SCIENCES & BIOENGINEERING SCIENCES

The Research Group

Structural Biology Brussels

has the honor to invite you to the public defense of the PhD thesis of

Daniele RAIMONDI

to obtain the degree of Doctor of Bio-engineering Sciences

Joint PhD with Université Libre de Bruxelles

Title of the PhD thesis:

The effect of genome variation on human proteins: understanding variants and improving their deleteriousness prediction through extensive contextualisation.

Promotor:

Prof. dr. Wim Vranken (VUB)
Prof. Dr. Tom Lenaerts (ULB)
Prof. Dr. Marianne Rooman (ULB)

The defence will take place on

Monday May 15 2017 at 16:00h

in Auditorium D.2.01 at the Campus Humanities, Sciences and Engineering of the Vrije Universiteit Brussel, Pleinlaan 2 - 1050 Elsene, and will be followed by a reception.

Members of the jury

Prof. Dr. Gianluca Bontempi (chairman, ULB)
Prof. Dr. Ann Nowé (secretary)
Prof. Dr. Peter Tompa
Prof. Dr. Yves Moreau (KUL)
Prof. Dr. Rita Casadio (Univ. of Bologna)

Curriculum vitae

Daniele Raimondi (1988, Sondrio Italy) studied Computer Science and Bioinformatics at the University of Bologna. After a Master thesis at Stockholm University, he started an IWT-funded Ph.D. at the VUB-ULB $(IB)^2$ Interuniversity Institute of Bioinformatics in Brussels to work on the prediction of deleterious variants in human proteins. He was also involved in Machine Learning-driven research to predict structural features of proteins, such as the formation of disulfide bonds. He is (joint) first author of 7 papers and co-author of 1 other paper in international peer-reviewed journals.

Abstract of the PhD research

Rapid technological advances are providing unprecedented insights in the biological sciences, with massive amounts of data generated on genomic and protein sequences. These data continue to grow exponentially, and they are extremely powerful when used in combination with computational approaches. A highly relevant subset of these approaches attempts to predict whether genomic variants will likely affect human health based on the corresponding amino acid variant (AAV). The state-of-the-art tools in this field give varying results and only tend to agree in the case of single variants that are strongly correlated to disease.

The aim of this work is to increase the reliability of such methods, as well as our understanding of the underlying biological mechanisms that lead to disease. We first developed machine learning (ML) based tools that are able to predict molecular features of proteins from their sequence alone. We then used these tools for *in silico* analysis of the molecular effects of known variants on the affected proteins, and integrated these data with other heterogeneous sources of information, such as the essentiality of a gene, that put AAVs into their broader biological context.

With this information, we created DEOGEN, a novel predictor that deals with the two most common forms of genomic variation, namely Single Nucleotide Variants (SNVs) and short Insertions and DELetions (INDELs). DEOGEN performs at least on par with other state of the art methods on different datasets. The method was then extended with additional contextual data and is now available as DEOGEN2 via a web server (http://deogen2.mutaframe.com/), which visualizes the predicted results for all variants in most human proteins through an interactive interface targeted to bio-informaticians and clinicians.