Groningen Bioinformatics Centre

Projects for students/Bsc/Msc

Please find below a list of possible projects at GbiC. The projects are suitable for either students of Hanzehogeschool, Bsc, Msc or PHd.

For more information about the projects please contact r.breitling@rug.nl

- **Machine learning to identify biochemically plausible empirical formulae for FT-MS mass peaks, either learning a classifier for novelty detection (decision tree or SVM), or try a recursive molecule construction kit to exhaustive enumerate everything that looks reasonable based on standard building blocks**
- **De novo network construction for FT-MS data: detecting common mass differences, detect common mass difference motifs in a graph, label common transitions based on chemical knowledge, develop hyperbolic tree-based layout algorithm for network navigation, implement web version as a community service**
- **Bayesian inference method to determine the most likely metabolite network based on prior probabilities for empirical formulae, chemical transformations and measurement accuracies**
- **Web implementation of GiGA method based on up-to-date collection of annotations (incl. literature, interactomes, ontologies) and flexible probe-to-gene mapping for common organisms as a free competitive alternative to Ingenuity, including useful visualization**
- **Natural language for formulating systems biology queries as an interface to model simulators and property checkers. Explore graphical languages, too**
- **Sort genes based on expression information or literature as cancer or disease associated (or various other interesting classes, e.g. species origin, phylogenetic profile or conservation, localization, crystallizability, “scientific importance” [the hypothesis being that due to physicochemical bias only certain protein types have received proper attention so far]) and try to create successful machine learning classifiers based purely on primary sequence. Also try a novelty detection classifier that defines the boundary of the biologically feasible protein universe**
- **iGA or GiGA kernel for classification of microarray results by PCA and SVM; also include cancer/disease prognosis and diagnosis and meta-marker discovery at the iGA/GiGA level [see below]**
- **More flexible entry point to PubMed from microarray results – including identification of most relevant papers describing a subset or graph of genes, automated summary generation,**

graphical overview of relevant literature based on citations, co-occurrence of keywords, and topicality

- **Descriptive language for protein complexes that can describe their compositional dynamics in a concise, yet unambiguous manner and generate model equations from this. Also consider “selective” descriptions or simulators that allow only a subset of the prohibitively large state space to be explored**
- **Modelling signal pathway localization: What’s the impact of trafficking, scaffolding and localization on properties like stability, robustness, specificity, cross-talk, evolvability**
- **Modelling expressional instability: What are plausible mechanisms to explain the observed “loss of co-regulation” (R. Spang), “loss of consistency” (K. Vass), and “disorderly transcription” (J.W.I.M. Simons) of specific genes in cancerous states? Is the phenomenon real and reproducible at all?**
- **The possibility of reconstructing metabolic networks from genetical genomics data: explore correlation structure (at pure expression level, QTL profile, genetically “purified” expression, binary marker vector, ranked marker vector, more complex models with multimarker and interaction effects) using stochastic or deterministic simulations of realistic biochemical pathway models (Steuer et al., Bioinf. 2003; Mendes et al., Biochem Soc Trans. 2005; Zhu et al. Cytogenet Genome Res 2004; Khanin & Wit CAMDA 2004)**
- **Multiplicity calculations of protein coding sequences: how well do multiplicity extremes correlate with structural features and multiple alignments? Can we use this approach to annotate genomes? Can we identify hot proteins? Is there correlation with regulation sites?**