Genome **Biology**

MEETING REPORT

The rise of translational bioinformatics

Yves A Lussier[1-5],* and Haiquan Li[1,2]

Abstract

A report on the 20th International Conference on Intelligent Systems for Molecular Biology (ISMB), held at Long Beach, California, USA, July 15-17, 2012.

Keywords Biomarkers, complex diseases, computational medicine, drug repositioning, mechanism classifiers, next-generation sequencing, off-target mechanisms, translational bioinformatics.

The Intelligent Systems for Molecular Biology conference is the largest conference for bioinformaticians and computational biologists. It promotes presentations grounded on meaningful algorithm innovations and is motivated by the analytic problems arising from advances in biotechnologies. Translational bioinformatics, a key theme of this conference, has radically progressed due to the recent advent of next-generation sequencing technologies (for example, DNA-Seq, RNA-Seq), which provide the first affordable opportunity to measure the individual differences in therapeutic response and disease mechanisms at the nanoscale level. As a direct consequence, novel algorithms for big data (larger than terabytes) as well as those for personal polymorphism (a few bytes) have become mandatory for understanding the genomic architecture of an individual's therapeutic response in a disease of complex inheritance that is beyond conventional single-gene approaches. Here, we highlight three groups of translational bioinformatics presentations addressing mechanisms and clinical challenges associated with the treatment of diseases of complex inheritance in the era of increasingly affordable personal 'omics.

Complex inheritance of disease polymorphisms, genetic aberrations or biomarkers

The substantial investment in genome-wide associations studies (GWASs) has revealed approximately 7,000 new SNPs associated with complex inheritance traits. Unsurprisingly, the majority of these polymorphisms account for a <1% increase in the odds of developing a disease, and together they account for a small proportion of the inheritance of complex diseases. A number of speakers provided original efforts aimed at discovering buried inheritability. Gregory Darnell (University of California, USA) proposed a novel association study model that incorporated prior information as a weight into GWASs and demonstrated improved statistical power and detection resolution on simulated data. Gunter Klambauer (Johannes Kepler University, Austria) developed a mixture of Poisson statistics with Bayesian estimates of parameters and hidden Markov modeling (named cn.MOPS) to discover copy number variants in next-generation sequencing data. Experiments on simulated databases derived from different sources demonstrated the sensitivity of this method. Similarly, Atushi Niida (University of Tokyo, Japan) employed a Poisson-binomial statistic to model the count of recurrence aberrations in unpaired cancer samples (where paired normal tissue is absent). This approach is remarkable since all current methodologies require paired normal tissue to identify genetic aberrations such as loss of heterozygosity and copy number variants. In addition, this group provided a superior evaluation using both simulated and retrospective biological experiments, where the gold standard uses paired cancer-normal tissue to confirm the validity of the method. The DELISHUS method developed by Derek Aguiar (Brown University, USA) leverages graph theoretic maximum cliques to identify deletion polymorphisms in next-generation sequencing, while current approaches misinterpret the data as homozygosity. The evaluation consisted of a simulation with autism data. Using multitask regression with LASSO regularization, Seunghak Lee (Carnegie Mellon University, USA) improved expression quantitative trait loci (eQTL) studies to identify the contextual structural versus marginal effect of each SNP. The regression model is inclusive of epistatic interactions among genetic variations and it was applied to a yeast dataset. A concordance statistic between predicted SNP-SNP interactions and their associated Gene Ontology annotations provided biological evidence to support the accuracy of the results. In an ambitious and

*Correspondence: ylussier@uic.edu
[1]Department of Medicine, University of Illinois at Chicago, Chicago, IL 60612, USA
Full list of author information is available at the end of the article

BioMed Central

well-conducted study incorporating three-dimensional imaging of the brain (positron emission tomography, computed tomography) with SNP array data, Hua Wang (University of Texas at Arlington, USA) identified candidate biomarkers of complex diseases. The methodology consists of sparse multimodal multitask machine and was evaluated with a fivefold cross-validation in clinical samples.

Discovery of biological mechanisms underpinning diseases of complex inheritance

The causal molecular interplay underpinning the pathophysiological mechanisms of complex diseases remains elusive despite the recent discovery of thousands of associated polymorphisms. The following presentations addressed this phenome-genome gap. Chirag J Patel (Stanford University, USA) identified gene-environment interactions contextual to diabetes type 2 with an original integration of three datasets: gene-environment interactions, disease-SNP associations and disease-environment associations. The evaluation remains qualitative: among the six predictions, one had been characterized previously. Using mutual information of the transcriptome, Pavel Sumazin (Columbia University, USA) reported a new regulatory mechanism at the miRNOME scale, known as sponge modulators, which consists of 248,000 micro-RNA-mediated RNA interactions that collectively regulate canonical oncogenic pathways in glioblastoma. He confirmed experimentally that the sponge modulators are mechanistically responsible for loss of *PTEN* expression. A sponge modulator is a miR program-mediated post-transcriptional regulatory (mPR) network consisting of multiple interacting RNAs (including both mRNA and non-coding RNA) via sharing binding sites of their common microRNAs. mPR genes regulate the expression of each other via network interactions leading to robust co-expression modules. Ohad Balaga (The Hebrew University of Jerusalem, Israel) proposed a network theoretical model to disrupt human pathways with the minimum number of microRNAs to demonstrate their susceptibility to fragmentation. He computationally imputed that more than 75% of the biological pathways expressed can be disrupted with two or three correctly selected microRNAs, which are pathway specific. He provided a description of the disruption in Alzheimer's disease as a biological rationale supporting the predictions.

Structural biology provides further insights into disease mechanisms. Wilson Wen Bin Goh (Imperial College London, UK) has developed a highly original algorithm to score the expression of biomodules consisting of protein complexes using mass spectrometry data and protein interaction networks. Using liver cancer data, he further provided a proof of concept that unsupervised clustering over all assessable protein complexes (proteome-wide) can discriminate between late- and moderate-stage tumors. In other words, he showed that biological mechanisms can substitute gene-level classifiers and provide more interpretable results for biologists and clinicians. The approach could potentially identify more biologically informative signatures because the expression level of a single protein may be marginal, which prevents traditional approaches from identifying significant signals. In a comprehensive translational report, Haiyuan Yu (Cornell University, USA) integrated three databases - three-dimensional protein structures, protein-protein interactions and disease-associated mutations - to infer candidate disease-associated interactions. He showed that in-frame mutations (including missense, insertion and deletion) are enriched at the interfaces of disease-associated protein interactions. Conclusively, the observed protein interactions were abrogated when the mutations were experimentally introduced into yeast two-hybrid assays.

Drug repositioning and drug toxicities

Repositioning drugs can yield new patentable drugs with medical uses valuable to the pharmaceutical industry, while *in silico* compound toxicity assays can motivate the design of animal experiments, thus preventing hazardous clinical trials. Lei Xie (The City University of New York, USA) employed protein-ligand docking and molecular dynamics simulations to search for weak-binding off-targets for nelfinavir, a drug originally used as a HIV protease inhibitor. Xie discovered kinase-like proteins and computed the mechanism by which nelfinavir inhibits carcinogenesis and metastasis pathways via the phosphoinositide 3-kinase/Akt signaling pathways, as confirmed by homogeneous time resolved fluorescence (HTRF®) inhibition assays. In related work, Lun Yang (GlaxoSmithkline, USA) conducted a comparative study between the chemical-protein interactome (docking) of clozapine and its analog olanzapine, because the latter has a lower incidence of life-threatening agranulocytosis. He thus identified an off-target of clozapine: HSPA1A (heat shock 70 kDa protein 1). In the connectivity map data of the HL60 promyelocytic leukemia cell line (myelocytes are granulocytes), he observed downregulation of oxidoreductase production among HSPA1A-interacting proteins, consistent with agranulocytosis. Indeed, quinone oxidoreductase 2 (NQO2) has previously been reported to be associated with clozapine-induced agranulocytosis.

This meeting provided the opportunity for translational bioinformaticians to substantially improve the understanding of complex diseases and their therapies. The microscope has provided momentous advancements in cellular biology and pathology, and next-generation sequencing (RNA-Seq and DNA-Seq) seems poised to

propel translational bioinformatics towards paradigm shifting discoveries, particularly in the 'omics mechanisms of personal therapeutic response.

Abbreviations

mPR, miR program-mediated post-transcriptional regulatory; SNP, single nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Medicine, University of Illinois at Chicago, Chicago, IL 60612, USA. [2]Institute for Interventional Health Informatics, University of Illinois at Chicago, Chicago, IL 60612, USA. [3]University of Illinois Cancer Center, Chicago, IL 60612, USA. [4]Center for Clinical and Translational Science, University of Illinois at Chicago, Chicago, IL 60612, USA. [5]Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60612, USA.