Title of the PhD project:

Rationalizing the use of prior knowledge in systems modeling and molecular neurosciences

PhD Supervisor

Name, first name : NERI, Christian
Email : christian.neri@inserm.fr
Phone : 0144276045
Title/Employer : DR INSERM
Address : CNRS UMR8256, Sorbonne Université, Building B, 6th floor, Boite 12
9 Quai St Bernard 75005 Paris - France
Laboratory : Biology of Adaptation and Aging
Title of the team : Brain-C
Team leader (*if different*) *: NA*
Doctoral School : ED3C

Overview of the scientific projects of the team

Notre objectif principal est de comprendre comment la capacité des neurones à maintenir leur fonction et à résister aux maladies neurodégénératives (compensation) est régulée au niveau moléculaire, cellulaire et inter-cellulaire, et quel en est l'impact sur l'évolution de ces maladies. Nous étudions la maladie de Huntington, une maladie héréditaire dont l'étude a permis de nombreuses avancées sur les maladies neurodégénératives, ainsi que la maladie d'Alzheimer. Notre approche inclue modélisation mathématique des données génomiques, biologie cellulaire (cellules souches humaines), génétique de *C. elegans*, et recherche pré-clinique. Nous collaborons depuis 5 ans avec les laboratoires de mathématiques du campus de Jussieu (LJLL, LPSM, SCAI), et avec plusieurs laboratoires en neurosciences à l'IBPS (unité NPS) ou ailleurs en France (CEA/Mircen en IdF, UNISTRA à l'Université de Strasbourg, GIN à Grenoble). Cf https://www.ibps.upmc.fr/fr/Recherche/umr-8256/brainc

Main publications since January 1er, 2016

1. J Voisin, F Farina, S Naphade, M Fontaine, KT Tshilenge, CG Aguirre, A Lopez-Romirez, J Dancourt, A Ginisty, S Sasidharan Nair, KL Madushani, N Zhan, FX Lejeune, M Verny, J Campisi, LM Ellerby, **C Neri**. (2020). FOXO3 targets are reprogrammed as Huntington's disease neural cells and striatal neurons face senescence with $p16^{INK4a}$ increase. *Aging Cell* doi.org/10.1111/acel.13226
2. Megret L, Nair SS, Dancourt J, Aaronson J, Rosinski J, and **Neri C**. (2020) Combining Feature Selection and Shape Analysis Uncovers Precise Rules for miRNA Regulation in Huntington's Disease Mice. *BMC Bioinformatics* **2**:75. doi: 10.1186/s12859-020-3418-9.

3. Bigan E, Nair SS, Lejeune FX, Fragnaud H, Parmentier F, Megret L, Verny M, Aaronson J, Rosinski J, and **Neri C**. (2019) Genetic cooperativity in multi-layer networks implicates cell survival and senescence in the striatum of Huntington's disease mice synchronous to symptoms. *Bioinformatics*, doi: 10.1093/bioinformatics/btz514.
4. Vergallo, A., Megret, L., Lista, S., Cavedo, E., Zetterberg, H., Blennow, K., Vanmechelen, E., De Vos, A., Habert, M.O., Potier, M.C., Dubois B, **Neri C**, Hampel H. (2019) Plasma amyloid beta 40/42 ratio predicts cerebral amyloidosis in cognitively normal individuals at risk for Alzheimer's disease. *Alzheimers Dement*, **15**, 764-775.
5. Melentijevic I, Toth M, Arnold M, Guasp R Harinath G, Parker JA, **Neri C**, C Gabel, DH Hall, Driscoll M. (2017). *C. elegans* neurons jettison protein aggregates and mitochondria into the extracellular environment in response to neurotoxic stress. *Nature*, **542**, 367-371.

PhD Co-Supervisor

Name, first name : Gris Barbara
Email : bgris.maths@gmail.com
Phone : 0603082658
Title/Employer : Chargée de recherche CNRS
Address : LJLL, 4 place Jussieu, 75004 Paris
Laboratory : Laboratoire Jacques Louis Lions, UMR 7598
Title of the team : Laboratoire Jacques Louis Lions
Team leader *(if different) : Emmanuel Trélat*

Overview of the scientific projects of the team

Le laboratoire Jacques-Louis Lions est un laboratoire de mathématiques appliquées. Ses axes de recherche recouvrent l'analyse, la modélisation et le calcul scientifique haute performance de phénomènes représentés par des équations aux dérivées partielles. Le LJLL collabore avec le monde économique et avec d'autres domaines scientifiques à travers un large spectre d'applications : médecine et biologie ; dynamique des fluides ; physique, mécanique et chimie théoriques ; contrôle, optimisation et finance ; traitement du signal et des données. Barbara Gris collabore en particulier avec l'équipe Brain-C depuis 2018.

Main publications since January 1er, 2016

[1] L. Younes, **B. Gris** and A. Trouvé. (2020). Sub-Riemannian methods in shape analysis. *Chapter of Handbook of Variational Methods for Nonlinear Geometric Data. Springer.* (pp. 463-495).

[2] **B. Gris**. (2019). Incorporation of a deformation prior in image reconstruction. *Journal of Mathematical Imaging and Vision*. 10.1007/s10851-018-0868-z

[3] C. Chen, **B. Gris**, O. Öktem. (2018). A New Variational Model for Joint Image Reconstruction and Motion Estimation in Spatiotemporal Imaging. *SIAM Journal on Imaging Science. 12(4), 1686-17.*

[4] **B. Gris**, C. Chen, O. Öktem. (2020) Image reconstruction through metamorphosis. *Inverse Problems. 36(2), 025001*.

[5] **B. Gris**, S. Durrleman and A. Trouvé. (2018). A sub-Riemannian modular framework for diffeomorphism based analysis of shape ensembles. *SIAM Journal of Imaging Sciences. 10.1137/16M1076733.*

Doctoral Project

Title: Rationalizing the use of prior knowledge in systems modeling and molecular neurosciences

Abstract :
Huntington's disease (HD), a disease caused by CAG expansion in huntingtin, is recognized as a model to understand the role of cellular resilience over senescence systems in neurodegenerative diseases (ND). As large omic datasets become available to study these systems in HD and other NDs, there is a critical need for biologically-precise systems-modeling methods. To this end, we developed a new and promising metric for precisely modeling the responses to mutant huntingtin on molecular and functional levels. This metric is based on a design matrix that contains the prior knowledge on HD that is introduced in the analysis. Understanding the influence of such prior knowledge on the information provided by *in silico* models is of major interest in systems modeling. The aim of the PhD project is to rationalize and automatize the use of prior knowledge in our metric. This will make our new method generalizable and easy to use for a wide range of applications while obtaining new insights into the dynamics of disease resistance systems in HD.

Context and objective :
Huntington's disease (HD) is a prototypical neurodegenerative disease caused by CAG expansion in the huntingtin (Htt) gene. HD is recognized as an insightful model for understanding how the brain may compute the molecular responses to neurodegenerative insults and resist disease, as it is a genetic disease that shares several mechanisms with other NDs, notably gene deregulation and acceleration of cellular aging associated with the production of misfolded mRNAs/proteins. The **aim** of this PhD project is to identify the temporal and cell-type-specific features of HD on a system level while prioritizing targets genes for early-stage (before clinical conversion, before neurons are committed to die) intervention in HD via gene therapy. The PhD candidate will address the following **question**: what is the temporal dynamics of cellular resilience over senescence systems in HD on molecular and cellular levels? To this end, he/she will analyze *multi-dimensional data* as primarily obtained in the allelic series of HD knock-in mice (*Hdh* mice) [1-3]. These data belong to the *most comprehensive dataset* available to date to study the molecular pathogenesis of a neurodegenerative disease on a systems level. In *Hdh* mice, these data cover up to 5 mutant *Htt* alleles, up to 3 age points, several brain areas (e.g. striatum, cortex), several cell types (e.g. striatal neurons, striatal astrocytes), and several layers of molecular regulation (e.g. miRNAs, transcriptome, proteome, genome-wide functional screen data). The Brain-C Lab, a team expert in neurosciences at Sorbonne Université, has developed new machine learning pipelines that are biologically precise and that have identified interesting and previously-undetected targets in HD [4-8]. Along these lines, the Brain-C Lab and the Laboratoire Jacques-Louis Lions (LJLL), a laboratory expert in applied mathematics at Sorbonne Université, have developed an innovative approach, namely a metric well-adapted to the analysis of multi-point profiles (e.g. time-series) in complex genomic datasets. This approach shows superior ability to characterize the dynamics of molecular responses to mutant *Htt* in *Hdh* mice compared to using classical metrics, such as Euclidean distances (Megret L, Gris B // and Neri C, submitted). Although this method was developed for the analysis of complex data obtained in the allelic series of *Hdh* mice, it can be generalized to all types of multipoint data by providing an answer to the essential question on the effect of using prior knowledge in building in silico models of biological processes. However, this approach is based on a design matrix that contains the prior information (*i.e.* the prior knowledge on HD) introduced in the analysis. Understanding the effect of using prior knowledge on the performance of a machine learning algorithm is a major challenge in systems modeling [9]. Additionally, the need, and difficulty, to choose a proper design matrix prior to analysis represents a major obstacle to the generalization and use of this new metric. The PhD

candidate will perform research on defining an automatic process to detect an optimal design matrix and he/she will analyze the effects of variations in this matrix on pattern and rule discovery in the resulting *in silico* model. The PhD candidate will also develop a library making our approach for analysis of dimensional omic datasets a method that is easy to use by the scientific community. The development of this library will rely on the in-depth analysis of the effects of the parameters that account for prior knowledge. This part will directly benefit from the expertise of B. Gris at LJLL. Detailed information on the project will be communicated to the PhD candidates upon application. The PhD candidate will benefit from *biological validation studies* of top *in silico* hypotheses thanks to collaborations with the biologists that in the Brain-C Lab use experimental models of HD (e.g. in human cells, *C. elegans*, mice). In addition to first co-authorship on this research (focused on transcriptomic and functional data), productivity (co-authorship) may also come from collaborations covering the analysis of other types of HD genomic datasets. **Expected results** are (i) the generalization and accessibility of the new approach that we are developing, and (ii) a deep understanding of how the distances between genes that account for prior knowledge may influence the information that is generated by machine learning, and (iii) precise *in silico* models of HD, precisely highlighting the dynamics of cellular repair and resilience systems in neurodegenerative disease conditions.

References

1. Langfelder P, Cantle JP, Chatzopoulou D, et al. Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. Nat Neurosci. 2016;19(4):623-33. doi: 10.1038/nn.4256.
2. Wertz MH, Mitchem MR, Pineda SS, et al. Genome-wide In Vivo CNS Screening Identifies Genes that Modify CNS Neuronal Survival and mHTT Toxicity. Neuron. 2020;106(1):76-89 e8. Epub 2020/02/01. doi: 10.1016/j.neuron.2020.01.004.
3. Lee H, Fenster RJ, Pineda SS, et al. Cell Type-Specific Transcriptomics Reveals that Mutant Huntingtin Leads to Mitochondrial RNA Release and Neuronal Innate Immune Activation. Neuron. 2020. Epub 2020/07/19. doi: 10.1016/j.neuron.2020.06.021.
4. Lejeune FX, Mesrob L, Parmentier F, et al. Large-scale functional RNAi screen in C. elegans identifies genes that regulate the dysfunction of mutant polyglutamine neurons. BMC Genomics. 2012;13:91. doi: 10.1186/1471-2164-13-91.
5. Bigan E, Sasidharan Nair S, Lejeune FX, et al. Genetic cooperativity in multi-layer networks implicates cell survival and senescence in the striatum of Huntington's disease mice synchronous to symptoms. Bioinformatics. 2020;36(1):186-96. Epub 2019/06/23. doi: 10.1093/bioinformatics/btz514.
6. Megret L, Nair SS, Dancourt J, Aaronson J, Rosinski J, Neri C. Combining feature selection and shape analysis uncovers precise rules for miRNA regulation in Huntington's disease mice. BMC Bioinformatics. 2020;21(1):75. Epub 2020/02/26. doi: 10.1186/s12859-020-3418-9.
7. Tourette C, Farina F, Vazquez-Manrique RP, et al. The Wnt receptor Ryk reduces neuronal and cell survival capacity by repressing FOXO activity during the early phases of mutant huntingtin pathogenicity. PLoS Biol. 2014;12(6):e1001895. doi: 10.1371/journal.pbio.1001895.
8. Voisin J, Farina F, Naphade S, et al. FOXO3 targets are reprogrammed as Huntington's disease neural cells and striatal neurons face senescence with p16(INK4a) increase. Aging Cell. 2020:e13226. Epub 2020/11/07. doi: 10.1111/acel.13226.
9. Bzdok D, Nichols TE, Smith SM. Towards Algorithmic Analytics for Large-scale Datasets. Nat Mach Intell. 2019;1(7):296-306. Epub 2019/11/09. doi: 10.1038/s42256-019-0069-5.

Justification of suitability for *i-Bio*:
The proposed PhD project is a frontier research project. This project requires theoretical development to fully understand the behavior of our method in terms of reliability and robustness, and it lies at the interface of biology and mathematics, building up on motivation of biology for mathematics (Brain-C Lab: C. Neri, L. Mégret) and mathematics for biology (LJLL: B. Gris). It directly addresses a pressing and difficult question in machine learning, that is how to best use prior knowledge in the analysis of highly dimensional data, the answer to which will drive change and make a better practice of machine learning in biological research as it has direct implications on the biological precision of *in silico* models. In addition, the proposed PhD project aims to make a new machine learning paradigm accessible to the non-specialized scientific community, fostering advances in molecular systems biology.

Role of each supervisor / skills provided:
C. Neri will provide mentorship in genome sciences, cell biology, and research on neurodegenerative diseases. The PhD candidate will benefit from training in bioinformatics provided by researchers of the Brain-C Lab (L. Mégret).

B. Gris will provide mentorship in mathematics applied to modeling complex datasets in biology.

Profile of the desired student:
The PhD candidate should have a master degree in biology with a specialization in bioinformatics and mathematical modeling. He/she must be able to justify a training in computer science including an understanding of algorithmics and programming (implementation of usual statistical analysis and machine learning algorithms) as well as a good understanding of mathematics for modeling.