Early Prediction of Clinical Response to Anti-TNF Treatment using Multi-omics and Machine Learning in Rheumatoid Arthritis

Niyaz Yoosuf (✉ niyaz.yoosuf@ki.se)

Karolinska Institutet, Department of Medicine, Solna <https://orcid.org/0000-0002-3792-5766>

Mateusz Maciejewski

Pfizer Inc

Daniel Ziemek

Pfizer Inc

Scott A. Jelinsky

Pfizer Inc

Lasse Folkersen

Danish National Genome Center, Copenhagen, Denmark

Malin Müller

Karolinska Institutet, Department of Medicine, Solna

Peter Sahlström

Karolinska Institutet, Department of Medicine, Solna

Nancy Vivar

Karolinska Institutet, Department of Medicine, Solna

Anca Catrina

Karolinska Institutet, Department of Medicine, Solna

Louise Berg

Karolinska Institutet, Department of Medicine, Solna

Lars Klareskog

Karolinska Institutet, Department of Medicine, Solna

Leonid Padyukov

Karolinska Institutet, Department of Medicine, Solna

Boel Brynedal

Karolinska Institutet, Department of Medicine, Solna

Research Article

Keywords: Rheumatoid arthritis, anti-TNF, Methotrexate, treatment response, inflammation, peripheral blood mononuclear cells

Posted Date: March 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-60573/v2>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Rheumatology on June 27th, 2021. See the published version at <https://doi.org/10.1093/rheumatology/keab521>.

Abstract

Objectives

Advances in immunotherapy by blocking TNF have remarkably improved treatment outcomes for rheumatoid arthritis patients. Although treatment specifically targets TNF, the downstream mechanisms of immune suppression are not completely understood. The present study was aimed to detect biomarkers and expression signatures of treatment response to TNF inhibition.

Methods

Peripheral blood mononuclear cells were obtained from 39 patients collected before anti-TNF treatment initiation (day 0) and after three months. Response to treatment was defined based on the EULAR criteria and classified 23 patients as responders and 16 as non-responders. We investigated differences in gene expression in peripheral blood mononuclear cells, the proportion of cell types and cell phenotypes in peripheral blood using flow cytometry, and the level of proteins in plasma. Finally, using biological measurements, we run machine learning models to predict non-response.

Results

The gene expression analysis in baseline samples revealed notably a higher expression of the gene EPPK1 in future responders. We also detected the suppression of genes and proteins following treatment, including suppression of expression of the gene, T-cell inhibitor CHI3L1, and its protein YKL-40. The gene expression results were replicated in an independent cohort. Finally, machine learning models mainly based on transcriptomics data showed high predictive utility (ROC AUC $\pm$ SEM: 0.81 ± 0.17) in classifying non-response to anti-TNF treatment in RA.

Conclusions

Our integrative multi-omics analyses identified new biomarkers for prediction of response, found pathways influenced by treatment and suggested new predictive models of anti-TNF treatment in RA patients.

Rheumatology Key Messages

- Expression of the CHI3L1 gene and its protein YKL-40 that regulates T cell activation was suppressed in responders following anti-TNF treatment.
- Machine learning models using transcriptomics data at baseline showed high predictive utility in classifying response to future anti-TNF treatment.

Background

Rheumatoid arthritis (RA) is a systemic autoimmune disease characterized by chronic inflammation in symmetric joints that leads to pain and eventually bone destruction. RA is one of the most common autoimmune diseases which affects approximately 0.5 -1% of the world's population(1). Currently, there is no cure for RA, but several disease-modifying anti-rheumatic drugs (DMARDs) are used to treat patients with the disease. During a successful treatment course, inflammation in the joints decreases, resulting in disease remission or low disease activity (2,3). Methotrexate (MTX) is recommended as the first-line treatment of early RA, however, at least 30 % of patients do not respond to the treatment where significant disease activity remains (4,5). The patients who do not respond to first-line treatment are recommended for additional treatment, in most cases with drugs that inhibit tumor necrosis factor (TNFi). TNF is a pro-inflammatory cytokine secreted mainly by monocytes and macrophages, but also by other immune and non-immune cells, including fibroblasts and endothelial cells, involved in systemic inflammation.

The anti-TNF therapy has been used to treat RA for two decades, but one-third of treated patients do not respond or have poor responses(6). Hence the prediction of the efficacy of treatments before initiating treatment would help patients to start effective treatment and decrease the number of adverse effects.

In the present study, we selected a well-characterized cohort that allowed us to investigate gene expression differences in peripheral blood mononuclear cells (PBMCs), the proportion of different cell types and cell phenotypes in peripheral blood measured using flow cytometry and the level of several proteins in serum. In addition, we developed models using machine learning techniques to predict non-response for the anti-TNF treatment.

Materials And Methods

Patient Cohort

In this study, patient samples were obtained from the COMBINE cohort that includes individuals treated with methotrexate (89 patients), treated with anti-TNF drugs (59 patients), or treated with a second biologic agent (31 patients) and healthy controls 60 individuals (7). Healthy controls were recruited from the Swedish Blood Centre service in Uppsala closely matched with age and sex as closely matched with patient groups. The patients who did not respond adequately to MTX treatment (based on local rheumatologist's judgment) were prescribed anti-TNF treatment in combination with MTX. Patients included in this study donated blood at the clinic before starting anti-TNF therapy and at the follow-up visit after approximately three months. Clinical assessments and routine blood sampling were made at both visits, and were used to calculate the disease activity score DAS28 CRP (8). Of the 59 patients who started anti-TNF therapy, three patients dropped out before the scheduled three-month visit and therefore lacked clinical assessment. For RNA-seq analysis, we did not include patient samples with low quality or low quantity RNA samples. Due to the low number of male non-responders in our study cohort we, therefore decided to include only female patients (n=39) in our study (Supplementary Fig. S1). The clinical and demographic variables at baseline are shown in Table 1.

	RA (n=39)
Age, median (range)	57 (19-76)
Swedish (%)	34 (82.9)
Current smoker (%)	11 (28.2%)
HLA-DR shared epitope positive (%)	26 (66.6%)
ACPA positive (%)	29 (74.3%)
Bone erosions (%)	18 (46.1%)
DAS28, median (range)	4.79 (2.49-7.48)
	6 (1-25)
28-joint swollen joint count, median (range)	8 (1-28)
28-joint tender joint count, median (range)	23 (58.9%)
Prednisolone treatment (%)	
Anti-TNF Drugs	16
Remicade	8
Enbrel	11
Humira	2
Simponi	2
Cimzia	2 (0.5-59)
CRP, median (range)	50 (5-100)
Patient global health assessment, median (range)	0.75 (0-2.6)
Physical function (HAQ), median (range)	45 (11-82)
Health professional global health assessment, median (range)	

Table 1: Baseline demographic characteristics of female rheumatoid arthritis patients treated with anti-TNF. HAQ = Health Assessment questionnaire disability index; ACPA = anti-citrullinated peptide antibody; DAS28 = Disease Activity Score in 28 joints; CRP = C-reactive protein.

Response measures

We used the European League Against Rheumatism (EULAR) response criteria to classify patient response to treatment (9,10). In our analysis, we considered Good and Moderate EULAR responders as “responders”, and compared these to the EULAR “non-responders”.

RNA sequencing

The RNA was extracted from PBMCs, freshly isolated using CPT tubes (BD Biosciences), and sequenced as previously described (7). Of the 39 female RA patients, we obtained high-quality RNA-seq data from 28 patients at baseline (responders = 10; moderate-responders = 9; non-responders = 9) and 32 patients after treatment (responders = 11; moderate-responders = 9; non-responders = 12) that leads to a total of 25 patients with paired RNA-seq samples (both anti-TNF naïve and treated) (Supplementary Fig. S1). The sequencing reads were trimmed using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and then mapped to the GRCh38 human reference genome and subsequently gene counts were generated using Star v.2.5.3a (11).

Results

Gene expression signatures of response to anti-TNF treatment at baseline

The gene expression analysis identified 59 differentially expressed genes between future responders and non-responders, but most of these genes achieved significance ($FDR \leq 0.1$) due to a single outlier sample showing different gene expression profile (Supplementary Fig. S2A-B). Therefore, we decided to exclude this patient sample (both anti-TNF naïve and treated) and a subsequent analysis (without the outlier sample) yielded 192 differentially expressed genes ($FDR \leq 0.1$), including 103 genes with higher expression, and 89 genes with lower expression, in future responders. The top differentially expressed genes are represented in Fig. 1A (also see Table 2 and Supplementary Table. S1A).

To assess possible heterogeneity and detect consistent differential gene expression, we employed a leave-one-out (LOO) approach, where one sample is removed in each iteration and the association analysis was repeated. The genes are considered if they meets statistical significance ($FDR < 0.1$) in each iteration. The LOO approach yielded two genes, Epiplakin (EPPK1) and FosB Proto-Oncogene, AP-1 Transcription Factor Subunit (FOSB) with higher expression in responders in all 28 iterations (Table 2). An additional five genes, EGR1, EGR2, BCL6-AS1, IGLV10-54, and IGKV1D-39 showed significance ($FDR < 0.1$) in 27 LOO iterations. The genes EGR1, EGR2, and BCL6-AS1 were expressed higher in future responders, whereas immunoglobulin light chain genes IGLV10-54 and IKV1D-39 had lower expression in future responders. The genes EPPK1, BCL6-AS1 and CDC20 showed a clear expression difference between responders and non-responders (Fig. 1B). We also noticed 6 immunoglobulin light chain genes (IGLV10-54, IGKV1D-39, IGKV3-20, IGLV3-1, IGKV1-17, IGKV2-24) and 1 heavy chain gene (IGHV5-10-1) expressed lower in responders compared to non-responders, however these genes did not pass the LOO analysis criteria (Fig. 1A).

Genes	Description	log2FoldChange	p-value	Iteration count
FOSB	FosB Proto-Oncogene, AP-1 Transcription Factor Subunit	3.88	6.25E-09	28
EPPK1	Epiplakin 1	1.89	6.06E-07	28
EGR2	Early Growth Response 2	3.98	1.63E-07	27
BCL6-AS1	BCL6 Antisense 1	2.14	2.95E-07	27
EGR1	Early Growth Response 1	3.68	5.91E-06	27
IGLV10-54	Immunoglobulin Lambda Variable 10-54	-2.73	5.46E-06	27
IGKV1D-39	Immunoglobulin Kappa Variable 1D-39	-2.68	2.13E-05	27
PDIA4	Protein Disulfide Isomerase Family A Member 4	-0.45	5.59E-06	26
HSP90B1	Heat shock protein 90kDa beta member 1	-0.60	1.17E-05	26
FAM46C	family with sequence similarity 46, member C	-1.03	1.42E-05	26
KDM6B	Lysine Demethylase 6B	0.61	1.92E-05	26
FBXO7	F-Box Protein 7	-0.37	2.57E-05	26
PSAT1	Phosphoserine Aminotransferase 1	-0.78	2.49E-05	26
CDC20	Cell Division Cycle 20	-1.95	8.70E-06	25
NDC80	NDC80 Kinetochore Complex Component	-0.80	1.46E-05	25
CHEK1	Checkpoint Kinase 1	-0.80	1.91E-05	25
ITM2C	Integral Membrane Protein 2C	-0.89	4.86E-05	25
SOGA1	Suppressor Of Glucose, Autophagy associated 1	0.66	5.61E-05	25
TXNDC15	Thioredoxin Domain Containing 15	-0.42	6.42E-05	25
IGLV3-1	Immunoglobulin Lambda Variable 3-1	-1.80	8.19E-05	25
MTCO2P12	MT-CO2 Pseudogene 12	2.76	7.17E-04	25

Table 2: Differentially expressed genes in future responders and non-responders before anti-TNF treatment. The iteration count is the number of leave-one-out iterations where the gene remained significant.

The gene set enrichment analysis based on ranking of differentially expressed genes (log2fold change) identified a total of 127 regulated pathways (Supplementary Table. S1B) between response groups. The pathway analysis reveal that responders were preferentially characterized by higher expression of genes involved in graft versus host disease, antigen processing and presentation, and AP-1 transcription factor network, whereas non-responders were characterized by higher expression of genes involved in cell cycle pathways, mainly - cell cycle mitotic activity, cell cycle checkpoints. The top 15 most upregulated and downregulated pathways are represented in Supplementary Fig. S3A; Supplementary Table. S1B).

Effect on gene expression in PBMCs during anti-TNF treatment

Since biological and technical confounders between individuals may significantly affect downstream analyses, using paired samples before and after treatment is the most preferable approach to address changes related to the treatment. We analyzed treatment effects on gene expression using 25 paired RA

patients without considering the response. The analysis identified 25 differentially expressed genes, of which 14 genes were suppressed whereas 11 genes showed a slight increase in expression (Supplementary Table. S1C). The genes BHLHE40 that controls cytokine production by T cells and CHI3L1, Chitinase-3-like protein were suppressed during treatment whereas the B cell novel protein 1 (alias FAM129C) and TTC21A were induced by treatment (Fig. 2A). The pathway analysis of differentially expressed genes did not show any enrichment of gene sets, but when using a less conservative threshold of $FDR < 0.1$, we detected 114 gene sets regulated during the treatment (Supplementary Table. S1D). Induced genes are significantly enriched for genes involved in the RNA and protein metabolism, and interferon signaling, whereas suppressed genes are predominantly enriched for genes involved in hemostasis and signaling by GPCR (Fig. 2B).

Distinct gene expression signatures for response during anti-TNF treatment

The trajectory of gene expression changes may correlate with measured clinical outcomes. Therefore, we investigated the transcriptional changes in paired samples of responders ($n=17$) and non-responders ($n=8$) to anti-TNF treatment separately. Our analysis identified five genes regulated in responders, whereas no significant regulation was identified in non-responders. Of the five regulated genes in responders, C-X-C Motif Chemokine Receptor 2 (CXCR2), Myeloperoxidase (MPO), Myeloid Associated Differentiation Marker (MYADM), TNF Alpha Induced Protein 6 (TNFAIP6) were suppressed by treatment, whereas gene Low-Affinity Immunoglobulin Gamma Fc Region Receptor II-B (FCGR2B) was induced during anti-TNF treatment. The gene expression plots for all five genes using normalized counts show a clear difference in responders before and after anti-TNF treatment (Fig. 3A). Interestingly, for these genes we observed a similar trend of regulation in non-responders (same directionality), but it does not reach statistical significance (Supplementary Fig. S4). The gene set enrichment analysis of differentially expressed genes ranked based on foldchange identified 78 regulated pathways in responders (Supplementary Table. S1E). We noticed induction of pathways involved in the regulation of cell cycle mitotic and protein metabolism, whereas genes involved in the extracellular matrix organization, signaling by GPCR and Toll-like receptor cascades were downregulated in responders (Fig. 3B; Supplementary Table. S1E).

Replication of gene expression associations at baseline

To validate our study, we sought to replicate our findings in an independent cohort (BiOCURA) of gene expression on PBMCs with 80 RA patients before they began adalimumab or etanercept anti-TNF treatment (12). From the replication cohort, we selected only female patients gene expression data for the analysis. At baseline, we replicate the gene CDC20 that showed lower expression in responders compared to non-responders ($p\text{-value} < 0.087$). None of the remaining 20 genes (Table 2) achieved statistical significance with $p\text{-value} < 0.1$. In the study cohort, we reported the suppression of the gene, CHI3L1 upon anti-TNF treatment in responders, we observed similar regulation in replication cohort with lower expression in responders compared to non-responders ($p\text{-value} < 0.096$) at baseline.

Changes in cell phenotypes and protein blood plasma during anti-TNF treatment

We studied the changes in 422 immune phenotypes measured by flow cytometry during anti-treatment using paired samples. When analyzing the effect of treatment in responders and non-responders, we observed differences in seven cell phenotypes in responders and no significant differences in non-responders (Supplementary Table. S1F). With treatment, responders showed a strong suppression of the proportion of granulocytes among leukocytes (defined as CD45+ cells), and the overall concentration of neutrophils in whole blood. The proportion of T cells (defined as CD3+ cells) and B cells (CD3-CD19+) among leukocytes and the proportion of NKG2A+ NKp44+ NK cells out of all NKp44+ NK cells were instead up-regulated (Fig. 3C).

The longitudinal analysis using paired samples identified regulation of 12 proteins in plasma in responders and one protein in non-responders. The proteins were predominantly downregulated by anti-TNF treatment in responders, including CRP, IL-6, MMP-1, MMP-3, SAA, TNF-RI, VEGF, YKL-40, MIG, and MIP-1 beta, whereas TNFR2 is upregulated (Fig. 3D). Moreover, we identified the downregulation of protein matrix metalloproteinase-3 (MMP-3) measured using two different methods (Supplementary Data S1), whereas protein adiponectin is induced during anti-TNF treatment in non-responders, while showing a trend for induction in responders (FDR: 0.11). The list of proteins that were regulated during anti-TNF treatment is shown in Supplementary Table. S1G.

Further, our association analysis of immune phenotypes and plasma protein levels to clinically defined response to anti-TNF treatment did not show any significant association between responders and non-responders before treatment.

Classifier performance for anti-TNF response data

We investigated the utility of machine learning methods to predict anti-TNF response using clinical data, flow cytometry measurements, protein measurements, and transcriptomic data. At baseline, the linear model predicted response with high accuracy for transcriptomics data (ROC AUC $\pm$ SEM: 0.81 ± 0.17) (Fig. 4) whereas the kernel based method predicted response with highest accuracy for clinical data (ROC AUC: 0.73 ± 0.17), for flow cytometry (ROC AUC: 0.72 ± 0.18), and for protein data (ROC AUC: 0.72 ± 0.15) (Fig. 4). We further studied the classifier performance for the anti-TNF treated data and observed limited classifier utility for models based on flow cytometry, protein as well as transcriptomics data. In contrast, as expected, the linear model based on clinical data showed high classifier performance with ROC AUC: 0.85 ± 0.15 . For the FACS data, we found maximum classifier utility of ROC AUC: 0.68 ± 0.17 while using linear model, whereas the non-linear models based on proteins and transcriptomics data showed maximum classifier utility with ROC AUC: 0.73 ± 0.15 and ROC AUC: 0.72 ± 0.18 respectively.

Discussion

This study represents a comprehensive analysis of transcriptomics, proteomics, and flow cytometry data analysis of female RA patients treated with anti-TNF. Our analysis was directed towards finding transcriptional and translational regulations during anti-TNF treatment and predict response using blood

samples collected at two-time points. We identified genes that are differentially regulated between responders and non-responders at baseline and reported changes induced by treatment in gene expression, in protein measurements, and in cell phenotypes. Finally, using biological measurements, we employed machine learning models to predict non-response to anti-TNF treatment. Our integration of multi-omics data revealed the gene CHI3L1 and subsequent protein product YKL-40 expressions were suppressed upon anti-TNF treatment. We also observed an increase in the proportion of B cells, T cells, and NK cells in responders during treatment, whereas the proportion of granulocytes was strongly suppressed.

The identification of potential biomarkers for RA with prognostic value for response to a given therapy is challenging as RA is a heterogeneous disease by its clinical characteristics and pathological processes. This also may be due to the high clinical heterogeneity in RA patient samples and strong confounding (covariates) effects, for example, variation in cell subsets is one of the strong covariates of gene expression (13). Also, RNA sequencing experiments can often generate outlier read counts in one or few RNA samples and this considerably limits the power of differential testing. Therefore we performed extensive transcriptomics analysis using leave-one-out approach to find genes with stable association to response.

The gene expression results at baseline highlighted cell adhesion gene Epiplakin1, EPPK1 showed strong association with response and was expressed higher in future responders compared to non-responders. Importantly, the difference we observed in the gene, EPPK1(log2 fold change:1.95, FDR:5.5E-03) showed similar regulation following the adjustment to CRP measurements in the gene expression model suggesting that EPPK1 is not a surrogate marker for inflammation in patients. Previous studies have reported the mechanism of infiltration of inflammatory cells into the synovial lining due to deregulation of several adhesion receptors, therefore association of response to corresponding genes suggested its role in the pathogenesis of rheumatoid arthritis and hemophilic arthropathy (14,15). When findings from the discovery analysis were investigated in an independent cohort, we replicate the lower expression of the gene cell division cycle 20 gene (CDC20) in future responders compared to non-responders (p-value, $p = 0.087$). Interestingly, non-responders at baseline showed an increase in cell cycle pathways mainly related to mitotic activity and cell cycle checkpoints. This may partly correspond to the hyperproliferation of cells that leads to the accumulation of pro-inflammatory cytokines in the RA inflamed joints (16,17). Next, we extended our transcriptomics analysis using paired patient's samples to study the changes induced by treatment with two different approaches: i) changes in gene expression patterns in samples from all RA patients not considering response status; ii) treatment effects in responders and non-responders separately to understand differences in gene expression regulation. Our first approach identified twenty-five genes differentially expressed in RA after treatment, whereas in the second approach yielded five genes differentially expressed in responders and no significant changes in non-responders. Interestingly, both approaches found four common genes CXCR2, MPO, and MYADM that were downregulated and gene, FCGR2B that was upregulated upon treatment. Also, we observed TNFAIP6 gene (TNF Alpha Induced Protein 6, also called tumor necrosis factor-stimulated gene 6, TSG-6) was suppressed in responders upon treatment. Previous studies have shown that the expression of TSG-6 has a strong

correlation with disease severity and is a potential biomarker of inflammation (18–20). The gene TSG-6 plays a key role in the remodeling of the extracellular matrix, regulation of leukocyte migration, and stimulation of cell proliferation during inflammation (21,22). These findings correspond well with the pathway analysis that showed downregulation of genes involved in extracellular matrix organization and signaling by interleukins, and an upregulation of genes mitotic cell cycle regulation. These results substantially extend our understanding of the transcriptomic profile of responders to anti-TNF treatment.

Our longitudinal studies on cell phenotypes revealed significant changes in seven cell populations in responders, indicating a marked change in major cell type proportions during treatment (Fig. 3C). The proportions of B cells and T cells among leukocytes were increased and the proportions of granulocytes were suppressed, upon treatment in responders. There was also a decrease in the overall concentration of neutrophils in whole blood in responders as well as in non-responders. The reduction of neutrophils seen by flow cytometry was also associated with a significant reduction in peripheral blood neutrophil count seen by routine blood analysis, leading to 19% of patients becoming neutropenic by clinical judgement. This finding corroborates a recent study that showed that patients treated with anti-TNF in combination with methotrexate decreased the blood neutrophil counts regardless of their clinical response to therapy (23,24). Additionally, our results also corroborate with earlier studies showing induction of B cells and NK cells in whole blood following anti-TNF treatment (25).

Similarly, the longitudinal analysis of protein data showed changes in plasma protein levels in responders. The expression of the protein YKL-40 was suppressed upon treatment (beta: -0.26) and the gene encoding for YKL-40, *CHI3L1*, was also down-regulated (Fig. 2A, Fig. 3D). The model including only responders showed similar suppression of *CHI3L1* (log2 fold change: -1.07) with a p-value <1.73E-04 and FDR <0.21. Interestingly, we also observed that the gene *CHI3L1* was expressed lower in responders compared to non-responders (p-value<0.096) at baseline in the replication cohort. The *CHI3L1* gene may negatively regulate T cell activation and may also inhibit Th1 differentiation via the IFN γ -STAT1 pathway (26). We also showed that treatment with TNF inhibitors leads to a larger proportion of T cells among responders (beta: 8.16).

Interestingly, two proteins, soluble TNFR2, and adiponectin showed induced plasma levels in responders and non-responders during treatment. The result of our study replicates previously shown similar induction of adiponectin after treatment with anti-IL-6R, Tocilizumab and also highlights its multifaceted role in RA with both pro-inflammatory and anti-inflammatory functions (27,28). Similarly, an increase in TNFR2 signaling leads to the activation and proliferation of Tregs and promotes tissue regeneration in RA patients (29–31).

Our machine learning models showed high predictive utility in classifying non-response prior to anti-TNF treatment. The linear model based on transcriptomics data at baseline found suggestive good predictability using the presently applied gene expression classifier. To our surprise, the models with transcriptomics data predict response with higher accuracy compared to models with clinical data. Interestingly, we also found clinical variables add less further utility to the transcriptomics based

predictive models at baseline (data not shown). Previously, a similar outcome was observed in a study where models using transcriptomics data alone predicted fibrous cap thickness response to statin treatment with better accuracy compared to the models with clinical data or transcriptomics plus clinical variables (32). Our prediction model should be used with caution and these findings need to be validated with a larger patient independent cohort, before its potential use in clinical settings.

There are some limitations to this study. This study addressed multi-omics profiling of female RA patients treated with anti-TNF and with relatively few patients included. Also, we observed genes that are regulated upon treatment in responders showed similar direction of regulation in non-responders with low significance. This might be partly explained due to the low number of non-responders included in the current study. Extended future studies with a larger patient cohort with similar frequency of responders and non-responders are thus warranted to validate these results.

In summary, the integrative multi-omics study expanded our growing knowledge of biology of anti-TNF treatment in RA patients by finding a number of changes in gene expression, protein and cell phenotypes during treatment in anti-TNF responders. We reported genes that showed distinct expression in responders and non-responders before treatment initiation. Our analyses demonstrate that treatment causes a major regulation of cell subsets, which are also mirrored in the protein analysis. Also, our study highlights the machine learning predictive utility of transcriptomics data at baseline in stratifying patients and/or predicted response before anti-TNF treatment initiation. We envision that in the future machine learning models based on multi-omics data could predict response to anti-TNF treatment and support rheumatologist's decision towards personalized treatment.

Abbreviations

ACPA: Anti-citrullinated peptide

AUC: Area under curve

CRP: C-reactive protein

DAS28: Disease activity score in 28 joints

DMARDs: Disease modifying anti-rheumatic drugs

EULAR: European League Against Rheumatism

FDR: False discovery rate

GSEA: Gene set enrichment analysis

HAQ: Health Assessment questionnaire disability index

LOO: Leave one out approach

MTX: Methotrexate

NK cells: Natural Killer cells

PBMCs: Peripheral blood mononuclear cells

RA: Rheumatoid arthritis

ROC: Receiver operating characteristic

SEM: Standard error of the mean

Th1: T helper1

TNF: Tumor necrosis factor

Tregs: Regulatory T cells

Declarations

Acknowledgements

We wish to thank all the participating patients and healthy controls. We acknowledge support and resources provided by SNIC through Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) and NGI Sweden.

Funding

The parts of the study were supported by Novo Nordisk A/S and Pfizer Inc. LP was supported by the Swedish Council grant.

Conflicts of interest

The authors have declared no conflicts of interest

Author information

Affiliations

Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden

Niyaz Yoosuf, Malin Müller, Peter Sahlström, Nancy Vivar, Anca Catrina, Louise Berg, Lars Klareskog, Leonid Padyukov

Translational Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Niyaz Yoosuf, Boel Brynedal

Pfizer, Cambridge, Massachusetts, USA

Mateusz Maciejewski, Daniel Ziemek, Scott Jelinsky

Danish National Genome Center, Copenhagen, Denmark

Lasse Folkersen

Authors contributions

NY, BB, AC, LB, LP and LK conceived the study. MMU, PS and NV performed wet lab experiments. NY, BB, AC, LF and LP contributed to the data preparation. NY, BB and MM performed the data analysis. NY, BB, MM, DZ, LP and LK interpreted the data. NY prepared figures. NY, BB, MM, SJ LB, LP, LK wrote the manuscript. All authors read and approved the final version of the manuscript.

Corresponding author

Correspondence to Niyaz Yoosuf.

Ethics declarations

Ethics approval and consent to participate

The COMBINE biobank was generated after written informed consent from all participants had been obtained according to the declaration of Helsinki and with approval by the Stockholm (number 2010-351-31-2) and Uppsala (2009-013) Regional Ethics Committees.

References

1. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet*. 2009 Feb 21;373(9664):659–72.
2. Schett G, Emery P, Tanaka Y, Burmester G, Pisetsky DS, Naredo E, et al. Tapering biologic and conventional DMARD therapy in rheumatoid arthritis: current evidence and future directions. *Ann Rheum Dis*. 2016 Jun 3;75(8):1428–37.
3. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis*. 2016 Jan;75(1):3–15.
4. Smolen JS, Landewé R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. *Ann Rheum Dis*. 2014 Mar;73(3):492–509.

5. Singh JA, Furst DE, Bharat A, Curtis JR, Kavanaugh AF, Kremer JM, et al. 2012 update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2012 May;64(5):625–39.
6. Mewar D, Wilson AG. Treatment of rheumatoid arthritis with tumour necrosis factor inhibitors. *Br J Pharmacol*. 2011 Feb;162(4):785–91.
7. Folkersen L, Brynedal B, Diaz-Gallo LM, Ramsköld D, Shchetynsky K, Westerlind H, et al. Integration of known DNA, RNA and protein biomarkers provides prediction of anti-TNF response in rheumatoid arthritis: results from the COMBINE study. *Mol Med*. 2016 Sep;22:322–8.
8. Fransen J, van Riel PLCM. The Disease Activity Score and the EULAR response criteria. *Rheum Dis Clin North Am*. 2009 Nov;35(4):745–57, vii.
9. van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis*. 1990 Nov;49(11):916–20.
10. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum*. 1995 Jan;38(1):44–8.
11. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15–21.
12. Tao W, Concepcion AN, Vianen M, Marijnissen ACA, Lafeber FPGJ, Radstake TRDJ, et al. Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. *Arthritis Rheumatol*. 2020 Sep 10;
13. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, et al. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA*. 2003 Feb 18;100(4):1896–901.
14. Burkhardt J, Blume M, Petit-Teixeira E, Hugo Teixeira V, Steiner A, Quente E, et al. Cellular adhesion gene SELP is associated with rheumatoid arthritis and displays differential allelic expression. *PLoS ONE*. 2014 Aug 22;9(8):e103872.
15. McMurray RW. Adhesion molecules in autoimmune disease. *Semin Arthritis Rheum*. 1996 Feb;25(4):215–33.
16. Weyand CM, Zeisbrich M, Goronzy JJ. Metabolic signatures of T-cells and macrophages in rheumatoid arthritis. *Curr Opin Immunol*. 2017 Jun;46:112–20.
17. Goronzy JJ, Weyand CM. Successful and maladaptive T cell aging. *Immunity*. 2017 Mar 21;46(3):364–78.
18. Nagyeri G, Radacs M, Ghassemi-Nejad S, Trynieszewska B, Olasz K, Hutas G, et al. TSG-6 protein, a negative regulator of inflammatory arthritis, forms a ternary complex with murine mast cell tryptases and heparin. *J Biol Chem*. 2011 Jul 1;286(26):23559–69.

19. Milner CM, Day AJ. TSG-6: a multifunctional protein associated with inflammation. *J Cell Sci.* 2003 May 15;116(Pt 10):1863–73.
20. Mahoney DJ, Swales C, Athanasou NA, Bombardieri M, Pitzalis C, Kliskey K, et al. TSG-6 inhibits osteoclast activity via an autocrine mechanism and is functionally synergistic with osteoprotegerin. *Arthritis Rheum.* 2011 Apr;63(4):1034–43.
21. Bayliss MT, Howat SL, Dudhia J, Murphy JM, Barry FP, Edwards JC, et al. Up-regulation and differential expression of the hyaluronan-binding protein TSG-6 in cartilage and synovium in rheumatoid arthritis and osteoarthritis. *Osteoarthr Cartil.* 2001 Jan;9(1):42–8.
22. Wisniewski HG, Maier R, Lotz M, Lee S, Klampfer L, Lee TH, et al. TSG-6: a TNF-, IL-1-, and LPS-inducible secreted glycoprotein associated with arthritis. *J Immunol.* 1993 Dec 1;151(11):6593–601.
23. Hastings R, Ding T, Butt S, Gadsby K, Zhang W, Moots RJ, et al. Neutropenia in patients receiving anti-tumor necrosis factor therapy. *Arthritis Care Res (Hoboken).* 2010 Jun;62(6):764–9.
24. Farutin V, Prod'homme T, McConnell K, Washburn N, Halvey P, Etzel CJ, et al. Molecular profiling of rheumatoid arthritis patients reveals an association between innate and adaptive cell populations and response to anti-tumor necrosis factor. *Arthritis Res Ther.* 2019 Oct 23;21(1):216.
25. Conigliaro P, Triggianese P, Perricone C, Chimenti MS, Di Muzio G, Ballanti E, et al. Restoration of peripheral blood natural killer and B cell levels in patients affected by rheumatoid and psoriatic arthritis during etanercept treatment. *Clin Exp Immunol.* 2014 Jul;177(1):234–43.
26. Kim D-H, Park H-J, Lim S, Koo J-H, Lee H-G, Choi JO, et al. Regulation of chitinase-3-like-1 in T cell elicits Th1 and cytotoxic responses to inhibit lung metastasis. *Nat Commun.* 2018 Feb 5;9(1):503.
27. Fioravanti A, Tenti S, Bacarelli MR, Damiani A, Li Gobbi F, Bandinelli F, et al. Tocilizumab modulates serum levels of adiponectin and chemerin in patients with rheumatoid arthritis: potential cardiovascular protective role of IL-6 inhibition. *Clin Exp Rheumatol.* 2019 Apr;37(2):293–300.
28. Liu D, Luo S, Li Z. Multifaceted roles of adiponectin in rheumatoid arthritis. *Int Immunopharmacol.* 2015 Oct;28(2):1084–90.
29. Yang S, Wang J, Brand DD, Zheng SG. Role of TNF-TNF Receptor 2 Signal in Regulatory T Cells and Its Therapeutic Implications. *Front Immunol.* 2018 Apr 19;9:784.
30. Ahmad S, Azid NA, Boer JC, Lim J, Chen X, Plebanski M, et al. The Key Role of TNF-TNFR2 Interactions in the Modulation of Allergic Inflammation: A Review. *Front Immunol.* 2018 Nov 9;9:2572.
31. Yang N, Huang J, Frits M, Iannaccone C, Weinblatt ME, Rifai N, et al. Interference of tumor necrosis factor inhibitor treatments on soluble tumor necrosis factor receptor 2 levels in rheumatoid arthritis. *Practical Laboratory Medicine.* 2019 Aug;16:e00122.
32. Johnson KW, Glicksberg BS, Shameer K, Vengrenyuk Y, Krittanawong C, Russak AJ, et al. A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin treatment: Validation by serial intracoronary OCT imaging. *EBioMedicine.* 2019 Jun;44:41–9.

Figures

Figure 1

A) Volcano plot representation of differentially expressed genes in PBMCs between future responders and non-responders before anti-TNF treatment. Top regulated genes are marked in blue (upregulated genes) and red (downregulated genes). The vertical lines correspond to log2 fold change of 1 (genes are represented in black) and the horizontal line represents a p-value of 0.001. B) The box plot of normalized log2 expression values for the differentially expressed genes, EPPK1, BCL6-AS1 and CDC20 in PBMCs before treatment.

Figure 2

A) The box plots showing the expression levels of differentially expressed genes (all RA patients) in PBMCs for baseline vs. treated patient samples. The expression levels of responders for selected genes are plotted. Abbreviations: BHLHE40, Basic Helix-Loop-Helix Family Member E40; CHI3L1, Chitinase 3 Like 1; FAM129C, Family With Sequence Similarity 129 Member C; TTC21A, Tetratricopeptide Repeat Domain 21A. B). The enrichment plot from GSEA represents functional gene sets enriched between baseline and treated RA patients.

Figure 3

A) The box plots showing the expression levels of genes in responders at baseline vs. treated patient samples. Abbreviations: CXCR2, C-X-C Motif Chemokine Receptor 2; MPO, Myeloperoxidase; MYADM, Myeloid Associated Differentiation Marker; TNFAIP6, TNF Alpha Induced Protein 6; FCGR2B, Fc Fragment Of IgG Receptor IIb. B) The enrichment plot from GSEA representing functional gene sets enriched in PBMCs for baseline vs. treated RA patients for responders. C) The bar plot showing the percentage of granulocytes, B cells, T cells, NK cells, and monocytes of peripheral blood leukocytes before and after anti-TNF treatment in responders. D) The box plots showing the YKL-40 protein expression levels in responders before and after anti-TNF treatment.

Figure 4

Statistical machine learning models to predict response (evaluated after three months) at baseline and after anti-TNF treatment using clinical variables, flow cytometry measurements, protein measurements and gene expression data. The Y-axis represents the area under the receiver operating characteristic (ROC) curves (AUCs) calculated for estimating the predicted performance.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryDataS1.docx](#)
- [SupplementaryDataS1.docx](#)
- [SupplementaryTableS1.xlsx](#)