

**Date:** 12 October 2012

**Headline:** IMPROVER-ing Data Verification for Systems Biology

**Byline:** Alison Proffitt

There was CASP (see, On the CASP of a DREAM). Then there was DREAM (see, DREAM Project Vision Expands). And now, IMPROVER. The latest algorithm competition is tackling data verification in systems biology research.

IMPROVER—Industrial Methodology for PROcess VERification in Research—is a project of IBM Research and Philip Morris International (PMI) R&D. The two organizations wanted to find a way to enable more rigorous evaluation of large and complex datasets and verify conclusions reached by peer-review.

“Think of [IMPROVER] as a very critical peer, almost as a peer reviewer, that uses challenges in order to see if what you are saying is right or wrong,” Gustavo Stolovitzky, manager, Functional Genomics and Systems Biology, IBM Computational Biology Center, explained to Bio-IT World.

Biosciences are generating huge quantities of data at unprecedented speeds, added Hugh Browne, director, research & development, PMI. “Getting value from those data is becoming increasingly more difficult and complex. What IBM and PMI are trying to do in collaboration is come up with an approach which allows scientists from around the world to collaborate and get more from these very big datasets than perhaps they would using the resources they have individually within their own laboratories.”

The collaboration began in 2009 and the two groups published the IMPROVER approach in Nature Biotechnology in 2011 (Nature Biotechnology 29, 811–815 (2011)).

“Applying this methodology first requires identifying the building blocks of a research workflow,” the authors explained in the commentary. “Some might involve generating biological measurements, others analyzing data... The idea behind IMPROVER is to test each key method at crucial junctures of a research workflow by posing challenges designed to see whether or not the process works at the necessary level of accuracy.”

After a small, introductory symposium in 2011, in March 2012, the IBM-PMI IMPROVER team launched the first of a series of public challenges to “road test this methodology and see how it can be applied in the scientific community,” said Browne.

The Diagnostic Signatures Challenge aimed to assess and verify computational approaches that classify clinical samples across four disease areas: psoriasis, multiple sclerosis (MS), chronic obstructive pulmonary disease (COPD) and lung cancer.

Participants received a list of publically-available datasets to use as training sets and an unpublished, completely independent test set, Stolovitzky explained. “We created a panel of microarray data in [four] areas of disease... and we invited the community to predict which samples came from someone with the disease and which samples came from a control.”

The first challenges were designed to see the extent to which the best methods depended on the endpoint, or disease condition, said Stolovitzky. He admits that the IMPROVER team did not know what to expect, but the entrants did surprisingly well.

Fifty-five teams participated in the Diagnostic Signature Challenge, said Manuel Peitsch, VP, biological systems research, PMI. Submissions were scored by the IBM Computational Biology Centre and independently reviewed by the IMPROVER Scoring Review Panel.

The overall best performing team was Adi L Tarca and Roberto Romero of Wayne State University, Detroit. Second and third best performances went to Mario Lauria (Computational Systems Biology, Rovereto, Italy) and Michael Unger, Preetam Nandy, Kushal Kumar Dey, Christoph Zechner & Heinz Koepl (ETH, Zurich, Switzerland) respectively.

Sub-challenge awards were given in each disease area to Sol Efroni & Rotem Ben-Hamo (Bar-Ilan University, Israel ) for lung cancer; Kai Wang, Ji-Hoon Cho & Alan Lin (Institute for Systems Biology, Seattle) for psoriasis; Steve Horvath & Lin Song (University of California, Los Angeles) for COPD; and Mario Lauria (Computational Systems Biology, Rovereto, Italy) for COPD.

The vast majority of the teams came from academia, Peitsch pointed out, and the overall winners and best performers came from all over the globe. “What’s interesting is that overall, it is those places that put—either at the national or the institutional level—a heavy emphasis on systems biology that had the best performers.”

### **Any Industry**

“We meant to be as relevant to an industry where IMPROVER could be deployed as possible,” said Stolovitzky. “We are creating a methodology that can be deployed in any industry that applies systems biology. We are doing this pilot and showing its value within the context of the research and development in systems biology being done at Philip Morris International.”

The best performers were named at the beginning of October at the IMPROVER Symposium\*, but the event was not just a symposium; it was a workshop too. The IMPROVER team listened to entrants discuss their algorithms—what worked and what didn’t—in order to extract lessons for other applications. “We had lengthy sessions of discussion that were very, very active and lively!” said Stolovitzky.

Although no single approach was consistently better than the rest, some methods seemed to have an edge over others. The IMPROVER team is still working on distilling the best practices. The findings revealed in the first symposium will serve as a foundation for subsequent challenges to some extent, though each challenge over the next two to three years will have a different focus.

A grand challenge will focus on COPD and will serve to bring together all of the pieces. The team also plans to publish their findings.

“The competition provides a framework for scientists from all around the world to collaborate, to have common problems, to bring them together at a symposium to try to get consensus on what the

right way to approach that particular problem is, which they can then take back into their own environments whether that be industry or academia, to drive forward a verified approach to solving that particular problem,” said Browne. “It’s about using the wisdom of the crowds to solve big challenges in biology.”

The second IMPROVER challenge is due to launch in the second quarter of 2013. The Species Translation Challenge will explore the predictive value of data from different species in the quest to understand more about biological processes.

\* The IMPROVER Symposium, 2-3 October 2012, Boston, Mass.

-- ENDS --