2008;6:e159.
- [63] Kaiser M. The potential of the human connectome as a biomarker of brain disease. *Front Hum Neurosci* 2013;7:484.
- [64] Dai Z, He Y. Disrupted structural and functional brain connectomes in mild cognitive impairment and Alzheimer's disease. *Neurosci Bull* 2014;30:217–32.
- [65] Dickerson BC, Brickhouse M, McGinnis S, Wolk DA. Alzheimer's disease: the influence of age on clinical heterogeneity through the human brain connectome. *Alzheimers Dement (Amst)* 2017;6:122–35.
- [66] Wang J, Zuo X, Dai Z, Xia M, Zhao Z, Zhao X, et al. Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biol Psychiatry* 2013;73:472–81.
- [67] Surmeier DJ, Obeso JA, Halliday GM. Selective neuronal vulnerability in Parkinson disease. *Nat Rev Neurosci* 2017;18:101.
- [68] Ingalhalikar M, Smith A, Parker D, Satterthwaite TD, Elliott MA, Ruparel K, et al. Sex differences in the structural connectome of the human brain. *Proc Natl Acad Sci U S A* 2014;111:823–8.
- [69] Markram H. The human brain project. *Sci Am* 2012;306:50–5.
- [70] Eliasmith C, Stewart TC, Choo X, Bekolay T, Dewolf T, Tang C, et al. A large-scale model of the functioning brain. *Science* 2012;338:1202.
- [71] Eliasmith C, Trujillo O. The use and abuse of large-scale brain models. *Curr Opin Neurobiol* 2014;25:1–6.
- [72] Hsu J. IBM's new brain [news]. *IEEE Spectr* 2014;51:17–9.
- [73] Izhikevich EM, Edelman GM. Large-scale model of mammalian thalamocortical systems. *Proc Natl Acad Sci USA* 2008;105:3593–8.
- [74] Sporns O. The human connectome: origins and challenges. *Neuroimage* 2013;80:53–61.
- [75] Sporns O, Tononi G, Kötter R. The human connectome: a structural description of the human brain. *PLoS Comput Biol* 2005;1:e42.
- [76] Vu AT, Auerbach E, Lenglet C, Moeller S, Sotiropoulos SN, Jbabdi S, et al. High resolution whole brain diffusion imaging at 7 T for the human connectome project. *Neuroimage* 2015;122:318–31.

- [77] Essen DCV, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, et al. The human connectome project: a data acquisition perspective. *Neuroimage* 2012;62:2222–31.
- [78] Smith SM, Beckmann CF, Andersson JLR, Auerbach EJ, Bijsterbosch JD, Douaud G, et al. Resting-state fMRI in the human connectome project. *Neuroimage* 2013;80:144–68.
- [79] Landhuis E. Neuroscience: big brain, big data. *Nature* 2017;541:559.
- [80] Horn A, Ostwald D, Reisert M, Blankenburg F. The structural-functional connectome and the default mode network of the human brain. *Neuroimage* 2014;102:142–51.
- [81] Ugurbil K, Xu J, Auerbach EJ, Moeller S, An TV, Duarte-Carvajalino JM, et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* 2013;80:80–104.
- [82] Szalkai B, Kerepesi C, Varga B, Grolmusz V. The budapest reference connectome server v2.0. *Neurosci Lett* 2015;595:60–2.
- [83] Szalkai B, Kerepesi C, Varga B, Grolmusz V. Parameterizable consensus connectomes from the human connectome project: the budapest reference connectome server v3.0. *Cogn Neurodyn* 2016;11:1–4.
- [84] Gewaltig MO, Diesmann M. NEST (neural simulation tool). *Scholarpedia J* 2007;2:1430.
- [85] Izhikevich EM. Simple model of spiking neurons. *IEEE Trans Neural Netw* 2003;14:1569–72.
- [86] Sterling T, Lusk E, Gropp W. *Beowulf cluster computing with linux*. Cambridge: MIT Press; 2002.
- [87] Marcus DS, John H, Timothy O, Michael H, Glasser MF, Fred P, et al. Informatics and data mining tools and strategies for the human connectome project. *Front Neuroinform* 2011;5:4.
- [88] Hodge MR, Horton W, Brown T, Herrick R, Olsen T, Hileman ME, et al. ConnectomeDB—sharing human brain connectivity data. *Neuroimage* 2016;124:1102–7.
- [89] Glasser MF, Smith SM, Marcus DS, Andersson JLR, Auerbach EJ, Behrens TEJ, et al. The human connectome project's neuroimaging approach. *Nat Neurosci* 2016;19:1175.
- [90] Elam JS, Van Essen D. Human Connectome Project. In: Jaeger D, Jung R, editors. *Encyclopedia of computational neuroscience*. New York: Springer New York; 2013. p. 1–4.
- [91] Hines ML, Carnevale NT. The NEURON simulation environment. *Neural Comput* 2014;9:1179–209.
- [92] Davison BD. NCS: network and cache simulator – an introduction. In: Technical report DCS-TR-444 [Internet] New Brunswick: The State University of New Jersey; 2011, <https://rucore.libraries.rutgers.edu/rutgers-lib/59031/>.
- [93] Goddard N, Hood G, Howell F, Hines M, De Schutter E. NEOSIM: portable large-scale plug and play modelling. *Neurocomputing* 2001;38:1657–61.
- [94] Delorme A, Gautrais J, Van Rullen R, Thorpe S. SpikeNET: A simulator for modeling large networks of integrate and fire neurons. *Neurocomputing* 1999;26:989–96.
- [95] Migliore M, Cannia C, Lytton WW, Markram H, Hines ML. Parallel network simulations with NEURON. *J Comput Neurosci* 2006;21:119.
- [96] Amunts K, Bücker O, Axer M. Towards a multiscale, high-resolution model of the human brain. *Braincomp* 2013;8603:3–14.
- [97] Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell* 2015;163:456.
- [98] Jordan J, Ippen T, Helias M, Kitayama I, Sato M, Igarashi J, et al. Extremely scalable spiking neuronal network simulation code: from laptops to exascale computers. *Front Neuroinform* 2018;12:2.