



**Date:** 21 December 2015

**Headline:** SBV IMPROVER Project Launches New Systems Toxicology Challenge

**Byline:** Uduak Grace Thomas

The organizers of the Systems Biology Verification: Industrial Methodology for Process Verification in Research (SBV IMPROVER) challenges have begun accepting submissions for a new systems toxicology challenge launched last week which focuses on identifying blood gene expression signatures that could serve as markers of smoking exposure.

The IMPROVER challenges, which are led and funded by Philip Morris International's research and development arm, are designed to provide a robust methodology for verifying systems biology methods and results in the context of industrial and academic research.

This particular challenge offers computational scientists who are developing predictive modelling technologies an opportunity to pit their methods against each other on data drawn from human clinical and mouse inhalation studies. It features two sub-challenges: For the first, participants will be expected to derive models that can predict gene signatures for smoking exposure using human whole blood gene expression data. They'll compare data from smokers and non-smokers as well as categorize non-current smokers as former smokers and never smokers. In the second sub-challenge, participants will also have access to blood gene expression data from mice and will attempt to find species-independent gene signatures that can predict smoking exposure.

Participants will be provided with training, test, and verification microarray-based gene expression datasets from studies run by researchers at PMI's research and development arm and elsewhere. They can also use additional public blood gene expression datasets to train their computational models.

This challenge provides an opportunity to test available prediction methodologies. By receiving robust datasets, computational scientists should be able to develop models that are "sensitive enough and specific enough that it doesn't matter [where] the data comes from," Julia Hoeng, director of systems toxicology for PMI R&D and PMI project leader of the SBV IMPROVER project, told GenomeWeb. "The computational signature basically will be strong enough [so that] when you give me another dataset generated anywhere else ... the signature will be able to identify that," she said. If the models work the way they should, they could be applied more generally to classify subjects based on their exposure to other toxic substances besides tobacco, for example asbestos or exhaust.

It's also a chance to acquaint members of the toxicology community with the benefits of using big data and associated technologies in their projects, Hoeng added. "Toxicologists tend to just ... look at some very specific markers, or maybe they collect blood and look at a few specific proteins," she said. Challenges like this are an opportunity to say to toxicologists "don't be afraid of large scale data, we know how to produce it in a reliable manner and there are plenty of sophisticated people in the world of machine learning who can actually take the data

and make sense of it ... and the results of their evaluation can actually be very valuable in the future for risk assessment."

Submissions will be accepted until April 29, 2016 from researchers in industry and academia. Entries will be scored against gold-standard datasets by an independent panel made up of systems biology researchers from the National Technical University of Athens, ETH Zurich, and the Leibniz Institute for Farm Animal Biology. To help with the evaluation, the organizers are working with SciPinion, a PMI partner that provides forums for scientists to present their research to groups of experts who then anonymously provide feedback on the studies. It's an approach PMI has used to obtain independent verification of some results of its own studies to prevent potential biases and conflicts of interests in its research efforts aimed at assessing reduced risk product candidates for commercialization.

Best performers in the challenge will have the opportunity to contribute to peer-reviewed scientific publications and will present their work at the 2016 Intelligent Systems for Molecular Biology conference, which will be held in Orlando, Florida next July.

This latest challenge is the fourth from the group. Previous challenges included one which asked participants to assess and verify computational approaches for classifying clinical samples across four disease areas — psoriasis, multiple sclerosis, chronic obstructive pulmonary disease, and lung cancer. The second challenge sought to explore how well biological processes in mouse or rat models translate to humans. A third challenge, which ran for two rounds, asked the community to review computational network models for use in toxicological assessment, drug discovery, and research in the context of respiratory diseases.

The IMPROVER organizers have also begun mulling a second systems toxicology challenge, in partnership with the OpenTox consortium, that will focus on computational methods for evaluating high content screening data, Hoeng said. If they move forward with it, the new challenge could launch in the latter half of 2016.

-- ENDS --