DR. YUANFANG GUAN (Orcid ID : 0000-0001-8275-2852)
DR. DIMITRIOS A PAPPAS (Orcid ID : 0000-0001-8338-027X)

Article type : Full Length

Machine learning to predict anti-TNF drug responses of rheumatoid arthritis patients by integrating clinical and genetic markers

Yuanfang Guan, Ph.D.[1*†,] Hongjiu Zhang, Ph.D.[1†^], Daniel Quang, Ph.D.[1], Ziyan Wang[1], Stephen C.J. Parker, Ph.D.[1], Dimitrios A. Pappas, M.D.[2,3] Joel M. Kremer, M.D.[3,4], Fan Zhu, Ph.D.[5†]

[1] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA 48109

[2] Columbia University College of Physicians and Surgeons, New York, USA

[3] Corrona LLC Waltham, MA, USA

[4] Albany Medical College and The Center for Rheumatology. Albany, USA

[5] Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China 400000

^ Current affiliation: Microsoft, Inc, Seattle, WA, USA

† YG, FZ, and HZ equally contributed to the work.

* To whom correspondence should be addressed: gyuanfan@umich.edu

FZ, DXQ, DAP, JMK, and SCJP have no financial disclosures. HZ is an employee of Microsoft, Inc and holds stock of Microsoft, Inc. YG receives personal payment from Eli Lilly and Company, Genentech Inc, F. Hoffmann-La Roche AG, and Cleerly Inc; holds equity shares at Cleerly Inc and Ann Arbor Algorithms Inc; and receives research support from Merck KGaA as research contracts and Ryss Tech as unrestricted donations.

This work was supported by the National Science Foundation 1452656.

Abstract

Objective: Accurate prediction of the responses of rheumatoid arthritis patients can provide valuable information for effective drug selection. An important second-line treatment after methotrexate, the classical first-line treatment, is anti-TNF drugs. However, patient heterogeneity hinders identification of predictive biomarkers and accurate modeling of anti-TNF drug responses.

Methods: We present the best-performing model in predicting anti-TNF response in the DREAM Rheumatoid Arthritis Responder Challenge. Given demographic, baseline disease assessment, treatment, and SNP array data of a patient, our Gaussian process regression model predicts changes in disease activity scores for the patient and classifies the patient into the responder or nonresponder group. The model was developed and cross-validated on 1892 patients. It was evaluated on an independent dataset of 680 patients. We examined the effectiveness of the similarity modeling and the contribution of individual features.

Results: In the cross-validation tests, our method predicts changes in disease activity scores with a correlation coefficient of 0.406. It correctly classified responses of 78% of subjects. In the independent test, the method achieved a Pearson correlation coefficient of 0.393 in predicting ΔDAS, *i.e.*, delta disease activity score. The method won first place in the DREAM Challenge. Gaussian process regression effectively re-mapped the feature space and identified subpopulations that do not respond well to anti-TNF treatments. Genetic SNP biomarkers show small additional contribution in the prediction on top of the clinical models.

Conclusions: The model shows promise in guiding drug selections in clinical practice based on primarily clinical profiles with additional genetic information.

Keywords

Rheumatoid arthritis, anti-TNF drugs, Gaussian process regression, drug responses, patient heterogeneity

Background

Rheumatoid arthritis (RA) patients show great heterogeneity in their responses to treatments [1], and accurate prediction of their responses would provide valuable information for optimal drug selection [2]. In current practice, patients who respond inadequately to conventional therapies usually receive anti-tumor-necrosis-factor (anti-TNF) drugs [3]. These expensive drugs are mainly chosen on a "trial-and-error" basis [4], and about 30% of the patients respond poorly [5]. These nonresponders suffer from drug expenses [6], non-improving disease conditions [6,7], side effects [8,9], and infection risks [10]. In order to provide effective treatments for these patients, physicians need to predict the patients' responses to different anti-TNF drugs in advance.

However, patient heterogeneity hinders identification of predictive biomarkers and accurate modeling of anti-TNF drug responses. Early studies reported that demographic and clinical markers such as gender and baseline disease activity are related to drug responses [11,12], but these markers are not predictive enough to identify nonresponders by themselves [13]. Recent genome-wide association studies found multiple genetic markers that are associated with poor drug response, and these variants were used to facilitate the identification of nonresponders [5,14–16]. Yet these variant markers are often confounded with cohort or ethnic information [17]. RA patients exhibits great genotypic and phenotypic heterogeneity, and markers found in one ethnic population or cohort may not be applicable to others [18]. Thus, effective modeling of patient heterogeneity is the key to accurate drug response prediction.

Here we present a Gaussian process regression (GPR) model for predicting anti-TNF drug responses, the first-place winner in the Dialogue on Reverse Engineering Assessment and Methods (DREAM): Rheumatoid Arthritis Responder Challenge [19,20]. The model combines demographic, clinical, and genetic markers, predicts patients' changes in disease activity scores 24 months after their baseline assessment, and identifies nonresponders to anti-TNF treatments. Specifically, the pipeline predicts the changes in DAS of patients who have taken 12 months of anti-TNF treatments, and also classifies the patients' responses based on the EULAR response metric. Based on the model, we examined the transformation that the GPR kernel applied to the patient data as well as the distribution of the patients in the transformed space. Through independent testing dataset, we aimed at evaluate the potential of the GPR model in improving anti-TNF drug selection and the effect of genetic markers across multiple cohorts. Based on the model, we also investigated the effect of genetic markers across multiple cohorts.

Methods

Data acquisition

The training and testing datasets were provided by the DREAM Challenge organizers (**Table 1**). The training data consist of 2706 individuals, 1892 of which were chosen randomly and released to the participants of the competition before the final evaluation. The remaining samples were withheld for real-time submission evaluation, so that participants can estimate their models throughout the competition. None of these data were included in the final evaluation. The samples in the training dataset consists of individuals of European ancestry from 13 cohorts: Autoimmune Biomarkers Collaborative Network (ABCoN) from the U.S. [21]; the Genetics Network Rheumatology Amsterdam (GENRA); the Dutch Behandelstrategieen voor Rheumatoide Arthritis (BeSt) [22]; the U.K. Biological in Rheumatoid arthritis Genetics and

Genomics Study Syndicate (BRAGGSS) [23,24]; the U.S. Brigham Rheumatoid Arthritis Sequential Study (BRASS) [25]; the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) [26]; the Immunex Early Rheumatoid Arthritis study (eRA) [27]; the Swedish Karolinska Institutet study (KI); the Netherlands collection from Leiden University Medical Center (LUMC); the U.S. Treatment of Early Aggressive RA (TEAR); the Dutch Rheumatoid Arthritis Monitoring registry (DREAM) in the Netherlands; the ApotheekZorg (AZ) database [28,29]; and the French Research in Active Rheumatoid Arthritis (ReAct) [30]. All subjects were either diagnosed by a board-certified rheumatologist or met 1987 American College of Rheumatology criteria (Arnett, et al. 1988). All patients had a baseline DAS > 3.2.

The testing data consist of 680 subjects from the Consortium of Rheumatology Researchers of North America (CORRONA) [31], who participated the Comparative Effectiveness Registry to study Therapies for Arthritis and Inflammatory Conditions (CERTAIN) study [32]. The study is within the CORRONA registry and involves adult RA patients diagnosed by certified rheumatologists, having at least moderate disease activity defined by a clinical disease activity index (CDAI) score >10 who are starting or switching biologic agents.

For all subjects in both training and testing datasets, gender, age, methotrexate, and a baseline disease activity score (DAS28) were collected and provided. Post-treatment disease activity scores of all subjects in the training datasets are available for all participants, whereas those of testing subjects are withheld by the DREAM Challenge organizers until the end of the challenge. For each subject, a panel of genotype imputation is provided. See https://www.synapse.org/#!Synapse:syn1734172/wiki/62201 for details.

Drug response prediction

The proposed model in this study adopted GPR to predict ΔDAS. GPR is designed to predict the unknown dependent variable for any given independent variables based on known but noisy observations of the dependent and independent variables. Gaussian process regression does not match its target function to some specific models (e.g. linear, quadratic or cubic models). In this study, the GPR model took the input ΔDAS of known subjects as noisy observations and predict the drug responses of incoming patients based on clinical and genetic features. See **Additional**

File S1 for its full formulation. The final model is available at https://www.synapse.org/#!Synapse:syn2368045/wiki/64596 .

In details, the kernel function of the GPR model accepts the difference in demographic, treatment, and genetic features between two patients as input variables. For each input variable, the kernel function performs an squared exponential transformation and takes the summation of the transformed values. The distance of two patients is then jointly determined by the nonlinear transformed difference across all features.

The genetic features included in the model were chosen via literature mining or statistical analysis (**Table S2**). Genetic features from literature were collected based on their reported association with either RA risks or anti-TNF drug responses; those from statistical analysis were collected based on the correlation between the genotype dosages and the ΔDAS. During the DREAM challenge, the organizing team provided a leaderboard set to test the models. Both sets of candidate features were then tested through cross-validation and forward feature selection, and only features that improved the cross-validation performance and on the leaderboard were included. For a SNP collected in literature, even though it does not show substantial correlation to the drug response phenotype, we retrain it if it adds on top of the prediction performance. In the end, the model included from 10 to 15 SNPs (depends on the drug types) from literature and about 5 SNPs from the correlation test. In addition, they were also tested using Fast-LMM [33], a Linear Mixed Model-based program for performing both single-SNP and SNP-set genome-wide significance test. Unfortunately, none of the genetic features passed the $5x10^{-8}$ threshold in the Fast-LMM test when clinical variables are available.

All clinical and demographic features provided in the datasets were included in the model. Separate models were developed for different anti-TNF drugs. Each model was trained with only patients who take the corresponding anti-TNF drug. For predicting responses of Certolizumab users, a general model was developed using all patients in the training dataset, but without genetic features.

Models that classify patients into responders or nonresponders were developed by giving a binary cutoff for 0 (responders) or 1 (nonresponders) based on the predicted values, so that the proportion of the non responders is consistent with the training set.

Evaluation

The prediction models were evaluated through 5-round 2-fold cross-validation [34]. Each round of the repeated tests started with randomly splitting the training dataset into two halves. A model was trained on one half and scored on the other, and again with both halves swapped. Five rounds of tests gave a total of 10 scores, and their average becomes the estimated performance of the model. The repeated cross-validation tests were performed on the training dataset provided by the DREAM Rheumatoid Arthritis Responder Challenge. Because we had three different drugs, we carried out evaluation on each of the drugs, as well as the entire set.

The predicted ΔDASes were evaluated in terms of Pearson correlation coefficient. The performance of nonresponder identification is measured in terms of correctly predicted percentage and area under the receiver operating characteristic curve (AUC). The receiver operating characteristic curve is created by plotting the true positive ratio (the ratio of correctly identified patients out of all AD/MCI subjects) against false positive ratio (the ratio of incorrectly predicted “patients” out of all normal subjects) at various threshold settings.

In order to assess the improvement of the GPR models over other algorithms, models of different algorithms were tested through bootstrap tests. The tests were repeated for 100 times. In each round, the training data were of the same size as the original challenge training dataset (N=1892) and sampled from the original challenge training data with replacement. The unsampled data were merged as the testing data. The performance was evaluated as how many rounds of bootstrap tests that GPR outperformed other methods. The bootstrap test is similar to the test method used in the original DREAM Challenge to assess the robustness of the models.

Results

Gaussian process regression accurately predicts anti-TNF drug responses

Given demographic, baseline disease assessment, treatment, and SNP array data of a patient, our GPR model predicts ΔDAS for the patient and classifies the patient into the responder or nonresponder group (Figure 1). It relies on a custom kernel function to weight collected individuals proportionally to their similarity to the new patient in terms of their clinical and genetic data. The model predicts treatment outcomes by leveraging the outcomes and features of training patients. The model estimates ΔDAS of the patient 24 months after his or her initial disease assessment and classifies the patient as a responder or a nonresponder. By intuition, the estimation of our method would be close to the outcomes of patients with similar demographic, treatment, and SNP array data. We developed separate GPR models for different anti-TNF treatments.

A major difficulty in this challenge is to deal with the heterogeneity in the datasets. The challenge took different cohorts for training and testing, requiring the participating models to deal with the cohort effects. Post-challenge analysis on the two datasets showed that the on DAS (Mann-Whitney U test $p < 10^{-23}$) and delta DAS (Mann-Whitney U test $p < 10^{-17}$) are drastically different. We chose GPR to deal with the heterogeneity. GPR takes similar patients from the training data to guide the prediction without explicitly specifying the feature distribution across cohorts. We then moved on to compare our GPR model to other alternative models.

We evaluated the GPR models through repeated cross-validation tests as described in Methods. GPR was compared against linear models, classification and regression tree models, and a support vector machine (SVM) model. For ΔDAS prediction, GPR achieved the best average correlation (0.406) between predicted and observed ΔDASes, followed by ridge regression, SVM, and regression tree models (**Figure 2B**). For response classification, GPR, the overall best performer, correctly classified ~78% of subjects, with an area under receiver operating characteristic curve (AUC) of ~0.66 (**Figure 2C&D**). Interestingly, another well-performing model, SVM, is also a kernel-based method, which shares similar properties to GPR. These methods can be used in clinical settings to inform the decision-making process. Compared to random assignment (50-50 chance to be correct), we will be able to correctly make 28% more correct treatment to patients. To evaluate the margin of the improvement from our GPR models, we performed 100-time bootstrap tests on the original dataset and found GPR showed substantial improvement over all other methods (**Table S1**). We also evaluated the models for individual drugs separately (**Figure S1**). The performance rankings varied, but GPR achieved an overall best performance, closely followed by linear and SVM models. We further optimized the hyperparameter in the model (**Figure S2**). The GPR model was then submitted for the DREAM Challenge evaluation.

Evaluation on an independent hidden cohort:

The DREAM Rheumatoid Arthritis Responder Challenge also evaluated the GPR model on an independent testing cohort [20]. The independent dataset, released after the competition, consists of 680 patients from the Comparative Effectiveness Registry to study Therapies for Arthritis and Inflammatory Conditions (CERTAIN) study, conducted by the Consortium of Rheumatology Researchers of North America (CORRONA) [32]. The GPR model achieved a Pearson correlation coefficient of 0.393 in predicting ΔDAS ($p < 1e\text{-}6$ compared to random hypothesis) and an AUC of 0.615 in classifying anti-TNF nonresponders. This represents only a ~0.01 drop in correlation from the cross-validation, indicating limited over-fitting and batch effects. in The GPR model showed consistent prediction performance over both our cross-validation dataset and the independent testing dataset.

Similarity modeling enhances treatment response prediction

To investigate how GPR effectively modeled patient heterogeneity, we inspected the properties of individual features. Among all the features, baseline disease activity scores (DAS) have the highest correlation coefficient against ΔDAS (**Table S3**). Across all the samples in the training dataset regardless of their anti-TNF drugs, their baseline DAS have a correlation coefficient of 0.370 against their ΔDAS. In the CORRONA CERTAIN cohort, the correlation coefficient is 0.351. Yet the high correlation of baseline DAS against ΔDAS alone does not fully explain the performance of the GPR model. The difference between the performance of the GPR model and that of a naive baseline DAS linear regression implies the contribution of other demographic, clinical, and genetic features.

GPR relies on its kernel function to transform input features. To study how the kernel transformation incorporates features to help predict drug responses, we projected the training sample in the feature spaces before and after the kernel transformation (**Figure 3** for adalimumab, **Figure S3** and **S4** respectively for etanercept and infliximab). Principal component analysis on the features reports major confounding factors such as geographic or cohort information. The major contributing features to the first two principal components are genetic features, which separate patients based on their cohort information instead of their drug responses. Conversely, principal component analysis on the kernel-transformed similarity matrix shows a clear gradient from anti-TNF drug responders to non-responders. Important features besides baseline DAS, such as age, methotrexate usage, and several genetic markers, correlated well with the first two principal components. The pattern demonstrates that in the kernel-transformed feature space, patient similarity correlates well to their similarity in disease activity changes rather than other confounding factors.

Cross-cohort evaluation shows clinical and genetic heterogeneity

During the inspection of above features, we noticed large variation in features across users of different anti-TNF drugs. Therefore, we developed specialized models for individual drugs in the original submission to the challenge. In the post-challenge analysis, we re-assessed the drug-specific approach on the testing cohort. Our original analysis in the training dataset showed that the baseline DAS more strongly correlated with ΔDAS among adalimumab and etanercept patients than infliximab ones; whereas in the CORRONA CERTAIN cohort, baseline DAS are substantially less predictive of responses to adalimumab and infliximab than those to etanercept. To accommodate with the variation, we excluded users of different drugs from kernel calculation for each drug-specific model. We also curated drug-specific genetic feature lists. We compared these models with two sets of genetic-free models, one trained on all patients, and the other on users of only corresponding anti-TNF drug. The performance of these approaches is similar (**Figure S5**). The negligible difference of sample exclusion suggests the variation across different drugs may be attributed to sampling biases instead of drug-specific characteristics. The conflict between aforementioned high contributions of genetic features in drug-specific models and high performance of clinical-only models suggests further analysis on the roles of genetic markers in the prediction.

We then investigated the contribution of genetic markers to drug response predictions. The genetic markers we curated were chosen based on either literature review or statistical analysis on the training dataset (**Table S2**). To assess these markers, we developed a baseline-only model, a genetic-only model, and an age+gender+methotrexate model. These models were compared against the original submission GPR models, which used all these features. We performed both cross-validation tests over the training dataset (**Figure 4**) and an independent test over the CORRONA dataset (**Figure 5**). The result shows that while clinical information and genetic markers have relatively low predictive power themselves, they can improve the accuracy of the GPR model on top of the baseline DAS feature, especially for ΔDAS prediction. Our analysis of genetic markers above has showed the strong cohort association (**Figure S6**). Considering that the training cohorts are of European descents only and associated with different geographic areas, it is challenging to apply these biomarkers to other cohorts.

Discussions

Here we demonstrated the state-of-the-art predictive power of the GPR model [35]. The similarity modeling approach of GPR complements the ongoing development of precision medicine efforts in RA [36]. The idea of similarity modeling has been widely used in social network analysis and other areas [37,38]. It is considered to be effective in dealing with heterogeneous datasets, which is commonly seen in cross-sectional studies [39]. Also, the heterogeneity of diseases often obstruct explicit modeling of underlying distributions of individual features, which can be even more problematic with a small population [40]. GPR circumvents the issue by matching patients to those with similar conditions. Specifically in this study, our GPR model can predict subpopulations that do not respond to the treatment.This can help physicians tailor treatments for individual patients based on their conditions.

The GPR model, as an interpretable method, has practical advantages in clinical application. Many sophisticated machine learning algorithms may make accurate predictions but lack interpretability for medical application [41]. Uninterpretable models are undesirable in many medical applications [42,43]. On the contrary, GPR is a well-studied statistical model. The similarity modeling approach is intuitive, and its results are easy to interpret [44]. Specifically for drug response prediction, the kernel function allows identification of known subjects with similar conditions. While the kernel function for both GPR and SVM provides information regarding the importance of the genetic and clinical features, GPR also bases its prediction on the most similar individuals in the training dataset. This allows physicians to inspect the conditions of known samples that have the highest weights in the GPR prediction. The additive design of our custom kernel function allows easy incorporation of new features with reduced parameter tuning. The model can also estimate confidence intervals for its predictions, allowing physicians to judge how confident the predictions are.

However, in comparison to many linear methods that come with feature penalty that allows built-in feature selection, GPR does not have this benefit. Given millions of genetic features, many genes that are by chance correlated with the clinical outcomes. Furthermore, because of the model difference, significance in linear model tests does not translate directly to accuracy improvement in a GPR model. To address these issues together, we chose a pre-selected set of genetic features in our model. The selection is based on cross-validation, with clinical features present in the training process. These genes chosen through this process often by themselves not informative enough to predict the outcomes, but can improve the performance of the clinical only models. Recent advances in deep learning may allow future development of methodologies that directly connect genome sequences to phenotypes and drug responses of RA, by using one-dimensional convolutional neural network which can be used to extract information from DNA sequences directly. With the increase in the amount of sequencing data, we foresee the growth of this area in the near future.

Clinical and genetic heterogeneity may pose a major challenge for predicting anti-TNF drug responses. On one hand, clinical and demographic markers worked well across different cohorts. The cross-validation showed that although clinical markers themselves possess predictive power, they improved accuracy on top of the baseline-DAS-only model regardless of cohorts. Previous studies also found several blood biomarkers that are informative of drug responses and are validated across different cohorts [45,46]. Genetic features used in our models have been reported to be immune-related, including insulin secretion (PDZD2), immunoresponse (CD84), eicosanoids synthesis (PLA2G4A). Previous studies reported ethnic differences in genetic markers for anti-TNF drug responses in treatment of both rheumatoid arthritis [15,47] and other related autoimmune diseases [48,49]. Our principal component analysis over the training dataset showed that the genetic markers are associated with cohort information. We showed that genetic markers could not improve the prediction accuracy on the CORRONA dataset as they did on the training dataset. While both training and testing data in our study involve only European-descent subjects, the variation across different geographic areas still obstruct modeling genetic markers. Accurate modeling of drug responses would require a larger panel of genetic features that covers multiple populations. We believe that extending the clinical feature panel to include blood markers or other clinical assays would be beneficial for cross-sectional predictions. On the other hand, genetic markers were found specific to populations in the context of sufficient genetic subtype modeling.

Compared to traditional trial-and-error practice, our model can help up to 40% of European-descent anti-TNF non-responders avoid ineffective treatments. The model performance is even comparable to some published models utilizing additional biomarker data, whose AUROC ranges from 55%~74% over various testing sets [46]. We caution the model users that the model is built upon European descendants. Considering the heterogeneity of the anti-TNF responses among rheumatoid arthritis patients, we do not expect the model to achieve a similar performance on other populations. Extension of the model over other populations requires new patient data and separate feature selection.

Conclusions

In this study, we developed a GPR model for predicting anti-TNF drug responses of rheumatoid arthritis patients and identifying nonresponders. The model interpretation shows promise in guiding drug selection. While we showed that the clinical features here are still the most predictive features, the prediction model allows researchers to assess the contribution of genetic markers over existing clinical information across cohorts. For the future work, various clinical markers may be potentially used for more accurate identification of non-responding subpopulations that carry predictive biochemical traits [46,50]. We caution that the model was developed for European descendants only. Transferring prediction models to other populations may face difficulties [51]. We envision future development involving more diverse, and bigger population will result in more robust models for predicting deltaDAS.

Acknowledgement

This work was supported by the National Science Foundation [1452656] and National Natural Science Foundation of China [61802360].

Author contributions:

FZ and YG developed the methods and carried out experiments. HZ wrote the manuscript. YG supervised the project and carried out revisions of the paper. DAP and JMK provided data access to CORRONA , ZW participated in Figure editing in revision, DXQ and SCJP provided insights into the genetic areas of the manuscript. All authors participated in the proof-reading of the manuscript.

References

1. Geiler J, Buch M, McDermott MF. Anti-TNF treatment in rheumatoid arthritis. Curr Pharm Des. 2011;17:3141–54.

2. Wijbrandts CA, Tak PP. Prediction of Response to Targeted Treatment in Rheumatoid Arthritis. Mayo Clin Proc. 2017;92:1129–43.

3. Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, et al. Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. N Engl J Med. 2000;343:1594–602.

4. Oliver J, Plant D, Webster AP, Barton A. Genetic and genomic markers of anti-TNF treatment response in rheumatoid arthritis. Biomark Med. 2015;9:499–512.

5. Prajapati R, Plant D, Barton A. Genetic and genomic predictors of anti-TNF response. Pharmacogenomics. 2011;12:1571–85.

6. de la Torre I, Valor L, Nieto JC, Hernández-Flórez D, Hernandez D, Martinez L, et al. Anti-TNF treatments in rheumatoid arthritis: economic impact of dosage modification. Expert Rev Pharmacoecon Outcomes Res. 2013;13:407–14.

7. Roda G, Jharap B, Neeraj N, Colombel J-F. Loss of Response to Anti-TNFs: Definition, Epidemiology, and Management. Clin Transl Gastroenterol. 2016;7:e135.

8. Aletaha D, Kapral T, Smolen JS. Toxicity profiles of traditional disease modifying antirheumatic drugs for rheumatoid arthritis. Ann Rheum Dis. 2003;62:482–6.

9. Antoni C, Braun J. Side effects of anti-TNF therapy: current knowledge. Clin Exp Rheumatol. 2002;20:S152–7.

10. Burmester GR, Landewé R, Genovese MC, Friedman AW, Pfeifer ND, Varothai NA, et al. Adalimumab long-term safety: infections, vaccination response and pregnancy outcomes in patients with rheumatoid arthritis. Ann Rheum Dis. 2017;76:414–7.

11. Hyrich KL, Watson KD, Silman AJ, Symmons DPM, British Society for Rheumatology Biologics Register. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. Rheumatology . 2006;45:1558–65.

12. Atzeni F, Antivalle M, Pallavicini FB, Caporali R, Bazzani C, Gorla R, et al. Predicting response to anti-TNF treatment in rheumatoid arthritis patients. Autoimmun Rev. 2009;8:431–7.

13. Umićević Mirkov M, Cui J, Vermeulen SH, Stahl EA, Toonen EJM, Makkinje RR, et al. Genome-wide association analysis of anti-TNF drug response in patients with rheumatoid arthritis. Ann Rheum Dis. 2013;72:1375–81.

14. Cui J, Saevarsdottir S, Thomson B, Padyukov L, van der Helm-van Mil AHM, Nititham J, et al. Rheumatoid arthritis risk allele PTPRC is also associated with response to anti-tumor necrosis factor alpha therapy. Arthritis Rheum. 2010;62:1849–61.

15. Sode J, Vogel U, Bank S, Andersen PS, Thomsen MK, Hetland ML, et al. Anti-TNF treatment response in rheumatoid arthritis patients is associated with genetic variation in the NLRP3-inflammasome. PLoS One. 2014;9:e100361.

16. Wu C, Wang S, Xian P, Yang L, Chen Y, Mo X. Effect of Anti-TNF Antibodies on Clinical Response in Rheumatoid Arthritis Patients: A Meta-Analysis. Biomed Res Int. 2016;2016:7185708.

17. Yamamoto K, Okada Y, Suzuki A, Kochi Y. Genetic studies of rheumatoid arthritis. Proc Jpn Acad Ser B Phys Biol Sci. The Japan Academy; 2015;91:410.

18. Weyand CM, Klimiuk PA, Goronzy JJ. Heterogeneity of rheumatoid arthritis: from phenotypes to genotypes. Springer Semin Immunopathol. 1998;20:5–22.

19. Plenge RM, Greenberg JD, Mangravite LM, Derry JMJ, Stahl EA, Coenen MJH, et al. Crowdsourcing genetic prediction of clinical utility in the Rheumatoid Arthritis Responder Challenge. Nat Genet. 2013;45:468–9.

20. Sieberts SK, Zhu F, García-García J, Stahl E, Pratap A, Pandey G, et al. Crowdsourced assessment of common genetic contribution to predicting anti-TNF treatment response in rheumatoid arthritis. Nat Commun. 2016;7:12460.

21. Liu C, Batliwalla F, Li W, Lee A, Roubenoff R, Beckman E, et al. Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. Mol Med. 2008;14:575–81.

22. Allaart CF, Goekoop-Ruiterman YPM, de Vries-Bouwstra JK, Breedveld FC, Dijkmans BAC, FARR study group. Aiming at low disease activity in rheumatoid arthritis with initial combination therapy or initial monotherapy strategies: the BeSt study. Clin Exp Rheumatol. 2006;24:S – 77–82.

23. Bluett J, Morgan C, Thurston L, Plant D, Hyrich KL, Morgan AW, et al. Impact of inadequate adherence on response to subcutaneously administered anti-tumour necrosis factor drugs: results from the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate cohort. Rheumatology . 2015;54:494–9.

24. Plant D, Bowes J, Potter C, Hyrich KL, Morgan AW, Wilson AG, et al. Genome-wide association study of genetic predictors of anti-tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci. Arthritis Rheum. 2011;63:645–53.

25. Iannaccone CK, Lee YC, Cui J, Frits ML, Glass RJ, Plenge RM, et al. Using genetic and clinical data to understand response to disease-modifying anti-rheumatic drug therapy: data from the Brigham and Women’s Hospital Rheumatoid Arthritis Sequential Study. Rheumatology . 2011;50:40–6.

26. Orellana C, Wedrén S, Källberg H, Holmqvist M, Karlson EW, Alfredsson L, et al. Parity and the risk of developing rheumatoid arthritis: results from the Swedish Epidemiological Investigation of Rheumatoid Arthritis study. Ann Rheum Dis. 2014;73:752–5.

27. Bathon JM, Genovese MC. The Early Rheumatoid Arthritis (ERA) trial comparing the efficacy and safety of etanercept and methotrexate. Clin Exp Rheumatol. 2003;21:S195–7.

28. Coenen MJH, Enevold C, Barrera P, Mascha M V A, Toonen EJM, Scheffer H, et al. Genetic Variants in Toll-Like Receptors Are Not Associated with Rheumatoid Arthritis Susceptibility or Anti-Tumour Necrosis Factor Treatment Outcome. PLoS One. 2010;5:e14326.

29. Toonen EJM, Coenen MJH, Kievit W, Fransen J, Eijsbouts AM, Scheffer H, et al. The tumour necrosis factor receptor superfamily member 1b 676T>G polymorphism in relation to response to infliximab and adalimumab treatment and disease severity in rheumatoid arthritis. Ann Rheum Dis. 2008;67:1174–7.

30. Miceli-Richard C, Comets E, Verstuyft C, Tamouza R, Loiseau P, Ravaud P, et al. A single tumour necrosis factor haplotype influences the response to adalimumab in rheumatoid arthritis. Ann Rheum Dis. 2008;67:478–84.

31. Kremer JM. The CORRONA database. Autoimmun Rev. 2006;5:46–54.

32. Pappas DA, Kremer JM, Reed G, Greenberg JD, Curtis JR. “Design characteristics of the CORRONA CERTAIN study: a comparative effectiveness study of biologic agents for rheumatoid arthritis patients.” BMC Musculoskelet Disord. 2014;15:113.

33. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. Nat Methods. 2011;8:833–5.

34. Kim J-H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Comput Stat Data Anal. 2009;53:3735–45.

35. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Mit Press; 2006.

36. Kłak A, Paradowska-Gorycka A, Kwiatkowska B, Raciborski F. Personalized medicine in rheumatology. Reumatologia. Termedia Publishing; 2016;54:177.

37. Liu C, Liu J, Jiang Z. A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks. IEEE Trans Cybern. 2014;44:2274–87.

38. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. PLoS One. 2011;6:e26752.

39. Huang Z, Zhang H, Boss J, Goutman SA, Mukherjee B, Dinov ID, et al. Complete hazard ranking to analyze right-censored data: An ALS survival study. PLoS Comput Biol. 2017;13:e1005887.

40. Mackay IR, Rose NR. The Autoimmune Diseases. Academic Press; 2013. p. 668.

41. Joseph A. Cruz DSW. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Inform. SAGE Publications; 2006;2:59.

42. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15 [Internet]. 2015. Available from: http://dx.doi.org/10.1145/2783258.2788613

43. Che Z, Purushotham S, Khemani R, Liu Y. Interpretable Deep Models for ICU Outcome Prediction. AMIA Annu Symp Proc. 2016;2016:371–80.

44. Caywood MS, Roberts DM, Colombe JB, Greenwald HS, Weiland MZ. Gaussian Process

Regression for Predictive But Interpretable Machine Learning Models: An Example of Predicting Mental Workload across Tasks. Front Hum Neurosci. 2016;10:647.

45. Cuchacovich M, Bueno D, Carvajal R, Bravo N, Aguillón JC, Catalán D, et al. Clinical parameters and biomarkers for anti-TNF treatment prognosis in rheumatoid arthritis patients. Clin Rheumatol. 2014;33:1707–14.

46. Thomson TM, Lescarbeau RM, Drubin DA, Laifenfeld D, de Graaf D, Fryburg DA, et al. Blood-based identification of non-responders to anti-TNF therapy in rheumatoid arthritis. BMC Med Genomics. 2015;8:26.

47. Honne K, Hallgrímsdóttir I, Wu C, Sebro R, Jewell NP, Sakurai T, et al. A longitudinal genome-wide association study of anti-tumor necrosis factor response among Japanese patients with rheumatoid arthritis. Arthritis Res Ther. 2016;18:12.

48. Adshead R, Tahir H, Bubbear J, Donnelly S, Chau I. 218. Ethnic Differences in the Response to Anti-TNF in Patients with Ankylosing Spondylitis. Rheumatology . Oxford University Press; 2015;54:i133–4.

49. Liu J, Dong Z, Zhu Q, He D, Ma Y, Du A, et al. TNF-α Promoter Polymorphisms Predict the Response to Etanercept More Powerfully than that to Infliximab/Adalimumab in Spondyloarthritis. Sci Rep. Nature Publishing Group; 2016;6:32202.

50. Hueber W, Tomooka BH, Batliwalla F, Li W, Monach PA, Tibshirani RJ, et al. Blood autoantibody and cytokine profiles predict response to anti-tumor necrosis factor therapy in rheumatoid arthritis. Arthritis Res Ther. 2009;11:R76.

51. Helliwell PS, Ibrahim G. Ethnic differences in responses to disease modifying drugs. Rheumatology . 2003;42:1197–201.

Figure 1. An overview of the drug response prediction model. The model accepts demographic, baseline disease assessment, treatment, and SNP array data of a patient, predicts changes in disease activity scores (ΔDAS) for the patient, and classifies the patient into the responder or non-responder group.

Figure 2. (A) An overview of the repeated cross-validation evaluation. All models went through both drug-specific and overall evaluations and were measured based on the listed three metrics. (B) Pearson correlation coefficients between the observed ΔDAS and predictions from tested regression methods. (C) Accuracy (the ratio of correct classification) from tested responder-vs-nonresponder classification methods. (D) Areas under receiver operating characteristic curve (AUC) of tested responder-vs-nonresponder classification methods. (The average scores are labeled above the X-axis. The final model is colored in red. GPR = Gaussian process regression, SVM = support vector machine, GB = gradient boosting regression/decision tree, RF = random forest

Figure 3. Feature space analysis of adalimumab users in the training dataset. (A) Principal component analysis of the original feature space (without kernel transformation, colored in ΔDAS) shows separation of several cohorts. (B) Principal component analysis of the original feature space (without kernel transformation, colored in cohort labels) does not show obvious separation of responders and nonresponders. (C) Principal component analysis of the kernel matrix (colored in ΔDAS) shows a clear gradient from responders to nonresponders. (D) Feature contributions to first two principal component in Subfigure C.

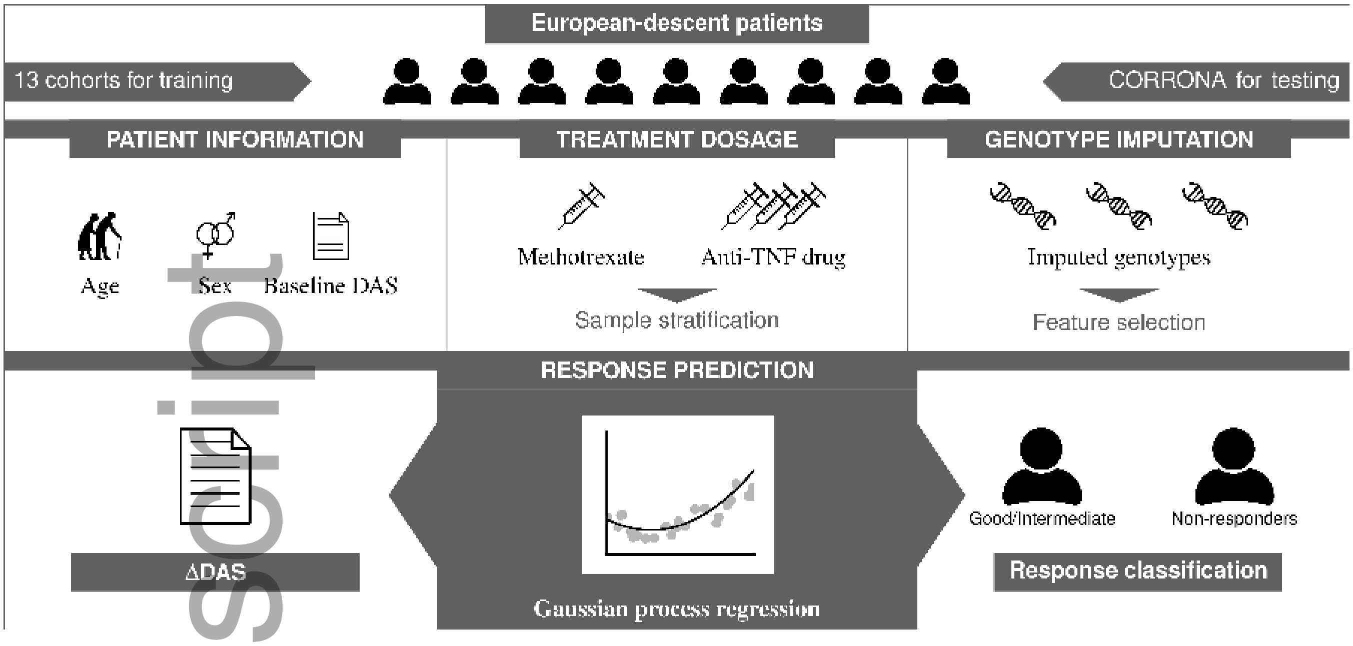
Figure 4. Repeated cross-validation tests of models with different feature sets on the training dataset. The average scores are labeled above the X-axis. The final model is colored in red. (A, D, G) Pearson correlation coefficients between the observed ΔDAS and predictions from tested regression methods for adalimumab, etanercept, and infliximab. (B, E, H) Correct classification ratio of predictions from tested responder-vs-nonresponder classification methods for adalimumab, etanercept, and infliximab. (C, F, I) Areas under receiver operating characteristic curve (AUC) of tested responder-vs-nonresponder classification methods for adalimumab, etanercept, and infliximab. (AGM = Age + Gender + Methotrexate, G-free: Model without Gender, Baseline: model using baseline DAS only)

Figure 5. Evaluation of drug-specific models and non-specific models on the CORRONA dataset. The average scores are labeled above the X-axis. The final model is colored in red. (A, D, G) Pearson correlation coefficients between the observed ΔDAS and predictions from tested regression methods for adalimumab, etanercept, and infliximab. (B, E, H) Correct classification ratio of predictions from tested responder-vs-nonresponder classification methods for adalimumab, etanercept, and infliximab. (C, F, I) Areas under receiver operating characteristic curve (AUC) of tested responder-vs-nonresponder classification methods for adalimumab, etanercept, and infliximab.

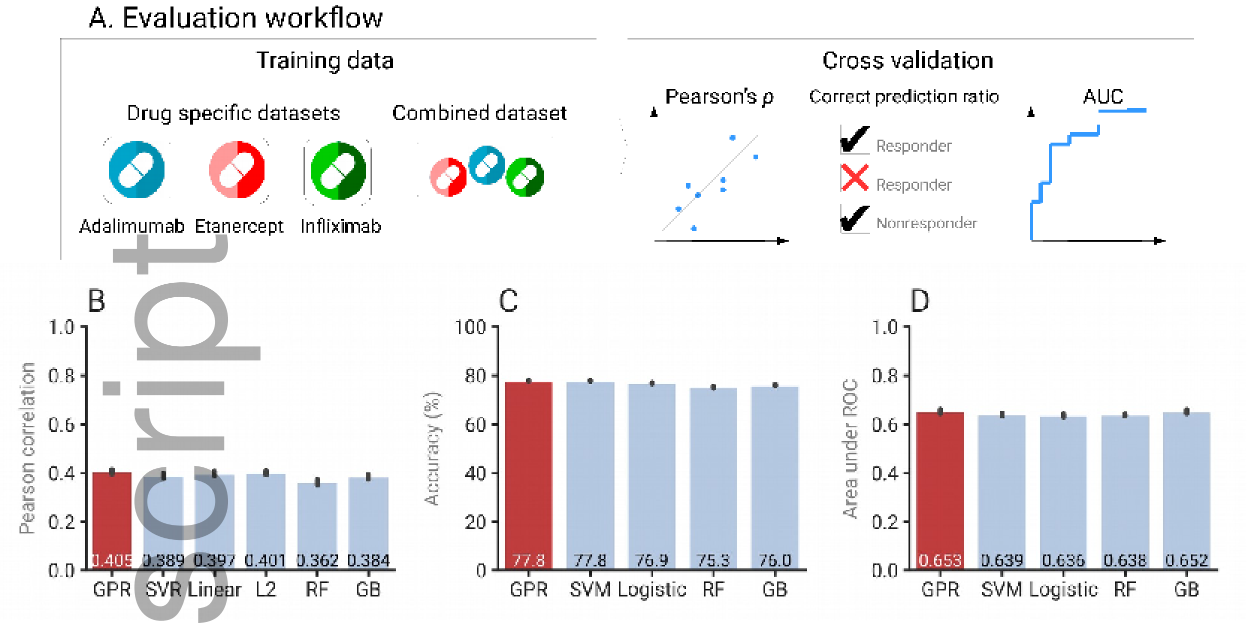
Table 1. A summary of training and testing datasets

	Training dataset (N=1892)	CORRONA dataset (N=680)
Demographic data		
Mean age (years)	54.9	55.6
Female %	75.1	77.9
Treatment information		
Methotrexate users	1332 (70.4%)	441 (64.9%)
Adalimumab	757 (40.0%)	210 (30.9%)
Etanercept	520 (27.5%)	179 (26.3%)
Infliximab	609 (32.2%)	177 (26.0%)
Certolizumab	0 (0%)	114 (16.8%)
Average baseline DAS	5.87	4.73

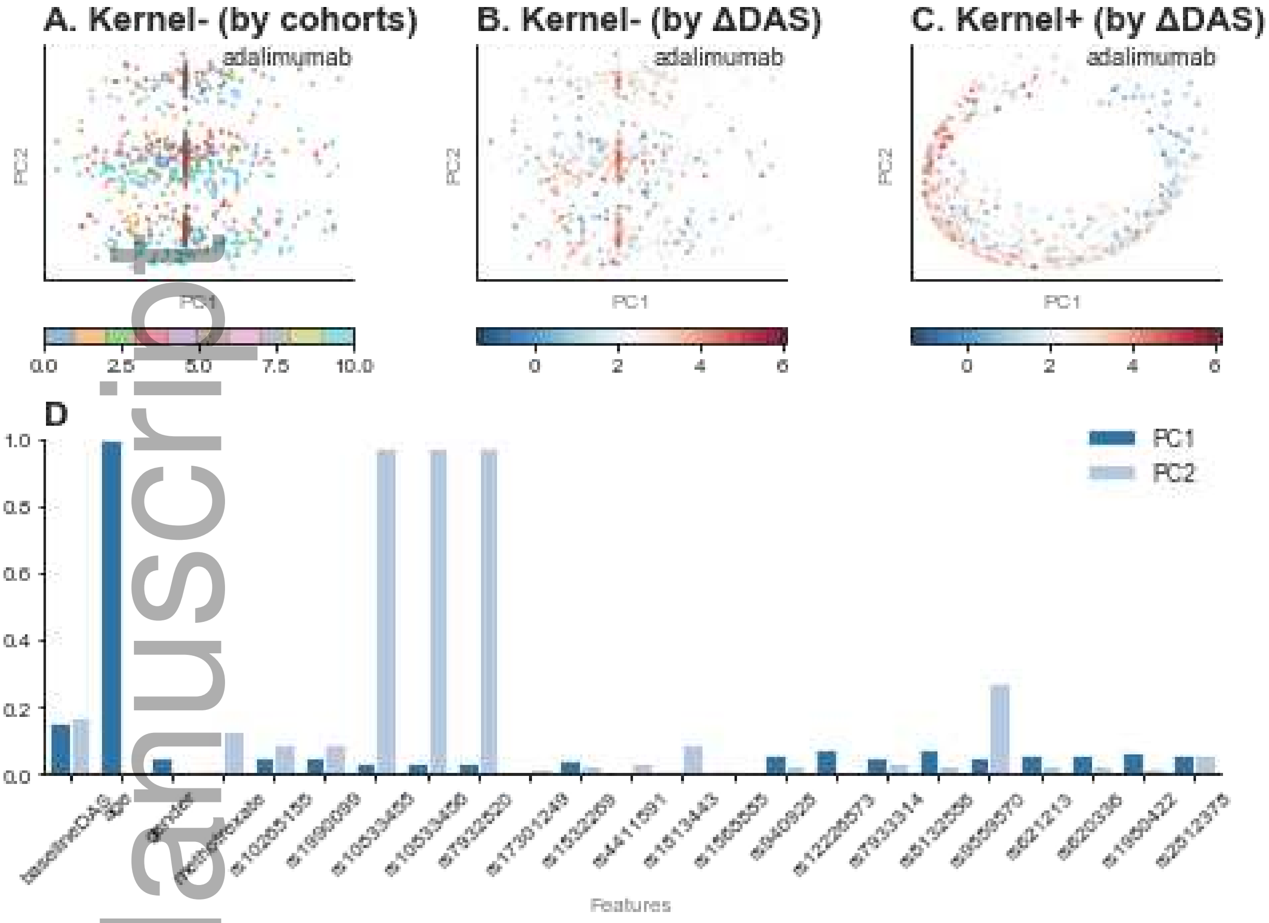
Response		
Nonresponders	436 (23%)	238 (35%)
Average ΔDAS	2.15	1.17



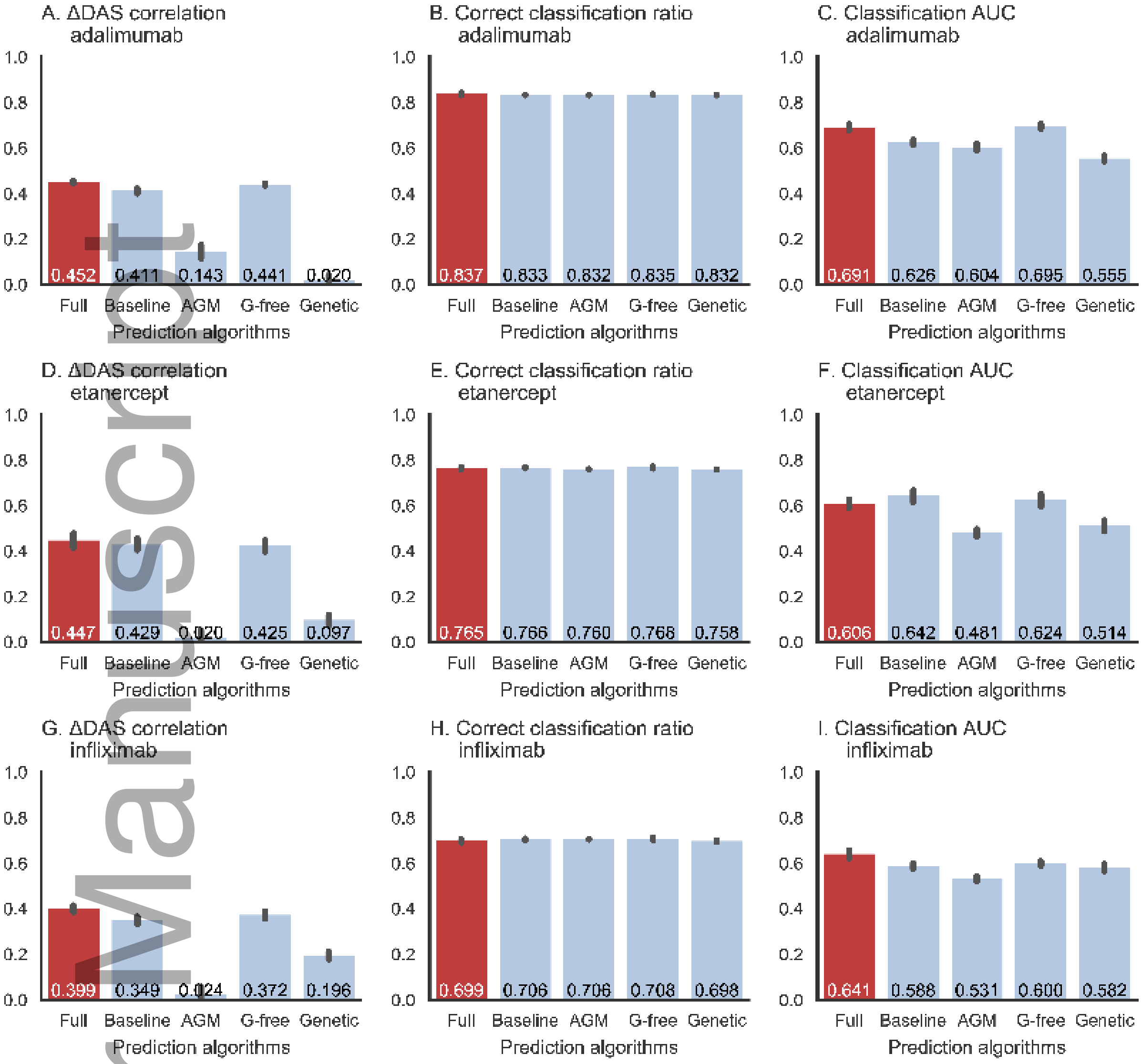
art_41056_f1.tif



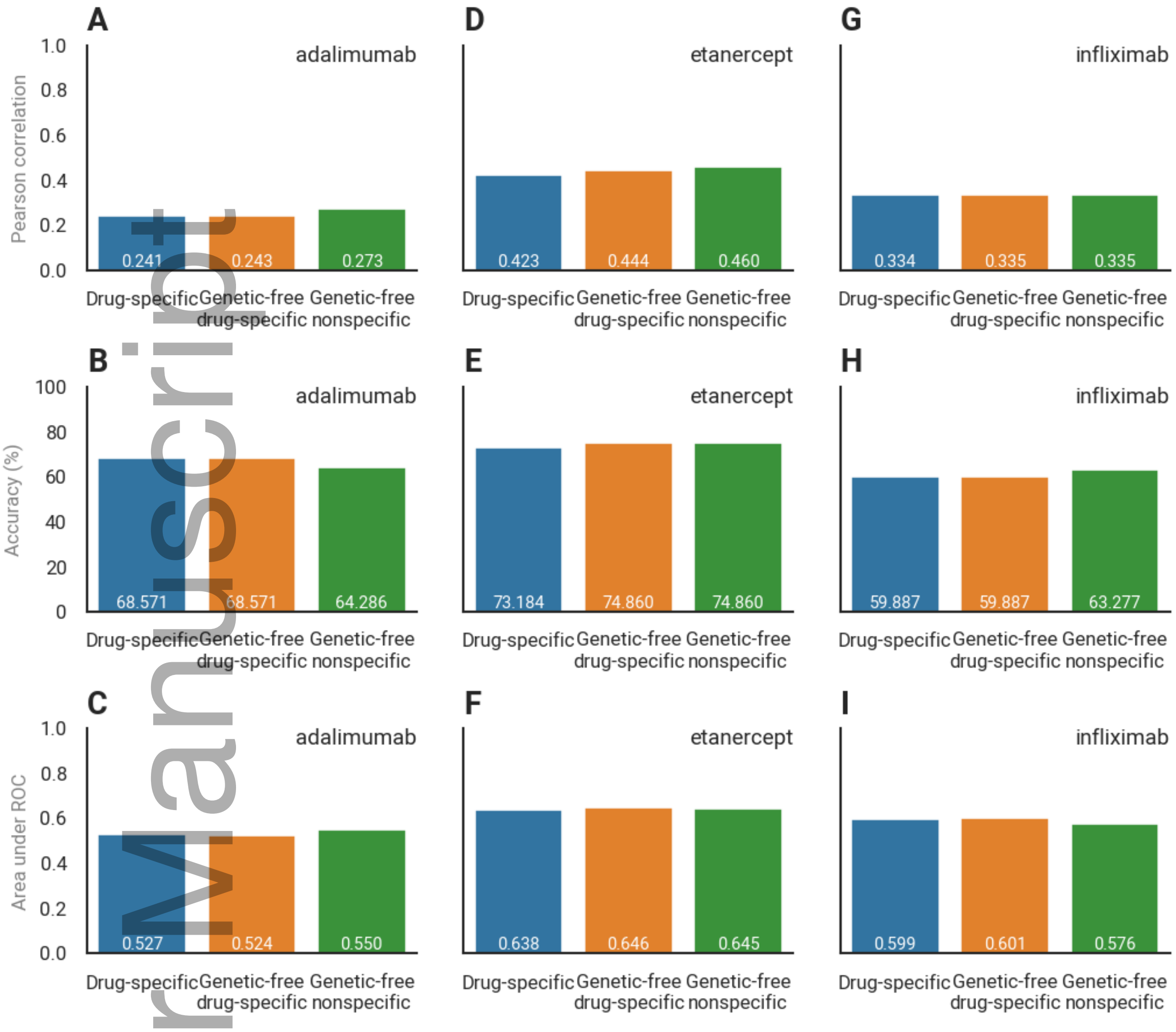
art_41056_f2.tif



art_41056_f3.png



art_41056_f4.tif



art_41056_f5.png