



## Post-doctoral researcher in Statistics/Machine Learning at the Brain and Spine Institute (ICM)

*The IHU-A-ICM is a Foundation of Scientific Cooperation that federates the scientific and medical expertise of four partners from the public sector (Public Hospitals of Paris (AP-HP), National Center for Scientific Research (CNRS), National Institute of Health and Medical Research (INSERM) and Pierre and Marie Curie University-Paris 6 (UPMC)). Our mission is to develop new diagnostic, therapeutic and teaching strategies, in order to improve the management of neurodegenerative diseases. The IHU-A-ICM is localized in the largest European hospital, in a top-level French university. Its research activity, grouped together in the ICM, takes place in a unique environment associating core technological facilities, space for industrial partners and a centre for clinical investigations. This environment has attracted the best international research groups.*

**Title:** A multimodal approach for the identification of biomarkers involved in the evolution of Spinocerebellar Ataxia disorder

**Start date:** From November 2015

**Contract duration:** 15 months

**Gross salary:** to be defined

**Primary Contact:** Arthur Tenenhaus ([arthur.tenenhaus@centralesupelec.fr](mailto:arthur.tenenhaus@centralesupelec.fr)) and Fanny Mochele ([fanny.mochel@upmc.fr](mailto:fanny.mochel@upmc.fr))

**Description of the proposal.** Neuroimaging enables the characterization of subtle alterations of brain structure and activity. Those characteristics can be considered as quantitative traits or endophenotypes that can be analyzed together with genomics. A well accepted hypothesis is that the modification of the brain structure/activity is modulated by genomic biomarkers. Therefore, the close study of correlations between genomics and neuroimaging datasets enables one to identify genes implicated in the alterations of a given neuroimaging structure and to understand the underlying biological mechanisms at work for a given disease. From this perspective, we propose to mine a cohort of patients with Spinocerebellar Ataxia disorder (SCA – a neurodegenerative disease) on which multimodal genomics datasets (expression and metabolomics data), multimodal neuroimaging (volumetry, magnetic resonance spectroscopy) and clinical information have been acquired. All participants (67 patients / 35 controls) came for a baseline assessment and then 2 years later for a follow-up visit.

The joint analysis of these complex and heterogeneous sources of information promises to significantly improve our understanding of the molecular mechanisms of SCA. The goal of this integrative analysis is primarily to identify a set of genomic biomarkers co-varying with a set of neuroimaging variables (e.g. cerebellar and brainstem tissue loss) which are associated with disease severity/progression. The main bottlenecks of such an integrative analysis are the high complexity and heterogeneity of the data that stems from: (i) various sources, (ii) the number of variants of SCA to be considered (iii) the high number of measurements in both genomics and neuroimaging data which involves the computation of million(s) of associations. Consequently, a successful investigation of such a structured dataset requires to develop/use statistical methods that fit both the peculiar structure of the data as well as their heterogeneous nature.

All the analyses might be conducted based on the recently published methods [Tenenhaus et al 2015, Lofstedt et al 2015, Tenenhaus & Tenenhaus 2014a, Tenenhaus et al 2014b, Tenenhaus & Tenenhaus 2011]. The methodological tools developed during this project will lead to prototype software (e.g. R packages).

All the datasets are already centralized at the bioinformatics and biostatistics platform of the ICM.

**Requirements (training/expertise) and profile.** PhD in statistics/machine learning. Past experience in multivariate data analysis applied to biological data and/or neuroimaging will be strongly appreciated. The post-doctoral researcher is expected to interact with the researchers of the bioinformatics/biostatistics platform (<http://icm-institute.org/fr/iconics-the-bioinformatics-and-biostatistics-core-facility/>), the investigators and the scientists involved in the SCA project at the ICM.

To apply, submit a cover letter indicating past research experience, motivation for the position, expected availability date and curriculum vitae directly to [arthur.tenenhaus@centralesupelec.fr](mailto:arthur.tenenhaus@centralesupelec.fr) and [fanny.mochel@upmc.fr](mailto:fanny.mochel@upmc.fr).

### Bibliography

Tenenhaus, A., Philippe, C., Frouin, V. (2015). Kernel Generalized Canonical Correlation Analysis. *Computational Statistics & Data Analysis*, 90, 114-131.

Löfstedt T., Hadj-Selem F., Guillemot V., Philippe C., Duchesnay E., Frouin V., Tenenhaus A. (2015). Structured variable selection for generalized canonical correlation analysis. In *The Multiple Facets of Partial Least Squares Methods*, Proceedings in Mathematics and Statistics, Springer New York

Tenenhaus A., Philippe C., Guillemot V., Lê Cao K.-A., Grill, J., Frouin V., (2014). Variable Selection for Generalized Canonical Correlation Analysis, *Biostatistics*, 15 (3) : 569-583

Tenenhaus A., Tenenhaus M., (2014). Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of Operational Research*, 238(2), 391-403.

Le Floch E., Guillemot V., Frouin V., Pinel P., Lalanne C., Trinchera L., Tenenhaus A., Moreno A., Zilbovicius M., Bourgeron T., Dehaene S., Thirion B., Poline J.-B., Duchesnay E., (2012). Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse Partial Least Squares, *NeuroImage*, Vol. 63, Issue 1, pp. 11-24

Tenenhaus A., Tenenhaus M., (2011). Regularized Generalized Canonical Correlation Analysis, *Psychometrika*, vol. 76, Issue 2, pp. 257-284