JOB OFFER
Postdoctoral fellow

Knowledge models for analysis and interpretation of genetic data in neurodegenerative diseases

Keywords: computational biology, bioinformatics, knowledge models, ontologies, genomic data

Context

This position is offered within the IPL (Inria Project Lab) Neuromarkers, a collaborative project between Inria and the Brain and Spinal cord Institute (ICM, www.icm-institute.org), which aims to identify markers of neurodegenerative diseases from joint analysis of neuroimaging and genetic data. This specific postdoctoral project will rely on a collaboration between INRIA ARAMIS, INRIA DYLISS, the ICM bioinformatics platform iCONICS, and the team of Martin Hoffman-Apitius at the Fraunhofer Institute (Germany) which has developed biological ontologies for neurodegenerative diseases. The post-doctoral fellow will be mainly based in Paris but will regularly travel to Rennes and Germany to interact with the other partners.

ARAMIS (www.aramislab.fr) is a joint research team between Inria, CNRS, Inserm and UPMC within the Brain and Spinal Cord Institute (ICM). The ICM is a recently created neuroscience research center within Pitié-Salpêtrière hospital in Paris. It gathers over 500 researchers covering the full spectrum of neuroscience. ARAMIS is the methodological research team of the ICM. It is a multidisciplinary research team gathering computer scientists and medical doctors. The team develops cutting-edge machine learning and image analysis approaches for neuroimaging data, in order to create new tools for diagnosis, prognosis and monitoring of brain disorders.

DYLISS (www.irisa.fr/dyliss) is a joint team between Inria, CNRS and University of Rennes 1. The team covers the field of knowledge representation and reasoning applied to molecular biology (bioinformatics and systems biology). The team develops methods based of formal systems to integrate, query and filter heterogeneous data in bioinformatics and confront it with knowledge stored in public repositories.

The Bioinformatics/Biostatistics iCONICS platform carries out development and offers access to computational methods and tools for the analysis of a variety of data, including genetics, genomics, transcriptomics, epigenomics, neuroimaging data and clinical observations. Its mission is (i) to perform the processing of omics data generated by scientific and clinical research teams; and (ii) to design and apply innovative biostatistics methods, especially integrative strategies for the analysis of multimodal and high-dimensional data.

ARAMIS Laboratory
Brain and Spine Institute (ICM)
CNRS UMR 7225 – Inserm U1127 – University Pierre and Marie Curie
Inria Paris Center
http://www.aramislab.fr
Twitter : @AramisLabParis

The team of Martin Hoffman-Apitius belongs to the Bioinformatics department at the Fraunhofer Institute for Algorithms and Scientific Computing (Germany). It has extensive experience in the field of knowledge representation for neurodegenerative diseases.

Project

Neurodegenerative diseases (such as Alzheimer's disease and Parkinson's disease) are major public health concerns. To develop new treatments for these diseases, it is crucial to identify at the earliest stage (ideally presymptomatic) the patients that will develop the disease. Genetic factors play an important role in these diseases. A major goal is to identify genetic variants and their combination that can influence disease evolution. To that aim, knowledge models of biological processes at play appear essential. First, such knowledge models could be used to inform the analysis of genetic variants (identified through sequencing and micro-array technologies), for instance by constraining statistical learning approaches. These models are also essential for the biological interpretation of the discovered variants.

The objective of this post-doctoral project is to **design approaches to integrate knowledge models of biological processes in neurodegenerative diseases in the analysis of genetic variants**. These will include both healthy and pathological metabolic and signaling pathway models. Pathways models can formalize the relationships between different gene activations in a given biological process or cellular cycle. The building of such models and their use with patient-specific data relies on approaches from the domains of ontologies, semantic web and graph-based representations. Different knowledge bases, such as that of the Gene Ontology (www.geneontology.org) for describing gene products, Reactome (www.reactome.org) for describing pathways, or OMIM and the Disease Ontology for describing pathologies have been developed by the scientific community. However, many of these models are either relatively generic or developed for other types of diseases (mainly cancer). Specific models of neurodegenerative disease have been proposed but the tools to automatically use these models for analysis of genetic data are still underdeveloped. Furthermore, knowledge about regional effects (such as effect on specific brain structures) needs to be added for better integration with imaging data. The present project will thus aim to propose knowledge models which are better adapted to these pathologies. These knowledge models will be based upon the increasing interoperability between specialized data repositories enabled by the Linked Open Data Initiative. Another important element is the ability to create a mapping between the knowledge model and the genetic data to be analyzed (such as for instance sets of Single Nucleotide Polymorphisms or structural variants). Such a mapping is non-trivial, in particular in non-coding regions and because of distant regulations. The second aim of the project will thus be to develop mapping strategies that can map knowledge models to genetic data. To address both issues, we propose to use query building tools such as the Askomics (https://github.com/askomics/askomics) tool in development by Dyliss. Askomics supports both the integration of tabulated data into an RDF triplestore, and an intuitive interface for generating SPARQL queries in order to analyze them in combination with domain ontologies. Based on this approach, the first step of the

project will be to integrate and standardize all genomic data produced in the project, and to link these datasets with external disease and pathway databases. The next step will be to extract for the local RDF database suitable gene-dependencies networks that will be used as a-priori knowledge for statistical methods. As a final step, the post-doc will represent the mapping between variants and regulated genes by taking into account additional genomic information.

Your profile

- PhD in Computer Science, Bioinformatics, Computational Biology or a related field.
- Previous work on ontologies or semantic web technologies for genomic data would be a plus. Alternatively, an expertise in genomic sequence analysis (SNP, variants) would be highly appreciated.
- Strong relational skills to interact with professionals from various backgrounds.
- Ability to synthesize informations from different sources
- Excellent written and oral communication skills

Starting date: Around November 2017
Duration: 18 months

Contacts

Olivier Colliot - Olivier.Colliot@upmc.fr
Ivan Moszer – i.moszer-ihu@icm-institute.org