DREAM challenge uses crowd sourcing to test the state of the art in systems biomedicine

August 10 2015

An international study published in *Nature Biotechnology* presents the combined results of a 2013 DREAM Challenge: a crowd-sourcing initiative to test how well the effects of a toxic compound can be predicted in different people. The study, which is relevant to public and occupational health, shows that computational methods can be used to predict some toxic effects on populations, although they are not yet sensitive enough to predict such effects in individuals. It also presents algorithms useful for environmental risk assessment.

If we could use computers to predict whether a compound would have a toxic effect on people, chemical safety testing would be a lot simpler. In a community-based challenge led and organised by scientists from EMBL-EBI, Sage Bionetworks, IBM, the University of North Carolina, and the NIH's National Institute of Environmental Health Sciences (NIEHS) and National Center for Advancing Translational Sciences (NCATS), hundreds of computational biologists from all over the world tried their hand at predicting the toxicities of environmental compounds that had potential adverse health effects.

The organisers used 884 lymphoblastoid cell lines that had SNP and gene-expression data available through the 1000 Genomes Project. They measured cellular toxicity of 156 compounds on these cell lines, which represented individuals from nine subpopulations throughout Europe, Africa, Asia and the Americas. Participants were challenged to develop

algorithms that could predict toxic response in different individuals and across populations, all based on the structural attributes of the compounds.

"Our partners in the US took 1000 Genomes Project cell lines and treated them with different compounds, so we knew which compound had a toxic effect for each cell line," explains Julio Saez-Rodriguez, former Research Group Leader at EMBL-EBI now at RTWH Aachen University. "So we wanted to know, can you predict that? For a given compound, how will it affect people? For a given person, what compounds will they be sensitive to? This is really important for things like manufacturing, where people might be exposed to a new compound that hasn't been tested yet."

Dozens of teams submitted 179 predictions based on state-of-the-art computational models, and the organisers compared them against the experimental results. In the great tradition of crowd sourcing in bioinformatics, the organisers integrated the results, taking the best of each and forming a new tool to predict toxicity.

Predictions were slightly better than random for individuals, but the combined results could roughly predict population-level response to different compounds. However, improved accuracy is needed before it is possible to predict health risks associated with unknown compounds accurately.

One key benefit of the study is that it offers new methodologies for improvements in some areas of hazard evaluation and assessment.

"This partnership and challenge offer a way to provide both powerful scientific insights and meaningful public health impact by accelerating the pace of toxicity testing," says Allen Dearry, Director of the NIEHS Office of Scientific Information Management. "The winning

computational models provide significant advances in our ability to predict toxicity risk for environmental chemicals and set the stage for future data-driven challenges and competitions in environmental health science."

"The ability of the top teams to predict population-level toxicity for unknown compounds - based on similarities in chemical structure to known compounds - far surpassed our anticipations," says Lara Mangravite, Director of Systems Biology at Sage Bionetworks. "This was a true case where crowd-sourcing the problem provided answers that would otherwise never have been found."

"We had hundreds of people from all over the world participating, from prestigious labs to people who don't even work in biology," says Federica Eduati, who carried out the analyses and is an EMBL interdisciplinary postdoctoral fellow (EIPOD) at EMBL-EBI. "You don't need to be at a top-tier institute to play with great data - if you've got a good idea, you can share it."

More information: Eduati, F., et al. (2015). Predicting response of toxic compounds in human populations: a crowdsourcing study. *Nature Biotechnology*. Published online 10 August 2015; DOI: 10.1038/nbt.3299

Provided by European Molecular Biology Laboratory