

MEETING REPORT

Open Access



# “Big Data” in neuroscience: open door to a more comprehensive and translational research

Audald Lloret-Villas<sup>1†</sup>, Rachel Daudin<sup>2,4†</sup> and Nicolas Le Novère<sup>3\*</sup>

\* Correspondence: n.lenovere@gmail.com

†Equal contributors

<sup>3</sup>The Babraham Institute, Cambridge, UK

Full list of author information is available at the end of the article

## Abstract

The International Symposium “Advances in Systems Biology in Neurosciences” was held in February 2015 in Geneva. A hundred scientists with a variety of expertise gathered around the theme of human brain complexity and cognitive disorders. Through a series of lectures and poster sessions, the symposium showcased state-of-the-art high-throughput biotechnologies, supercomputers and neuroimaging, and illustrated the latest advances in systems approaches to tackle Neurosciences and Neurodegenerative disorders. The meeting highlighted the power of big data to understand complex pathologies and also the need for more open and integrated data.

**Keywords:** Systems biology, Imaging, Data integration, Modeling, Alzheimer, Parkinson

Over the last two decades, research in life sciences experienced a paradigm shift, adding to the very successful reductionist analysis of isolated entities (molecules, cells, organs) the study of interacting entities, called “Systems Biology”. Neuroscience pioneered the systems approaches and is now pushing the boundaries of multi-scale descriptions, from the molecule to the disease [1]. The human brain contains about 100 billions neurons [2], connected by 1 quadrillion synapses. Understanding brain function, and disorders, as well as designing treatments requires gathering enormous amounts of heterogeneous data. The challenges faced range from collection and annotation to storage and integration (Fig. 1).

On Friday, February 6, 2015 in Geneva was held the International Symposium “Advances in Systems Biology in Neurosciences”, funded by the European Commission through the AgedBrainSYSBIO project and jointly sponsored by the Vital-IT/Swiss-Prot groups of the Swiss Institute of Bioinformatics (SIB), Novartis and Roche. The meeting brought together clinicians, biologists, bioinformaticians, and statisticians from different European-funded consortia such as AgedBrainSYSBIO, SynSys, the European flagship Human Brain Project (HBP), and also US projects such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and the Allen Institute for Brain Science. This international assembly of scientists was gathered in the high-tech building that hosts the HBP and SIB on the Campus Biotech of Geneva (formerly Merck-Serono). Through a series of lectures and poster sessions, the symposium showcased



**Fig. 1** Geneva Biotech campus

to the attendees state-of-the-art approaches, including high-throughput biotechnologies, supercomputers and neuroimaging, and illustrated the latest advances in systems approaches to tackle Neurosciences and Neurodegenerative disorders. This was an opportunity for more than one hundred researchers and students from academia and industry to gather and share their knowledge in the field. The first authors of this report are students in partners of AgedBrainSYSBIO. Attendance to this meeting allowed them to be exposed to the foremost research in the systems biology of neuronal systems. Professor Michel Simonneau (Centre de Psychiatrie et Neurosciences, INSERM U894, coordinator of AgedBrainSYSBIO) and Ioannis Xenarios (SIB Group Leader, head of Vital-IT and Swiss-Prot) opened the meeting with welcoming remarks (Fig. 2).

The first session, moderated by Dr Le Novère, from the Babraham Institute, Cambridge, UK), began with a keynote presentation by Dr. Sean Hill, co-Director of the Blue Brain Project [3] and co-Director of Neuroinformatics in the European Union funded Human Brain Project (HBP) at the École Polytechnique Fédérale de Lausanne (EPFL). This large project (involving more than 250 researchers), supported by EU information technology funding, aims to provide innovative tools to the global Neuroscience community in order to model and simulate the human brain. Sean Hill, one of the coordinators of this flagship European program, introduced the audience with an impressive data infrastructure built over the first 10 years of work. Electrophysiological recordings, gene expression patterns, morphological studies, among others, are stored and integrated with rich semantic metadata in order to inform mathematical models of neuronal circuits. He then presented preliminary simulations of the electrical behavior of neocortical columns, built upon the compilation of the large amount of data described above [4]. Such complex models require large amount of storage and computing power (the simulations are currently using a Blue Gene computer). The HBP still faces significant challenges when developing a platform to generate and make accessible for the whole community the huge amount of data required for understanding the entire brain function.

Dr. Yann Hérault, head of the Institut Clinique de la Souris (ICS) and research group leader at the Institut de Génétique Biologie Moléculaire et Cellulaire (IGBMC) in Illkirch, France, presented the second lecture about a large mouse aneuploidy zoo. His talk described efforts to decipher the physiopathology of Intellectual Disabilities [5] using mouse models of human trisomy 21. A complete and standardized



**Fig. 2** Sean Hill, from the EPFL (Lausanne, Switzerland)

characterization of each model is provided (behavior, gene expression, pathways manipulations) in order to understand the role of different chromosome 21 regions implicated in cognitive defects. These mouse models help understand the pathophysiology of trisomy 21, and also of related neurological disorders.

Dr. Maksym V. Kopanitsa, from the biotechnology company Synome, a spin-off from the Wellcome Trust Sanger Institute (Cambridge, UK) presented his company's work. Expert on high throughput analysis of synaptic function and cognition, he has been involved in different European projects (EUROSPIN, SynSys, GENCODYS and PharMEA). The company specializes in multiple electrophysiological recordings and has developed useful scoring algorithms describing the importance of each protein in synapse activity. Using mouse models mutated for genes encoding key proteins of pre- and post-synaptic compartments, Synome aims to describe the role of each of these proteins [6].

The last speaker of the first session was Dr Le Novère. To understand the function of the synapse, and in particular its plasticity, at the molecular level, his group develops mathematical models and numerical simulation. His presentation focused on the response of synaptic protein kinases to calcium signals. He described in particular mechanistic allosteric models of calmodulin, based on the thermodynamic equilibrium between the two conformations of calmodulin's lobes [7]. These models were then placed in the context of the synapse, to obtain a proper simulation of calcium signal

dynamics. Finally, biochemical and electrophysiological models can be integrated in large detailed models of entire neurons [8]. Such models permit to integrate existing knowledge and test our hypothesis about neuronal function.

Dr Jane Ann Driver from Brigham and Women's Hospital (Boston, MA, USA) opened the second session, chaired by Prof. Michel Simonneau. Dr. Driver's talk shed light on the correlation between different diseases and their impact. Epidemiological evidences suggest an inverse association - that is a negative correlation - between the presence of cancer and the most common neurodegenerative conditions, such as Alzheimer's disease (AD) and Parkinson's Disease (PD) [9]. She underlined how difficult it was to assess such association, due to many biases induced by treatments, lack of reporting and cancer-induced death. She then presented results supporting the involvement of Pin1, a protein known to promote cellular health by restoring phosphorylated Tau and amyloid precursor protein to a functional state [10]. Dr. Driver also described a mouse knock-out model for the gene encoding the Pin1 protein, cancer-resistant and more sensitive to neurodegenerative and aging disorders. Understanding inverse comorbidities and predicting them is crucial to choose treatments. Their detection relies on large epidemiological data. The more we have, the better the detection, understanding and treatment (Fig. 3).

Following Dr. Driver, prof. Tal Pupko from the Department of Cell Research & Immunology, Tel-Aviv University (Tel-Aviv, Israel) focused his talk on the differential



**Fig. 3** Jane Ann Driver, from Brigham and Women's Hospital (Boston, MA, USA)

selection of genes related to neurodegenerative diseases. Using a bioinformatics pipeline, they found (in collaboration with researchers from the Centre Psychiatrie & Neurosciences, INSERM U894) an important enrichment of positively selected genes in those pathologies. Positive selection is mainly found when genes drift due to pressures in new environments. Such work on human-specific gene evolution may point to limitations of using distant animal models, which might have experienced different selection pressures.

Dr. Hervé Rhinn from the Department of Pathology & Cell Biology, Columbia University Medical Center (Columbia, NY, USA) then talked about variants in the human genome linked to the risk of developing non-familial AD and PD. Application of bioinformatics techniques, such as Differential Co-expression Analysis to whole transcriptome gene expression data, enables the study of expressed and non-expressed genetic components. These tools revealed the role of genetic components in key processes of neurodegeneration including alpha-synuclein regulation in PD [11] and the impact of RNF219 gene as a mediator of the effect of APOE-E4 in AD [12].

Prof. Simonneau concluded the second session with a presentation centered on using protein-protein interactions to understand neuronal diseases. He presented regulatory networks obtained from large-scale hypothesis-free analysis such as yeast two-hybrids, whole-genome sequencing and epistatic linkage analysis. These networks linked several genes identified in Schizophrenia and Late-Onset Alzheimer's disease – Genome-wide association studies (LOAD-GWAS). He then described his group's efforts to validate this network on mouse brain tissue and primary neurons. They used endogenous immunoprecipitation, proximity ligation assays, electrophysiology and behavior on a novel transgenic model that overexpresses human BIN1 and thus mimicks BIN1 changes found in LOAD patients. Neurodegenerative pathologies are “network diseases”. The analysis of large networks of physical or functional interactions between genes and proteins permits to discover new molecular causes to the diseases, better understand their development and also suggest targets for treatments.

Dr. James Adjaye from the University of Düsseldorf (Düsseldorf, Germany) moderated the third and last session of the Symposium.

The opening talk was given by Prof. Arthur W. Toga from the Center for Computational Biology, USC (Los Angeles, CA, USA). Prof. Toga leads the Alzheimer's Disease Neuroimaging Initiative (ADNI) that produced a unique set of open-access databases of genetics, biomarkers and imaging [13]. Because of the ever increasing size of existing and emerging databases, Prof. Toga emphasized the importance of the treatment and description of the data in order to enable their meaningful comparison and integration. Shared and unified ontologies are becoming key to validate analysis results obtained from combining different sources. Integration of genetic and neuroimaging data will be key to understand Alzheimer's disease. However the linkages are often complex, and more data is needed, from more patients, for all effects to be detectable. The talk thus ended with the presentation of the Global Alzheimer's Association Interactive Network (GAAIN), a set of analytical tools and computational resources to connect AD scientists worldwide [14]. Such a large scale sharing of data across the world promises to speed up research on the disease.

Dr. Jean-François Demonet from the Leenaards Memory Centre CHUV, Lausanne University (Lausanne, Switzerland) then gave a presentation on the role of neuropsychology

of Ageing-Brain Cognitive Diseases (ABCDs) as a complement to the promising era of biomarkers. He highlighted the lack of data on diagnosis of ABCDs and called for a set of reliable cognitive tests to detect and identify the main conditions. Being able to detect such diseases early will improve treatment and care.

Dr. Jérôme Dauvillier, from the SIB (Lausanne, Switzerland), discussed the analysis of epistatic effects - the effect of a genetic variant that depends on the presence of another genetic variant - to understand the physiopathology of Alzheimer's Disease. He first described an epistatic analysis of two patient cohorts (TGen and HBTRC) based on genotyping data linked with the Braak Score (distribution of neurofibrillary tangles). A Gene Ontology enrichment analysis showed that the epistatic genes are linked to synapses, axonal guidance, and cell-cell adhesion. Dr. Dauvillier also presented an epistasis analysis based on whole-genome sequencing and ADNI's data consortium, providing many perspectives, taking advantage of their unique computing facilities (ViTal-IT). This kind of analysis brings information on gene-gene interactions in AD physiopathology by complementing the more classical approaches of GWAS. Although crucial to understand the complex genetic basis of Alzheimer disease, the number of statistical tests required by epistatic analyses explodes with the number of variants involved and requires large amounts of computing power.

This last session ended with a lecture from Dr. Adjaye, who presented his group's efforts on investigating disease mechanism using induced Pluripotent Stem Cells (iPSC) from AD patients [15]. Indeed, direct biochemical observation of living patients is impossible. Producing neurons from those patient using reprogrammed cells (generally blood or skin cells) provided an alternative. Dr Adjaye was able to differentiate the iPSCs into neuronal cells and to detect the expression of p-Tau and GSK3 $\beta$ . Transcriptome analysis of AD-iPSC derived neuronal cells revealed significant changes in the expression of genes associated with AD, such as APOE, APP or PSEN. Dr. Adjaye finally described his group's work on neuronal cells derived from patient having rare variants of susceptibility genes such as TREM2. Such human cellular models present an alternative to animal models to understand molecular and cellular aspects of the disease.

Ioannis Xenarios concluded the meeting thanking all the participants of the symposium (Fig. 4).

Breaks for lunch and refreshments took place around posters, allowing everyone to ask questions and provide feedback to their authors. The posters covered a large variety



**Fig. 4** Around 100 researchers attended the meeting symposium

of topics, from fluorescence microscopy in exocytosis to mathematical modelling of neurodegeneration or protein overexpression on pathophysiology. Three poster authors were selected to present their works to the audience. PhD student Wenjia Wang from PHARNEXT (Issy-Les-Moulineaux, France) took the opportunity to present a multi-marker genetic association test based on the Rasch model that provides new insights into the genetics of Alzheimer's Disease. Claire Lesieur from the École Normale Supérieure Systèmes Complexes de Lyon IXXI (Lyon, France) talked about natural "Smart" material from proteins. Computer prediction of conformational changes of proteins involved in neuronal diseases will lead to engineered folded mutant proteins able to interact with misfolded forms and avoid their aggregation. Finally, Dr. Lavinia Alberi from the Unit of Anatomy Department of Medicine - UNIFR (Fribourg, Switzerland) presented her results suggesting the implication of Notch1 in neural plasticity and neurodegeneration.

This symposium brought together scientists with a variety of expertise around the theme of human brain complexity and cognitive disorders. Despite aiming towards the same goal - a better understanding of neurodegenerative conditions -, the speakers provided the audience with very different but promising approaches. The only commonalities were the need to consider the role of parts within a system, and the interplay between experimental and computational investigations. For years, the scientific community has tried to explain complex and multifactorial pathologies like Alzheimer's disease by exploring a unique aspect of the pathology (neurodegeneration, immune defects, metabolism disorders) or the effect of given genes and proteins. Whereas a more comprehensive approach based on large-scale heterogeneous dataset is certainly needed. Big Data has entered biologists everyday life, and they tend to generate and use an every growing amount of such information. This symposium showed how such high-throughput data, from genotypes to imaging, including gene expression, proteomics, physiological recordings, can be useful to develop new biological hypothesis and test them in mathematical models. It also showed the huge work that represents data curation and management. Even if, as Pr. Toga underlined it, one should keep in mind that these approaches are providing associations and not causalities, Big Data is opening the door to a more translational science, a better communication between the different fields (animal models, clinical trial, mathematics and bioinformatics). This day was a good representation of such an evolution of scientific research and the associated promises, as recognized by international organization such as OECD [16]. However, despite the existence of a so-called data deluge, we do not in fact have enough data to understand complex pathologies and in particular neurodegenerative diseases. We need more data, more open, and better integration.

#### Abbreviations

ABCDs, Ageing-Brain Cognitive Diseases; AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; EPFL, École Polytechnique Fédérale de Lausanne; GAAlN, Global Alzheimer's Association Interactive Network; HBP, Human Brain Project; IGBMC, Institut de Génétique Biologie Moléculaire et Cellulaire; iPSC, induced Pluripotent Stem Cells; LOAD-GWAS, Late-Onset Alzheimer's disease – Genome-wide association studies; PD, Parkinson's Disease; SIB, Swiss Institute of Bioinformatics

#### Acknowledgements

ALV and RD are supported by the European Union Seventh Framework Programme AgedBrainSYSBIO grant agreement 305299 <http://www.agedbrainsysbio.eu/>.

**Authors' contributions**

All authors participated the workshop and contributed to the writing of this document. All authors read and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**Consent for publication**

Sean Hill and Jane Driver gave their consent to the publication of their picture.

**Author details**

<sup>1</sup>EMBL – European Bioinformatics Institute, Wellcome-Trust Genome Campus, Cambridge, UK. <sup>2</sup>Centre Psychiatrie et Neurosciences, INSERM U894 and Université Paris Descartes, 75014 Paris, France. <sup>3</sup>The Babraham Institute, Cambridge, UK. <sup>4</sup>Hybrigenics Services SA, Paris, France.

Received: 4 August 2015 Accepted: 4 August 2016

Published online: 30 August 2016

**References**

1. Le Novère N. The long journey to a systems biology of neuronal function. *BMC Systems Biology*. 2007;1:28. doi:10.1186/1752-0509-1-28.
2. Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, et al. Equal numbers of neuronal and non-neuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*. 2009;513(5):532–41. doi:10.1002/cne.21974.
3. Markram H. The blue brain project. *Nature Reviews Neuroscience*. 2006;7:153–60. doi:10.1038/nrn1848.
4. Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell*. 2015;163(2):456–92. doi:10.1016/j.cell.2015.09.029.
5. Sahun I, Marechal D, Pereira PL, Nalesso V, Gruart A, Garcia JM, Antonarakis SE, Dierssen M, Hérault Y. Cognition and hippocampal plasticity in the mouse is altered by monosomy of a genomic region implicated in down syndrome. *Genetics*. 2014;197:899–912. doi:10.1534/genetics.114.165241.
6. Ryan TJ, Kopanitsa MV, Indersmitten T, Nithianantharajah J, Afinowi NO, Pettit C, et al. Evolution of GluN2A/B cytoplasmic domains diversified vertebrate synaptic plasticity and behavior. *Nature Neuroscience*. 2013;16(1):25–32. doi:10.1038/nn.3277.
7. Lai M, Brun D, Edelstein SJ, Le Novère N. Modulation of calmodulin lobes by different targets: an allosteric model with hemiconcerted conformational transitions. *PLoS Computational Biology*. 2015;11:e1004063. doi:10.1371/journal.pcbi.1004063.
8. Mattioni M, Le Novère N. Integration of biochemical and electrical signaling-multiscale model of the medium spiny neuron of the striatum. *PLoS One*. 2013;8(7):e66811. doi:10.1371/journal.pone.0066811.
9. Driver JA. Inverse association between cancer and neurodegenerative disease: review of the epidemiologic and biological evidence. *Biogerontology*. 2014;15(6):547–5. doi:10.1007/s10522-014-9523-2.
10. Driver JA, Zhou XZ, Lu KP. Pin1 dysregulation helps to explain the inverse association between cancer and Alzheimer's disease. *Biochimica et Biophysica Acta*. 2015. doi:10.1016/j.bbagen.2014.12.025.
11. Rhinn H, Qiang L, Yamashita T, Rhee D, Zolin A, Vanti W, Abeliovich A. Alternative  $\alpha$ -synuclein transcript usage as a convergent mechanism in Parkinson's disease pathology. *Nature communications*. 2012;3:1084. doi:10.1038/ncomms2032.
12. Rhinn H, Fujita R, Qiang L, Cheng R, Lee JH, Abeliovich A. Integrative genomics identifies APOE  $\epsilon$ 4 effectors in Alzheimer's disease. *Nature*. 2013;500(7460):45–50. doi:10.1038/nature12.
13. Toga AW, Crawford KL. The Alzheimer's disease neuroimaging initiative informatics core: a decade in review. *Alzheimers Dementia*. 2015;11(7):832–29. doi:10.1016/j.jalz.2015.04.004.
14. Toga AW, Neu SC, Bhatt P, Crawford KL, Ashish N. The global Alzheimer's association interactive network. *Alzheimers Dementia*. 2015. doi:10.1016/j.jalz.2015.06.1896.
15. Hossini AM, Megges M, Prigione A, Lichtner B, Toliat MR, Wruck W, Schröter F, Nuernberg P, Kroll H, Makrantonaki E, Zouboulis CC, Adjaye J. Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks. *BMC Genomics*. 2015;16(1):84. doi:10.1186/s12864-015-1262-5.
16. OECD. Unleashing the Power of Big Data for Alzheimer's Disease and Dementia Research: Main Points of the OECD Expert Consultation on Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia. *OECD Digital Economy Papers*. 2014; No. 233, OECD Publishing. doi:10.1787/5jz73kvmvbw-en.