American Laboratory®

Date: November 2014

Headline: Success of the Third sbv IMPROVER Network Verification Challenge

Byline: Larisa Fedorova

The accomplishment of the Human Genome Project in 2001 created the illusion that the time would soon come when everyone would go to the doctor's office, read his or her genome, and get treatment suited exactly to his or her individual needs. From May 2014 the cost of population scale whole genome sequencing with the new Illumina's HiSeq X Ten machine is down to $1000. However, the hopes of easily identifying disease-related mutations and providing personal medical treatment for a number of complex multifactorial disorders have faded.

The functional genomics approach taken by big pharma in the last decade resulted in the development of several drugs that are very efficient for a limited number of patients with specific mutations underlying certain diseases. For the rest of patients, these promising drugs either do not provide adequate treatment or are potentially dangerous.

Numerous genome-wide association studies (GWAS) identified multiple loci involved in the development of many human chronic diseases. However, weak correlations between particular single nucleotide polymorphism sets and disease phenotypes demonstrated that a majority of autoimmune and chronic diseases are multifactorial. Predisposition to these diseases to some extent is determined by genetic variations, but disease phenotype is dependent on thousands of interrelated molecular interactions, transcriptional regulations, intricate biochemical pathways, physiological adaptations, and environment. Thus, the future of personalized medicine seems to be dependent on a complex systemic analysis of genetic, epigenomic, transcriptomic, proteomic, and metabolomic data, which, in some circumstances, would allow identification of major disease-driving mechanisms for a particular patient.

Systems medicine methodology

Such a holistic approach for predicting and treating diseases requires the application of systems medicine methodology to elucidate convoluted biological mechanisms and their dynamic behavior. Systems biology methodologies rely on both experimental quantitative, imaging, and phenotyping data and computational algorithms for data analysis. The recent proliferation of publications with high variations in experimental settings and analytical techniques has already compromised the efficiency of the peer-review process of academic papers (marked by a large number of retractions even from high-ranking scientific journals) and, consequently, may result in decreasing quality of the high-throughput data that is in use in systems biology. Thus, mechanistic models and predictions, generated by systems biology

approaches, have to be evaluated and verified prior to the applications of this knowledge for basic and clinical research and trials. Still, there are very few proper systems biology/medicine studies published in the first place since the biomedical community has not yet agreed on how to best share the large-scale data sets with the community.

sbv IMPROVER challenge

To resolve this problem in academia, a methodology of engaging interested researchers to collaboratively and objectively verify the results and claims in system biology, called crowdsourcing, has been used to assess the quality of models and predictions from high-throughput data. Recently, scientists at IBM Research (Yorktown Heights, NY) and Philip Morris International R&D (Neuchâtel, Switzerland) developed a methodology for verifying models generated by systems biology, adjusted specifically for the industrial setting, called sbv IMPROVER (System Biology Verification: Industrial Methodology for PROcess VERification; https://sbvimprover.com). The project is designed as a series of challenges, with critical blocks of research workflows made available to participants. Members of the sbv IMPROVER team have been working intensively to increase the size of the crowd participating in the project and three challenges have already been successfully accomplished. All were open to an unlimited number of independent participants around the world.

The latest challenge—Network Verification Challenge—is devoted to the analysis of a complex biological model of chronic obstructive pulmonary disease (COPD). Biological network models generated for verification have been built based on the manual curation of biomedical literature and on data-driven reverse causal reasoning methodology. The project is formalized and encoded using the Biological Expression Language (BEL). Both the network models and the BEL were developed by Selventa, a pioneering personalized healthcare company.

I decided to participate in the Network Verification Challenge despite the fact that I did not have any previous experience in the field of lung physiology and pathology. Three reasons and motivations led me to take part.

First, participation in sbv IMPROVER is an exciting opportunity to study the basic principles, methodology, and tools of systems biology. With an explosive flow of scientific publications and the lack of profound verification of systems biology predictions, participation in this challenge-based approach is a great educational experience. For me it was relatively easy because I had strong experience in the related fields of bioinformatics and genomics. The project provides a clear vision as to how to increase efficiency of accumulated scientific data, add scrutiny, and eliminate the subjective bias in the interpretation of results. The Network Verification Challenge allows us to work on network models using a high-performance online platform and the BEL syntax is easy to learn and fun to use.

One of the main characteristics of the challenge was the manual curation of data used since the entire biomedical literature database was mined for the networks creation. Specifically, all papers presenting tumorigenic cell lines, species other than human, mouse and rat, and cells/cell lines not presented in the lung were excluded. Manual evaluation of papers, which helped to produce network edges and nodes, allowed an independent and critical analysis of the methods, approaches, and data supporting outcome of the research, and its significance for the models. All in all, 50 biological networks divided into 15 groups comprised of around 2000 nodes and 2500 edges were created for verification.

Secondly, my special research interests in medicine have long been chronic kidney disease and preeclampsia. Studying these complex disorders I focused on the molecular networks related to oxygen utilization/oxidative stress/hypoxia. In this respect, lung is a unique tissue compare to all others in multicellular organisms. Lung cells are in direct contact with the atmospheric oxygen tension while in the inner organs the oxygen tension is much lower. In humans, oxygen tension in the arterial blood range from 75 mmHg to 100 mmHg at sea level.

Moreover, some organs are normally hypoxic. This is especially true for the kidney where, in spite of high blood flow (20% of cardiac output), the oxygen tension in the cortex is 30 mmHg. In the medulla, oxygen saturation is even lower, not higher than 10 mmHg, because of oxygen shunt diffusion between arterial and venous vessels. Thus the oxygen handling by the renal epithelial cells in that region is particularly adapted to such levels of oxygen.

Compromised coordination of oxidative phosphorylation, antioxidant response by NRF2 systems, HIF1 signaling pathways, and mitochondrial damage may result in bioenergetics imbalance, inflammation, and renal cell loss. Interestingly, fibrosis and inflammation in almost all kidney models for chronic renal disease most often are presented in corticomedullary junction of the kidney.

Regarding preeclampsia, adequate oxygen handling by the maternal placenta is absolutely critical for embryonic development and untimely modulation of oxygen delivery by maternal blood can cause miscarriages and presumably preeclampsia. Paradoxically, in the vast majority of in vitro studies with use of cell culture we often do not pay attention to the fact that ambient oxygen tension in the cell culture incubator is very different from the in vivo environment. In contrast, lung epithelial cells, fibroblasts, and immune cells can be studied most adequately in vitro. From this point of view I analyzed most intensively the NRF2 and hypoxia molecular network models in the lung.

The third reason for my participation in this project is that I am in the middle of transition from academia to industry. Currently I am establishing my own genomic/biomics company (GEMAbiomics LLC). In this respect, the sbv IMPROVER approach has been very helpful in demonstrating the specific needs of industry, such as speed and protection of proprietary data constraints, as well as market considerations and consumer protection.

As one on the most actively contributing participants in the Network Verification Challenge, I was invited to participate in an international networking jamboree session to discuss controversial edges and review added evidences. These sessions were open, unbiased, and empowered by the presence of recognized experts in areas of particular networks. For example, during verification of hypoxic networks, the consensus was reached that hypoxia per se is not present in the normal lung and in early stages of COPD development. However, the network models of hypoxic signaling still are open for analysis and verification since it is possible that hypoxia-inducible factors could be activated in immune cells even under normal oxygen levels by the Warburg effect mechanisms.

Overall, I found the Network Verification Challenge to be a unique and very important initiative, empowering biomedical scientists and healthcare providers with an integrative systemic methodology for understanding the complex nature of progressive chronic diseases and developing personalized therapeutic intervention.

Larisa Fedorova, Ph.D., is with GEMAbiomics LLC, 4121 Halifax Rd., Toledo, OH 43606, U.S.A.; tel.:+11 419 536 0688; e-mail: lvfedorova3@gmail.com

-- ENDS --