Expert Review of Clinical Immunology

ISSN: 1744-666X (Print) 1744-8409 (Online) Journal homepage: https://www.tandfonline.com/loi/ierm20

RNA sequencing to predict response to TNF-α inhibitors reveals possible mechanism for nonresponse in smokers

Bart V.J. Cuppen, Marzia Rossato, Ruth D.E. Fritsch-Stork, Arno N. Concepcion, Suzanne P. Linn-Rasker, Johannes W.J. Bijlsma, Jacob M. van Laar, Floris P.J.G. Lafeber & Timothy R. Radstakeon behalf of all SRU investigators

To cite this article: Bart V.J. Cuppen, Marzia Rossato, Ruth D.E. Fritsch-Stork, Arno N. Concepcion, Suzanne P. Linn-Rasker, Johannes W.J. Bijlsma, Jacob M. van Laar, Floris P.J.G. Lafeber & Timothy R. Radstakeon behalf of all SRU investigators (2018) RNA sequencing to predict response to TNF-α inhibitors reveals possible mechanism for nonresponse in smokers, Expert Review of Clinical Immunology, 14:7, 623-633, DOI: 10.1080/1744666X.2018.1480937

To link to this article: https://doi.org/10.1080/1744666X.2018.1480937

EXPERT REVIEW OF CLINICAL IMMUNOLOGY
2018, VOL. 14, NO. 7, 623–633
https://doi.org/10.1080/1744666X.2018.1480937

Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH

RNA sequencing to predict response to TNF-α inhibitors reveals possible mechanism for nonresponse in smokers

Bart V.J. Cuppen[a*], Marzia Rossato[b,c*], Ruth D.E. Fritsch-Stork[a,d,e], Arno N. Concepcion[a], Suzanne P. Linn-Rasker[f], Johannes W.J. Bijlsma[a], Jacob M. van Laar[a], Floris P.J.G. Lafeber[a] and Timothy R. Radstake[a,b]on behalf of all SRU investigators

[a]Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands; [b]Laboratory of Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands; [c]Department of Biotechnology, University of Verona, Verona, Italy; [d]1st Medical Department & Ludwig Boltzmann Institute of Osteology at the Hanusch Hospital of WGKK and AUVA Trauma Centre Meidling, Hanusch Hospital, Vienna, Austria; [e]Sigmund Freud University, Vienna, Austria; [f]Rheumatology, Meander Medisch Centrum, Amersfoort, Netherlands

ABSTRACT

Background: Several studies have employed microarray-based profiling to predict response to tumor necrosis factor-alpha inhibitors (TNFi) in rheumatoid arthritis (RA); yet efforts to validate these targets have failed to show predictive abilities acceptable for clinical practice.

Methods: The eighty most extreme responders and nonresponders to TNFi therapy were selected from the observational BiOCURA cohort. RNA sequencing was performed on mRNA from peripheral blood mononuclear cells (PBMCs) collected before initiation of treatment. The expression of pathways as well as individual gene transcripts between responders and nonresponders was investigated. Promising targets were technically replicated and validated in $n = 40$ new patients using qPCR assays.

Results: Before therapy initiation, nonresponders had lower expression of pathways related to interferon and cytokine signaling, while also showing higher levels of two genes, GPR15 and SEMA6B ($p = 0.02$). The two targets could be validated, however, additional analyses revealed that GPR15 and SEMA6B did not independently predict response, but were rather dose-dependent markers of smoking ($p < 0.0001$).

Conclusions: The study did not identify new transcripts ready to use in clinical practice, yet GPR15 and SEMA6B were recognized as candidate explanatory markers for the reduced treatment success in RA smokers.

ARTICLE HISTORY
Received 21 January 2018
Accepted 22 May 2018

KEYWORDS
mRNA; prediction; qPCR; response; rheumatoid arthritis; RNA sequencing; TNF-alpha

1. Background

Rheumatoid arthritis (RA) is a chronic, disabling disease that mainly affects the synovial joints. With a large arsenal of treatments including tumor necrosis factor-alpha inhibitors (TNFi), RA disease activity decreases sufficiently in most, but not all cases. The identification of responders and nonresponders before initiation of therapy would therefore aid in making strategic treatment decisions and improve clinical outcomes. Thus far, however, no biomarker to predict response to TNFi treatment is robust enough to be used in clinical practice [1].

Although TNF does not act alone, it is considered to play an instrumental role in the pathogenesis of RA [2]. Indeed, TNFα plays a pleiotropic role in the immune system, including stimulating cytokine production (including its own), enhancing the expression of adhesion molecules and in neutrophil activation, and it is also a costimulator of T cell activation and antibody production by B cells [3]. From a scientific and commercial point of view, TNFis are undoubtedly one the biggest successes of rational drug design [4]. All TNFis (such as adalimumab, infliximab, golimumab, certolizumab) specifically ligate with transmembrane TNF and soluble TNF, thereby blocking TNFα from binding with the TNF-receptor (TNFR) 1 and 2 on TNF responsive cells, and preventing downstream signaling leading to the transcription of inflammatory genes [2]. Besides TNF, etanercept also blocks lymphotoxin alpha (LTα), a member of the TNF family with the potency to activate the downstream pathways after binding with TNFR1, TNFR2, tumor necrosis factor receptor superfamily member 14 (TNFRSF14/HVEM receptor) and, when bound to LTβ, with the lymphotoxin beta receptor (LTBR/TNFR3) [2,4,5]. Any biomarkers of response are therefore expected to have a role in any of the downstream pathways of these mechanisms. However, although TNF is considered to play a major role in RA, the therapeutic efficacy is comparable with therapeutics that specifically target other inflammatory pathways (e.g. interleukin 6 receptor, T cell costimulation blockade, and B cell depletion [6]. These observations in clinical practice suggest a common final pathway that is inhibited by all of these drugs: i.e. proinflammatory cytokine production [6,7]. Biomarkers of TNFi response might therefore

CONTACT Bart V.J. Cuppen b.v.j.cuppen@gmail.com Rheumatology & Clinical Immunology, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, The Netherlands

*These authors contributed equally to this work

Supplemental data for this article can be accessed here.

also arise from more general immunologic pathways. These biomarkers would then predict a general biological treatment refractory RA, rather than a treatment-specific marker of non-response. This knowledge is also of interest, as many more treatment options are available, including combinations of several drugs. In addition, an expected poor prognosis in terms of therapeutic response may aid the patient and physician in making long-term decisions. Therefore, discovering biomarkers of drug response remains to be one of the major challenges for RA.

Messenger RNAs (mRNAs) are of potential interest as biomarkers of response since they constitute the bridge between the genetic and protein formation. In order to identify mRNA gene transcripts able to predict response to biological treatment, the transcriptome has been studied extensively by microarrays [8,9]. However, the results generated by these studies are heterogeneous and efforts to replicate models in external cohorts failed to reach predictive abilities acceptable for clinical practice [10–12]. The inability to validate results in multiple cohorts might be partially related to the technical drawbacks of microarrays. RNA sequencing (RNA-seq) is a newer technique that overcomes multiple limitations characterizing microarrays and therefore will eventually been used more routinely for the measurement of mRNA levels [13,14]. Whereas microarray profiling is a chip-based method using a selected set of gene transcripts, RNA-seq is not based on a predefined platform, and it is thus able to identify modulated genes without any a-priori selection [15]. RNA-seq is also able to quantify genes with very low and high expression more accurately, because it is characterized by less background noise and has no (cross) hybridization issues [15,16]. In addition, due to its bigger dynamic range, RNA-seq is more suitable for the quantification of absolute gene expression, thus resulting in more reproducible data [17–19]. One recent study employed RNA-seq-based transcriptome analysis of neutrophils to predict the response to TNFi and found that a combination of >100 interferon regulated transcripts at baseline predicted response [20]. In a follow-up study, the authors refined their selection further and additionally found neutrophil granule protein (NGP)-related transcripts to be of importance [21]. These were further narrowed down in a new cohort, in which response could be predicted by expression of 2 IFN-related genes and 1 NGP-related gene.

In this study, we sought to extend the search for gene transcripts predictive of therapeutic response, by employing RNA-seq on baseline mRNA of peripheral blood mononuclear cells (PBMCs) in RA patients treated with two distinct TNFi treatments, namely adalimumab (ADA) and etanercept (ETN). Additionally, we aimed to validate the identified predictors in a separate cohort and investigated the association of gene expression with PBMC cell subsets and clinical characteristics of patients.

2. Patients and methods

2.1. Clinical data collection

Patients initiating ADA or ETN therapy were selected from the 'Biologicals and Outcome Compared and predicted Utrecht region in Rheumatoid Arthritis' (BiOCURA) study. BiOCURA is an observational cohort, in which RA patients eligible for biological treatment according to regular clinical practice were enrolled and followed after start of treatment, in one academic hospital and seven regional hospitals in the Netherlands. Re-inclusion after switching to a different biological treatment was possible, at which patients re-entered baseline again. The study was approved by the local ethics committee of the University Medical Center Utrecht and the institutional review boards of the participating centers (see Acknowledgement). Informed consent was obtained from each patient.

Trained nurses collected all data, which included all clinical parameters, joint counts and collection of blood. Visits were scheduled at baseline (before initiation) and after 3, 6, and 12 months of treatment. Disease activity was assessed each visit using the disease activity score based on a 28-joint count (DAS28) and subsequently the European League Against Rheumatism (EULAR) response compared with baseline was calculated [22]. Thereby, this study design allowed the determination of a clinical response of each patient, on the basis of three responses over the course of one year.

2.2. Patient selection

Two separate cohorts were composed: a discovery cohort to select potentially predictive gene transcripts, and a validation cohort to test if the results found in the discovery phase were reproducible. The discovery cohort was formed by selecting the most extreme patients regarding clinical response, among all ADA and ETN treated patients included between June 2009 (start of BiOCURA) and October 2012. As patients switching between biologicals were also included, patients did not always initiate a TNFi when having a high disease activity (DAS28 >5.1 [6]). Yet, patients with a baseline DAS28 <2.6, usually due to involvement of joints that are not calculated in the 28-joint count such as in the foot, were excluded from the analysis in order to reduce the chance that limited improvement in DAS28 resulted in misclassification as EULAR nonresponders. Of all remaining patients (n = 74 ADA and n = 68 ETN), the top responding patients were identified after ranking of patients according to the best three EULAR responses over the course of 1 year (n = 20 for both ADA and ETN, from now on called 'responders'). The selection of poor responders, was based on ranking of patients according to the least optimal EULAR responses over the course of one year and/or (early) discontinuation of TNFi treatment due to inefficacy (n = 20 for both ADA and ETN, 'nonresponders'). The selected 80 patients were divided over four subgroups (ADA n = 20 responders versus n = 20 nonresponders, and ETN n = 20 responders versus 20 nonresponders), the top ten patients in each subgroup with most extreme responses were named 'extreme responders' and 'extreme nonresponders' to analyze groups of patients with even clearer treatment outcomes. Supplementary Figure 1 shows the DAS28 over time of all selected patients.

For the validation cohort, responders (n = 10 each for ADA and ETN) and nonresponders (n = 10 each for ADA and ETN) were selected using the same criteria as in the discovery cohort, among patients included from October 2012 until June 2015 (n = 25 ADA and n = 40 ETN). Due to the smaller size of the cohort from which patients were selected, the

differences in clinical outcome between responders and non-responders in the validation cohort was less extreme than in the discovery cohort. The baseline characteristics for responders and nonresponders are shown in Table 1 and for responders and nonresponders split per cohort in Supplementary Table 1.

2.3. Blood processing and RNA extraction

Blood was collected in a 70 ml lithium heparin tube and PBMCs were isolated using Ficoll Paque Plus (GE Healthcare, Uppsala, Sweden). $5*10^6$ PBMCs were lysed for total RNA isolation using RNeasy kit following the manufacturer's instructions (Qiagen). Quantification of RNA and purity was assessed using Nanodrop (Thermo Fischer Scientific, Waltham, Massachusetts, USA) and the quality with Bioanalyzer (Agilent, Santa Clara, California, USA). RNA was stored per 2 µg at −80°C until use. All RNA samples analyzed had a RNA-integrity score (RIN) higher than 8 according to the bioanalyzer assessment.

2.4. mRNA analyses

2.4.1. RNA sequencing

RNA-seq of the discovery cohort was performed at the Beijing Genomics Institute (BGI, Hong Kong). RNA-seq libraries were prepared starting from 100 ng total RNA using the TruSeq kit (Illumina). Briefly, mRNA molecules were enriched by using the oligo(dT) magnetic beads and fragmented into short fragments (about 200 bp). After conversion to double-stranded cDNA using random hexamer-primer, end reparation and 3′-end single-nucleotide A (adenine) addition were performed. Finally, sequencing adaptors were ligated to the fragments that were subsequently enriched by PCR amplification. Quality and quantity of RNA-seq libraries were determined using, respectively, the Agilent 2100 Bioanaylzer and the ABI StepOnePlus Real-Time PCR System. The library products were sequenced on an Illumina HiSeq™ 2000 sequencer using 50bp single-end read, generating 20 million clean reads per sample. After quality filtering according to the BGI pipeline, reads were aligned to the GrCh38 reference human genome (Genome Reference consortium) and the H. sapiens transcriptome (Ensembl, version 78) using SOAPaligner/SOAP2 [23] obtaining an average map rate of 87.76% and 79.78%, respectively (Supplementary File 1). Summed exon read counts per gene were estimated using HTSeq-count [24]. Differential expression analysis was performed using the negative binomial distribution-based method implemented in DESeq on the summed exon read counts per gene [25]. Gene expression levels were calculated as Reads per kilo base per million mapped reads (RPKM), according to the formula: numReads/(geneLength/1000 * totalNumReads/1,000,000). A filter was applied to exclude all genes with a mean reads per kilobase per million mapped reads (RPKM) lower than 0.5, as genes with very low expression levels are less reliably measured by RNA-seq [14], and therefore are less suitable as biomarkers [26]. Group comparisons were performed between six subgroups: nonresponders versus responders to ADA, ETN and all TNFi (pooled ADA and ETN), and extreme nonresponders versus extreme responders to ADA, ETN and all TNFi.

2.4.2. RT-qPCR analyses

For the technical replication and validation experiments, Taqman gene expression qPCR kits for SEMA6B (Hs00220339_m1), GPR15 (Hs00922903_s1), ACTB (Hs99999903_m1), and GAPDH (Hs99999905_m1) were used (ThermoFisher). ACTB and GADPH were selected as reference genes since they were both highly expressed and stable

Table 1. Baseline characteristics of patients, split for treatment, and response.

	ADA (n = 60)			ETN (n = 60)		
Item	Non-resp (n = 30)	Resp (n = 30)	p-value	Non-resp (n = 30)	Resp (n = 30)	p-value
Female gender, n (%)	21 (70)	21 (70)	1.00	25 (83)	21 (70)	0.36
Age, mean years ±sd	54.4 ± 10.9	53.5 ± 12.7	0.76	58.3 ± 9.2	55.1 ± 10.5	0.22
Current smoker, n (%)	16 (53)	8 (26.7)	0.06	8 (26.7)	7 (23.3)	1.00
RF positivity, n (%)	16 (53)	21 (70)	0.29	20 (67)	22 (73)	0.78
ACPA positivity, n (%)	19 (63)	19 (63)	1.00	19 (63)	26 (87)	0.07
CRP, median (IQR)	5.2 (1.6–10.5)	5.5 (2.0–12.3)	0.78	4.0 (2.0–9.0)	8.5 (4.0–18.3)	**0.03**
No. of previously used bDMARDs			0.48			1.00
0	20 (67)	23 (78)		22 (73)	22 (73)	
1	9 (30)	7 (23)		7 (23)	7 (23)	
2	1 (3)	0 (0)		1 (3)	1 (3)	
Concomitant treatment, n (%)	29 (97)	29 (97)	1.00	27 (90)	29 (97)	0.61
MTX, n (%)	21 (70)	27 (90)	0.10	18 (60)	25 (83)	0.08
SSZ, n (%)	2 (7)	4 (13)	0.67	4 (13)	2 (7)	0.67
HCQ, n (%)	8 (27)	7 (23)	1.00	10 (33)	11 (37)	1.00
GC, n (%)	15 (50)	4 (13)	0.01	11 (37)	6 (20)	0.25
Baseline DAS28, mean ±sd	3.9 ± 1.4	4.7 ± 0.9	0.01	4.3 ± 1.2	4.6 ± 0.9	0.21
TJC, median (IQR)	5.0 (1.0–13.0)	7.0 (4.0–14.3)	0.35	6.5 (2.8–11.3)	5.0 (2.8–11.3)	0.87
SJC, median (IQR)	0.0 (0.0–4.0)	2.0 (0.0–4.0)	**0.03**	1.0 (0.0–3.3)	2.0 (0.8–4.0)	0.20
VAS-GH, mean ±sd	55.2 ± 23.8	63.8 ± 22.0	0.15	55.5 ± 22.8	55.1 ± 10.5	0.76
ESR, median (IQR)	11.0 (3.8–26.0)	16.5 (9.0–32.0)	0.14	13.0 (5.8–33.8)	21.0 (14.3–39.5)	0.07

Patients were selected from the observational BiOCURA cohort based on treatment outcome over the course of one year after start of either ADA or ETN. The presented clinical characteristics for responders (resp) and nonresponders (non-resp) are all before treatment initiation. P-values for comparisons were calculated by means of an independent sample t-test, Mann–Whitney U test, fisher exact test (2*2) or chi-square (>2*2) based on distribution of the clinical parameter.

ACPA: anti-citrullinated protein antibody, ADA: adalimumab, bDMARDs: biological disease-modifying antirheumatic drugs, CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, ETN: etanercept, GC: glucocorticoid, HCQ: hydroxychloroquine, IQR: interquartile range, MTX: methotrexate, RF: rheumatoid factor, SJC: swollen joint count, SSZ: sulfasalazine, TJC: tender joint count, VAS-GH: visual analogue scale general health.

between responders and nonresponders (log2FC <0.03). Briefly, 0.3 µg PBMC-derived RNA was used as starting material for both the technical replication (n = 80) and validation analysis (n = 40). cDNA was synthesized from RNA using the iScript^{tm} synthesis kit (Biorad), according to the manufacturer's instructions, while using the following thermal cycler conditions: 5 min, 25°C, 30 min 42°C, 5 min 85°C. Quantification of gene transcripts was performed in duplicate from 2 µl 6-times diluted cDNA, TaqMan-specific primers of the gene expression assay and TaqMan Fast Advance Master Mix in a final volume of 15 µl, on a Quantstudio 12kflex Real-Time PCR system (LifeTechnologies) using the following thermal cycler conditions: 2 min, 50°C; 20 sec, 95°C; 45 cycles of 1 sec, 95°C; 20 sec, 60°C. Data were analyzed according to the comparative threshold cycle method [27], after normalization by the selected housekeeping genes. The fold change (FC) was calculated based on one randomly chosen sample that was set as 1 and analyzed on each qPCR plate, thus allowing proper comparison of data across different plates.

2.5. Flow cytometry

After isolation from whole blood, PBMCs were stained by using specific monoclonal antibodies and analyzed by fluorescence-activated cell sorting (FACS) using an FACS-LSRII (BD) in order to determine the abundance of specific mononuclear cell subsets: anti-CD45 (FITC, Beckman Coultier) was used to identify leukocytes, anti-CD3 (FITC, Beckman Coultier) for T cells, anti-CD4 (PECy5, Dako) for CD4 T cells, anti-CD8 (PE, BD) for CD8 T cells, anti-CD14 (PE, Beckman Coultier) for monocytes, anti-CD16 + 56 (PE, Beckman Coultier) for NK cells, anti-CD19 (FITC, BD) for B cells. Per sample, 30,000 events were registered, after which these data were further analyzed using FlowJo software (Tree star, Ashland, OR, USA).

2.6. Statistical analyses

2.6.1. Pathway enrichment analysis

Pathway enrichment analyses was performed on the complete set of expressed genes (i.e. with mean expression >0.5 RPKM), using the online available tool Gene Set Enrichment Analyses (GSEA [28]). GSEA is a computational method that determines whether a predefined set of genes, e.g. a pathway, shows statistically significant and concordant differences between two biological states, in this case responders and nonresponders. GSEA analysis was performed on the relative expression of genes in all TNFi nonresponders versus responders using the REACTOME pathway database [29] as source of gene lists. Enriched pathways with a Benjamini & Hochberg adjusted p-value (B&H adj. p-val.) lower than 0.20 were considered significantly enriched [28].

A second pathway enrichment analysis was performed on all protein-coding genes identified as differentially expressed in responders versus nonresponders via the web-portal Toppfun (https://toppgene.cchmc.org/enrichment.jsp [30],).

To verify whether the baseline expression of genes downstream of TNFα and lymphotoxin-alpha (LTα) were differentially expressed in responders versus nonresponders, intracellular pathways activated by the TNFα and LTα receptors were retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG [31], ID 04668 + 04064 (http://www.genome.jp/kegg/pathway.html); the expression of genes belonging to these pathways was derived from the differential gene expression analysis, without the application of a restriction based on minimal expression (i.e. RPKM>0.5).

2.6.2. Selection of relevant genes

In order to identify the most interesting individual targets in the context of TNFi response, we selected genes that significantly predicted response in multiple subgroups. Genes with absolute log2(FC)>0.58 (i.e. a corresponding FC<0.67 or >1.5) and nominal p-value<0.05 were considered significantly different and relevant in the context of TNFi response. These cut-offs were set low to not miss any important genes, whereas the validation phase was supposed to test if the targets are reliable and replicable. We considered genes that emerged in multiple comparisons as most promising targets, especially when seen in both nonresponders versus responders and extreme nonresponders versus extreme responders, as this can be considered a small-scale internal validation. The genes of particular interest were plotted in GraphPad Prism software (GraphPad, Lo Jolla, CA, USA) as the reads RPKM of (excellent) nonresponders versus (excellent) responders. Sensitivities and specificities (i.e. the proportion of, respectively, responders and nonresponders to TNFi that are correctly identified as such) for genes were calculated on the log(RPKM)-values with the cut-off based on the Youden's index [32].

2.6.3. Influencing parameters

We investigated whether the expression of selected genes (as measured by single qPCR) was related to the relative subset composition of PBMCs (as measured by flow cytometry). Plots for gene expression versus percentage of cells on all 120 patients (development + validation cohort) were made in Graphpad, and correlations and p-values were calculated by a Spearman's r.

We additionally investigated whether the gene expression of the selected genes (as measured by qPCR) in the complete cohort (n = 120) associated with several patient characteristics and RA (treatment)-specific clinical parameters. We performed a linear regression on the ddCt values of the genes with these parameters as independent variables and performed a backward selection procedure (p < 0.05), so that the most explanatory parameters remained.

3. Results

3.1. Multiple immune-related pathways are modulated at baseline in nonresponders to TNFi

In order to identify gene transcripts and pathways differentially expressed in responders versus nonresponders before the initiation of ADA and ETN treatment, transcriptome analysis by RNA-seq was performed on baseline PBMCs from a cohort of 80 RA patients: 40 for each treatment, including 20 responders and 20 nonresponders. To enhance the discovery power, the patients selected for this analysis had the best and

poorest responses over the course of one year after therapy (Material and Methods and Supplementary Figure 1). Within these groups, the top-10 in ranking based on EULAR response were named 'extreme responders' and 'extreme nonresponders.' These subgroups were also compared to push the discovery power even further thanks to the enhanced difference in the clinical outcome. Supplementary Table 2 reports in details the total number of genes that had different baseline levels in responders to treatment vs. nonresponders, in the different comparisons tested. Overall, 178 unique genes were expressed higher or lower at baseline in nonresponders.

To identify the pathways that were different at baseline in nonresponders, we performed GSEA enrichment analysis on the set of measured transcripts with a minimal level of expression (RPKM >0.5, i.e. 12,716 genes). In total, 26 pathways were significantly differentially expressed in nonresponders compared to responders prior to treatment. Among these, seven pathways could be directly linked to immune response: three interferon related (all expressed lower), two virus (influenza) life-cycle related (both expressed higher), one cytokine signaling related (expressed lower), and one MHC class II related (expressed lower) (Supplementary Table 3). Further investigation on the gene content of these seven pathways demonstrated that around 67% (228 of 338) of enriched genes were annotated to multiple immunological pathways, thus constituting a core enrichment gene set (Supplementary File 2).

GO-term enrichment analysis on all 178 selected unique targets confirmed that, also among the most robustly differentially expressed genes, pathways related to immune processes were over-represented, such as 'antigen-binding' and 'immune response' (Table 2).

Manual inspection of the genes involved in the downstream pathways of TNFa and LTα, the targets of ADA and ETN, demonstrated that none of them was significantly differentially expressed in more than one subgroup prior to start of treatment (data not shown).

3.2. Responders and nonresponders have different expression levels of GPR15 and SEMA6B

In order to identify specific targets that could be used for prediction of TNFi response, we further narrowed down those gene transcripts differentially expressed in multiple comparisons, out of the six considered (Supplementary Table 4). Overall we observed a small number of genes predictive of response to both TNFi treatment: six genes were differentially expressed in all nonresponders versus responders, and 22 genes in all extreme nonresponders versus extreme responders. The overlap in predictive genes between these two comparisons consisted of two targets: GPR15 and SEMA6B. Since SEMA6B was also predictive in the two ADA-specific comparisons and GPR15 in ADA subgroup with extreme responses, we considered these the most interesting targets and focused further analyses on these two genes. Moreover we observed that GPR15 and SEMA6B showed similar trends in the ETN comparisons, with a FC (*p*-value) for GPR15 of 1.63 (0.09) and for SEMA6B 1.94 (0.20) for ETN responders versus nonresponders. The expression of GPR15 and SEMA6B as

Table 2. Pathways and gene ontology (GO) biological processes related to the 178 selected genes. All 178 genes with a minimal expression (RPKM >0.5) and a significant difference between responders and nonresponders (nominal $p < 0.05$) with an absolute log2FC >0.58, were used as input for a Toppgene pathway enrichment analysis. 146/178 gene could be annotated to functional pathways, among which especially immune-related pathways were overrepresented. Shown are the top 10 gene ontology (GO) molecular functions (MF)/biological processes (BP) and pathways, as based on nominal *p*-values (not shown). After a Benjamini & Hochberg correction of the *p*-value (B&H adj. P-val), all pathways in the top 10 were still significantly enriched.

Source	Pathway	B&H adjusted p-value	Genes total in pathway	No. of gene hits	Enriched genes
GeneRIF	New genetic associations detected in a host response study to hepatitis B vaccine.	2.05E-06	826	20	SLAMF8, COL4A3, FCRL2, HLA-B, HLA-C, FPR3, FCRLA, TNFRSF10C, ADORA3, ORM1, FCRL1, COCH, CD19, MS4A1, CD22, SIGLEC6, TNFRSF13C, C1QC, CD79A, CLEC10A
GeneRIF	Association study of B cell marker gene polymorphisms in European Caucasian patients with systemic sclerosis.	2.05E-06	4	4	CD19, MS4A1, CD22, CD24
GO: BP	Immune response	7.06E-06	1572	33	MRC1, LILRB3, HLA-A, HLA-B, HLA-C, FOS, TNIP3, VSIG4, POU2AF1, TNFRSF10C, CD209, BLK, ADORA3, JUN, CD1E, THBS1, COCH, CD19, MS4A1, ERAP2, IGHD, IGHG1, IGHM, IGKC, CD24, TNFRSF13C, TNFRSF13B, C1QC, CCL3L1, CD79A, PAWR, PAX5, CLEC10A
GO: MF	Antigen binding	3.91E-06	120	10	HLA-A, HLA-B, HLA-C, CD209, CD1E, MS4A1, IGHD, IGHG1, IGHM, IGKC
GeneRIF	Association of maternal histocompatibility at class II HLA loci with maternal microchimerism in the fetus.	1.10E-05	7	4	HLA-A, HLA-B, HLA-C, GSTM1
GeneRIF	Functional characterization of the human immunodeficiency virus type 1 Nef acidic domain.	1.10E-05	10	4	HLA-A, HLA-B, HLA-C, JUN
GO: BP	Adaptive immune response	1.20E-04	402	15	LILRB3, HLA-A, HLA-B, CD209, CD1E, ERAP2, IGHD, IGHG1, IGHM, IGKC, TNFRSF13C, TNFRSF13B, C1QC, CD79A, CLEC10A
GO: BP	B cell receptor signaling pathway	1.98E-04	65	7	BLK, CD19, IGHD, IGHG1, IGHM, IGKC, CD79A
GO: BP	Antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP independent	1.98E-04	3	3	HLA-A, HLA-B, HLA-C
GO: BP	Regulation of immune response	3.27E-04	899	21	HLA-A, HLA-B, HLA-C, FOS, TNIP3, VSIG4, CD209, BLK, ADORA3, JUN, COCH, CD19, IGHD, IGHG1, IGHM, IGKC, CD24, TNFRSF13C, C1QC, CD79A, PAWR

measured by RNA-seq is displayed in Figure 1. GPR15 showed a sensitivity of 55% and specificity of 75% to distinguish nonresponders from responders, whereas SEMA6B showed a sensitivity of 37.5% and specificity of 92.5%, (as determined by the Receiver Operating Characteristic (ROC) curve, Supplementary Figure 2). These high specificities indicate that the implementation of SEMA6B and GPR15 measurement in a potential clinical test would be especially suitable to accurately identify (mainly ADA) nonresponders, with the clinical implication of withholding treatment when the risk of response is extremely low. Combining the genes did not result in better discriminating abilities, with a slightly improved sensitivity of 60.0%, yet a lower specificity (67.5%).

3.3. GPR15 and SEMA6B are markers for smoking, rather than independent predictors

In order to validate GPR15 and SEMA6B as predictors of response to TNFi therapy, the expression of GPR15 and SEMA6B was analyzed using quantitative PCR in the discovery cohort (i.e. technical replication, $n = 80$) as well as in a new cohort of patients (validation, $n = 40$). Technical replication in the discovery cohort showed comparable differences for all comparisons, yet in all cases the magnitude and statistical significance decreased to some extent (Table 3). Nevertheless, the observed differences were small enough to assume that they emerged from intrinsic variations between the two techniques, as demonstrated by the high positive correlation between the RNA-seq and qPCR measurements ($r = 0.883$ ($p < 0.001$) and $r = 0.857$ ($p < 0.001$) for GPR15 and SEMA6B, respectively)(Supplementary Figure 3). In the validation cohort, the expression levels of SEMA6B still showed a significant upregulation in ADA nonresponders (FC 6.18, $p = 0.01$), although in ETN, still a not significant and now inverse direction as compared to the discovery phase was observed (FC = 0.68, $p = 0.63$) (Figure 2). GPR15 was also replicated as being expressed higher at baseline in nonresponders to ADA (although not significant, with FC = 1.67, $p = 0.30$), whereas in ETN also an inverse direction was seen (FC = 0.52, $p = 0.20$).

The ability to validate SEMA6B and (to some extend) GPR15 as predictors of ADA response, raised the question whether

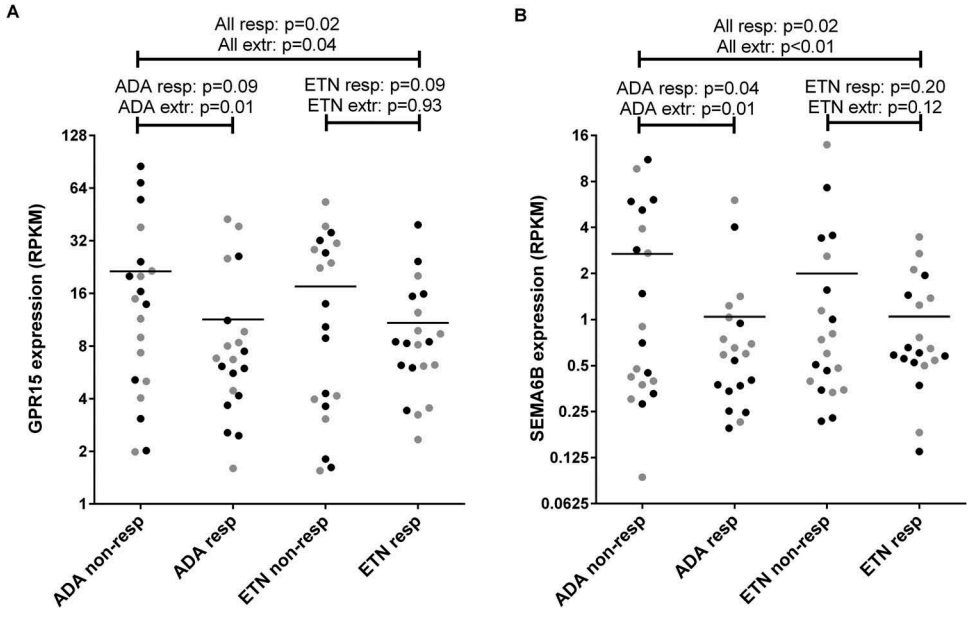

Figure 1. GPR15 and SEMA6B expression in the discovery cohort, split for ADA and ETN nonresponders versus responders. Shown are the baseline gene expression levels of (a) GPR15 and (b) SEMA6B as measured by RNA sequencing, in reads per kilobase per million mapped reads (RPKM), for nonresponders and responders to treatment ($n = 40$ ADA and $n = 40$ ETN) in the discovery cohort. Black dots indicate the extreme (non)responders among all patients, whereas the gray dots represent the remaining (non)responders. The horizontal bar indicates the geometric mean.

Table 3. Technical replication of GPR15 and SEMA6B using single qPCR. In the discovery cohort of 80 patients, GPR15 and SEMA6B were selected as potential predictors for response to ADA and ETN. In order to test the technical replicability of these genes, all 80 samples were re-measured using single RT-qPCR assays. Shown are the fold changes (FC) of nonresponders versus responders, and the corresponding *p*-value (based on DESeq-analysis for discovery, and t-test on ddCt in the replication). The differential expression in the technical replication was in the same direction, however, the FCs were in general less extreme, with a parallel increase in *p*-value.

		GPR15				SEMA6B			
		Discovery		Tech. replication		Discovery		Tech. Replication	
Subgroups	***n***	**FC**	***p*-val**	**FC**	***p*-val**	**FC**	***p*-val**	**FC**	***p*-val**
All resp	**80**	1.752	**0.02**	1.494	0.08	2.190	**0.02**	1.432	0.27
All extr	**20**	2.073	**0.04**	1.514	0.27	3.603	**0.00**	2.723	**0.02**
Ada resp	**40**	1.873	0.09	1.559	0.18	2.431	**0.04**	1.476	0.41
Ada extr	**20**	3.891	**0.01**	2.977	**0.05**	4.516	**0.01**	4.238	**0.04**
Etn resp	**40**	1.628	0.09	1.432	0.28	1.941	0.20	1.390	0.37
Etn extr	**20**	1.086	0.93	0.770	0.59	2.681	0.12	1.750	0.30

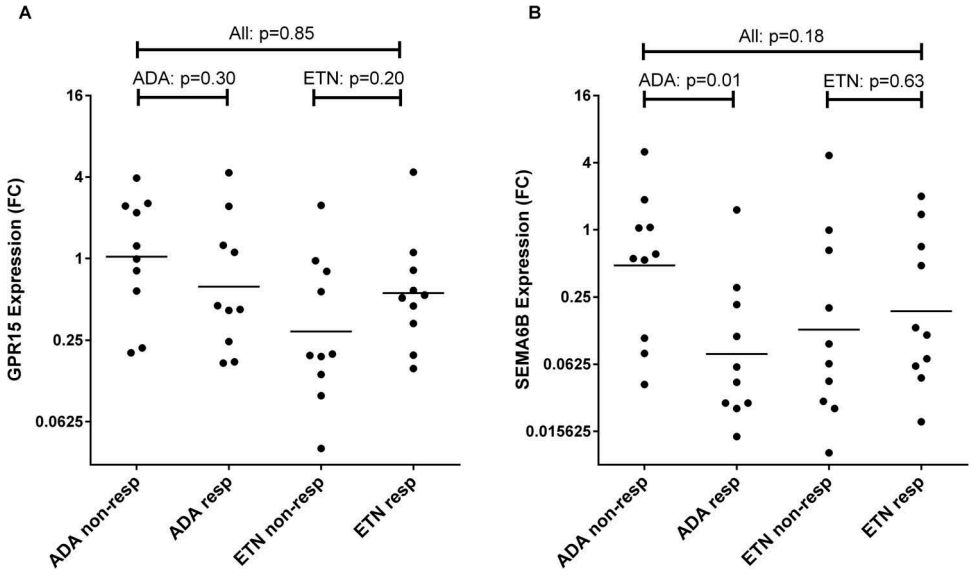

Figure 2. GPR15 and SEMA6B expression in the validation cohort, split for ADA and ETN nonresponders versus responders. For no-responders and responders to ADA (n = 20) and ETN (n = 20) treatment in the validation cohort, the baseline relative gene expression levels of (a) GPR15 and (b) SEMA6B as measured by RT-qPCR, is shown as the fold change (FC). The horizontal bar indicates the geometric mean.

other parameters influenced the relation between the expression levels of these transcripts and the response to therapy. To verify whether SEMA6B and GPR15 could be a reflection of the PBMC cell-subset composition before mRNA extraction, we correlated gene expression (n = 120 in discovery and validation) with the percentage of cell subsets as measured by FACS (Supplementary Figures 4 and 5). SEMA6B expression did not significantly correlate with any cell-subset, whereas GPR15 expression only very weakly negatively correlated with the relative number of monocytes (r = − 0.20, p = 0.04) and B cells (r = 0.25, p = 0.01), indicating that PBMC composition probably had no influence on the observed results. Additionally, we verified whether the expression of these genes was associated with any baseline clinical characteristics of the patients enrolled in the study. Higher expression SEMA6B was most notably associated with current smoking (B = 2.41, p < 0.001) and the (log-transformed) C-reactive protein levels (B = 0.32 p = 0.029), whereas higher expression of GPR15 was also associated with current smoking (B = 2.18, p < 0.001) and lower expression with concomitant glucocorticoid use (B = -0.551, p = 0.015). The association between smoking and high levels of GPR15 and SEMA6B explained why these genes were identified as robustly differentially expressed only in the discovery phase: in the discovery cohort, indeed, nonresponders included 15% more current smokers (Supplementary Table 1). In line with this observation, the ability to validate SEMA6B and (to some extent) GPR15 in ADA can be explained by a 30% higher proportion of smokers among nonresponders. We further investigated the relationship between the expression of these two genes and smoking, by stratifying the patients into never smokers (n = 33), past smokers (n = 48) and current smokers (n = 39), while concomitantly accounting for the number of daily smoked cigarettes and total smoked pack years (PY) (Figure 3). While focusing on current smokers, we observed positive correlations with extremely significant p-values (p < 0.0001) between the number of daily smoked cigarettes and PY with the expression of GPR15 and SEMA6B, indicating that there is a dose-response relationship for the exposure to smoke and upregulation of these genes. Additionally, the highly significant positive correlation between PY and expression of GPR15 seen in past smokers, indicates that there is a certain prolonged and cumulative effect of (heavy) smoking and the upregulation of this gene, whereas for SEMA6B these effects were not observed.

4. Discussion

In this study, we investigated the baseline transcriptome of PBMC mRNA to identify predictors of response to TNFi. Although in the discovery phase we used loose cut-offs (i.e. selection of patients with extreme clinical response and no correction for multiple testing) to decrease chances of excluding potentially interesting targets, only a limited number of genes were differentially expressed at baseline in nonresponders versus responders and very few were identified in common between the two treatment considered, ADA and ETN. This could be related to multiple factors, such as the heterogeneity of the disease, or within the cohort selected, or due to concomitant/previous additional treatments. Also, the analysis at a single time point (baseline) and the focus on PBMCs rather than synovial tissue of the affected joints, may have impacted the results. Thus, to overcome false-positive findings in the discovery phase, we narrowed down the selection to genes differentially expressed in multiple comparisons only, and aimed to validate the findings in an additional cohort of patients. Several additional analyses were performed to assess the biomedical relevance and the generalizability of our findings.

Our approach led to the selection of GPR15 and SEMA6B as the most promising gene transcripts for prediction of TNFi-response, of which SEMA6B could be validated in a separate

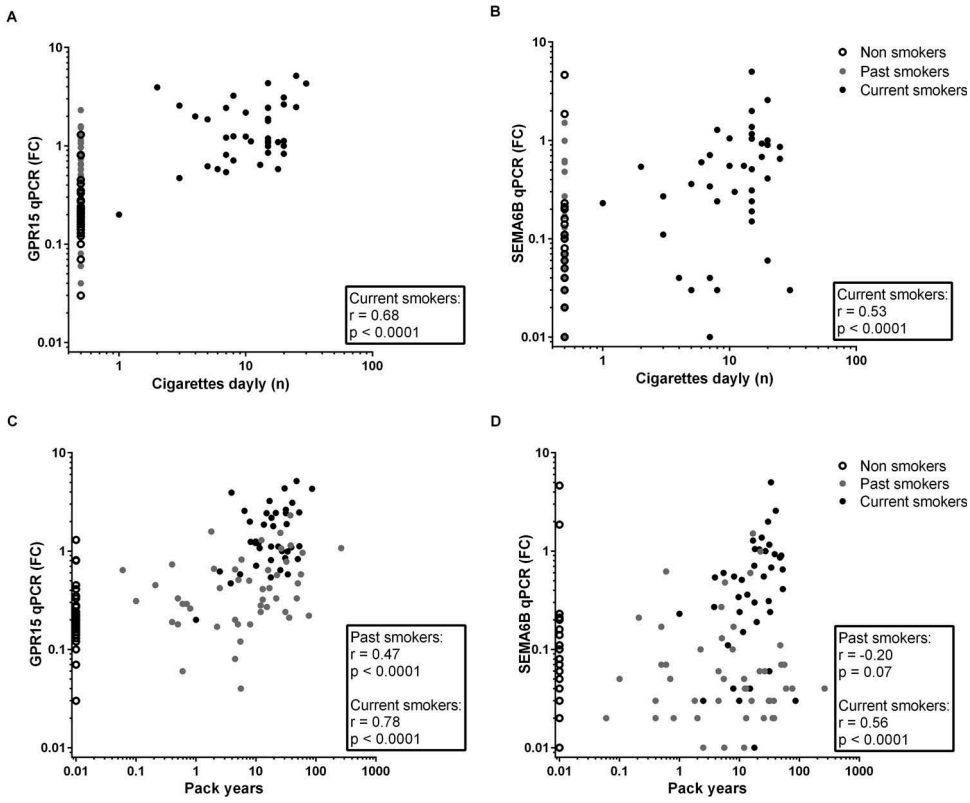

Figure 3. Correlations between smoking and GPR15 and SEMA6B expression. Graphs show the correlation between the relative gene expression levels of GPR15 and SEMA6B as measured by single qPCR assay and (a,b) the number of cigarettes smoked per day and (c,d) pack years (PY) of smoking (i.e. 1 PY = 20 cigarettes per day for one year). Patients were categorized as current smokers, discontinued smokers and never-smokers, the latter of which were given arbitrarily the lowest value for daily cigarettes or pack years. All FC values >1 were in (discontinued) smokers, except for three patients who declared to have never smoked. These three patients were contacted in retrospect to ask if there was any possibility of smoke inhalation at the time of the baseline visit. Two patients declared that they did not inhale smoke around treatment initiation, whereas one (with GPR15 FC = 0.8 and SEMA6B FC = 4.6) declared to be occasionally in a room with a smoker. A clear positive correlation for daily cigarettes and pack years with GPR15 and SEMA6B expression was seen in current smokers, and for pack years and GPR15 expression in discontinued smokers (all $p < 0.0001$).

cohort of consecutively included ADA treated patients. However, additional analyses revealed that the expression of both genes was strongly associated with smoking, which was unequally distributed between responders and nonresponders. It is know that smoking increases RA susceptibility and disease severity [6,33], and it has been frequently reported, although not consistently, that smoking negatively impacts treatment effects in RA [34–39]. The exact mechanism underlying these differences between smoking and non-smoking RA patients remains unclear [33,39]; however, it is possible that SEMA6B and GPR15 play a role in the mechanism(s) that leads to a reduced therapeutic effect of TNFi. This hypothesis is not unlikely, given that SEMA6B and GPR15 are strongly and dose dependently related to smoking and probably play a relevant role in the pathophysiology of RA. SEMA6B is a protein with pro-proliferative effects via its receptor Plexin-A4, which forms complexes with fibroblast growth factor-receptor 1 and 2 (FGFR1-2) and vascular endothelial growth factor-receptor 2 (VEGFR2), thereby inducing the signal transduction of these receptors [40]. VEGF and bFGF are the ligands of these receptors, and are known to have instrumental roles in RA: VEGF is a critical angiogenetic factor responsible for vascular proliferation and blood vessel invasion of the synovial lining membrane in RA, whereas acidic FGF (FGF-1) and basic FGF (FGF-2) have also been implicated in synovial hyperplasia and apoptosis resistance in adult RA [41]. Considering that SEMA6B has the ability to influence these important proinflammatory pathways in RA pathogenesis, its upregulation may explain the insensitivity to treatment. Additionally and in line with this study, upregulation of SEMA6B in smokers has been described before in a genome-wide study on PBMC mRNA ($p < 0.0001$) [42].

In turn, GPR15 or Brother of Bonzo (BOB) is expressed on the cell surface of monocytes and neutrophils, and is a chemo attractant for T cells [43,44]. GPR15 mRNA expression is increased in the synovium and in the peripheral blood leukocytes of RA patients compared to non-RA controls [44,45]; similarly, GPR15 protein is increased in synovial tissue macrophages and in circulating monocytes and neutrophils [44]. We found a clear dose–response effect of smoking on the expression of GPR15, which also showed a cumulative and prolonged effect, as the expression of GPR15 was still higher in past smokers. The strong positive correlation between smoking and a higher expression of GPR15 is also supported by other studies, showing DNA hypomethylation of the GPR15 promoter and subsequent increased mRNA expression in smokers [46–51]. Because of its biological role and the strong relation with smoking, GPR15 is a good candidate to explain the health hazards of smoking with regard to chronic inflammatory disease [48], and in particular the reduced therapeutic effect to TNFi.

Taken together, our results and those from other studies, it appears that smoking dose dependently increases GPR15 and SEMA6B expression in PBMCs and thereby directly or indirectly induces, respectively, T cell involvement and synovial vascular proliferation, leading to a more severe and treatment refractory RA. However, as the exact upstream and downstream pathways related to these genes are largely unknown, the possibility to investigate if the aberrant expression of SEMA6B and GPR15 reflects a larger dysregulation of the transcriptome using these data is limited.

Future functional *in vivo* studies should address whether the mRNA and protein expression of SEMA6B and GPR15 are modulated in the synovial tissue of a joint before and after (heavy) smoke exposure. The specific link between SEMA6B and GPR15 on T cell recruitment and synovial vascular proliferation, and subsequent RA severity and nonresponse to treatments, represent another but more challenging topic to investigate.

Interestingly, neither the TNFα nor LTα-pathway self, but rather general immunologic pathways at baseline were found to be related to a decreased response to ADA and ETN. This finding supports the hypothesis, already suggested by Smolen et al. [6], that TNFis (and non-TNFis) eventually mediate their efficacy by interfering with a common final pathway, namely proinflammatory cytokine production. Among the most altered pathways, we identified three interferon (IFN)-related pathways as being expressed significantly lower in nonresponders at baseline. In line with this observation are the results Wright and colleagues [20], who compared the expression of genes in neutrophils of RA patients (before start of treatment) and non-RA control patients by RNA-seq, and found the IFN-signaling pathway to be as most specific for RA. Subsequently, upregulation of an IFN-score (calculated on the expression levels of these and other IFN-related genes) predicted EULAR good response to TNFi (ADA, ETN and golimumab) in these RA patients, with an area under the ROC curve of 0.76. In our selection of 178 genes predictive of response, five genes overlapped with those included in the IFN-score: CCL3L3, THBS1, HLA-A, HLA-B, and HLA-C. However, the proposed IFN-score by Wright et al. calculated on the basis of our PBMC RNA-seq data-set did not predict the TNFi response (AUC = 0.55, 95% CI = 0.42–0.68) (Supplementary Figure 6). The predictive ability of the three genes selected in the second study of Wright et al. [21], was also weak when applied to our cohort of responders and nonresponders, even after refitting the regression coefficients for each gene (AUC = 0.56, 95% CI = 0.34–0.69). The differences between the results of our study and those produced by and Wright et al. could be related to the starting cell source (PBMC versus neutrophils), with neutrophils having been identified as biggest contributors in type I IFN signature seen in RA [52,53]. Therefore, the (partially) replicable results of Wright et al. remain of major interest, and the validity of their model based on the expression of only three genes on neutrophil RNA should be established in an independent cohort with fixed parameters.

5. Conclusions

Efforts of the present and previous studies have not provided gene transcripts that are able to independently and consistently predict response to TNFi treatment. However, we demonstrated that pathways altered in nonresponders at baseline are rather linked to general immune functions and cytokines than to TNFα and LTα specifically, indicating that all biological treatments target a common cytokine related pathway, as previously suggested by others. Furthermore, the identification of GPR15 and SEMA6B expression as markers of response and (substitute) dose-dependent indicators of smoking, opens new venues for the identification of the molecular mechanisms underlying TNFi refractory RA.

Key issues

- The identification of responders and nonresponders before initiation of therapy with tumor necrosis factor-alpha inhibitors (TNFi) would aid in making strategic treatment decisions and improve clinical outcomes in RA patients.
- To date, no biomarker to predict response to TNFi treatment has been shown to be robust enough to use in clinical practice.
- This study searches for gene transcripts predicting therapeutic response, by employing RNA-seq on baseline mRNA of peripheral blood mononuclear cells (PBMCs) in RA patients treated with two distinct TNFi treatments, namely adalimumab (ADA) and etanercept (ETN).
- This study did not identify new transcripts ready to use in clinical practice, yet GPR15 and SEMA6B were recognized as candidate explanatory markers for the reduced treatment success observed in RA smokers.

Acknowledgments

We thank Katja Coeleveld, Kim van der Wurff-Jacobs and Dorien van de Berg, for bio-banking, Karin A.L. Schrijvers-te-Brake, Annemiek Sloeserwij, Joke Nijdeken and Marieke Vianen who facilitated in the acquisition of data, Cornelis Bekker for his technical assistance, Aridaman Pandit for his aid in statistical analyses and the Society for Rheumatology Research Utrecht (SRU) for including patients. We are grateful to all of the patients who have participated in this study.

Contributing SRU hospitals and their contact persons: University Medical Center Utrecht, dr. J. Tekstra; Antonius Hospital Nieuwegein/Utrecht, dr. E.J. ter Borg; Diakonessen Hospital Utrecht, drs. Y. Schenk Meander Medical Center Amersfoort, dr. S. Linn-Rasker; Sint Maartenskliniek Woerden, dr. W.H. van der Laan; Hospital St. Jansdal Harderwijk, drs. D.G. Kuiper-Geertsma; Tergooi Hospital Hilversum, dr. M. Nabibux; Flevo Hospital Almere, dr. C.M. Verhoef.

Authors' contributions

Bart Cuppen collected clinical data, performed experiments, performed statistical analyses and drafted the manuscript, Marzia Rossato coordinated the experiments, aided in statistical analyses and interpretation of the results and drafted the manuscript, Ruth Fritsch-Stork included patients, coordinated the study and helped to draft the manuscript, Arno N. Concepcion performed qPCR experiments and helped to draft the manuscript, Suzanne Linn-Rasker included patients and helped to draft the manuscript, Jaap van Laar included patients and helped to draft the manuscript, Floris P.J.G. Lafeber conceived and coordinated the study and helped to draft the manuscript, Johannes W.J. Bijlsma conceived and coordinated the study, included patients and helped to draft the manuscript, and Timothy R.D.J. Radstake conceived the study and helped to draft the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by AbbVie, ZonMW: The Netherlands Association for Health Research and Development under Grant number 152001008, Foundation Friends of the University Medical Center Utrecht based on a private gift and unrestricted grants from Pfizer, BMS, UCB, Roche and Schering-Plough. Marzia Rossato was funded by IEF Marie Curie fellowship under Grant number 622811 (MicroSCAP). These funders provided financial support to this research but did not have any additional role in study design, data collection and analysis, or preparation of the manuscript.

Declaration of interest

JWJ Bijlsma received research grants (to his department) and consultancy fees from AbbVie, BMS, Crescendo, MSD, Mundipharma, Pfizer, Roche, Sun and UCB. TRD Radstake received grants and consultancy fees from AbbVie, BMS, Pfizer, Roche, Lilly and UCB. JM van Laar received grants from Roche, MSD, Eli Lilly and Pfizer. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or ofconsiderable interest (••) to readers.

1. Cuppen BV, Welsing PM, Sprengers JJ, et al. Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. Rheumatology (Oxford). 2016 May;55 (5):826–839. PubMed PMID: 26715775.
   •• **Besides the numerous reviews on this topic, this is the only systematic review on prediction of response in RA.**
2. Tracey D, Klareskog L, Sasso EH, et al. Tumor necrosis factor antagonist mechanisms of action: a comprehensive review. Pharmacol Ther. 2008 Feb;117(2):244–279. PubMed PMID: 18155297.
3. Kooloos WM, de Jong DJ, Huizinga TW, et al. Potential role of pharmacogenetics in anti-TNF treatment of rheumatoid arthritis and Crohn's disease. Drug Discov Today. 2007 Feb;12(3–4):125–131. PubMed PMID: 17275732.
4. Tansey MG, Szymkowski DE. The TNF superfamily in 2009: new pathways, new indications, and new drugs. Drug Discov Today. 2009 Dec;14(23–24):1082–1088. S1359-6446(09)00337-7 [pii].
5. Sedger LM, McDermott MF. TNF and TNF-receptors: from mediators of cell death and inflammation to therapeutic giants - past, present and future. Cytokine Growth Factor Rev. 2014 Aug;25(4):453–472. PubMed PMID: 25169849.
6. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet. 2016 May 3;388:2023–2038. PubMed PMID: 27156434.
7. Smolen JS, Aletaha D. Forget personalised medicine and focus on abating disease activity. Ann Rheum Dis. 2013 Jan;72(1):3–6. annrheumdis-2012-202361 [pii].
   •• **A critic view on the topic of personalized medicine in RA.**
8. Smith SL, Plant D, Eyre S, et al. The potential use of expression profiling: implications for predicting treatment response in rheumatoid arthritis. Ann Rheum Dis. 2013 Jul;72(7):1118–1124. PubMed PMID: 23486412.
9. Burska AN, Roget K, Blits M, et al. Gene expression analysis in RA: towards personalized medicine. Pharmacogenomics J. 2014 Apr;14 (2):93–106. PubMed PMID: 24589910; PubMed Central PMCID: PMCPMC3992869.
10. Toonen EJ, Gilissen C, Franke B, et al. Validation study of existing gene expression signatures for anti-TNF treatment in patients with rheumatoid arthritis. PLoS One. 2012;7(3):e33199.
    • **Well caried out validation study of several earlier published micro-array models.**
11. MacIsaac KD, Baumgartner R, Kang J, et al. Pre-treatment whole blood gene expression is associated with 14-week response assessed by dynamic contrast enhanced magnetic resonance imaging in infliximab-treated rheumatoid arthritis patients. PLoS One. 2014;9(12):e113937. PONE-D-14-28467 [pii].
12. Thomson TM, Lescarbeau RM, Drubin DA, et al. Blood-based identification of non-responders to anti-TNF therapy in rheumatoid arthritis. BMC Med Genomics. 2015;8:26.
13. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol. 2011;9:34. 1741-7007-9-34 [pii].
14. Sirbu A, Kerr G, Crane M, et al. RNA-Seq vs dual- and single-channel microarray data: sensitivity analysis for differential expression and clustering. PLoS One. 2012;7(12):e50986. PONE-D-11-19570 [pii].
15. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Brief Funct Genomic Proteomic. 2009;8(3):174–183. elp013.
16. Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008 Jul;5 (7):621–628. nmeth.1226 [pii].
17. Fu X, Fu N, Guo S, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics. 2009;10:161. 1471-2164-10-161 [pii].
18. 't Hoen PA, Ariyurek Y, Thygesen HH, et al. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. 2008 Dec;36(21):e141. gkn705[pii].
19. Marioni JC, Mason CE, Mane SM, et al. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008 Sep;18(9):1509–1517. gr.079558.108 [pii].
20. Wright HL, Thomas HB, Moots RJ, et al. Interferon gene expression signature in rheumatoid arthritis neutrophils correlates with a good response to TNFi therapy. Rheumatology (Oxford). 2015 Jan;54 (1):188–193. keu299 [pii].
21. Wright HL, Cox T, Moots RJ, et al. Neutrophil biomarkers predict response to therapy with tumor necrosis factor inhibitors in rheumatoid arthritis. J Leukoc Biol. 2017 Mar;101(3):785–795. PubMed PMID: 27733572.
22. van Gestel AM, Prevoo ML, van 't Hof MA, et al. Development and validation of the European league against rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American college of rheumatology and the world health organization/international league against rheumatism criteria. Arthritis Rheum. 1996;39(1):34-40.
23. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009 Aug 1;25(15):1966–1967. PubMed PMID: 19497933.
24. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166–169. PubMed PMID: 25260700; PubMed Central PMCID: PMCPMC4287950.
25. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106. PubMed PMID: 20979621; PubMed Central PMCID: PMCPMC3218662.
26. Mayeux R. Biomarkers: potential uses and limitations. NeuroRx. 2004 Apr;1(2):182–188. PubMed PMID: 15717018; PubMed Central PMCID: PMCPMC534923.
27. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001;25(4):402–408.
28. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005 Oct 25;102 (43):15545–15550. PubMed PMID: 16199517; PubMed Central PMCID: PMCPMC1239896.

29. Fabregat A, Sidiropoulos K, Garapati P, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2016 Jan 4;44(D1):D481–D7. PubMed PMID: 26656494; PubMed Central PMCID: PMCPMC4702931.
30. Chen J, Bardes EE, Aronow BJ, et al. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W305–W11. PubMed PMID: 19465376; PubMed Central PMCID: PMCPMC2703978.
31. Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D457–62. PubMed PMID: 26476454; PubMed Central PMCID: PMCPMC4702792.
32. Youden W. Index for rating diagnostic tests. Cancer. 1950 Jan;3 (1):32–35.
33. Chang K, Yang SM, Kim SH, et al. Smoking and rheumatoid arthritis. Int J Mol Sci. 2014 Dec 3;15(12):22279–22295. PubMed PMID: 25479074; PubMed Central PMCID: PMCPMC4284707.
34. Abhishek A, Butt S, Gadsby K, et al. Anti-TNF-(alpha) agents are less effective for the treatment of rheumatoid arthritis in current smokers. J Clin Rheumat. 2010;16(1):15–18.
35. Forslind K, Hafstrom I, Ahlmen M, et al. Sex: a major predictor of remission in early rheumatoid arthritis? Ann Rheum Dis. 2007 Jan;66(1):46–52. PubMed PMID: 17158139; PubMed Central PMCID: PMCPMC1798403.
36. Hyrich KL, Watson KD, Silman AJ, et al. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British society for rheumatology biologics register. Rheumatology (Oxford). 2006 Dec;45(12):1558–1565.
37. Saevarsdottir S, Wedren S, Seddighzadeh M, et al. Patients with early rheumatoid arthritis who smoke are less likely to respond to treatment with methotrexate and tumor necrosis factor inhibitors: observations from the epidemiological investigation of rheumatoid arthritis and the Swedish rheumatology register cohorts. Arthritis Rheum. 2011;63(1):26–36.
38. Mattey DL, Brownfield A, Dawes PT. Relationship between pack-year history of smoking and response to tumor necrosis factor antagonists in patients with rheumatoid arthritis. J Rheumatol. 2009 Jun;36(6):1180–1187. doi:10.3899/jrheum.081096
39. Cuppen BV, Jacobs JW, Ter Borg EJ, et al. Necessity of TNF-alpha inhibitor discontinuation in rheumatoid arthritis is predicted by smoking and number of previously used biological DMARDs. Clin Exp Rheumatol. 2017 Mar-Apr;35(2):221-228. PubMed PMID: 27749223
40. Kigel B, Rabinowicz N, Varshavsky A, et al. Plexin-A4 promotes tumor progression and tumor angiogenesis by enhancement of VEGF and bFGF signaling. Blood. 2011 Oct 13;118(15):4285–4296. PubMed PMID: 21832283.
41. Malemud CJ. Growth hormone, VEGF and FGF: involvement in rheumatoid arthritis. Clin Chim Acta. 2007 Jan;375(1–2):10–19. PubMed PMID: 16893535
42. Charlesworth JC, Curran JE, Johnson MP, et al. Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. BMC Med Genomics. 2010;3:29. PubMed PMID: 20633249; PubMed Central PMCID: PMCPMC2911391
43. Kim SV, Xiang WV, Kwak C, et al. GPR15-mediated homing controls immune homeostasis in the large intestine mucosa. Science. 2013 Jun 21;340(6139):1456–1459. PubMed PMID: 23661644; PubMed Central PMCID: PMCPMC3762262.
44. Cartwright A, Schmutz C, Askari A, et al. Orphan receptor GPR15/BOB is up-regulated in rheumatoid arthritis. Cytokine. 2014 Jun;67(2):53–59. PubMed PMID: 24725539; PubMed Central PMCID: PMCPMC3996549.
45. Schmutz C, Hulme A, Burman A, et al. Chemokine receptors in the rheumatoid synovium: upregulation of CXCR5. Arthritis Res Ther. 2005;7(2):R217–R29. PubMed PMID: 15743468; PubMed Central PMCID: PMCPMC1065316.
46. Dogan MV, Xiang J, Beach SR, et al. Ethnicity and smoking-associated DNA methylation changes at HIV co-receptor GPR15. Front Psychiatry. 2015;6:132. PubMed PMID: 26441693; PubMed Central PMCID: PMCPMC4585036.
47. Gao X, Jia M, Zhang Y, et al. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics. 2015;7:113. PubMed PMID: 26478754; PubMed Central PMCID: PMCPMC4609112.
48. Koks G, Uudelepp ML, Limbach M, et al. Smoking-induced expression of the GPR15 gene indicates its potential role in chronic inflammatory pathologies. Am J Pathol. 2015 Nov;185(11):2898–2906. PubMed PMID: 26348578.
49. Tsaprouni LG, Yang TP, Bell J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics. 2014 Oct;9(10):1382–1396. PubMed PMID: 25424692; PubMed Central PMCID: PMCPMC4623553.
50. Wan ES, Qiu W, Baccarelli A, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Hum Mol Genet. 2012 Jul 1;21 (13):3073–3082. PubMed PMID: 22492999; PubMed Central PMCID: PMCPMC3373248.
51. Breitling LP, Yang R, Korn B, et al. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet. 2011 Apr 8;88(4):450–457. PubMed PMID: 21457905; PubMed Central PMCID: PMCPMC3071918.
52. de Jong TD, Lubbers J, Turk S, et al. The type I interferon signature in leukocyte subsets from peripheral blood of patients with early arthritis: a major contribution by granulocytes. Arthritis Res Ther. 2016;18:165. PubMed PMID: 27411379; PubMed Central PMCID: PMCPMC4944477.
53. Wright HL, Makki FA, Moots RJ, et al. Low-density granulocytes: functionally distinct, immature neutrophils in rheumatoid arthritis with altered properties and defective TNF signalling. J Leukoc Biol. 2017 Feb;101(2):599–611. PubMed PMID: 27601627.