DREAM Challenges: A Global Crowdsourcing Platform for Biomedical Challenges and Education

Justin Guinney, PhD, Sage Bionetworks (justin.guinney@sagebase.org)
Gustavo Stolovitzky, PhD, IBM and Icahn School of Medicine at Mount Sinai School (gustavo@us.ibm.com)

Introduction. The Dialogue on Reverse Engineering Assessment and Methods (DREAM) -- better known as DREAM Challenges -- is an open science, collaborative competition and education framework that has operated for over 10 years. Today, the DREAM mission is recognized as a successful model for motivating research teams to solve complex biomedical problems, where individuals and groups collaborate openly so that the "wisdom of the crowd" can provide the greatest impact on science and human health[1]. DREAM has now successfully run over 40 Challenges in multiple disease and biological areas, including Alzheimer's[2], rheumatoid arthritis, amyotrophic lateral sclerosis[3], olfaction, toxicology[4], and cancer[5–12].

The DREAM Challenges foster innovation in several ways. First, DREAM draws together a global and diverse set of experts to collaborate on important biomedical problems. With a solver pool of around 20,000, DREAM's Challenge participants have diverse backgrounds spanning the biological and quantitative sciences, including engineering and physics. Second, DREAM provides a systematic framework for evaluating and comparing methodologies to establish benchmarks and best-practices. Finally, DREAM encourages data providers to generate and publicly share important and unique data sets. As the concept of challenges increases in popularity as a way to motivate solutions to complex problems, DREAM continues to innovate and develop an ever wider community. Here, we highlight several current and future directions of DREAM.

Open clinical trial data. In the biomedical community, a vigorous discussion around clinical trial data sharing has emerged, largely instigated by a New England Journal of Medicine commentary from its editor-in-chief entitled "Data Sharing". One of the arguments posited in this article *against* clinical trial data sharing is that there is a paucity of examples demonstrating the value of trial data sharing. In contradiction to this article, the recently completed Prostate Cancer DREAM Challenge has successfully demonstrated how open clinical trial data can be used to extract new insight and to develop new computational approaches (Guinney, et al, Lancet Oncology, In Press). The Prostate Cancer DREAM Challenge asked participants to develop prognostic models that could exceed existing benchmarks, utilizing 5 phase III clinical trials in metastatic castration-resistant prostate cancer (N=2394). Fifty teams comprised of over 300 individuals participated in the challenge. The top-performer, based on an ensemble of penalized Cox regression models, uniquely identified predictive interaction effects with immune biomarkers and markers of hepatic and renal function. Overall, it significantly outperformed all other methods (AUC=0.791) and surpassed the Challenge benchmark model (AUC=0.743). We believe the results from this Challenge strongly affirm the value of open clinical trial data, and will encourage the field to accelerate access to these important data sets.

Moving models to data. A core principle of DREAM is to advance biomedicine through education and cumulative learning. We therefore regard model reproducibility as an important component of the learning process and Challenge outcomes. While reproducibility includes access to code and documentation, we have also been experimenting with new technology that supports the *re-runnability* of participants' models. The recent emergence of cloud computing as a widely available resource combined with lightweight virtualization technologies have made it feasible to host challenges that require participants to submit fully functional virtual machine images. In this approach, tools and methods are moved to the data rather than visa-versa. In several recent DREAM Challenges (SMC-RNA, SMC-Het and ALS-2), we required teams to submit an executable, re-runnable algorithm packaged as a Docker container. This allowed Challenge administrators to make

predictions on a team's behalf, even on data to which the Challenge team lacked direct access. Furthermore, we automated the execution and evaluation of the submitted models such that successful execution of the model became part of the submission process, instead of a post-challenge evaluation. This shifted the burden of producing a reproducible solution to a problem from the organizers to the Challenge participants, and did so in a manner where this approach became embedded in participants' active work streams, instead of a burdensome post-challenge activity.

As part of the Digital Mammography (DM) DREAM Challenge - a deep learning challenge that seeks to improve cancer detection using mammography images - we have taken this approach of virtualization even further. Our data provider, Group Health, has donated over 650k digital images representing 86k women to use in this Challenge. However, their data sharing policy required that all data remain private and not directly accessible to Challenge participants. We have therefore structured the Challenge whereby model validation *AND* training occur in the cloud. In partnership with two cloud providers, Amazon and IBM, participants must submit a Docker model which is first trained on a subset of the data, and then validated on a hold-out portion. Given the large size of the data (over 10TB) and the expected deep learning model approaches, we have improved our technical infrastructure to accommodate this format and scale within our challenge platform (synapse.org).

Reducing Overfitting. Over-fitting is a dreaded foe in challenge-based competitions. Because participants rely on public leaderboards to evaluate and refine their models, there is always the danger they might over-fit to the holdout data supporting the leaderboard. The recently published Ladder algorithm aims to address this problem by preventing the participants from exploiting willingly or inadvertently minor fluctuations in public leaderboard scores during model refinement. We report a vulnerability of the Ladder that induces severe over-fitting of the leaderboard when the sample size is small. To circumvent this attack, we propose a variation of the Ladder that releases a bootstrapped estimate of the public leaderboard score instead of providing participants with a direct measure of performance. We also extend the scope of the Ladder to arbitrary performance metrics by relying on a more broadly applicable testing procedure based on the Bayesian bootstrap. Our method makes it possible to use a leader- board, with the technical and social advantages that it provides, even in cases where data is scant.

1. Saez-Rodriguez, J. *et al.* Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nat. Rev. Genet.* **17,** 470–486 (2016).
2. Allen, G. I. *et al.* Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. *Alzheimers. Dement.* (2016). doi:10.1016/j.jalz.2016.02.006
3. Küffner, R. *et al.* Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nat. Biotechnol.* **33,** 51–57 (2015).
4. Eduati, F. *et al.* Prediction of human population responses to toxic compounds by a collaborative competition. *Nat. Biotechnol.* **33,** 933–940 (2015).
5. Hill, S. M. *et al.* Inferring causal molecular networks: empirical assessment through a community-based effort. *Nat. Methods* (2016). doi:10.1038/nmeth.3773
6. Abdallah, K., Hugh-Jones, C., Norman, T., Friend, S. & Stolovitzky, G. The Prostate Cancer DREAM Challenge: A Community-Wide Effort to Use Open Clinical Trial Data for the Quantitative Prediction of Outcomes in Metastatic Prostate Cancer. *Oncologist* **20,** 459–460 (2015).
7. Ewing, A. D. *et al.* Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat. Methods* **12,** 623–630 (2015).
8. Bansal, M. *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* **32,** 1213–1222 (2014).
9. Margolin, A. A. *et al.* Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. *Sci. Transl. Med.* **5,** 181re1 (2013).
10. Cheng, W.-Y., Ou Yang, T.-H. & Anastassiou, D. Development of a prognostic model for breast cancer survival in an open challenge environment. *Sci. Transl. Med.* **5,** 181ra50 (2013).
11. Altman, R. B. Predicting cancer drug response: advancing the DREAM. *Cancer Discov.* **5,** 237–238 (2015).
12. Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32,** 1202–1212 (2014).