**Summary of ISMB/ECCB 2017 – 21-25 July, Prague, Czech Republic**

report by Ron Zeira, reviewed by Ron Shamir, August 2017

**About the conference:** ISMB 2017 (Intelligent Systems for Molecular Biology) is the 25th annual meeting of the International Society for Computational Biology (ISCB) and is the world’s largest bioinformatics/computational biology conference. This year it joined forces with the European Conference on Computational Biology (ECCB). ISMB/ECCB brings together scientists from a large range of fields and its principal focus is on the development and application of advanced computational methods for biological problems. About 1600 participants attended the conference this year.

The conference included between four and eight parallel sessions on vast array of subjects. This year, the conference structure was changed, and the various tracks were focused around ISCB COSIs (Communities of Special Interest). COSIs include interest groups such as microbiome, network biology, regulatory genomics, high throughput sequencing, translational medicine and more. In addition, the conference included workshops and special sessions on machine learning, computational immune oncology and more. Talks and keynotes were very interesting although the conference structure made it hard to follow from time to time.

The conference proceedings are available as a special issue of *Bioinformatics*. Videos of the talks should appear in the conference [website](#) soon.

I will describe some highlights and talks I think may be relevant/interesting to the group.

---

**Topics: cancer, multi-omics, drug response**

**Modeling pharmacogenomic interactions in cancer - Lodewyk Wessels, Netherlands Cancer Institute.**

The CAMDA keynote speaker discussed methods to predict the drug response based on various molecular features, such as mutation status, copy number alterations, methylation and gene expression profiles. Using “flat” classifiers that combine all molecular features result in models that are almost exclusively based on gene expression. Their new approach has two stages; the first stage explains response using upstream features (mutations, copy number, methylation and cancer type) and the second stage explains the remainder using downstream features (gene expression). This part is presented in the following paper. See picture below.

To discover the hierarchy between data types, they used the RV coefficient to measure the similarity between data matrices. Then, the PC algorithm was used to infer a DAG structure between data types. Their newest approach uses tree phases of classification with upstream data types, proteomics and gene expression.
Topics: cancer, multi-omics, precision medicine

**Genome-guided Framework for Personalized Cancer Treatment - Krishna Kalari, Mayo Clinic**

This talk presented **PANOPLY - Precision Cancer genOmic rePort: single sampLe inventory**, a computational approach to integrate both germline and somatic data obtained from multi-omics platforms for an individual of interest, and the analysis of the data in the context matched-control samples (see picture). PANOPLY takes pre-processed data (from TCGA and neo-adjuvant data sets) such as germline DNA sequence, SNVs, small INDELs, CNVs, fusion transcripts, along with RNA seq gene expression data to build an integrated network that ranks the cancer gene-sets for a given patient. From these data, the PANOPLY system predicts the most promising drug targets. Available as R package upon request.
**Topics:** cancer, multi-omics, precision medicine, networks

KnowEnG: Knowledge Network-guided analysis of genomics data on the Cloud - Saurabh Sinha, University of Illinois Urbana-Champaign

This talk introduced KnowEng, a cloud based portal for genomic data mining. It enables the user to upload its processed data (e.g., gene X samples matrix) and perform data mining and machine learning tasks on those data, with results being displayed in a highly intuitive and user-friendly fashion. KnowEnG’s uses prior knowledge in the form of a massive heterogeneous network called the Knowledge Network, which aggregates information from nearly 100 externally curated databases. Example pipelines include: sample clustering, gene prioritization, gene-set characterization, literature mining, signature analysis, phenotype prediction, regulatory network reconstruction, multi-omic analysis and more (see image below).
**Topics: cancer, structural variation, copy number, enhancer-promoter interaction**

**Decoding genome structure - Jian Ma, Carnegie Mellon University**

The first part of the talk was devoted for *Weaver*, an algorithm for the quantification and analysis of allele-specific copy numbers of structural variations (SVs). Weaver uses a Markov random field to estimate joint probabilities of allele-specific copy numbers of SVs and their inter-connectivity based on paired-end whole-genome sequencing data. However, one major limitation is that not all SVs identified by Weaver are phased. Their recent *RECOMB* paper develops a general convex programming framework that predicts the interconnectivity of unphased SVs with possibly noisy allele-specific copy number estimations as input.

The second part of the talk discussed how to exploit sequence-based features for predicting enhancer–promoter interactions. They develop a new computational method (called *PEP*) to predict enhancer–promoter interactions based on sequence-based features only, when the locations of putative enhancers and promoters in a particular cell type are given. PEP is built on two modules, one based on known TF binding motif and the other is based on embedding enhancer/promoter regions into a new feature space. The method is effective in predicting enhancer–promoter interactions as compared to the state-of-the-art methods that use functional genomic signals.

**Topics: cancer, autism, multi-omics, precision medicine, patient classification, networks**

**Pathway analysis of genomics data. From correlation to causation to drug discovery - Gary Bader, University of Toronto**

The NetBio keynote presented *netDx*, a framework for patient classifiers based on similarity networks. *netDx* provides workflows to build patient class predictors from known examples. *netDx*
first converts patient data into a set of patient similarity networks for each data type. Second, all the networks are integrated into a single network or data base that can be used for prediction. In addition, feature selection is performed by scoring each network based on its information content. A classifier is then built for each class, and a new patient is assigned to the highest rank class. The first example they give is classifying autism based on copy number variants. The second example identifies luminal A breast cancer subtype by integrating gene expression, CNV and methylation data (see schematic outline below).

**Topics:** cancer, phylogeny, heterogeneity, sampling

**Tumor Phylogeny Inference Using Tree-Constrained Importance Sampling - Gryte Satas, Ben Raphael's group, Brown University**

Cancer is an evolutionary process, and bulk tumor tissue contains heterogeneous mixtures of cells. Many methods have been proposed for clustering cell populations and/or reconstructing a phylogenetic tree based on SNV frequencies in each sample. Some of these methods are based on purely probabilistic models while other rely on combinatorial constraints.

This paper presents Probabilistic Algorithm for Somatic Tree Inference (PASTRI), a new algorithm for bulk-tumor sequencing data that clusters somatic mutations into clones and infers a phylogenetic tree that describes the evolutionary history of the tumor. PASTRI first uses a probabilistic model to find likely cell fractions. Then it uses a combinatorial algorithm to enumerate possible trees and their likelihood (see method outline below). PASTRI outperforms other cancer phylogeny algorithms in terms of runtime and accuracy on simulated data. Furthermore, it was tested on CLL data.
**Topics: cancer, phylogeny, single cell, MCMC**

Reconstructing tumor evolution from single-cell sequencing data - Katharina Jahn, Niko Beerenwinkel's group, ETH Zurich

Single cell measurement technologies suffer from noise and incomplete data. This paper, recently published, presents SCITE, a stochastic search algorithm to identify the evolutionary history of a tumor from noisy and incomplete mutation profiles of single cells (see picture below). Unlike most phylogenetic reconstruction methods, SCITE does not make the “infinite sites” assumption and has a probability for back mutations. Their recent unpublished work assessed the probability of reoccurring mutations at $10^{-4} - 10^{-6}$, which makes the “infinite sites” assumption questionable.

**Topics: cancer, phylogeny, clustering, optimization**

Deconvolution of heterogeneous bulk tumor genomic data via structured mixed membership models - Russell Schwartz, Carnegie Mellon University
Genomic deconvolution provides a strategy for reconstructing models of cancer tumor progression in the face of limited amount of single cell data. The method, presented in the following paper, use an interpretation of the mixture problem as that of reconstructing geometric objects called simplices. Furthermore, they search for structured unions of simplices called simplicial complexes that one would expect to emerge from mixture processes describing branches along an evolutionary tree. The method outline is presented in the following picture. They have applied their method to TCGA breast cancer samples on RNAseq to recover known subtypes. In their latest work, a newer version of the method was applied to BRCA CNV data.

---

**Topics:** cancer, variant calling, multi-omics, networks

**Network-based integration of multi-omics data for prioritizing cancer genes - Niko Beerenwinkel, ETH Zurich**

Calling mutations for cancer sequencing data is a challenging task with a lot of methods proposed for it. Using simulations based on kidney tumor data, their recent work compared the performance of nine state-of-the-art variant callers. Then, they have integrated the ranks of the variants called by the other tools which yielded the best performance, outperforming all individual tools.

The second part of the talk reviewed different methods for detection of cancer driver mutations and pathways. Some tools rely purely on prior knowledge like known pathways and public data bases, while some try to learn the pathways de novo based on their combinatorial patterns of occurrence. Their new method, NetICS, integrates different tumor omic profiles, predicts how the same pathway is affected in different ways in different patients, and prioritizes mediator genes.

---

**Topics:** cancer, drug response, precision medicine, machine learning, regression

**Ask the doctor - Improving drug sensitivity predictions through active expert knowledge elicitation - Iiris Sundin, Samuel Kaski and Pekka Marttinen's group, Helsinki Institute for Information Technology**

In precision medicine, we want to predict the efficacy of a drug for a given individual based on genomic data. However, identifying features on which to base the predictions remains a challenge. Using expert knowledge may improve prediction but is very demanding. This work introduces a probabilistic model that can incorporate expert feedback about the impact of genomic measurements on the sensitivity of a cancer cell for a given drug (see picture below). The model is based on sparse linear regression with expert knowledge plugged into the model as prior. They
suggest two schemes of asking the expert, trying to balance between improving the model prediction (exploration) and use of the expert's time (exploitation).