ORIGINAL RESEARCH ARTICLE   **Open Access**

Clinical Validation of a Blood-Based Predictive Test for Stratification of Response to Tumor Necrosis Factor Inhibitor Therapies in Rheumatoid Arthritis Patients

Theodore Mellors,[1] Johanna B. Withers,[1] Asher Ameli,[1] Alex Jones,[1] Mengran Wang,[1] Lixia Zhang,[1] Helia N. Sanchez,[1] Marc Santolini,[2] Italo Do Valle,[3] Michael Sebek,[3] Feixiong Cheng,[3] Dimitrios A. Pappas,[4,5] Joel M. Kremer,[5,6] Jeffery R. Curtis,[7] Keith J. Johnson,[1] Alif Saleh,[1] Susan D. Ghiassian,[1] and Viatcheslav R. Akmaev[1,*]

Abstract

Objectives: For rheumatoid arthritis (RA) patients failing to achieve treatment targets with conventional synthetic disease-modifying antirheumatic drugs, tumor necrosis factor (TNF)-α inhibitors (anti-TNF therapies) are the primary first-line biologic therapy. In a cross-cohort, cross-platform study, we developed a molecular test that predicts inadequate response to anti-TNF therapies in biologic-naive RA patients.

Materials and Methods: To identify predictive biomarkers, we developed a comprehensive human interactome—a map of pairwise protein/protein interactions—and overlaid RA genomic information to generate a model of disease biology. Using this map of RA and machine learning, a predictive classification algorithm was developed that integrates clinical disease measures, whole-blood gene expression data, and disease-associated transcribed single-nucleotide polymorphisms to identify those individuals who will not achieve an ACR50 improvement in disease activity in response to anti-TNF therapy.

Results: Data from two patient cohorts ($n = 58$ and $n = 143$) were used to build a drug response biomarker panel that predicts nonresponse to anti-TNF therapies in RA patients, before the start of treatment. In a validation cohort ($n = 175$), the drug response biomarker panel identified nonresponders with a positive predictive value of 89.7 and specificity of 86.8.

Conclusions: Across gene expression platforms and patient cohorts, this drug response biomarker panel stratifies biologic-naive RA patients into subgroups based on their probability to respond or not respond to anti-TNF therapies. Clinical implementation of this predictive classification algorithm could direct nonresponder patients to alternative targeted therapies, thus reducing treatment regimens involving multiple trial and error attempts of anti-TNF drugs.

Keywords: rheumatoid arthritis; anti-TNF; gene expression; precision medicine; drug response prediction

[1]Scipher Medicine, Waltham, Massachusetts, USA.
[2]Center for Research and Interdisciplinarity (CRI), University Paris Descartes, Paris, France.
[3]Center for Complex Network Research, Department of Physics, Northeastern University, Boston, Massachusetts, USA.
[4]Division of Rheumatology, College of Physicians and Surgeons, Columbia University, New York, New York, USA.
[5]CORRONA, LCC, Waltham, Massachusetts, USA.
[6]Albany Medical College, The Center for Rheumatology, Albany, New York, USA
[7]Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA.

*Address correspondence to: Viatcheslav R. Akmaev, PhD, Scipher Medicine, 221 Crescent Street, Suite 103A, Waltham, MA 02453, E-mail: slava.akmaev@ sciphermedicine.com

Introduction

Rheumatoid arthritis (RA) is a complex autoimmune disease with no cure. However, many potent treatment options are available to mitigate the symptoms and arrest progression of the disease. Treatment guidelines recommend early therapeutic intervention to forestall permanent functional debilitation associated with structural joint damage.[1–3] Antitumor necrosis factor-α (anti-TNF) therapeutics are the first-line targeted therapy for nearly 90% of biologic-naive RA patients whose disease is not adequately controlled with conventional synthetic disease modifying antirheumatic drugs (csDMARDs), such as methotrexate.[4,5] However, ∼70% of these RA patients do not achieve meaningful clinical change on anti-TNF therapy.[6–16] Therefore, there is a clinical need for a test that predicts which patients will not respond to anti-TNF agents before the initiation of therapy.

RA patient treatment targets are defined by the American College of Rheumatology (ACR) as disease activity scores indicating either disease remission or low disease activity.[17–19] Treatment response relative to baseline is measured as ACR20, ACR50, and ACR70 where the number refers to the percent improvement in a standard set of measures.[19] Factors used to calculate disease activity scores and ACR values include the number of swollen joints, tender joints, and both patient- and physician-reported assessments of pain and global health, as well as blood biomarker levels.[20] ACR20 is reported as the benchmark in clinical trials for approval of new therapies; however, an ACR20 response is often insufficient to reach the clinically meaningful disease activity scores outlined in RA treatment guidelines.[21–24] Rather, patients typically need to achieve an ACR50 response to reach disease activity scores corresponding to remission or low disease activity.[25]

Complex multifactorial disorders such as RA have traditionally been difficult to study.[26] The lack of a clear pattern of inheritance indicates that genetic information is integral to disease biology, but not sufficient for diagnosis or to guide treatment choices.[27,28] Classic approaches to understand disease biology rely on either genetic analysis of mutations that drive disease-associated phenotypes or high-throughput molecular expression methodologies that are limited in their ability to sufficiently represent the complexity of the disease pathology.[29] Advanced computational approaches that combine clinical data with integrative genetics are a necessary progression to bring modern precision medicine to autoimmune diseases. Network-based approaches to understanding human disease biology,[30–33] collectively known as network medicine, have the ability to reveal underlying molecular patterns among features such as DNA markers and transcripts. Commonly, RNA sequencing (RNAseq) data have been used with the single purpose of analyzing gene expression features. However, RNAseq data also contain critical information on functionally active single-nucleotide polymorphisms (SNPs) that, when combined with transcript expression, have the potential to dramatically enhance the predictive power of a molecular signature. This study describes a multifaceted computational approach to develop a clinically useful tool to guide the treatment of RA patients.

Materials and Methods

Study populations

Discovery cohort. Patient microarray data (accession GSE15258) were obtained from the Gene Expression Omnibus. Details of sample collection and cohort information were previously reported[34]; sample collection study protocols were approved by the local ethics committees and patients provided informed consent. Briefly, RA patients naive to anti-TNF therapy were enrolled and blood samples collected in PAXgene tubes. Therapeutic response was evaluated 14 weeks after initiation of treatment according to the DAS28-CRP EULAR response definition.[35] Fifty-eight female patient samples were arbitrarily selected for this study.

Training cohort and validation cohort. RA patient whole-blood samples and clinical measurements were prospectively collected in the CERTAIN trial by the Consortium of Rheumatology Researchers of North America (CORRONA).[36] The CERTAIN study was designed as a prospective comparative effectiveness study involving 43 sites and 117 rheumatologists. Institutional review board or ethics committee approvals were obtained before sample collection and study participation, and patients provided informed consent. Samples selected for the present study were from patients who were biologic-naive at the time of sample collection. Patients were treated with adalimumab, certolizumab pegol, etanercept, golimumab, or infliximab at the discretion of the treating physician and followed longitudinally for at least 6 months. In addition to a medical history, clinical assessments collected at 0 and 6 months post-therapy initiation included tender and swollen joint counts, physician and patient global disease activity scores, csDMARD dose, patient pain

evaluation, and quality-of-life surveys. Laboratory studies performed at a central laboratory included a complete blood count, C-reactive protein (CRP) levels, rheumatoid factor titer, and anti-cyclic citrullinated protein (anti-CCP) titer. Training ($n = 143$) and validation trial ($n = 175$) patient cohorts were balanced for response rate, age, and gender. Patients were included in the independent validation trial if they had a Visual Analog Scale pain score[37] of >15 out of a maximum score of 100. Consistent with the inclusion criteria of the CERTAIN study, all patients in the validation trial had a Clinical Disease Activity Index greater than 10.

Evaluation of clinical response to anti-TNF therapy

Among the CERTAIN study samples, response at 6 months after anti-TNF therapy initiation was defined by ACR50. An ACR50 responder was defined as an individual exhibiting ≥50% improvement in 28 tender joint count, ≥50% improvement in 28 swollen joint count, and ≥50% improvement in at least three out of five clinical variables (Health Assessment Questionnaire disability index, patient pain, patient global assessment, physician global assessment, and CRP level).[19]

RNA isolation and quality control

Total RNA was isolated from blood collected in PAXgene Blood RNA Tubes using the PAXgene Blood miRNA Kit (PreAnalytiX) according to the manufacturer's instructions. RNA quality was assessed using the Agilent Bioanalyzer, and samples were quantitated using a NanoDrop ND-8000 spectrophotometer.

RNAseq analysis and gene expression preprocessing

RNA was processed using the GLOBINclear (Thermo Fisher), Ribo-Zero Magnetic Gold (Epidemiology), and TruSeq Stranded Total RNA (Illumina) kits according to the manufacturer's instructions. Libraries were processed on a NextSeq 550 DX or a NovaSeq 6000 sequencer for 75 cycles. An average of 42.4 million reads were captured per patient, with a range of 33.7–58.6 million. Fifty-nucleotide reads were mapped to the GRCh37 human genome with STAR alignment software.[38,39] Per gene abundance in fragments per kilobase of transcript per million mapped reads was calculated with the RSEM software package.[40] Samples with an RNA integrity score of >4, and >7 million protein-coding reads were analyzed. Samples were processed in seven sequencing batches over a 1-year period. No detectable batch effect was observed between the NextSeq and NovaSeq processed libraries based on a principal component analysis.[41] Hierarchical cluster analysis of RNA expression data for 37 genes is as previously described.[42]

SNP analysis

Samples were aligned to the GRCh38 human genome with STAR alignment software.[38] SNPs were called using a modified version of the Genome Analysis ToolKit Best Practices workflow for SNP and indel calling on RNAseq data.[43–45] Thirty-nine RA-associated SNPs were evaluated (Supplementary Table S2).[46]

Selection of gene expression biomarkers of nonresponse

Gene expression biomarkers of nonresponse to anti-TNF therapy were selected from the 58-patient discovery cohort. Throughout 100 repeats, 80% of samples were randomly selected. Mann–Whitney U test was used to eliminate any gene expression features not significantly different in expression between responders and nonresponders ($p > 0.05$). Random Forest from Scikit-learn[34,47,48] was then used to rank the remaining features based on mean decrease impurity, a metric of feature importance. Features that ranked in the top 100 in at least 30 out of the 100 repeats were selected and considered for further evaluation in the RNAseq training cohort. The anti-TNF therapy response signal was observed in the female subpopulation, but was cross-validated in the entire training data set that included males and females. The all-patient cross-validation performance of the classifier based on the feature set selected from the female data was consistently higher.

Model building, validation, and statistical analyses

For model training, 143-patient samples were evaluated. Before final model building, a second round of feature selection took place to further evaluate the biomarker panel among the RNAseq training cohort. A total of 70 features were evaluated, consisting of gene expression biomarkers selected from the discovery cohort, SNPs, and clinical features (Supplementary Tables S1–S3). Each of the 70 features was ranked by assessing the decrease in cross-validated model performance (96 repeats of 20% withheld cross-validation) when a given feature was removed. Features that caused the largest drops in cross-validated model performance upon removal were considered to be the most important. The top 25 ranked features were used to build a finalized Random Forest model, which was subsequently

applied to the withheld 175-patient validation set samples. Model performance was evaluated using area under the receiver operating curve,[49] positive predictive value, and specificity.[50] All statistical analyses were performed using Python 3.7.6. Odds ratios and confidence intervals (CI) were calculated as previously described.[51,52] Chi-squared test was used to determine the significance of differences in model performance stratified by ethnicity.

Building the human interactome and RA disease module, and performing network medicine analyses of molecular features

The human interactome was assembled as previously described[30] from 21 public databases (Supplementary Table S4) containing different types of experimentally derived protein/protein interaction (PPI) data: (1) binary PPIs, derived from high-throughput yeast-two hybrid experiments (HI-Union),[53] three-dimensional protein structures (Interactome3D,[54] Instruct,[55] Insider[56]), or literature curation (PINA,[57] MINT,[58] LitBM17,[53] Interactome3D, Instruct, Insider, Bio-Grid,[59] HINT,[60] HIPPIE,[61] APID,[62] InWeb[63]); (2) PPIs identified by affinity purification followed by mass spectrometry present in BioPlex2,[64] QUBIC,[65] CoFrac,[66] HINT, HIPPIE, APID, LitBM17, and InWeb; (3) kinase/substrate interactions from KinomeNetworkX[67] and PhosphoSitePlus[68]; (4) signaling interactions from SignaLink[69] and InnateDB[70]; and (5) regulatory interactions derived by the ENCODE consortium. We used the curated list of PSI-MI IDs provided by Alonso-López et al.[62] for differentiating binary interactions among the several experimental methods present in the literature-curation databases. All proteins were mapped in their corresponding NCBI Entrez ID and the proteins that could not be mapped were removed. The resulting human interactome includes 18,505 proteins and 327,924 interactions.

The DIAMOnD approach[31] was used to generate an RA disease module. Proteins used to seed the disease module were linked to RA by at least two of five databases: GWAS Catalog,[71] HuGE Navigator Phenopedia,[72] ClinVar,[73] OMIM,[74] and MalaCards.[75] DIAMOnD identified proteins that were significantly enriched in the same Gene Ontology biological process terms as the disease-associated proteins.

Proximity of the molecular features to each other on the human interactome map was calculated as previously described.[76] Briefly, the closest distance was defined as the minimum path length between each protein and the other proteins in the set. Significance of the observed closest distance, reported as a *z*-score, was evaluated in comparison with the expected closest distance determined from 10,000 random protein sets of the same size. Randomizations were performed as previously described.[76]

Pathway enrichment analysis

KEGG, BioCarta, Reactome, and Signal Transduction pathway annotations were obtained from the Molecular Signatures Database (MSigDB), Version 6.2.[77] Fisher's exact test was used to identify biological pathways. Pathways with a Bonferroni corrected *p*-value of <0.05 were considered enriched. IL10,[78] POMC,[79] JAK1,[80] ICOSLG,[81] TNF, TNFSF11,[82] NR3C1,[83] P2RY12[84] (NCT02874092), PTGER4,[85] GGPS1,[86] FDPS,[86] TNFRSF13B (NCT03016013), IL6,[87] ESR1,[88] ESR2,[88] ITK[89] (NCT02919475), BTK,[90] TLR4[91] (NCT03241108), IRAK4,[92] JAK2,[80] JAK3,[80] HDAC1 (NCT02965599), PSMB5,[93,94] ADORA3,[95] ITGA9[96] (NCT02698657, NCT03257852), IFNB1[97] (NCT02727764; NCT03445715), and CX3CL1[98] were the approved drug targets in RA.

Results

Building the human interactome and a map of RA disease biology

To begin developing the network medicine tools necessary to evaluate human disease biology, we created a map of cellular components and their physical interactions. By amalgamating publicly available data (Supplementary Table S4) of 327,924 pairwise PPIs between a total of 18,505 proteins, a comprehensive map of biology called the human interactome was created (see the Materials and Methods section).[30] There are ~20,000 proteins encoded by the human genome, and therefore, this human interactome incorporates interaction data from more than 90% of human proteins.

Disease-associated proteins tend to interact with each other in a subnetwork on the human interactome called a disease module.[30] Using the DIAMOnD approach[31] that aggregates potential disease-associated proteins based on their network proximity to known disease-associated proteins, an RA disease module was generated that contains ~200 proteins. Of these, 66% were linked to RA in genome-wide association study databases and DIAMOnD identified the remaining proteins that are significantly enriched in the same Gene Ontology biological process terms as the known disease-associated proteins.

Using information from the human interactome and the RA disease module, we sought to identify a blood-based molecular signature that integrates clinical variables and molecular features to predict which RA patients will not respond to anti-TNF therapies (Fig. 1). Briefly, molecular features that discriminate between responders and nonresponders to anti-TNF therapies were selected from a publicly available microarray data set. In a cross-platform analysis, these features were combined with RA disease module-associated SNPs and clinical factors. Then, a machine-learning algorithm was trained using these features and RNAseq data. Finally, performance of the predictive drug response algorithm was validated in an independent validation trial.

Cross-platform identification of discriminatory gene expression features in whole blood predictive of inadequate response to anti-TNF therapy in RA patients

To maximize the clinical utility of a test that predicts nonresponse to therapy, a routine noninvasive or min-imally invasive sample source that does not require specialized specimen collection procedures is ideal. For this reason, we analyzed gene expression data derived from whole blood.[99,100] Gene expression that is discriminatory between patients considered responders and nonresponders to anti-TNF therapies was selected from a publicly available microarray discovery cohort data set of 58 biologic-naive RA patients using the Random Forest machine-learning algorithm (see the Materials and Methods section).[34,47] Of the 21,818 genes in the discovery data set for which gene expression was assessed, 37 were selected as discriminatory biomarkers (Supplementary Table S1). Hierarchical cluster analysis illustrates that the gene expression profiles of these 37 biomarker transcripts can distinguish responders ($n=17$) and nonresponders ($n=41$) to anti-TNF therapies (Fig. 2). Two main clusters were observed, one predominantly nonresponders and the other responders, thereby substantiating the discriminatory nature of this multivariate molecular signature for response prediction. Transcriptional profiling by
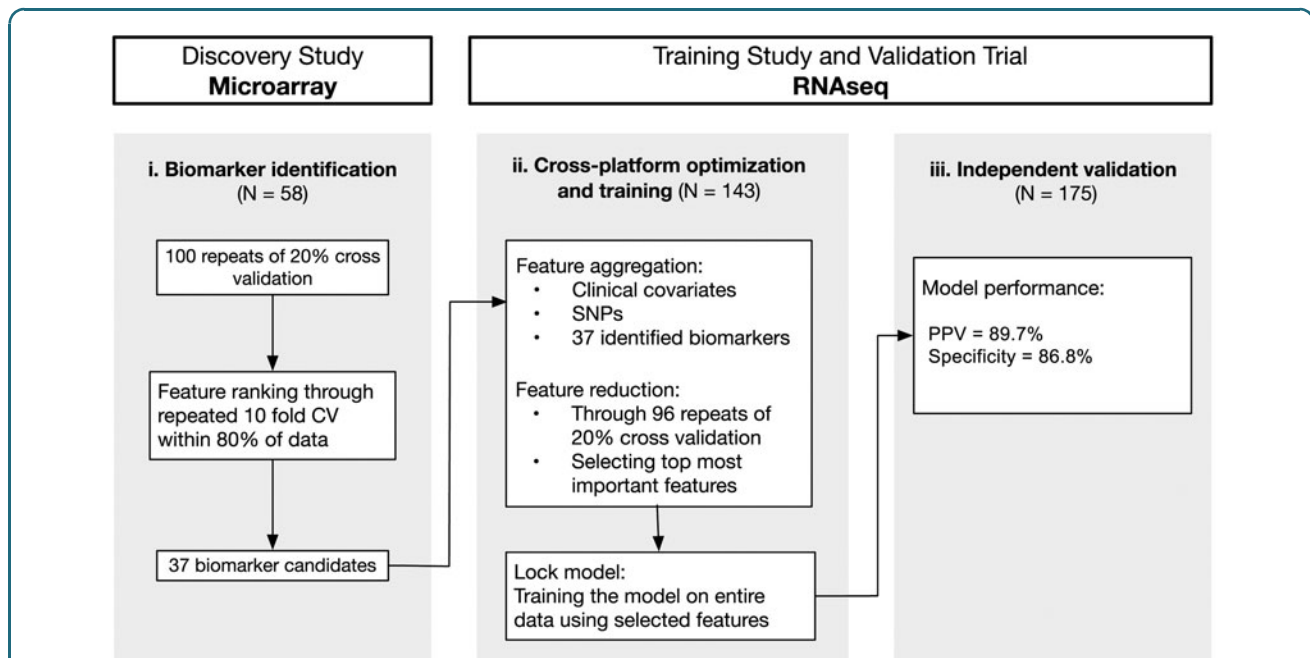

FIG. 1. Flowchart describing development of anti-TNF drug response algorithm in RA. Gene expression that discriminates between responders and nonresponders to anti-TNF therapies was selected from a publicly available microarray data set. In a cross-platform analysis, these features were combined with network disease module-associated SNPs and clinical factors, and then used to train a machine-learning algorithm using RNAseq data. Finally, performance of the predictive drug response algorithm was validated in an independent validation trial. CI, confidence interval; CV, cross-validation; PPV, positive predictive value; RA, rheumatoid arthritis; SNPs, single-nucleotide polymorphisms; TNF, tumor necrosis factor.

microarray and RNAseq varies in dynamic range and exhibits some discordance in the number and extent of differential gene expression observed.[101–107] Nonetheless, the majority of the transcripts (19/37; 51.4%) identified as discriminatory of anti-TNF drug response in microarray data also differentiated between responders and nonresponders in RNAseq data (Supplementary Fig. S1).

Evaluation of disease-associated SNPs from RNAseq data

RNAseq provides information on nucleotide sequence that is lacking from microarray analyses. In addition to gene expression, variations in RNA sequence may be predictive of nonresponse to anti-TNF therapy in RA patients. To this end, a training data set was generated from clinical data and whole-blood RNAseq data obtained from 143 RA patients in the CORRONA CERTAIN study.[36] Characteristics and demographics of the patient populations are summarized in Table 1. Although SNP analysis is traditionally performed on whole-genome sequencing data, the majority of the genome is transcribed.[108,109] Thus, most SNP variants can be detected in ribosomal RNA-depleted RNAseq data.[110] Because these RNAseq data were derived from blood samples, only SNPs that are associated with RA, which have been functionally linked to gene expression changes in peripheral blood mononuclear cells through expression quantitative trait loci (eQTL) analysis, were examined (Supplementary Table S2).[46] The genetic loci associated with the selected SNPs have a significant overlap with the RA disease module (Fig. 3B). Twenty-two such disease module-associated SNPs were above the limit of detection in the patient RNAseq data and thus included in further analyses (Supplementary Fig. S2).

Integration of SNPs, gene expression data, and clinical variables to develop a multifactorial predictive drug response algorithm

Gene expression indicative of drug response (Supplementary Table S1), RA-associated SNPs (Supplementary Table S2), and clinical factors (Supplementary Table S3) represent a set of 70 biomarker features
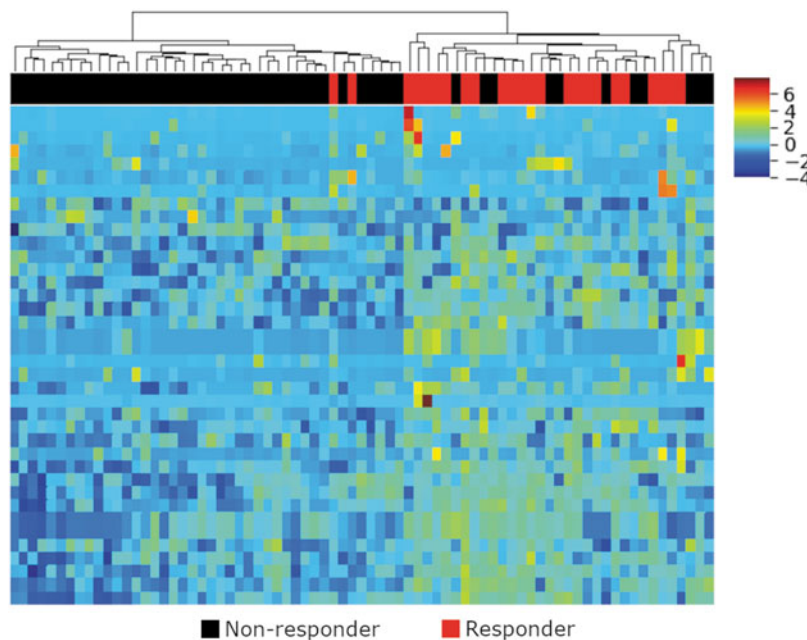

■ Non-responder　■ Responder

FIG. 2. Discriminatory genes indicative of nonresponse to anti-TNF therapy in RA patient blood samples. Hierarchical cluster analysis[42] of RNA expression data for 37 genes illustrates two main groupings, one predominantly nonresponders and the other responders, thereby substantiating the discriminatory nature of these genes for anti-TNF response prediction. The heatmap represents the relative RNA expression level in arbitrary units.

Table 1. Patient Demographics and Disease Characteristics at Baseline

Characteristics	Training cohort (*n* = 143)	Validation trial cohort (*n* = 175)
Age, years (mean ± SD)	55.0 ± 12.9	53.8 ± 11.9
Female, %	72.7	73.1
Duration of disease, years (mean ± SD)	4.7 ± 6.7	5.0 ± 7.5
Positive for anti-cyclic citrullinated peptide, %	66.9	60.9
Positive for rheumatoid factor, %	73.9	70.9
Race		
White	88.1	84.6
Black	6.3	6.3
Other	5.6	9.1
Current csDMARD use, %		
Methotrexate	65.0	60.0
Hydroxychloroquine	3.5	4.0
≥2 csDMARDs	13.3	13.7
None	15.4	16.0
Concomitant prednisone, %	34.3	22.9
Baseline prednisone dose, mg (mean ± SD)	7.4 ± 4.7	8.5 ± 5.3
Anti-TNF use, %		
Adalimumab	38.5	38.9
Etanercept	33.6	30.9
Infliximab	16.8	20.0
Certolizumab pegol	8.4	7.4
Golimumab	2.8	2.9
CDAI (mean ± SD)	28.5 ± 13.5	31.0 ± 12.6
DAS28-CRP (mean ± SD)	4.9 ± 1.1	5.0 ± 1.0
Swollen joint count (mean ± SD)	7.2 ± 6.0	8.1 ± 5.5
Tender joint count (mean ± SD)	10.8 ± 7.3	12.0 ± 7.3
ACR50 responders, %	30.8	30.3

CDAI, Clinical Disease Activity Index; csDMARDs, conventional synthetic disease modifying antirheumatic drugs; SD, standard deviation; TNF, tumor necrosis factor.

that were used to train and develop a drug response algorithm that is predictive of nonresponse to anti-TNF therapies. Using ACR50 at 6 months as a benchmark, the training cohort population had a response rate to anti-TNF therapies of 30.8% (44/143). This is representative of the general population[111] and reflects the real-world prospective collection approach of the CORRONA CERTAIN study.[36] Random Forest was used to generate predictive models with 80% of the RNAseq training data set using features from the discriminatory gene expression set, SNPs, and clinical factors. The remaining 20% of the data set was withheld for performance testing during the model building process. Although 70 candidate biomarkers were considered in this analysis, not all were required to predict the anti-TNF therapy response. The biomarker panel used to develop the final model predictive of inadequate response to anti-TNF therapies included 10 SNPs, 8 transcripts, 2 laboratory tests (CRP and anti-CCP), and 3 clinical metrics (sex, body mass index [BMI], patient disease assessment).

Independent validation trial of a drug response biomarker panel predictive of nonresponse to anti-TNF therapy in RA patients

To confirm that the drug response biomarker panel is generalizable, an independent group of prospectively collected samples (*n* = 175) were used to conduct a validation trial. The samples included in the validation cohort were not used for any stage of the biomarker panel development, and the algorithm has no information derived from the gene expression data or clinical outcomes from these patients.

Independent validation of the drug response biomarker panel stratified the validation cohort into predicted nonresponders and responders, with a highly statistically significant odds ratio of 6.57 (95% CI 2.75–15.70) of being a nonresponder in the respective subgroup. The odds ratio is a statistic that quantifies the strength of association between two events, in this case the signal from the biomarker panel and inadequate response to anti-TNF therapies. This means that a patient identified by the drug response biomarker panel to be a nonresponder is 6.57 times more likely to inadequately respond to an anti-TNF therapy than if that patient was a responder. The drug response biomarker panel identified patients who are unlikely to have an adequate response to anti-TNF therapies with a positive predictive value (PPV) of 89.7% (95% CI 79.0–95.7%), specificity of 86.8% (95% CI 72.4–94.1%), and sensitivity of 50.0% (95% CI 40.8–58.7%) (Table 2). Patients predicted to be nonresponders have an observed ACR50 response rate of 10.3% (7/68) with anti-TNF therapies, significantly lower than the overall response rate of 30.3% (53/175). Conversely, the predicted responders had an observed ACR50 response rate of 43.0% (46/107), which is a 41.9% improvement from that of the unstratified patient population.

Alternatively, using the more stringent threshold of response, ACR70, which requires patients to exhibit a 70% improvement in the ACR response criteria, 81.7% (143/175) of patients in the validation cohort were nonresponders. The drug response biomarker panel identified patients who are unlikely to achieve an ACR70 response to anti-TNF therapies with a PPV of 93.5% (95% CI 84.0–97.9%) and specificity of 84.4% (95% CI 62.7–96.3%). The drug response biomarker panel predicted 50.3% of these nonresponders (72/143) in addition to misclassifying five ACR70 responders.
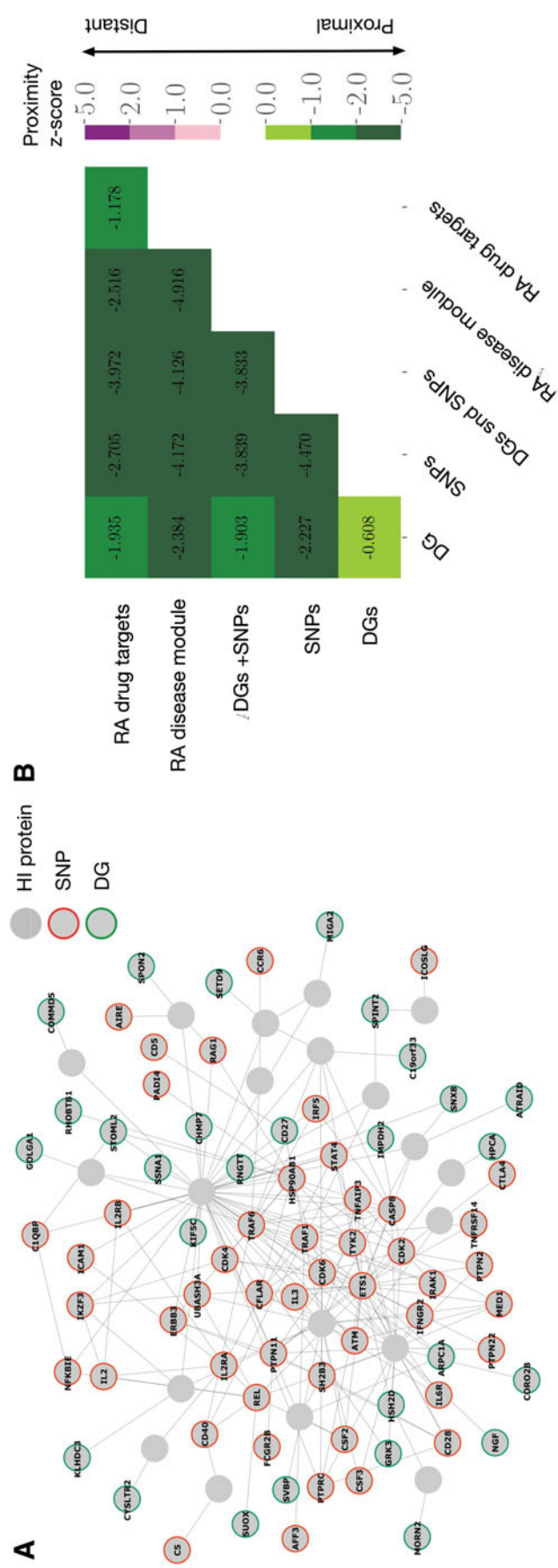
FIG. 3. Molecular features are closely associated with each other and disease-associated proteins on the HI. **(A)** Visualization of a subset of the HI protein/protein network. Proteins included on the network are indicated as gray circles. Those outlined in red represent proteins encoded by SNP eQTL transcripts and those outlined in blue represent proteins associated with RA (disease module proteins), and molecular features included in the drug response biomarker panel (proteins encoded by SNP-associated transcripts and proteins encoded by DGs). Proximity scores (z-scores) reflect whether two sets of proteins are significantly proximal to each other (z-score < −1.6). DG, discriminatory genes; eQTL, expression quantitative trait loci; HI, human interactome.

Table 2. Predictive Drug Response Biomarker Panel Validation Performance

	ACR50 (R)	ACR50 (NR)	Sum
Drug response biomarker panel (R)	46	61	107
Drug response biomarker panel (NR)	7	61	68
Sum	53	122	

NR, non-responder; R, responder.

The drug response biomarker panel includes SNPs, and SNP allele frequencies can vary between ethnic groups.[112,113] Although the training and validation set patient populations were predominantly Caucasian (88.1% and 84.6%, respectively), no significant differences ($p > 0.23$) were observed in the ability of the biomarker panel to predict an inadequate response to anti-TNF therapies among patients of other ethnicities. Future work using population-specific studies will address transethnic prediction performance.

Biological interpretation of gene products that discriminate between responders and nonresponders to anti-TNF therapy

To characterize the applicability of the predictive drug response biomarker panel to RA disease biology, the protein products of the discriminatory genes and SNP eQTLs[46] were analyzed using the human interactome and pathway enrichment analyses. The proteins encoded by the discriminatory genes and SNP eQTLs included in the drug response biomarker panel were mapped onto the human interactome map (Fig. 3A and Supplementary Fig. S3). In total, 42 proteins mapped onto the human interactome: 24 are contributed by the discriminatory genes and 18 by the SNP eQTLs. These molecular features are significantly connected in the same network vicinity of the human interactome highlighting a small, yet cohesive biological network that unifies the molecular features that predict inadequate response to anti-TNF therapies. Quantification of this proximity (see the Materials and Methods section) indicates that these different molecular features are significantly close to each other in comparison with random expectation (z-score $= -3.83$). Furthermore, the RA disease module (z-score $= -4.92$) and RA drug targets such as JAK and TNF-α (z-score $= -3.97$) are proximal to the SNP eQTLs and discriminatory genes collectively (Fig. 3B).

Pathway enrichment analysis was performed to gain insight into the molecular pathways involved in the anti-TNF therapy response. T cell signaling was identi-fied as the most enriched pathway in the pathway analysis databases queried. The relevance and importance of T cell signaling to both anti-TNF therapy response and the disease biology of RA are well established.[114–117]

Discussion

By incorporating microarray gene expression data, RNAseq, biological network analyses, and machine learning to large patient cohorts, this study describes the development of a drug response biomarker panel that uses whole-blood gene expression data to identify a molecular signature that predicts nonresponse to anti-TNF therapies in biologic-naive patients with RA. Validation with a prospectively collected RNAseq data set demonstrated that the drug response biomarker panel could predict nonresponse to anti-TNF therapies in an independent cohort of biologic-naive RA patients. Anti-TNF therapies failed to help nearly 70% of the unstratified patient population reach an ACR50 response. The drug response biomarker panel could have prevented half of these individuals from taking a drug that did not ameliorate the signs and symptoms of their disease. The biomarker panel included 10 SNPs, 8 transcripts, 2 laboratory tests (CRP and anti-CCP), and 3 clinical metrics (sex, BMI, patient disease assessment).

A single, large-scale, high-throughput analysis approach is yet to capture the molecular signature of RA disease biology. Many studies have hypothesized that the biology of nonresponse to anti-TNF therapies is reflected in the transcriptome of whole blood.[34,118–123] However, none has been translated into the clinic, which is likely a reflection of both the complexity of RA disease biology and the varying methodologies used for algorithm development. Furthermore, limited sample sizes and the complexity of gene expression data analyses may have thus far prevented the development of an algorithm that is generalizable across patient cohorts and to the wider patient population. The drug response biomarker panel described in this work differs from previous attempts to predict response to anti-TNF therapies in that it integrated RNA transcript levels and RNA sequence information. SNPs can affect many aspects of cellular biology, including the propensity for regulatory elements to interact with their cognate protein partners, the ratio or identity of alternative splice variants produced from a gene locus, transcript levels, and protein sequence.[124–126] Therefore, the functional readout of disease-associated SNPs contributes to the propensity of an individual to develop disease as well

as the inclination for environmental factors to influence pathobiology.[127,128] Many regulatory elements and genomic regions that do not encode protein are transcribed, such as in the form of enhancer RNAs[129] and promoter-associated transcripts.[130] Thus, many SNPs that influence spatial- and temporal-specific changes in transcription can be evaluated from RNAseq data. Analyzed together into a single-molecular signature, SNP and gene expression analyses can capture phenotypic variation and pathway associations that may otherwise be missed.

The microarray and RNAseq gene expression analysis platforms differ in RNA detection methodologies and statistical tools to determine normalized gene expression values.[131,132] Despite these differences in technology, the cross-platform and cross-cohort universality of the molecular features identified in this study highlights the presence of a robust molecular signature underlying the biology of anti-TNF drug response. Examination of the molecular pathways that identify patients who will not respond to anti-TNF therapies demonstrated a connection between T cell signaling and RA disease biology.[133–138] Synovial inflammation results from leukocyte infiltration into and retention in the synovial compartment, as well as from insufficient apoptosis of chronic inflammatory cells.[139,140] This synovial infiltrate includes natural killer cells, CD4$^+$, and CD8$^+$ T cells.[141–145] Furthermore, large numbers of activated T regulatory cells can be detected in the joints of RA patients.[146] The remaining discriminatory genes that are not associated with T cell signaling likely represent different aspects of RA disease that differ between those patients who will or will not respond to anti-TNF therapies. The connection to RA disease biology speaks to the reliability and applicability of the drug response biomarker panel to be a powerful clinical tool for identification of anti-TNF nonresponders.

For patients predicted to inadequately respond to anti-TNF therapies, many alternative biologic and targeted synthetic therapies are available. Because predicted nonresponders had a 10% observed response rate to anti-TNF therapies, they may see a greater benefit from being prescribed an alternative treatment with a higher reported response rate. Clinical trials assessing the efficacy these alternative therapies report ACR50 response rates of 30–40% at 6 months following alternative treatment initiation for patients who had inadequately responded to an anti-TNF therapy.[147–150] The remaining patients—those lacking a molecular signature of nonresponse—would still have been prescribed an anti-TNF therapy and would have a response rate of 43% (Table 2). Inadequate responders to anti-TNF therapies who are not identified by the test likely would be directed to anti-TNF therapies, which is overwhelmingly the default first-line biologic for biologic-naive RA patients, and thus, their treatment path would be unaffected by the results of this molecular signature test.

Conclusion

Customization of treatment regimens to match the individualized disease biology of each patient is a goal of modern medicine. This personalized approach to medicine is used in oncology, where particular therapies are prescribed to patients with specific genomic markers.[151,152] Development and validation of a drug response algorithm that predicts nonresponse to a targeted therapy using this machine-learning and network medicine approach show great promise for advancing precision medicine in the treatment of RA and other complex autoimmune diseases where costly therapeutic interventions are met with inadequate patient response.

Authors' Contributions

T.M., A.A., A.J., M.W., L.Z., M.S., I.D.V., M.S., F.C., and S.D.G. performed computational analyses. J.M.K. and D.A.P. oversaw sample collection and clinical data curation. J.B.W., H.N.S., and K.J.J. aided in interpreting results. J.R.C. provided medical and clinical expertise. J.B.W., T.M., V.R.A., A.A., and S.D.G. wrote the article, with support and critical feedback from all authors. K.J.J., V.R.A., and A.S. devised and directed the project.

Acknowledgments

Author Disclosure Statement

T.M., A.A., A.J., M.W., L.Z., J.B.W., H.N.S., A.S., V.R.A., and S.D.G. are all full-time employees of and have stock ownership in Scipher Medicine. K.J.J. is a

former full-time employee of and has stock ownership in Scipher Medicine. The final predictive model will be proprietary to Scipher Medicine.

Funding Information

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4

References

1. Singh JA, Saag KG, Bridges SL, Jr., et al. 2015 American College of Rheumatology Guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol. 2016;68:1–26.
2. Tavakolpour S. Towards personalized medicine for patients with auto-immune diseases: opportunities and challenges. Immunol Lett. 2017; 190:130–138.
3. Quinn MA, Conaghan PG, Emery P. The therapeutic approach of early intervention for rheumatoid arthritis: what is the evidence? Rheumatology (Oxford). 2001;40:1211–1220.
4. Jin Y, Desai RJ, Liu J, et al. Factors associated with initial or subsequent choice of biologic disease-modifying antirheumatic drugs for treatment of rheumatoid arthritis. Arthritis Res Ther. 2017;19:159.
5. Curtis JR, Zhang J, Xie F, et al. Use of oral and subcutaneous metho-trexate in rheumatoid arthritis patients in the United States. Arthritis Care Res (Hoboken). 2014;66:1604–1611.
6. van de Putte LB, Atkins C, Malaise M, et al. Efficacy and safety of adali-mumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. Ann Rheum Dis. 2004;63:508–516.
7. Furst DE, Schiff MH, Fleischmann RM, et al. Adalimumab, a fully human anti tumor necrosis factor-alpha monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid ar-thritis: results of STAR (safety trial of adalimumab in rheumatoid arthri-tis). J Rheumatol. 2003;30:2563–2571.
8. Breedveld FC, Weisman MH, Kavanaugh AF, et al. The PREMIER study: a multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid ar-thritis who had not had previous methotrexate treatment. Arthritis Rheum. 2006;54:26–37.
9. Maini R, St Clair EW, Breedveld F, et al. Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. Lancet. 1999;354:1932–1939.
10. St Clair EW, van der Heijde DM, Smolen JS, et al. Combination of inflix-imab and methotrexate therapy for early rheumatoid arthritis: a ran-domized, controlled trial. Arthritis Rheum. 2004;50:3432–3443.
11. Schiff M, Keiserman M, Codding C, et al. Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, randomised, double-blind, placebo-controlled study in patients with rheumatoid ar-thritis and an inadequate response to methotrexate. Ann Rheum Dis. 2008;67:1096–1103.
12. Moreland LW, Schiff MH, Baumgartner SW, et al. Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. Ann Intern Med. 1999;130:478–486.
13. Bathon JM, Martin RW, Fleischmann RM, et al. A comparison of etaner-cept and methotrexate in patients with early rheumatoid arthritis. N Engl J Med. 2000;343:1586–1593.
14. Weinblatt M, Schiff M, Goldman A, et al. Selective costimulation mod-ulation using abatacept in patients with active rheumatoid arthritis while receiving etanercept: a randomised clinical trial. Ann Rheum Dis. 2007;66:228–234.
15. Klareskog L, van der Heijde D, de Jager JP, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. Lancet. 2004;363:675–681.
16. Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum. 2004;50: 1400–1411.
17. Aletaha D, Landewe R, Karonitsch T, et al. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. Ann Rheum Dis. 2008; 67:1360–1364.
18. Smolen JS, Landewe R, Breedveld FC, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis. 2014;73:492–509.
19. Felson DT, Anderson JJ, Boers M, et al. American College of Rheuma-tology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum. 1995;38:727–735.
20. Hobbs KF, Cohen MD. Rheumatoid arthritis disease measurement: a new old idea. Rheumatology (Oxford). 2012;51 Suppl 6:vi21–vi27.
21. Fagnani F, Pham T, Claudepierre P, et al. Modeling of the clinical and economic impact of a risk-sharing agreement supporting a treat-to-target strategy in the management of patients with rheumatoid arthritis in France. J Med Econ. 2016;19:812–821.
22. Johnson KJ, Sanchez HN, Schoenbrunner N. Defining response to TNF-inhibitors in rheumatoid arthritis: the negative impact of anti-TNF cy-cling and the need for a personalized medicine approach to identify primary non-responders. Clin Rheumatol. 2019;38:2967–2976.
23. Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League Against Rheumatism provisional defi-nition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum. 2011;63:573–586.
24. Lacroix BD, Karlsson MO, Friberg LE. Simultaneous exposure-response modeling of ACR20, ACR50, and ACR70 improvement scores in rheu-matoid arthritis patients treated with certolizumab pegol. CPT Phar-macometrics Syst Pharmacol. 2014;3:e143.
25. Pokharel G, Deardon R, Barnabe C, et al. Joint estimation of remission and response for methotrexate-based DMARD options in rheumatoid arthritis: a bivariate network meta-analysis. ACR Open Rheumatol. 2019; 1:471–479.
26. Mitchell KJ. What is complex about complex disorders? Genome Biol. 2012;13:237.
27. MacGregor AJ, Snieder H, Rigby AS, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum. 2000;43:30–37.
28. Yarwood A, Huizinga TW, Worthington J. The genetics of rheumatoid arthritis: risk and protection in different stages of the evolution of RA. Rheumatology (Oxford). 2016;55:199–209.
29. Hart GT, Ramani AK, Marcotte EM. How complete are current yeast and human protein-interaction networks? Genome Biol. 2006;7:120.
30. Menche J, Sharma A, Kitsak M, et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Sci-ence. 2015;347:1257601.
31. Ghiassian SD, Menche J, Barabasi AL. A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLoS Comput Biol. 2015;11:e1004120.
32. Kovacs IA, Luck K, Spirohn K, et al. Network-based prediction of protein interactions. Nat Commun. 2019;10:1240.
33. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12:56–68.

34. Bienkowska JR, Dalgin GS, Batliwalla F, et al. Convergent random forest predictor: methodology for predicting drug response from genome-scale data applied to anti-TNF response. Genomics. 2009;94:423–432.

35. Fleischmann RM, van der Heijde D, Gardiner PV, et al. DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. RMD Open. 2017;3:e000382.

36. Pappas DA, Kremer JM, Reed G, et al. Design characteristics of the CORRONA CERTAIN study: a comparative effectiveness study of biologic agents for rheumatoid arthritis patients. BMC Musculoskelet Disord. 2014;15:113.

37. Hawker GA, Mian S, Kendzerska T, et al. Measures of adult pain: visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken). 2011;63 Suppl 11:S240–S252.

38. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

39. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. Curr Protoc Bioinformatics. 2015;51:11.14.1–11.14.9.

40. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12: 323.

41. Espin-Perez A, Portier C, Chadeau-Hyam M, et al. Comparison of statistical methods and the use of quality control samples for batch effect correction in human transcriptome data. PLoS One. 2018;13:e0202947.

42. Köhn H-F, Hubert LJ. Hierarchical Cluster Analysis. Wiley StatsRef: Statistics Reference Online 2015. p. 1–13.

43. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–1303.

44. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–498.

45. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1–11.10.33.

46. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506:376–381.

47. Breiman L. Random forests. Mach Learn. 2001;45:5–32.

48. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2012;12:2825–2830.

49. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 2013;4:627–635.

50. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health. 2017;5:307.

51. Szumilas M. Explaining odds ratios. J Can Acad Child Adolesc Psychiatry. 2010;19:227–229.

52. Sperandei S. Understanding logistic regression analysis. Biochem Med (Zagreb). 2014;24:12–18.

53. Luck K, Kim D, Lambourne L, et al. A reference map of the human binary protein interactome. Nature. 2020;580:402–408.

54. Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks. Nat Methods. 2013;10:47–53.

55. Meyer MJ, Das J, Wang X, et al. INstruct: a database of high-quality 3D structurally resolved protein interaction networks. Bioinformatics. 2013;29:1577–1579.

56. Meyer MJ, Beltran JF, Liang S, et al. Interactome INSIDER: a structural interactome browser for genomic studies. Nat Methods. 2018;15:107–114.

57. Cowley MJ, Pinese M, Kassahn KS, et al. PINA v2.0: mining interactome modules. Nucleic Acids Res. 2012;40:D862–D865.

58. Licata L, Briganti L, Peluso D, et al. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. 2012;40:D857–D861.

59. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 2015;43: D470–D478.

60. Das J, Yu H. HINT: high-quality protein interactomes and their applications in understanding human disease. BMC Syst Biol. 2012;6:92.

61. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. Nucleic Acids Res. 2017;45(D1):D408–D414.

62. Alonso-Lopez D, Campos-Laborie FJ, Gutierrez MA, et al. APID database: redefining protein-protein interaction experimental evidences and binary interactomes. Database (Oxford). 2019;2019:baz005.

63. Li T, Wernersson R, Hansen RB, et al. A scored human protein-protein interaction network to catalyze genomic interpretation. Nat Methods. 2017;14:61–64.

64. Huttlin EL, Bruckner RJ, Paulo JA, et al. Architecture of the human interactome defines protein communities and disease networks. Nature. 2017;545:505–509.

65. Hein MY, Hubner NC, Poser I, et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell. 2015;163:712–723.

66. Wan C, Borgeson B, Phanse S, et al. Panorama of ancient metazoan macromolecular complexes. Nature. 2015;525:339–344.

67. Cheng F, Jia P, Wang Q, et al. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. Oncotarget. 2014;5: 3697–3710.

68. Hornbeck PV, Zhang B, Murray B, et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43:D512–D520.

69. Fazekas D, Koltai M, Turei D, et al. SignaLink 2—a signaling pathway resource with multi-layered regulatory networks. BMC Syst Biol. 2013;7: 7.

70. Breuer K, Foroushani AK, Laird MR, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res. 2013;41:D1228–D1233.

71. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). Nucleic Acids Res. 2017;45(D1):D896–D901.

72. Yu W, Clyne M, Khoury MJ, et al. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics. 2010;26:145–146.

73. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42:D980–D985.

74. Amberger JS, Bocchini CA, Scott AF, et al. OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res. 2019;47(D1):D1038–D1043.

75. Rappaport N, Twik M, Plaschkes I, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Res. 2017;45(D1):D877–D887.

76. Guney E, Menche J, Vidal M, et al. Network-based in silico drug efficacy screening. Nat Commun. 2016;7:10331.

77. Liberzon A. A description of the Molecular Signatures Database (MSigDB) web site. Methods Mol Biol. 2014;1150:153–160.

78. Galeazzi M, Sebastiani G, Voll R, et al. FRI0118 Dekavil (F8IL10) – update on the results of clinical trials investigating the immunocytokine in patients with rheumatoid arthritis. Annals of the Rheumatic Diseases. 2018;77:603–604.

79. Fischer PA, Rapoport RJ. Repository corticotropin injection in patients with rheumatoid arthritis resistant to biologic therapies. Open Access Rheumatol. 2018;10:13–19.

80. Jegatheeswaran J, Turk M, Pope JE. Comparison of Janus kinase inhibitors in the treatment of rheumatoid arthritis: a systemic literature review. Immunotherapy. 2019;11:737–754.

81. Zhang M, Lee F, Knize A, et al. Development of an ICOSL and BAFF bispecific inhibitor AMG 570 for systemic lupus erythematosus treatment. Clin Exp Rheumatol. 2019;37:906–914.

82. Chiu YG, Ritchlin CT. Denosumab: targeting the RANKL pathway to treat rheumatoid arthritis. Expert Opin Biol Ther. 2017;17:119–128.

83. Hendrickx R, Hegelund-Myrbäck T, Dearman M, et al. SAT0245 Azd9567: a novel oral selective glucocorticoid receptor modulator, demonstrated to have an improved therapeutic ratio compared to prednisolone in preclinical studies, is safe and well tolerated in first clinical study. Annals of the Rheumatic Diseases. 2018;77:984–985.

84. Husted S, van Giezen JJ. Ticagrelor: the first reversibly binding oral P2Y12 receptor antagonist. Cardiovasc Ther. 2009;27:259–274.

85. Caselli G, Bonazzi A, Lanza M, et al. Pharmacological characterisation of CR6086, a potent prostaglandin E2 receptor 4 antagonist, as a new

potential disease-modifying anti-rheumatic drug. Arthritis Res Ther. 2018;20:39.

86. Jarrett SJ, Conaghan PG, Sloan VS, et al. Preliminary evidence for a structural benefit of the new bisphosphonate zoledronic acid in early rheumatoid arthritis. Arthritis Rheum. 2006;54:1410–1414.

87. Kim GW, Lee NR, Pi RH, et al. IL-6 inhibitors for treatment of rheumatoid arthritis: past, present, and future. Arch Pharm Res. 2015;38:575–584.

88. Guiducci S, Del Rosso A, Cinelli M, et al. Raloxifene reduces urokinase-type plasminogen activator-dependent proliferation of synoviocytes from patients with rheumatoid arthritis. Arthritis Res Ther. 2005;7: R1244–R1253.

89. Deakin A, Duddy G, Wilson S, et al. Characterisation of a K390R ITK kinase dead transgenic mouse—implications for ITK as a therapeutic target. PLoS One. 2014;9:e107490.

90. Lv J, Wu J, He F, et al. Development of Bruton's tyrosine kinase inhibitors for rheumatoid arthritis. Curr Med Chem. 2018;25:5847–5859.

91. Elshabrawy HA, Essani AE, Szekanecz Z, et al. TLRs, future potential therapeutic targets for RA. Autoimmun Rev. 2017;16:103–113.

92. McElroy WT. Interleukin-1 receptor-associated kinase 4 (IRAK4) inhibitors: an updated patent review (2016–2018). Expert Opin Ther Pat. 2019; 29:243–259.

93. Oerlemans R, Franke NE, Assaraf YG, et al. Molecular basis of bortezomib resistance: proteasome subunit beta5 (PSMB5) gene mutation and overexpression of PSMB5 protein. Blood. 2008;112:2489–2499.

94. Lassoued S, Moyano C, Beldjerd M, et al. Bortezomib improved the joint manifestations of rheumatoid arthritis in three patients. Joint Bone Spine. 2019;86:381–382.

95. Silverman MH, Strand V, Markovits D, et al. Clinical evidence for utilization of the A3 adenosine receptor as a target to treat rheumatoid arthritis: data from a phase II clinical trial. J Rheumatol. 2008;35: 41–48.

96. Emori T, Hirose J, Ise K, et al. Constitutive activation of integrin alpha9 augments self-directed hyperplastic and proinflammatory properties of fibroblast-like synoviocytes of rheumatoid arthritis. J Immunol. 2017; 199:3427–3436.

97. Aalbers CJ, Bevaart L, Loiler S, et al. Preclinical potency and biodistribution studies of an AAV 5 vector expressing human interferon-beta (ART-I02) for local treatment of patients with rheumatoid arthritis. PLoS One. 2015;10:e0130612.

98. Tabuchi H, Katsurabara T, Mori M, et al. Pharmacokinetics, pharmacodynamics, and safety of E6011, a novel humanized antifractalkine (CX3CL1) monoclonal antibody: a randomized, double-blind, placebo-controlled single-ascending-dose study. J Clin Pharmacol. 2019;59: 688–701.

99. Tang R, She Q, Lu Y, et al. Quality control of RNA extracted from PAXgene blood RNA tubes after different storage periods. Biopreserv Biobank. 2019;17:477–482.

100. Chai V, Vassilakos A, Lee Y, et al. Optimization of the PAXgene blood RNA extraction system for gene expression analysis of clinical samples. J Clin Lab Anal. 2005;19:182–188.

101. Rao MS, Van Vleet TR, Ciurlionis R, et al. Comparison of RNA-Seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies. Front Genet. 2018;9:636.

102. Bottomly D, Walter NA, Hunter JE, et al. Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. PLoS One. 2011;6:e17820.

103. Merrick BA, Phadke DP, Auerbach SS, et al. RNA-Seq profiling reveals novel hepatic gene expression pattern in aflatoxin B1 treated rats. PLoS One. 2013;8:e61768.

104. Wang C, Gong B, Bushel PR, et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol. 2014;32:926–932.

105. Zhao S, Fung-Leung WP, Bittner A, et al. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014;9:e78644.

106. Marioni JC, Mason CE, Mane SM, et al. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18:1509–1517.

107. van Delft J, Gaj S, Lienhard M, et al. RNA-Seq provides new insights in the transcriptome responses induced by the carcinogen benzo[a]pyrene. Toxicol Sci. 2012;130:427–439.

108. Consortium EP, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447:799–816.

109. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. Nature. 2012;489:101–108.

110. Adetunji MO, Lamont SJ, Abasht B, et al. Variant analysis pipeline for accurate detection of genomic variants from transcriptome sequencing data. PLoS One. 2019;14:e0216838.

111. Alonso-Ruiz A, Pijoan JI, Ansuategui E, et al. Tumor necrosis factor alpha drugs in rheumatoid arthritis: systematic review and metaanalysis of efficacy and safety. BMC Musculoskelet Disord. 2008;9:52.

112. Li YR, Keating BJ. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. Genome Med. 2014;6:91.

113. Martin AR, Gignoux CR, Walters RK, et al. Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet. 2017;100:635–649.

114. Davignon JL, Rauwel B, Degboe Y, et al. Modulation of T-cell responses by anti-tumor necrosis factor treatments in rheumatoid arthritis: a review. Arthritis Res Ther. 2018;20:229.

115. Bystrom J, Clanchy FI, Taher TE, et al. TNFalpha in the regulation of Treg and Th17 cells in rheumatoid arthritis and other autoimmune inflammatory diseases. Cytokine. 2018;101:4–13.

116. Cope AP, Schulze-Koops H, Aringer M. The central role of T cells in rheumatoid arthritis. Clin Exp Rheumatol. 2007;25(5 Suppl 46):S4–S11.

117. Cope AP. T cells in rheumatoid arthritis. Arthritis Res Ther. 2008;10 Suppl 1:S1.

118. Julia A, Erra A, Palacio C, et al. An eight-gene blood expression profile predicts the response to infliximab in rheumatoid arthritis. PLoS One. 2009;4:e7556.

119. Lequerre T, Gauthier-Jauneau AC, Bansard C, et al. Gene profiling in white blood cells predicts infliximab responsiveness in rheumatoid arthritis. Arthritis Res Ther. 2006;8:R105.

120. Sekiguchi N, Kawauchi S, Furuya T, et al. Messenger ribonucleic acid expression profile in peripheral blood cells from RA patients following treatment with an anti-TNF-alpha monoclonal antibody, infliximab. Rheumatology (Oxford). 2008;47:780–788.

121. Stuhlmuller B, Haupl T, Hernandez MM, et al. CD11c as a transcriptional biomarker to predict response to anti-TNF monotherapy with adalimumab in patients with rheumatoid arthritis. Clin Pharmacol Ther. 2010;87:311–321.

122. van Baarsen LG, Wijbrandts CA, Rustenburg F, et al. Regulation of IFN response gene activity during infliximab treatment in rheumatoid arthritis is associated with clinical response to treatment. Arthritis Res Ther. 2010;12:R11.

123. Koczan D, Drynda S, Hecker M, et al. Molecular discrimination of responders and nonresponders to anti-TNF alpha therapy in rheumatoid arthritis by etanercept. Arthritis Res Ther. 2008;10:R50.

124. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. Genome Med. 2014;6:85.

125. Hsiao YH, Bahn JH, Lin X, et al. Alternative splicing modulated by genetic variants demonstrates accelerated evolution regulated by highly conserved proteins. Genome Res. 2016;26:440–450.

126. Stefl S, Nishi H, Petukh M, et al. Molecular mechanisms of disease-causing missense mutations. J Mol Biol. 2013;425:3919–3936.

127. Ottman R. Gene-environment interaction: definitions and study designs. Prev Med. 1996;25:764–770.

128. Wang H, Zhang F, Zeng J, et al. Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. Sci Adv. 2019;5:eaaw3538.

129. Wang D, Garcia-Bassets I, Benner C, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature. 2011;474:390–394.

130. Preker P, Nielsen J, Kammler S, et al. RNA exosome depletion reveals transcription upstream of active human promoters. Science. 2008;322: 1851–1854.

131. Koch CM, Chiu SF, Akbarpour M, et al. A beginner's guide to analysis of RNA sequencing data. Am J Respir Cell Mol Biol. 2018;59:145–157.

132. McLachlan GJ, Do K-A, Ambroise C. *Analyzing Microarray Gene Expression Data*. New York, NY: John Wiley & Sons, Inc., 2004.

133. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. 1987;30:1205–1213.
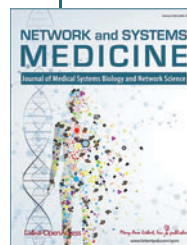
134. Okada Y, Kim K, Han B, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum Mol Genet. 2014;23:6916–6926.

135. Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012;44:291–296.

136. Okada Y, Suzuki A, Ikari K, et al. Contribution of a non-classical HLA gene, HLA-DOA, to the risk of rheumatoid arthritis. Am J Hum Genet. 2016;99:366–374.

137. Panayi GS, Lanchbury JS, Kingsley GH. The importance of the T cell in initiating and maintaining the chronic synovitis of rheumatoid arthritis. Arthritis Rheum. 1992;35:729–735.

138. Cope AP. Studies of T-cell activation in chronic inflammation. Arthritis Res. 2002;4 Suppl 3:S197–S211.

139. Raza K, Scheel-Toellner D, Lee CY, et al. Synovial fluid leukocyte apoptosis is inhibited in patients with very early rheumatoid arthritis. Arthritis Res Ther. 2006;8:R120.

140. Buckley CD. Michael Mason prize essay 2003. Why do leucocytes accumulate within chronically inflamed joints? Rheumatology (Oxford). 2003;42:1433–1444.

141. Morita Y, Yamamura M, Kawashima M, et al. Flow cytometric single-cell analysis of cytokine production by CD4+ T cells in synovial tissue and peripheral blood from patients with rheumatoid arthritis. Arthritis Rheum. 1998;41:1669–1676.

142. Kusaba M, Honda J, Fukuda T, et al. Analysis of type 1 and type 2 T cells in synovial fluid and peripheral blood of patients with rheumatoid arthritis. J Rheumatol. 1998;25:1466–1471.

143. Yamin R, Berhani O, Peleg H, et al. High percentages and activity of synovial fluid NK cells present in patients with advanced stage active rheumatoid arthritis. Sci Rep. 2019;9:1351.

144. Simon AK, Seipelt E, Sieper J. Divergent T-cell cytokine patterns in inflammatory arthritis. Proc Natl Acad Sci U S A. 1994;91:8562–8566.

145. Leipe J, Grunke M, Dechant C, et al. Role of Th17 cells in human autoimmune arthritis. Arthritis Rheum. 2010;62:2876–2885.

146. van Amelsfort JM, Jacobs KM, Bijlsma JW, et al. CD4(+)CD25(+) regulatory T cells in rheumatoid arthritis: differences in the presence, phenotype, and function between peripheral blood and synovial fluid. Arthritis Rheum. 2004;50:2775–2785.

147. Charles-Schoeman C, Burmester G, Nash P, et al. Efficacy and safety of tofacitinib following inadequate response to conventional synthetic or biological disease-modifying antirheumatic drugs. Ann Rheum Dis. 2016;75:1293–1301.

148. Bykerk VP, Ostor AJ, Alvaro-Gracia J, et al. Tocilizumab in patients with active rheumatoid arthritis and inadequate responses to DMARDs and/or TNF inhibitors: a large, open-label study close to clinical practice. Ann Rheum Dis. 2012;71:1950–1954.

149. Fleischmann R, van Adelsberg J, Lin Y, et al. Sarilumab and nonbiologic disease-modifying antirheumatic drugs in patients with active rheumatoid arthritis and inadequate response or intolerance to tumor necrosis factor inhibitors. Arthritis Rheumatol. 2017;69:277–290.

150. Genovese MC, Schiff M, Luggen M, et al. Efficacy and safety of the selective co-stimulation modulator abatacept following 2 years of treatment in patients with rheumatoid arthritis and an inadequate response to anti-tumour necrosis factor therapy. Ann Rheum Dis. 2008;67:547–554.

151. Mandal R, Chan TA. Personalized oncology meets immunology: the path toward precision immunotherapy. Cancer Discov. 2016;6:703–713.

152. Bode AM, Dong Z. Recent advances in precision oncology research. NPJ Precis Oncol. 2018;2:11.

Abbreviations Used

ACR = American College of Rheumatology
anti-CCP = anti-cyclic citrullinated protein
BMI = body mass index
CDAI = Clinical Disease Activity Index
CI = confidence interval
CORRONA = Consortium of Rheumatology Researchers of North America
CRP = C-reactive protein
csDMARDs = conventional synthetic disease modifying antirheumatic drugs
CV = cross-validation
DG = discriminatory genes
eQTL = expression quantitative trait loci
HI = human interactome
OR = odds ratio
PPI = protein/protein interactions
PPV = positive predictive value
RA = rheumatoid arthritis
RNAseq = RNA sequencing
SNPs = single-nucleotide polymorphisms
TNF = tumor necrosis factor