MEETING REPORT GSCN TRAVEL AWARD

Name: Nancy Mah

Event: 2nd Challenges Computational Biology: Gene Expression

Place: Institute of Molecular Biology, Mainz

Date: December 1, 2016 bis: December 2, 2016

Background:

The 2nd Challenges in Computational Biology meeting was inititated by Prof. Miguel Andrade at the Institute for Molecular Biology in Mainz. This year's theme was Gene Expression Data Analysis. The meeting participants included one or two invited speakers per session, and each session was complemented by short talks chosen from the registered participants.

Highlights:

Session 1: Genomics and transcriptomics. Alexandra Henrion-Caude (Necker Hospital, Paris). As the focus of the meeting was "Challenges", Dr. Henrion-Caude presented a number of challenges, with the last challenge being the building of genetic architecture and the inclusion of non-coding RNAs in these networks. As examples of non-coding RNAs with roles in disease, she presented the role of a LINE repeat within intronic sequence of SLC7A2, which mediates infantile anorexia, and potential roles of microRNAs in biliary atresia.

Session 2: Dynamical Modelling. Laurence Calzone (Insitute Curie, Paris). The Computational Systems Biology of Cancer hosts develops and hosts resources for cancer signaling pathways (https://acsn.curie.fr) and tools for modeling genetic regulatory networks (http://ginsim.org). Dr. Calzone presented work on the reconstruction of a discrete Boolean model of metastasis.

Session 3: Visualization. Wolfgang Huber (EMBL Heidelberg). Dr. Huber is well-known for his contributions to the BioConductor, the open source software repository (in R-language) for bioinformatics. Dr. Huber presented an improved method for multiple hypothesis testing, called independent hypothesis weighting (IHW), which is supposed to be more powerful than the popular Benjamini and Hochberg method.

Session 4: Chromatin. Jörn Walter (Saarland University, Saarbrücken). Prof. Walter is the coordinator of the BMBF-funded German Epigenome Program "DEEP", which also represents Germany's contribution to the International Human Epigenome Consortium (IHEC). Prof. Walter featured some of the analysis tools that have been developed within the DEEP project, such as DeepBlue (http://deepblue.mpi-inf.mpg.de/). These tools are freely available to anyone, without log-in restriction, and can be used interactively on the web interface or programmatically (API interface).

Session 5: Epigenomics. Junko Yamame (CiRA, Kyoto). Dr. Yamane presented an improved method for single-cell DNA methylome profiling (enhanced reduced representation bisulfite sequencing). Consensus DNA methylation patterns from single cells were suggested as a means to "barcode" cell types.

Session 6: Cell Identity. Waturu Fujibuchi (CiRA, Kyoto). Prof. Fujibuchi presented a machine learning method to measure drug toxicity. The method is based on the usage of human

embryonic stem cells, rather than neural differentiation cells, because the assay can be more easily standardized and run faster (4 days vs > 46 days in neural differentation cells). They used microarrays to profile the effect of 10 chemicals on hESC after 48h, and based on these microarray datasets, selected just 10 marker genes, which showed the greatest response to the chemicals. These 10 marker genes were then assayed by qPCR for 22 chemicals at 5 doses and 4 time points. The resulting qPCR data was then used to reconstruct Bayesian Networks and to train support vector machines to predict neurotoxins (NT), genotoxic carcinogens (GC), or non-genotoxic carcinogens (NGC) . The method was highly accurate (>97%) in predicting the toxicity of known chemicals. The assay was carried out on two uncharacterized chemicals to predict their potential toxicity: bisphenol-A had highest similarity to an NGC and permethrin was most similar to an NT and an NGC.

Following the conclusion of Day 1, participants of the meeting were invited to take part in a tour of the old city centre in Mainz and the Christmas Market.

Own Activities:

I gave a talk in Session 6 on the development of a molecular database for our cell-based on-line data repository, CellFinder (www.cellfinder.org), which we host at the BCRT. As I am interested in cell transitions (somatic cell reprogramming, directed differentiation and lineage reprogramming), I selected over 80 studies from public transcriptome databases for inclusion in the molecular database. The chosen datasets were manually annotated with ontology terms, which will enable efficient searches for molecular data associated to samples in the datasets. After expert manual annotation, the raw transcriptomic data were processed to yield normalized data for further analysis. These normalized datasets will become the source of data for analysis tools on the CellFinder website. As a further development of the CellFinder tools, I presented a scoring method, which aims to provide users with a "CellScore" to indicate the extent that an engineered cell (for example, an iPS cell) has transitioned from its donor cell type (e.g. fibroblast) to its desired target cell type (e.g. embryonic stem cell).

As Prof. Fujibuchi's group works on similar topics (iPS, regenerative medicine, ontology, information management) as the AG Kurtz, we could potentially work together on standardized control terms for MIACARM--minimum information about a cellular assay for regenerative medicine. Dr. Kunie Sakurai (poster presentation) from Prof Fujibuchi's group is developing MIACARM-3, to describe controlled vocabularies for engineered cell types. This ties in nicely with the sample annotation for the molecular database, and we hope to work closely with Prof. Fujibuchi's group in this direction.