Poster J-24

SymBioSys: K.U.Leuven Center for Computational Systems Biology

Authors:
Yves Moreau (*K.U.Leuven ESAT-SCD*)
Sigrid De Keersmaeker (*K.U.Leuven BIW-CMPG*)
Kathleen Marchal (*K.U.Leuven BIW-CMPG*)
Chantal Mathieu (*K.U.Leuven MED-Legendo*)
Frans Schuit (*K.U.Leuven MED-GEU*)
Jos Vanderleyden (*K.U.Leuven BIW-CMPG*)
Iven Van Mechelen (*K.U.Leuven PSY-QuantPsy*)
Geert Verbeke (*K.U.Leuven MED-BioStat*)
Joris Vermeesch (*K.U.Leuven MED-CME-UZ*)
Mieke Verstuyf (*K.U.Leuven MED-Legendo*)
Bart De Moor (*K.U.Leuven ESAT-SCD*)

Short Abstract: SymBioSys, the K.U.Leuven Center for Computational Systems Biology, is an interfaculty platform that integrates expertise in bioinformatics, statistics, and life sciences. The consortium targets three problems in computational systems biology: gene prioritization by genomic data fusion, regulatory module discovery, and network inference. http://www.kuleuven.be/symbiosys.

Long Abstract:

SymBioSys, the K.U.Leuven Center for Computational Systems Biology, is an interfaculty platform that brings together top expertise in bioinformatics, statistics, and life sciences from 7 K.U.Leuven partners.

MISSION

Molecular biology has been one of the main research frontiers for several decades. Yet, much of its progress is actually based on an extremely limited understanding of biology that is slowing further progress considerably (for example, in complex diseases, such as diabetes or cancer).

To break this complexity wall, biologists have been developing high-throughput techniques (sequencing, microarrays, proteomics, and so on) to collect as much data about a biological system as possible. As a result, these technologies have brought a deluge of data upon molecular biologists. Systems biology tackles head-on the challenges of analyzing such data by directly modeling or inferring the underlying biological processes and aiming for a global analysis of the system's behavior.

Computation plays a central role in this strategy. In our view, computational biology is much more than just providing information technology resources to biologists (although this is an important aspect). We regard computation as an essential method for doing molecular biology that should be intimately integrated with high-throughput genomics techniques.

Computational biology requires close interaction with expert molecular biologists, a thorough understanding of the data, and an estimate of what type of experiments are feasible, in time and effort. Moreover, insight into advanced mathematical methods and complex statistical tools is needed as well. The combination of complementary types of expertise and different viewpoints on the same methodologies has a strong leverage effect on research quality.

RESEARCH

The consortium targets three problems in computational systems biology: (1) gene prioritization by genomic data fusion, (2) regulatory module discovery, and (3) network inference. These three problems are targeted at improving the flow from high-throughput genomics to systems biology knowledge. They correspond to increasing levels of model complexity (genes, modules, networks), but conversely also to decreasing levels of biological complexity (patients, animal cell systems, prokaryotes), which explains why there is a trade-off as to what method can currently be applied to what system. The problems have been chosen because of their high level of innovation, the already established expertise, and the biological importance in the respective fields of research.

There is an urgent need for integrated centers for systems biology in the post-genomic era. Indeed, biologists have been rapidly adopting high-throughput technologies (e.g., sequencing, microarrays) that generate massive amounts of data. But data is not information and even less knowledge. When confronted to the flood of data they generate, biologists often reach for the safety of the methodologies they understand well. Often, this means selecting a few candidate genes (proteins, RNA molecules) for validation among the many candidates detected by the high-throughput method. This selection is most of the time highly arbitrary and driven by the “hotness” of the gene, the familiarity with it, or the surprise it generates. After a long and careful validation work, the few survivors become well-characterized genes or new functions thereof that are integrated into the existing body of knowledge (e.g., pathway description). There are many problems with this non-systematic approach. First, the arbitrary selection means a bias is introduced. Second, interesting candidates might be missed or ignored (e.g., genes with limited annotation). Third, many candidates will fail to be validated. The lack of a computational model of the biological body of knowledge also means that it is hard to use existing knowledge to decide which candidates are most promising. This proposal aims at bridging this gap by developing methods to prioritize candidates, to arrange candidates into regulatory modules, and to expand those modules into full blown networks.

To achieve this overall goal, partners are committed to a tight integration. The drive for this project comes from the computational methodology. This implies that the biological studies by the biological partners will be carried out with a constant feedback from and to the data analysis (instead of a post-mortem analysis after the data has been produced) and that biological data specifically necessary for methodological developments (e.g., calibration, benchmark) will be generated whenever appropriate. For the computational partners, this implies that the computational methods will be refined until they provide true biological results.