



Published in final edited form as:

Nat Genet. 2013 May ; 45(5): 468–469. doi:10.1038/ng.2623.

Crowdsourcing genetic prediction of clinical utility in the Rheumatoid Arthritis Responder Challenge

Robert M. Plenge^{1,2}, Jeffrey D. Greenberg³, Lara M. Mangravite⁴, Jonathan M. J. Derry⁴, Eli A. Stahl⁵, Marieke J. H. Coenen⁶, Anne Barton⁷, Leonid Padyukov⁸, Lars Klareskog⁸, Peter K. Gregersen⁹, Xavier Mariette¹⁰, Larry W. Moreland¹¹, S Louis Bridges Jr¹², Niek de Vries¹³, Tom W. J. Huizinga¹⁴, Henk-Jan Guchelaar¹⁵, International Rheumatoid Arthritis Consortium (INTERACT)¹⁶, Stephen H. Friend^{4,17}, and Gustavo Stolovitzky^{17,18}

Robert M. Plenge: rplenge@partners.org

¹Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA ²Medical and Population Genetics Program, Chemical Biology Program, Broad Institute, Cambridge, Massachusetts, USA ³Division of Rheumatology, New York University School of Medicine, New York, New York, USA ⁴Sage Bionetworks, Seattle, Washington, USA ⁵Division of Psychiatric Genomics and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mt. Sinai, New York, New York, USA ⁶Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands ⁷Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK ⁸Rheumatology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden ⁹The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA ¹⁰Université Paris-Sud, Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpitaux Universitaires Paris-Sud, Institut National de la Santé et de la Recherche Médicale (INSERM) U1012, Rheumatology, le Kremlin-Bicêtre, France ¹¹Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA ¹²Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, Alabama, USA ¹³Department of Clinical Immunology and Rheumatology, Academic Medical Center (AMC), University of Amsterdam, Amsterdam, The Netherlands ¹⁴Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands ¹⁵Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, The Netherlands ¹⁷Sage/DREAM Project, Sage Bionetworks, Seattle, Washington, USA ¹⁸IBM Computational Biology Center, T.J. Watson Research Center, Yorktown Heights, New York, USA

© 2013 Nature America, Inc. All rights reserved.

This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

¹⁶A full list of members is provided at <https://synapse.prod.sagebase.org/#!/Synapse:syn1734172>.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

To the Editor

The ability to translate large-scale genetics and genomics data into biological knowledge has not kept pace with our ability to generate these data sets. As a consequence, a major bottleneck in biomedical research has become access to data within a computational workspace that allows for robust, collaborative analyses. One innovative solution is to bring together scientific data, code, tools and disease models into an open commons or workspace, for example, the Synapse platform of Sage Bionetworks¹. This environment allows for real-time sharing of large genomic data sets, continuous peer review and rapid learning within a system constructed to provide data access in a manner aligned with the informed consent provided by patients and research participants.

This crowdsourcing approach has been used to predict breast cancer survival from clinical and omics data² and was suggested as a way to find new drugs³ by soliciting contributions from a large online community collaborating or competing to answer an inherently difficult but important question⁴. Researchers initiating an open challenge invite solutions but also incentivize the process by offering new data, a process in which the participants' methods can be assessed by testing their predictions against previously unseen data sets. This year, Sage and DREAM (Dialogue for Reverse Engineering Assessments and Methods) are running four open challenges (<http://www.sagebase.org/challenges-overview/2013-dream-challenges/>).

Here we announce the challenge to develop genetic predictors of response to immunosuppressive therapy in a common autoimmune disease, rheumatoid arthritis (RA). Disease-modifying antirheumatic drugs such as those that block the inflammatory cytokine tumor necrosis factor- α (known as anti-TNF therapy) are not effective in all patients with RA, with up to one-third of such patients failing to enter clinical remission after a standard course of therapy⁵. Moreover, the biological mechanisms underlying this failure are unknown, limiting the development of clinical biomarkers to guide either this therapy or the development of new drugs to target refractory cases.

The Rheumatoid Arthritis Responder Challenge is for teams to build the best genetic predictor of response to anti-TNF therapy. There are two phases to the challenge: discovery and validation (Fig. 1). In the discovery phase, teams will utilize genomic data sets—several of which will be generated for the purposes of this challenge—and a variety of analytical methods to build predictive polygenic models of treatment response. We recently published a genome-wide association study (GWAS) in ~2,700 patients with RA treated with anti-TNF therapy⁶. Our GWAS data indicate that the genetic architecture of the anti-TNF response is probably highly polygenic, similar to what has been observed for other complex traits, such as risk of RA⁷. Importantly, our challenge will incorporate a new GWAS data set, which will be used in the validation phase, in which models built in the discovery phase are tested. The data set of ~1,100 patients with RA treated with anti-TNF therapy will be made available through a public-private partnership between the Consortium of Rheumatology Researchers of North America, Inc. (CORRONA) and the Pharmacogenomics Research Network (PGRN) sponsored by the National Institute of General Medical Sciences (NIGMS) and the US National Institutes of Health (NIH).

A unique component of our Rheumatoid Arthritis Responder Challenge is the diversity of participation across a number of groups from academic institutions, private foundations and for-profit companies. In addition to support from CORRONA and PGRN, we received funding from pharmaceutical companies (see complete list on our website; link below) and a private foundation (the Arthritis Foundation) to support the public commons. We also received support from the Arthritis Internet Registry (AIR) and the Broad Institute to generate new genomic data sets, as well as in-kind support from a large number of academic collaborators from across the world to make GWAS data available in the discovery phase. We anticipate that a winning classifier could enable a follow-on prospective clinical trial within the group of appropriately consented patients in AIR.

Through Synapse, analysts who are inclined to establish collaborations will have the opportunity to see in real time the models that others are using so that each team can learn from the others (Fig. 1). A leaderboard will show the relative performance ranking of the different teams on the basis of a crossvalidation strategy designed to minimize overfitting. During the discovery phase, teams that choose to collaborate with each other will have the opportunity to check each other's algorithms for readability, speed and reproducibility. Then, during the validation phase, each team will submit computer code, which the SageDREAM team (<http://www.sagebase.org/>) will test in Synapse to establish whether it runs as expected to predict if a subject is an anti-TNF therapy responder or nonresponder on the basis of the GWAS data. Predefined performance metrics will be used to objectively determine the accuracy of the predictions, their statistical significance and the final performance ranking of the participating teams. The team that develops the most highly predictive model will be deemed the 'winner', with precise attribution of contributor roles going to all members of teams that contributed to building the final consensus model.

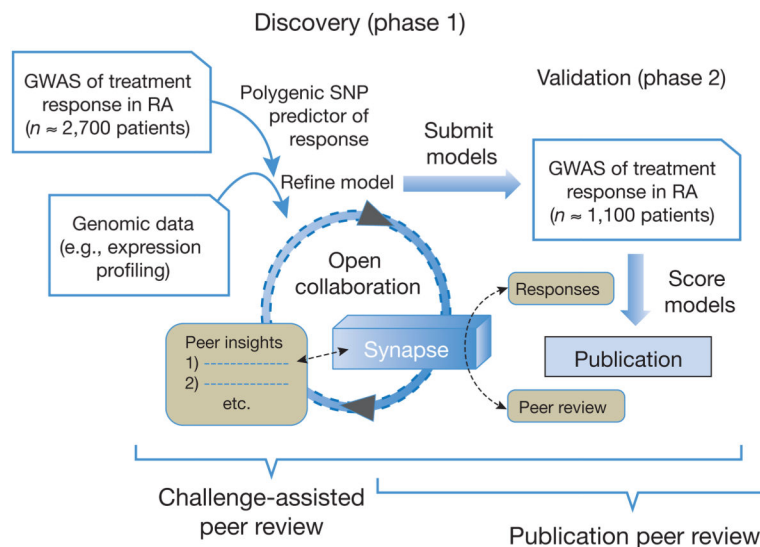
The best-performing models, therefore, will have passed a test of performance that is outside the realm of, and complements, traditional peer review. Indeed, this stringent test of method performance can be used as an enhanced way of publication vetting, what we call 'challenge-assisted peer review'. Traditional peer review is essential for ensuring the clarity, originality, contextualization and logical thread of a discrete set of work that is ready to be used by researchers in the form of a published article. However, the complexity of working with omics data—entailing multiple analytical decisions, computational simulations and statistical calculations—means that referees are challenged to follow and check the components of even a traditional research paper. In our Rheumatoid Arthritis Responder Challenge, we will explore the feasibility of enhancing the reliability and transparency of conventional peer review in partnership with *Nature Genetics*. This can be achieved if the referees and authors of the paper reporting on the best-performing methods in the challenge are willing to leave their comments openly (yet anonymously) on the Synapse platform (Fig. 1). We anticipate that the challenge-based assessment of accuracy will provide an objective metric of performance and a comparison with state-of-the-art analytical methodologies that will greatly enhance the task of refereeing a body of work with more quality control than is currently provided by conventional peer review.

In conclusion, we believe that the Rheumatoid Arthritis Responder Challenge is an apt use of crowdsourcing in human genetics to gain insight into clinical prediction and disease

biology. Details of the challenge, including the rules by which the models will be judged, can be found at <https://synapse.prod.sagebase.org/#!/Synapse:syn1734172>.

References

1. Derry MJ, et al. *Nat Genet.* 2012; 44:127–130. [PubMed: 22281773]
2. Margolin AA, et al. *Sci Transl Med.* 2013; 5:181re1.
3. Wadman M. *Nature.* 2012; 490:15. [PubMed: 23038441]
4. Costello JC, Stolovitzky G. *Clin Pharmacol Ther.* 2013; 93:396–398. [PubMed: 23549146]
5. McInnes IB, Schett G. *N Engl J Med.* 2011; 365:2205–2219. [PubMed: 22150039]
6. Cui, J., et al. *PLoS Genet.* Mar 28, 2013 published online, <http://dx.doi.org/10.1371/journal.pgen.1003394>
7. Stahl EA, et al. *Nat Genet.* 2012; 44:483–489. [PubMed: 22446960]

**Figure 1.**

Overview of the Rheumatoid Arthritis Responder Challenge. There are two phases to the challenge. In phase 1 (discovery), analysts build genetic models of response to anti-TNF therapy using SNP data from a GWAS of ~2,700 patients with RA. To facilitate model building, additional genomic data will be made available. In a model of open collaboration, participants will use Synapse to post code, share insights and engage in rapid learning prepublication. In phase 2 (validation), models will be posted, tested and scored in an independent GWAS data set of ~1,100 patients with RA treated with anti-TNF therapy. To complement challenge-assisted peer review (which occurs in both the discovery and validation phases), conventional peer review will have access to Synapse to understand the iterative process of model building. Synapse will allow study investigators to respond to peer-review critiques and resubmit versions of their models and studies.