Omics data integration identifies ELOVL7 and MMD gene regions as novel loci for adalimumab response in patients with Crohn’s disease

Mario Gorenjak1, Mateja Zupin1, Gregor Jezernik1, Pavel Skok1-2 & Uroš Potočnik1,3*

Response to anti-TNF therapy is of pivotal importance in patients with Crohn’s disease (CD). Here we integrated our and previously reported PBMC derived transcriptomic and genomic data for identification of biomarkers for discrimination between responders and non-responders to anti-TNF therapy. CD patients, who were naïve with respect to the treatment with biologicals, were enrolled in the study. DNA and RNA were extracted from peripheral blood mononuclear cells. RNA-seq was performed using BGISEQ-500. Genotyping was performed using Infinium Global Screening Array. Association regressions were carried out with 12 week response to adalimumab as an outcome variable. RNA-seq analysis confirmed 7 out of 65 previously suggested genes involved in anti-TNF response. Subsequently, analysis of single nucleotide variants in regions of confirmed genes identified 5 variants near MMD and two in ELOVL7 intronic regions associated with treatment response to anti-TNF. Functional analysis has shown that rs1465352, rs4422035 and rs78620886 are listed at H3K9ac_Pro histone modification epigenetic mark. The present study confirmed MMD and ELOVL7 involvement in anti-TNF response and revealed that the regulation of MMD and ELOVL7 gene regions in ADA response may be a part of a complex interplay extending from genetic to epigenetic and to transcriptomic level.

Inflammatory bowel disease (IBD) includes both Crohn’s disease (CD) and ulcerative colitis (UC) and results from improper inflammatory response to microbiota in a host with genetic susceptibility1. With tumor necrosis factor alpha (TNF-α) being the key cytokine in inflammation, many anti-TNF therapies emerged and dominated in the treatment of CD. On one hand, disease management was significantly improved with anti-TNF agents, such as adalimumab (ADA) and infliximab (IFX), but on the other hand, 30% of patients do not respond to the anti-TNF treatment and of those who initially benefit, up to 40% lose their response later. Moreover, unnecessary continuation of the anti-TNF therapy may severely impede patients’ quality of life and impose adverse effect risk without clinical justification with emerging of an additional challenge of loss of response management. Thus, identification of anti-TNF therapy response predictors is of pivotal importance and is posing an additionally challenge in the last decade. Prediction of response is also needed as we now have other therapies with different targets available. Several studies have focused on genetic background and gene expression as potential biomarkers for anti-TNF treatment response. Currently the genomics markers of anti-TNF response in CD are not reaching significant thresholds and practically no reproducibility between genetic and expression markers exists, as identified by the first integration of genomic and expression data prior to anti-TNF treatment in a systematic review. Prediction profiles have been mostly identified using epithelial gene expression data for which colonoscopy and biopsy sampling are needed. However, obtaining colon biopsies is a highly invasive procedure. Studies which would focus on non-invasive genetic expression biomarkers, such as gene
expression in peripheral blood mononuclear cells (PBMC) for anti-TNF response identification are scarce and non-homogenous in terms of anti-TNF therapy and disease classification.

Therefore, the aim of the present study is to integrate transcriptomic and genomic data obtained from blood of CD patients, who were naive with respect to the treatment with biologicals, using our and previously reported peripheral blood derived transcriptomic data and using a modified contemporary combination of omics approach as described previously. For that, we integrated our omics data based on 46 genes identified in CD and rheumatoid arthritis (RA) cohorts, which appeared the most significant to differentiate responders from non-responders to anti-TNF therapy (Table 1). We also included four genes that correlated with changes in endoscopic activity in UC patients with anti-TNF therapy and three differentially expressed genes between responding and non-responding IBD patients to anti-TNF therapy as identified by a cell-centered meta-analysis (Table 1). Additionally we included 12 independently confirmed and previously integrated genomic and expression (RNA and protein) markers of anti-TNF therapy response in CD patients (Table 1). The investigation gene panel built for the present study was altogether based on 65 previously suggested candidate genes in 2126 individuals with IBD and 34 individuals with RA, and treated with anti-TNF therapy.

Results

RNA-seq analysis. Using our RNA sequencing data we performed an expression study using previously suggested candidate genes discriminating anti-TNF response in CD patients. For that, we used two different approaches. In first approach, the analysis was adjusted to deconvoluted fractions of PBMCs and has identified 11 differentially expressed genes out of which MMD, ELVOL7 and BMP6 remained significantly down-regulated in responders relative to non-responders after correction (Table 2, Fig. 1). In the second approach, analysis was carried out unadjusted in order to obtain PBMC panel. The analysis has confirmed 10 differentially expressed genes out of which 4 (GPR84, EPSTI1, IFI6, MX1) remained significantly up-regulated after correction for multiple comparison in responders relative to non-responders (Table 2, Fig. 1).

Integration to genomics. Using an integrative transcriptomic-genomic approach we analyzed single nucleotide variants ranging ± 100 kb from previously identified differentially expressed genes in 84 patients with CD. Association analysis has identified 5 statistically significant variants near gene MMD (Fig. 2A) and 2 statistically significant variants in ELVOL7 gene region (Fig. 2B). For single nucleotide polymorphisms (SNPs) rs1465352, rs4422035, rs9892429, rs9893820 and rs11656799 near gene MMD the results have shown significant higher alternative allele frequency in non-responders in comparison to responders (Table 3). For the most significant rs4422035 the analysis has shown higher alternative allele frequencies in non-responders (Allele C: 0.69) in comparison to responders (Allele C: 0.45) (p = 0.018; OR 0.22; CI 0.09–0.55). Additional LD assessment has shown that that rs4422035 and rs1465352 are in linkage disequilibrium (D': 1.0; r2: 0.87; p < 10^-4). Identified rs9892429, rs9893820 and rs11656799 have shown higher linkage disequilibrium with rs1465352 (D': 1.0; r2: 0.96; p < 10^-4) as in comparison to most significant rs4422035, which is also evident from the Table 3. Similar effect was observed also for variants identified in ELVOL7 gene region. Both identified variants have shown the same trend of alternative allele frequencies (Table 3). For the most significant rs78620886 the analysis has shown higher alternative allele frequencies in non-responders (Allele A: 0.26) in comparison to responders (Allele A: 0.03) (p = 0.021; OR 0.05; CI 0.01–0.27). Additional LD analysis has also shown that rs78620886 is in linkage disequilibrium with rs9291695 (D': 1.0; r2: 0.95; p < 10^-4).

Table 1. Previously identified anti-TNF response genes in PBMC and integrated genomic markers. CD Crohn’s disease, RA rheumatoid arthritis, IBD inflammatory bowel disease, IBDu inflammatory bowel disease unclassified, UC ulcerative colitis.
Mendelian randomization analysis. An additional Inverse-variance random-effect Mendelian Randomization (MR) analysis was performed. First, variants ranging ± 100 kb from MMD and ELOVL7 were included into conditional regressions in order to detect independent signals conditioned to the identified most significant SNPs. Conditioning results have identified statistically significant rs12943795 besides rs4422035 for MMD, and rs1465352 and rs11656766 significant eQTLs with \( \log_2 \text{FC} \) responders relative to non-responders, 95 CI 95% confidence interval, \( L \) lower, \( U \) upper, Bold text indicates statistically significant genes after adjustment.

Validation of genes using RT-qPCR. In order to assess and validate the findings of RNA-seq analysis and subsequent genomic variants, RT-qPCR was performed. First, reference genes were checked for stability between responders and non-responders. Both genes, \( ACTB \) (\( p = 0.768 \)) and \( B2M \) (\( p = 0.782 \)) proved to be stably expressed in both groups. Subsequently gene expression of MMD and ELOVL7 was measured in PBMCs. Both genes were down-regulated in responders, but statistical significance was observed only at \( ELOVL7 \), which was a 1.43-fold down-regulated in responders relative to non-responders (\( p = 0.016; \beta = 0.535; \text{SE} 0.214 \)) (Fig. 4B). MMD was 1.10-fold down-regulated in responders relative to non-responders (\( p = 0.498; \beta = 0.141; \text{SE} 0.656 \)) (Fig. 4A). Additionally eQTLs in PBMCs were assessed for SNPs where chromatin state assignment was listed. No statistically significant eQTLs were observed for rs1465352 (\( p = 0.564; \beta = 0.002; \text{SE} 0.004 \)) (Fig. 4C), rs4422035 (\( p = 0.752; \beta = 0.001; \text{SE} 0.004 \)) (Fig. 4D) and rs78620886 (\( p = 0.448; \beta = 0.003; \text{SE} 0.004 \)) (Fig. 4E). In order to check the interplay of MMD and \( ELOVL7 \) a correlation analysis of expression was performed. Correlation analysis revealed a strong correlation between genes (\( p = 2.57 \times 10^{-15}; p = 0.799 \)) (Fig. 4F).

Table 2. Confirmation of candidate genes previously associated with anti-TNF response using RNA-seq analysis. \( \log_2 \text{Log}\text{-FC} \) Responders relative to non-responders, 95 CI 95% confidence interval, \( L \) lower, \( U \) upper, Bold text indicates statistically significant genes after adjustment.
Figure 1. RNA-seq volcano and smear plots of RNA-seq analyses. RNA-seq analysis. The plots were constructed using ggplot2 R package\textsuperscript{23}.

Figure 2. Regional Manhattan plots for MMD and ELOVL7 ± 100 kb gene region. (A) MMD; (B) ELOVL7. Regional Manhattan plots were constructed using LocusZoom\textsuperscript{24}. 
Table 3. SNPs significantly associated with anti-TNF response and located near genes differentially expressed in responders and non-responders according to our RNA-seq results. OR calculated for alternative allele and response, 95 CI 95% confidence interval, L lower, U upper, AF R alternative allele frequency responders, AF NR alternative allele frequency non-responders.

| SNP     | Chr | Position | Alleles | OR   | 95 CI L | 95 CI U | AF R | AF NR | p-value | Adj. p-value |
|---------|-----|----------|---------|------|---------|---------|------|-------|---------|--------------|
| rs1465352 | 17  | 53435734 | T/C     | 0.25 | 0.11    | 0.60    | 0.42 | 0.66  | 0.0017  | 0.0261       |
| rs4422035 | 17  | 53436116 | A/C     | 0.22 | 0.09    | 0.55    | 0.45 | 0.69  | 0.0012  | 0.0184       |
| rs9892429 | 17  | 53439439 | G/T     | 0.25 | 0.11    | 0.60    | 0.42 | 0.66  | 0.0017  | 0.0261       |
| rs9893820 | 17  | 53440395 | T/C     | 0.25 | 0.11    | 0.60    | 0.42 | 0.66  | 0.0017  | 0.0261       |
| rs11656766 | 5   | 60055644 | T/A     | 0.07 | 0.02    | 0.32    | 0.04 | 0.26  | 0.0005  | 0.0353       |
| rs78620886 | 5   | 60087336 | G/A     | 0.05 | 0.01    | 0.27    | 0.03 | 0.26  | 0.0003  | 0.0212       |

Table 4. Functional analysis of significant SNPs. NES Normalized effect size, NA not applicable, Bold text indicates statistically significant eQTLs.

| SNP                      | Epigenetics | Cells (Gene Tissue) | Gene  | Tissue       | p-value | NES |
|--------------------------|-------------|---------------------|-------|--------------|---------|-----|
| rs1465352 (near gene 3’) | H3K9ac_Pro  | Monocytes (RO 01746) | MMD   | EBV Lymphocytes | 0.2     | 0.093|
|                          |             |                     |       | Whole blood   | 0.4     | −0.018|
|                          |             |                     |       | Monocytes     | 0.028   | 0.003|
| rs4422035 (near gene 3’) | H3K9ac_Pro  | Monocytes (RO 01746) | MMD   | EBV Lymphocytes | 0.7     | 0.025|
|                          |             |                     |       | Whole blood   | 0.8     | −0.005|
|                          |             |                     |       | Monocytes     | NA      | NA  |
| rs9892429 (near gene 3’) | NA          | NA                  | MMD   | EBV Lymphocytes | 0.3     | 0.071|
|                          |             |                     |       | Whole blood   | 0.8     | −0.006|
|                          |             |                     |       | Monocytes     | NA      | NA  |
| rs9893820 (near gene 3’) | NA          | NA                  | MMD   | EBV Lymphocytes | 0.3     | 0.071|
|                          |             |                     |       | Whole blood   | 0.8     | −0.005|
|                          |             |                     |       | Monocytes     | NA      | NA  |
| rs11656766 (near gene 3’) | NA         | NA                  | MMD   | EBV Lymphocytes | 0.3     | 0.071|
|                          |             |                     |       | Whole blood   | 0.8     | −0.005|
|                          |             |                     |       | Monocytes     | 0.028   | 0.003|
| rs9291695 (intron)       | NA          | NA                  | ELOVL7| EBV Lymphocytes | 0.01    | 0.454|
|                          |             |                     |       | Whole blood   | 0.2     | 0.044|
| rs78620886 (intron)      | H3K9ac_Pro  | Th0 (PBMC)          | ELOVL7| EBV Lymphocytes | 0.02    | 0.553|
|                          |             |                     |       | Whole blood   | 0.1     | 0.062|

Figure 3. ROC curve of random forest classification of rs1465352, rs4422035 and rs78620886 genotypes for adalimumab response prediction. The ROC curve plot was constructed using ggplot2 R package23.
Discussion

Anti-TNF therapy is of pivotal importance in patients with CD and genomic biomarkers that would tailor personalized treatment are currently needed, in particular as new biologicals for CD treatment have recently emerged. In the recent systematic review of anti-TNF genomic biomarkers in CD, it was noted that almost no candidate gene could be confirmed in an independent association study and no overlap was detected among genes identified with association and gene expression studies. In the present study, we have adopted a new approach for the integration of PBMCs derived transcriptomic and genomic data. To our knowledge, this is the first time deconvolution was applied in order to account for PBMC composition in anti-TNF response transcriptomics, which identified two novel genetic loci on chromosomes 17 near gene MMD and 5 in gene ELOVL7, associated with anti-TNF response in CD patients.

For that, we included patients with CD, who were naïve with respect to the treatment with biologicals and who were subsequently treated with ADA and investigated PBMC derived transcriptomic-genomic data integration with targeting on previously identified gene expression panels for differentiation of responders from non-responders to anti-TNF therapy. Additionally, we investigated independently confirmed and previously integrated genomic and expression (RNA and protein) markers of anti-TNF therapy response in CD patients. Thus, the targeted expression panel consisted altogether of 65 genes, which were previously identified in 2126
individuals with IBD and 34 individuals with RA as potential predictors for different anti-TNF therapy (ADA, IFX, golimumab) responses in peripheral blood of CD, UC, unclassified IBD or RA patients.

First, using RNA-seq and using 3 responders and 3 non-responders to ADA we confirmed differentially expressed genes from previously identified gene panel. The model fitting for estimation of differentially expressed genes was carried out in two different ways—adjusted to deconvolution results and not adjusted in order to obtain PBMC expression picture. The results have shown that in deconvoluted analysis MMD, ELOVL7 and BMP6 remained significantly differentially expressed after correction for multiple comparisons. All three genes were down-regulated in responders relative to non-responders (up-regulated in non-responders relative to responders). Using unadjusted model fitting we observed the confirmation of GPR84, EPSTI1, IFI6, MX1, which remained significantly differentially expressed after the application of multiple comparison correction. All four were found to be up-regulated in responders relative to non-responders (down-regulated in non-responders relative to responders). Subsequently, confirmed genes were further analyzed using integration with genomic data obtained from 84 CD patients, who were naïve with respect to the treatment with biologics, using a genotype microarray and using an association analysis with 12 week response as an outcome variable. Imputed high quality single nucleotide variants were analyzed using a margin of ± 100 kb from confirmed genes in RNA-seq analysis in order to include variants in regulatory gene regions. The results have identified two significant signals, one near gene MMD and one in ELOVL7 gene region after application of Bonferroni-like correction using empirical autocorrelation estimation, but no statistically significant signals were observed in other 5 genes. Both MMD and ELOVL7 were identified in deconvoluted approach. Thus, obtained results are pointing to increased robustness of analysis when using deconvolution of PBMCs. As previously shown using deconvolution-meta-analysis paradigm, adjustment to estimation of cell proportions can capture differential expression otherwise masked due to cellular composition variation. Expression of MMD (Monocyte to macrophage differentiation-associated) gene can be detected in all tissues and its up-regulation initiated the production of TNF-α and nitrous oxide (NO). TNF was previously identified among key players of the transcription, leading to an induction of fatty-acid elongation, which can elucidate the interplay of these two genes and serves as additional supporting evidence of monocyte to macrophage differentiation and changes in lipid-related transcriptome. In addition, based on our results functional studies were carried out to confirm the strong correlation between MMD and ELOVL7, which further supports the interplay between these two genes and serves as additional supporting evidence of monocyte to macrophage differentiation and changes in lipid-related transcriptome. Nevertheless, we believe that using trimmed mean of M values method (TMM) normalization implemented in edgeR and the variance modeling at the observational level (VOOM) method implemented in the limma package before applying empirical Bayes over a linear model, made our analysis stringent enough to provide reliable results. Therefore, we hypothesize that locations of rs1465352, rs4422035 and rs78620886 might play a significant role in histone modifications in terms of H3K9 residue acetylation and up-regulation of gene expression. Using ENSEMBL and GTExportal database statistically significant eQTLs with MMD and ELOVL7 were also listed. Additionally, the implication of aforementioned SNPs in ADA response was also confirmed using random forest machine learning algorithm and ROC analysis, which has additionally confirmed the association of these three SNPs with ADA response in patients with CD. In order to fully address the implication of both genes in ADA response, a Mendelian randomization integrating eQTL and association analysis data was performed. Gene ELOVL7 was subsequently again confirmed as a significant candidate, however no analysis was performed for MMD due to unavailability of eQTL data for selected variants in corresponding tissues/cells in used databases. Furthermore, monocyte to macrophage differentiation was found to go along with profound changes in lipid-related transcriptome, leading to an induction of fatty-acid elongation, which can elucidate the interplay of MMD and ELOVL7 up-regulation in ADA non-responders. These findings are further supported by previously shown genetically influenced abnormalities in fatty acid profiles in IBD. The main limitation of the present study is the limited sample size in transcriptome analysis. Nevertheless, we believe that using trimmed mean of M values method (TMM) normalization implemented in edgeR and the variance modeling at the observational level (VOOM) method implemented in the limma package before applying empirical Bayes over a linear model, made our analysis stringent enough to provide reliable results. The aforementioned concern was also addressed using RT-qPCR analysis for estimation of MMD and ELOVL7 gene expression in PBMCs in a larger cohort in order to validate the findings. However, MMD and ELOVL7 were identified using an approach with deconvolution of PBMCs, thus differences between responders and non-responders may not have been fully revealed using RT-qPCR. Obtained results have shown that both, MMD and ELOVL7 tend to be up-regulated in non-responders relative to responders, but only ELOVL7 reached statistical significance, which is consistent with RNA-seq findings. Furthermore, for SNPs where chromatin state assignment was listed, no trend of association with MMD and ELOVL7 was observed in PBMCs, which is also consistent with reported eQTLs in whole blood. Moreover, using RT-qPCR we observed a strong correlation between MMD and ELOVL7, which further supports the interplay between these two genes and serves as additional supporting evidence of monocyte to macrophage differentiation and changes in lipid-related transcriptome. In addition, based on our results functional studies.
In summary, the present study integrated transcriptomic and genomic data obtained from CD patients’ PBMCs and identified two novel genetic loci, on chromosomes 17 and 5, associated with anti TNF response in CD patients and suggested MMD and ELOVL7 as best candidates for involvement in anti-TNF response. Moreover, the present study revealed that the implication and regulation of MMD and ELOVL7 gene regions in ADA response may be a part of a complex interplay extending from genetic to epigenetic and to transcriptomic level. This knowledge could be further translated into new clinical non-invasive baseline biomarkers for adalimumab response in patients with CD.

Methods
All methods were carried out in accordance with relevant guidelines and regulations.

Enrolled subjects. We included 84 Slovenian patients with CD who fulfilled the criteria for anti-TNF-α drug ADA (Humira, Abbott Laboratories, IL, USA), and who were naïve with respect to the treatment with biologicals. All enrolled subjects were of Slovenian (Caucasian—Central Europe) ethnicity. Baseline demographics of the patients are shown in Table 5. Inclusion criteria were refractoriness to corticosteroids, adverse effects to corticosteroids and lost response or intolerance to IFX as described previously. Exclusion criteria were defined as presence of CD complications as stenosis, abscesses, total colectomy, history of murine proteins allergy, active tuberculosis or a serious infection in the last 3 months, pregnancy or lactation and malignancy. First, a loading dose of 160 mg of ADA was administered, followed by 80 mg after 14 days and maintenance dose of 40 mg every other week. Azathioprine, 5-aminosalicylates, corticosteroids or antibiotic concomitant treatment was allowed if the dosage was stable in the last 3 months. The present study was evaluated and approved by Slovenian National Committee for Medical Ethics (KME 80/10/07, 21/p/12/07). Written informed consent was obtained from all enrolled subjects. In case of minors, written informed consent was obtained from parents and/or legal guardians. Response to treatment was measured using IBD questionnaire (IBDQ) after 12 weeks of treatment. Response was defined as an increase in score > 22 points in comparison of pre-treatment score or total score > 170 points.

Extraction of nucleic acids. DNA and RNA were extracted from peripheral blood mononuclear cells (PBMC) using TRI-reagent (Merck, Darmstadt, Germany) according to manufacturer’s instructions. Purity and concentration of nucleic acids was determined using Synergy 2 spectrophotometer (Biotek, Winoski, VT, USA). Integrity of RNA was checked using agarose gel electrophoresis and 2100 Bioanalyzer Instrument (Agilent, Santa Clara, CA, USA) with RNA 6000 Nanochip.

RNA-seq analysis. For RNA-seq analysis, 3 responders and 3 non-responders were selected. These patients were selected for RNA-seq based on available data for continuously measured treatment response, which was recorded at weeks 4, 12, 20 and 30. The selected 6 patients qualified as 3 were consistent responders and 3 consistent non-responders during all check point measurements. RNA RIN number was > 9.0 and 28S/18S ratio between 1.5 and 2.0 for all 6 samples. Both, lncRNA and mRNA 100 bp paired-end libraries were constructed using MGIEasy rRNA Depletion Kit (MGITech, Shenzen, China) and MGIEasy RNA Library Prep Set (MGITech) and using the BGISEQ-500 instrument (BGI, Hong Kong) at BGI facilities (BGI). Data analysis was performed using R 4.0.2 environment (R Core Team 2020, Vienna, Austria). Raw .fastq files were first assessed for quality using FastQC 0.11.9 software. Technical sequences and sequencing adapters were trimmed using Trimmomatic 0.39 tool. Paired-end reads were aligned to the hg19 reference genome using Rsubread 2.2.4 R package. Mapped reads were counted and assigned to genomic features using featureCounts with requirement of both ends to be mapped. Counts per million (CPMs) were calculated using edger 3.30.3 R package. Low expressed genes were filtered out based on CPMs corresponding to reads of 10 and retained genes were normalized using the trimmed mean of M values method (TMM). Subsequently, mean–variance modelling at the observational level transformation (VOOM) was applied. Differential expression of responders rela-

|                | Responders | Non-responders | p-value |
|----------------|------------|----------------|---------|
| Gender (M/F)   | 22/33      | 11/18          | 1.000*  |
| Age at Dg. (years) | 27.2 ± 12.3 | 29.2 ± 12.0 | 0.366** |
| Smoking (Y/N)  | 31/24      | 18/11          | 0.649*  |
| CRP (mg/L)     | 155.0 ± 28.9 | 145.3 ± 27.0 | 0.090** |
| Leukocytes (10⁹/L) | 8.0 ± 2.8   | 8.5 ± 4.3     | 0.989** |
| Hemoglobin (g/L) | 129.3 ± 21.4 | 120.3 ± 21.3 | 0.264** |
| CRP (mg/L)     | 21.4 ± 24.7 | 12.8 ± 11.5   | 0.531** |

Table 5. Baseline demographics of included patients. *Fisher exact test; **Mann–Whitney U-Test.
were done using HaploReg v4.157, GTExPortal34 and ENSEMBL33. Regional Manhattan plots were constructed of different white cell subtypes in PBMCs49. Using raw counts, transcripts per million (TPM) were calculated, abundances of member cell types in a mixed cell population, using gene expression data in terms of proportions peripheral blood mononuclear cell composition data. Deconvolution was used to obtain an estimation of the zero values were removed and data was deconvoluted using CIBERSORT19 and LM2249 signature matrix. Sample composition in terms of proportions of lymphocytes, monocytes/macrophages and neutrophils was used as a covariate in the model. Zero estimated cell proportions were excluded from analyses. In the second approach, the linear model was left unadjusted in order to obtain PBMC gene profile picture. Differential expression was considered for genes with adjusted p-value < 0.05.

Association analysis. Samples obtained from 84 enrolled individuals were genotyped using Infinium Global Screening Array (GSA_24v1) (Illumina, San Diego, California, USA). Quality control of the genotyped data was performed as described previously50. Further steps of analysis were based on an integrative transcriptomic-genomic approach19. Subsequently, imputation was performed using Michigan Imputation Server Minimac3 genotype imputation algorithm and using HRC r1.1 2016 reference panel and SHAPEIT v2.r790 phasing51. Association between responders at week 12 vs. non-responders at week 12 was tested using binary logistic Wald test implemented in EPACTS 3.2.652. Association analysis was adjusted to age at diagnosis, sex, azathioprine use, use of aminosalicylates, use of corticosteroids and first four principal components, which were calculated using PLINK 1.9 software53,54. Appropriate number of principal components was determined beforehand using gap v1.2.2 R package. Retained were only variants with imputation score (Rsq) > 0.3. Variants ± 100 kb from previously identified differentially expressed genes were further analyzed. Bonferroni-like correction was applied based on empirical autocorrelation determined using coda R package55. Statistically significant signal was considered for SNPs with adjusted p-value < 0.05. Linkage disequilibrium was calculated using LDlink software56.

In silico functional analysis, prediction value estimation and visualization. Functional analyses were done using HaploReg v4.157, GTExPortal34 and ENSEMBL33. Regional Manhattan plots were constructed using LocusZoom57. At SNPs, where chromatin state assignment was listed, predictive value of genotypes was further assessed using randomForest 4.6-14 R package58 and Receiver Operating Characteristics analysis using pROC R package59. All other plots were constructed using ggplot2 R package60.

Mendelian randomization analysis. Data was further analyzed using Mendelian randomization integration of eQTL and association analysis data. Previously retained variants ± 100 kb from MMD and ELOVL7 gene regions were further analyzed using GCTA-COJO conditional analysis60. GCTA-COJO conditional analysis was performed conditional on most significant SNP identified in previous association analysis. Reference data-set for GCTA-COJO was built based on previously imputed data. Variants with Rsq < 0.3 were removed from the reference dataset prior to analysis. Condition SNP and SNPs, which remained significant after conditioning, were included in subsequent Mendelian randomization analysis. Mendelian randomization was performed using MendelianRandomization v0.5.0 R package61 and inverse-variance weighted random-effect method60. Variant-gene association (eQTL) obtained from GTExportal34 and ENSEMBL33 in corresponding tissues/cells was considered as exposure and variant-outcome association as outcome for MR analysis. If no eQTL data was available in corresponding tissues/cells, the variant was excluded from MR analysis.

Validation of genes using RT-qPCR. A total of 1 µg of mRNA with RIN > 8 was obtained from 55 (21 non-responders and 34 responders) out of 84 enrolled individuals and was transcribed into cDNA using high capacity cDNA reverse transcription kit (Thermo Fisher, Waltham, MA, USA). mRNA nucleotide sequences of target genes MMD and ELOVL7 were retrieved from NCBI Nucleotide database (www.ncbi.nlm.nih.gov/nuccore/). Primers for reference genes ACTB and B2M were obtained from previous study62. Isomar non-specific primers were hand-picked and designed using IDT OligoAnalyzer Tool (eu.idtdna.com/calc/analyzer). Primer sequences and accession numbers are summarized in Table 6. All primers were synthesized by Sigma (Merck, Darmstadt, Germany). Reverse transcription quantitative polymerase chain reaction (RT-qPCR) gene expression assay was carried out using Lightcycler 480 SYBR Green I Master Mix and Lightcycler 480 real time thermocycler (Roche, Basel, Switzerland) according to manufacturer’s instructions. 2 µL of 20-fold diluted cDNA (2.5 ng/µL) was used as a template. Primer efficiency was > 90% for all primer pairs. Melting curves for each

| Gene      | Accession | Fw Primer 5'-3' | Rv Primer 5'-3' |
|-----------|-----------|-----------------|-----------------|
| MMD       | NM_012329.3 | TTAGTTTATCTGCGCTAGTGA | TGAAGTCCATGGGTGTGTT |
| ELOVL7    | NM_024930.3 | TTTGAAGTCCATGGGTGTGTT | TGAAGTCCATGGGTGTGTT |
| ACTB      | NM_001010.3 | CATGAGGCAGGCGAGGGTATCA | TAGCACGCCGTGATAGCAAC |
| B2M       | NM_004048.2 | TTTCTGGCCCTGGAGGCTATC | TCGAGAAATTGTGACCTTCCATC |

Table 6. Primer sequences and accession numbers of target and reference genes.
sample were analyzed after each run in order to confirm specificity of amplification. Raw Ct values were obtained from three run-independent technical replicates for each sample. Normalization of raw data was performed using geometric averaging of reference genes and relative expression was calculated using $2^{-\Delta\Delta Ct}$ method. Stability of gene expression was assessed using Spearman’s rank correlation. Stability of reference genes was statistically assessed using relative $2^{-\Delta Ct}$ calculation and Mann–Whitney U-Test. qRT-PLs based on the study data were estimated for SNPs where chromatin state assignment was listed. Estimation was performed using linear regression. Violation of linear regression assumption was considered if variance inflation factor > 10 and condition index > 30. Correlation between gene expressions was assessed using Spearman’s rank correlation.

Received: 14 September 2020; Accepted: 22 February 2021
Published online: 09 March 2021

References

1. Abraham, C. & Cho, J. H. Inflammatory bowel disease. N. Engl. J. Med. 361, 2066–2078 (2009).
2. Danese, S., Fiorino, G. & Reinisch, W. Review article: Causative factors and the clinical management of patients with Crohn’s disease who lose response to anti-TNF-alpha therapy. Aliment Pharmacol. Ther. 34, 1–10 (2011).
3. Van Deventer, S. J. Tumour necrosis factor and Crohn’s disease. Gut 40, 443–448 (1997).
4. Colombel, J. F. et al. Adalimumab for maintenance of clinical response and remission in patients with Crohn’s disease: The CHARM trial. Gastroenterology 132, 52–65 (2007).
5. Hanauer, S. B. et al. Maintenance infliximab for Crohn’s disease: The ACCENT I randomised trial. Lancet 359, 1541–1549 (2002).
6. Sandborn, W. J. et al. Adalimumab for maintenance treatment of Crohn’s disease: Results of the CLASSIC II trial. Gut 56, 1232–1239 (2007).
7. Sands, B. E. et al. Infliximab maintenance therapy for fistulating Crohn’s disease. N. Engl. J. Med. 350, 876–885 (2004).
8. Ben-Horin, S., Kopylov, U. & Chowers, Y. Optimizing anti-TNF treatments in inflammatory bowel disease. Autoimmun. Rev. 13, 24–30 (2014).
9. Restellini, S. et al. Therapeutic drug monitoring guides the management of Crohn’s patients with secondary loss of response to adalimumab. Inflamm. Bowel Dis. 24, 1531–1538 (2018).
10. Schoultz, I. & Ketta, A. V. Cellular and molecular therapeutic targets in inflammatory bowel disease-focusing on intestinal barrier function. Cells 8, 193 (2019).
11. Dezelak, M., Repnik, K., Koder, S., Ferkolj, I. & Potocnik, U. A prospective pharmacogenomic study of Crohn’s disease patients during routine therapy with anti-TNF-alpha drug adalimumab: Contribution of ATG5, NFKB1, and CRP genes to pharmacodynamic variability. OMICS 20, 296–309 (2016).
12. Koder, S. et al. Genetic polymorphism in ATG16L1 gene influences the response to adalimumab in Crohn’s disease patients. Pharmacogenomics 16, 191–204 (2015).
13. Repnik, K., Koder, S., Skok, P., Ferkolj, I. & Potocnik, U. Transferrin level before treatment and genetic polymorphism in HFE gene as predictive markers for response to adalimumab in Crohn’s disease patients. Biochem. Genet. 54, 476–486 (2016).
14. Barber, G. E. et al. Genetic markers predict primary non-response and durable response to anti-TNF biologic therapies in Crohn’s disease. Am. J. Gastroenterol. 111, 1816–1822 (2016).
15. Yoon, S. M. et al. Colonic phenotypes are associated with poorer response to anti-TNF therapies in patients with IBD. Inflamm. Bowel Dis. 23, 1382–1393 (2017).
16. Artis, I. et al. Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. PLoS ONE 4, e7984 (2009).
17. Gorenjak, M. et al. Genetic prediction profile for adalimumab response in Slovenian Crohn’s disease patients. Z. Gastroenterol. 57, 1218–1225 (2019).
18. Gole, B. & Potocnik, U. Pre-treatment biomarkers of anti-tumour necrosis factor therapy response in Crohn’s disease—A systematic review and gene ontology analysis. Cells 8, 515 (2019).
19. Hernandez-Pacheco, N. et al. Combined analysis of transcriptomic and genetic data for the identification of loci involved in glucocorticoid response in asthma. Allergy 90, 1–6 (2020).
20. Mesko, B. et al. Peripheral blood derived gene panels predict response to infliximab in rheumatoid arthritis and Crohn’s disease. Genome Med. 5, 59 (2013).
21. Planell, N. et al. Usefulness of transcriptional blood biomarkers as a non-invasive surrogate marker of mucosal healing and endoscopic response in ulcerative colitis. J. Crohns Colitis 11, 1335–1346 (2017).
22. Gaujoux, R. et al. Cell-centred meta-analysis reveals baseline predictors of anti-TNFalpha non-response in biopsy and blood of patients with IBD. Gut 68, 604–614 (2019).
23. Wickham, H. et al. ggplot2: Elegant graphics for data analysis (Springer, 2016).
24. Pruim, R. J. et al. LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).
25. Liu, Q. et al. Monocyte to macrophage differentiation-associated (MMD) positively regulates ERK and Akt activation and TNFalpha and NO production in macrophages. Mol. Biol. Rep. 39, 5643–5650 (2012).
26. Schmitt, H. et al. Expansion of IL-23 receptor bearing TNFR2+ T cells is associated with molecular resistance to anti-TNF therapy in Crohn’s disease. Gut 68, 814–828 (2019).
27. Luther, J. et al. Loss of response to anti-tumor necrosis factor alpha therapy in Crohn’s disease is not associated with emergence of novel inflammatory pathways. Dig. Dis. Sci. 63, 738–745 (2018).
28. Leonard, A. E., Pereira, S. L., Sprecher, H. & Huang, Y. S. Elongation of long-chain fatty acids. Proc. Soc. Exp. Biol. Med. 235, 36–54 (2004).
29. Janse, I. C. et al. Identification of clinical and genetic parameters associated with hidradenitis suppurativa in inflammatory bowel disease. Inflamm. Bowel Dis. 13, 106–113 (2007).
30. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837 (2007).
31. Park, I. A. et al. Deacetylation and methylation at histone H3 lysine 9 (H3K9) coordinate chromosome condensation during cell cycle progression. Mol. Cells 31, 343–349 (2011).
32. Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H. & Tora, L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genom. 13, 424 (2012).
33. Hunt, S. E. et al. Ensembl variation resources. Database (Oxford) 2018 (2018).
34. Consortium, G. Erratum: Genetic effects on gene expression across human tissues. Nature 553, 530 (2018).
35. Wallner, S. et al. Monocyte to macrophage differentiation goes along with modulation of the plasminogen pattern through transcriptional regulation. PLoS ONE 9, e94102 (2014).
Acknowledgements

The authors would like to thank the patients for participation in this study. The authors would like to thank Dr. Natalia Hernandez-Pacheco, MSc, PhD and Ms. Dr. Maria Pino-Yanes, PhD for advices and support. The authors would like to thank Dr. Boris Gole, PhD for a careful review and recommendations.

Author contributions

M.G. and U.P. were involved in the conception and design of the study. M.G., M.Z., G.J. and P.S. were involved in the data acquisition. M.G., M.Z. and G.J. contributed to the analysis of the data. M.G., P.S. and U.P. contributed to the interpretation of the results. All authors drafted the article. All authors approved the manuscript.

Funding

The authors acknowledge the project (Molecular genetic biomarkers and mechanisms of non-response to biological treatment with anti-TNF in patients with IBD, ID J3-9258) was financially supported by the Slovenian Research Agency. The Slovenian Research Agency had no role in the design, execution, interpretation, or writing of the study. M.Z. was founded by Ministry of Education, Science and Sport (C3330–19–952026).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to U.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
