SBV IMPROVER Challenge Uses DNA Methylation Data to Categorize Smoker Samples

Jun 21, 2017

Uduak Grace Thomas
Premium

NEW YORK (GenomeWeb) – The most recent iteration of the Systems Biology Verification: Industrial Methodology for Process Verification in Research (SBV IMPROVER) highlighted the challenge and complexity of using epigenetics data to classify and categorize biological samples.

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

This most recent one, which culminated in a multi-disciplinary symposium held in Tel Aviv, Israel in May, asked participants to classify samples from different systems toxicology studies based on the epigenomic impact of cigarette smoke cessation and the aerosol from so-called modified risk tobacco products or reduced-risk products — products that present, are likely to present, or have the potential to present less risk of harm to smokers who switch to these products versus continued smoking.

"One of the reasons we wanted to use epigenetics data for this challenge is … we wanted to add one layer of complexity," said Nicolas Sierro, manager, genomics at Philip Morris International and one of the creators of the challenge. "We had previously done challenges based on transcriptomics data only so we didn't want to repeat the same type of challenge or process." Furthermore, "epigenetic data is more challenging to interpret so we wanted to confront the community with that kind of data."

Moreover, "there are a lot of things that do not occur on [the] RNA level but occur on the epigenetic and genetic level," Mohamed Amin Choukrallah, a genomics scientist at PMI and a co-creator of the epigenetics challenge, added. "Sometimes you have epigenetic activation that occurs before the change in RNA expression and this is something you cannot see if you analyze only transcriptomic data."

Specifically, this challenge asked participants to extract signatures from internally generated DNA methylation data gleaned from mouse studies, said Stéphanie Boué, a senior computational biologist at Philip Morris International. They then had to try to use these signatures to classify new sets of samples from other studies. One task for participants was to see if there was enough biological signal in the DNA methylation data to separate the groups. Participants were also asked to identify genes in differentially methylated sites and then use transcription data from those genes to categorize samples into groups.

While several groups downloaded the data provided for the challenge, only two ultimately submitted entries. An entry from Hagit Philip, a faculty member in the systems biomedicine lab at Bar-Ilan University, was selected as the winner.

Overall, the results were comparable to what the organizers expected to see based on the data, but did not really improve on PMI's internally generated results, serving to highlight the complexity of using epigenetics data for classification, according to Boué. "Epigenomics is really tricky [and] unfortunately, it seems impossible to separate exposure groups based on epigenetic data alone. It is much easier to separate groups based on transcriptomic data," she said. "We do see different levels of methylation overall but its more random … and so it's difficult to assess what is the real impact."

Tamir Tuller, head of the laboratory of computational systems and synthetic biology at Tel Aviv University and a keynote speaker at the symposium, pointed out that the short time frame for the challenge was a factor for some groups. Participants had about a month from the launch of the challenge until the closing date on April 10 to generate their results. "The challenge was a little bit complicated and time consuming so some [researchers] decided not to submit because they didn't feel that they were ready," he said.

Still the fact that the teams were able to generate comparable results to PMI's within such a short time is worth noting, according to Choukrallah. Furthermore, some of participants used completely different approaches to get comparable results providing PMI with new computational approaches for gleaning results from epigenetic data.

This was a smaller IMPROVER challenge than usual and is a relatively new addition to the IMPROVER family of challenges, one that emphasizes and encourages local participation but may not yield the most optimal solutions for computational questions. "We want to reach out to scientists [but] it's also a way for us to raise awareness on the science we do," Boué said. However, "I personally believe that if you want to address computational question, its best to do it as wide as possible so maybe that's where the epigenetics one was not optimal."

The recently wrapped challenge, which was focused on Israeli researchers, is the second such micro challenge that the IMPROVER organizers have put together. Last year, they ran a similar challenge that called for participation from researchers in Singapore. Like the challenge in Israel, the Singapore endeavor also focused on epigenetics data, specifically DNA methylation data, but unlike the challenge in Israel, the Singapore teams were not given a specific research question.

Instead, they received a series of raw datasets and were asked to use different computational methods to extract insights from the data. Over the course of two days, participants were expected to analyze the data, try to identify correlations, and give feedback on their methods. "There was no expected answer against which we could rate the [responses] as in a traditional challenge," Sierro said. The focus of that program was "was more [about] how we could complement our approach to analyzing the data with new ideas."

The IMPROVER organizers used much of the same raw data from the Singapore event for the challenge in Israel although they provided additional datasets including publicly available literature and resources on the subject and added a clear research question. "We asked them questions that we are interested in and have already addressed here at PMI to see how other people with different backgrounds will answer the same question," Choukrallah said.

The IMPROVER organizers plan to run additional local challenges moving forward although that does not mean that they will stop planning more globally oriented computational contests. Exactly which mode the organizers pick will depend on the type of the questions that they want to address, according to Boué.

The next mini IMPROVER challenge is planned for Japan and will focus on RNA and protein data. "We [are] working on more general interpretations of transcriptomic and proteomic data for the next challenge," Boué said. "We are [also] integrating a little bit of epigenetics data and mechanisms in network models for different biological processes," she added. Like the Israel challenge, the one planned for Japan will focus on smoking and reduced risk products.

Moving forward, future computational challenges will address computational methods for analyzing microbiome data and links to disease as well as applications of DNA methylation in specific disease contexts such as cancer, Boué said.

The IMPROVER organizers are also exploring potential partnerships with organizers of existing challenges. Previously, the IMPROVER organizers partnered with IBM to organize systems biology challenges. Together, the partners launched the Species Translation challenge which aimed to better understand the limits of using rodent models to understand biological events in humans. They also organized the Network Verification challenge which asked participants to construct and improve biological network models for human lung disease. Independently, the IMPROVER organizers have launched challenges such as the Systems Toxicology challenge which focused on identifying blood gene expression signatures that could serve as markers of smoking exposure.

"There are communities that organize challenges around microbiome [for example] and we will certainly reach out to them so that we can complement each other and then run some challenges," Boué said. "But we have to establish ourselves initially."

Furthermore, questions related to smoking and reduced risk products may not be of interest to a more global audience, she added, and that could impact what partnerships the organizers pursue and form.

The incentives to participate in the mini IMPROVER computational challenges are quite modest. For the challenge in Israel, IMPROVER offered a $1,500 cash prize for the winning team, $1,000 for the second-place team and $750 for third place team. The group's larger global competitions typically have bigger prizes.

Although it's not much, "we've learned from the participants that [these challenges] allow them to [access] new types of data that they would not necessarily be working on," so that is one incentive to participate, Boué said. Also developers of the best-performing tools are invited to contribute to publications describing IMPROVER challenges, providing an outlet for them to share their work with a broader audience.

For his part, Tuller hopes more challenge development groups will be interested in partnering with Israeli institutions on future challenges besides IMPROVER. "We have a very large community of computational biologists [so] I hope that the future other types of challenges will also take place in Israel."

The IMPROVER organizers have made the data used for the Israel challenge available on the [Inhalation Toxicology Repository for Modified Risk Tobacco Products (INTERVALS) platform](#). INTERVALS was developed by PMI to allow relevant stakeholders share annotated datasets that they have provided in relation to the toxicity assessment of MRTPs and alternative products as well as the interpretation of the results they have obtained.

Although the challenge is now closed, researchers can continue to mine the data to further understand the role of epigenetics plays in human health as well as use the data to assess the effectiveness of their computational techniques. The IMPROVER researchers also hope to publish a paper discussing the findings from the study in the near future.