Supplementary Material

for

A cell-centered meta-analysis reveals baseline predictors of anti-TNF non-response in biopsy and blood of IBD patients

This document includes Supplementary Methods, Figures and Table legends.

Supplementary Methods.

Previously reported predictive gene signatures

We gathered 6 previously reported gene signatures from their respective original publication (Table 1) and mapped them to official gene symbols using annotation package *org.Hs.eg.db* (version 3.3.0). Altogether, the signatures counted 126 unique genes, and were named according to the patient cohort from which they were derived (Supplementary Table S1). Briefly, signatures UC-A and UC-B were defined as the top 20 differentially expressed genes found in two independent cohorts of UC patients, which we labelled A and B as per the original publication[1]; Signature UC-AB was defined as the overlap between all differentially expressed genes found in the same UC cohorts A and B, and comprised a total of 53 unique genes[1]; Signature UC-B-knn was also derived from UC cohort B, but using a different methodology based on a k-nearest-neighbor classifier[2]; Signature CDc was identified in CD patients from colon biopsies[3] (cohort CDc). The remaining signature named IRRAT was taken from the kidney transplant study[4].

Cell type expression pattern of predictive gene signatures

CEL files from sorted cell type samples from IRIS (GSE22886[5]) and the Human body index (GSE7307) were normalized separately using frma [6]. In GSE7307, we extracted the profiles from all immune cells (32 profiles from monocyte, T cell and B cell lineages) and colon tissues (2 profiles). We created a combined cell type gene expression matrix, corrected for dataset of origin effects using Combat [7], and subsequently averaged probesets into genes. This resulted in the creation of an expression matrix of 130 expression profiles (Supplementary Table S2), which we then standardized using z-scores and averaged into major cell lineages split into resting and activation/memory state.

The assign signature genes to the most likely contributing cell subpopulations they were detected from in the samples, we assigned each gene to the three most expressing cell subsets, and counted how many unique genes were assigned to each of 8 major functional cell lineages (Supplementary Table S3). Since we looked for enrichment of cell type expression, we restricted the analysis to the 122 signatures genes up-regulated in non-responders, of which 109 genes could be mapped to probesets in the sorted cell compendium data.

IBD cohorts' gene expression data

The gene expression data for each IBD cohort used in the deconvolution meta-analysis were obtained from 3 GEO datasets[8]: UC-A from GSE14580, UC-B form GSE12251 and CDc from GSE16879. These datasets contain biopsy gene expression profiles generated from 2 cohorts of UC patients (Cohort *A* and *B* in GSE14580 and GSE12251 respectively), and 1 cohort of CD patients (part of GSE16879). They were originally designed for the discovery of gene signatures that can predict, at baseline, if a patient is

likely to respond to an anti-TNFα treatment (Infliximab)[1,3]. In terms of signatures, all signatures were identified from baseline gene expression differential analysis between responders and non-responders to Infliximab treatment in the same set of 3 IBD cohorts of UC (cohorts UC-A and UC-B) or CD (cohort CDc) patients[1,3], exception being the IRRAT signature which, subsequent to the study it originated with, was found to correlate with anti-TNFα response at baseline in the UC-B cohort[9]. In addition, we obtained blood gene expression data of IBD patients for whom endoscopic activity was available (GSEXXX)[REF], in which we assessed the relation of the CCL7-CCR2-TREM1 axis to monitor disease activity.

Signature scores and ROC analysis

ROC analyses were performed on signature expression scores that summarize, for each sample, the expression level of all the genes in a predictive gene set[10]. Given a gene expression dataset (including data adjusted for proportion variations) and a gene signature/set, the signature score S_j for sample j was computed as:

$$S_j = \frac{1}{n} * \sum_1^n d_i * \log2(g_i)$$

where g_i is the expression level of the i-th gene of the signature in sample j, and d_i is the sign of the difference between its mean expression in nonresponders and responders. For adjusted data where negative expression values occurred, we shifted the data by $\widehat{g_i} = g_i - m_s + 1$, where m_s is the minimum expression value amongst the signature genes. ROC curve analysis of cellular biomarkers were computed either directly on estimated proportions for individual cell subsets. AUC values were computed using the R package pROC.

To assess whether the observed drops in AUC could result from the reduction in degrees of freedom incurred by the adjustment procedure itself, we repeatedly adjusted the data with random pairs of cell subset proportions and compared the derived "random" AUCs with the ones obtained using actual estimated proportions (Supplementary Figure S7B). This showed that all observed AUC differences were statistically significant (all p-values≤0.018). We also visualized the adjusted gene expression datasets ordered according to their respective signature score data, which confirmed that the association between signature scores and treatment response status was lost after adjustment

Response classification by a decision algorithm

In patients with CD, clinical remission was defined as cessation of diarrhea and abdominal cramping or, in the cases of patients with fistulas, cessation of fistula drainage and complete closure of all draining fistulas at week 14, coupled with a decision of the treating physician to continue IFX therapy at the current dosing and schedule. Partial response was defined as a reduction in the amount of diarrhea and abdominal cramping, or, in the case of fistula patients, a decrease in the drainage, size, or number of fistulas at last follow-up. In patients with UC, clinical remission was defined as cessation of diarrhea, rectal bleeding, and abdominal cramping at week 14 as indicated in the patient’s chart by the treating physician, coupled with a decision of the treating physician to continue IFX therapy at the current dosing and schedule, whereas partial response was defined as a reduction in the amount of diarrhea, rectal bleeding, and abdominal cramping. Outcomes not meeting one of the above definitions were classified as non-response.

To further stratify the response, patients deemed as partial responders was allocated to a decision tree following these steps: failure to withdraw steroid treatment at week 14 was deemed as therapeutic failure. In patients not treated with steroids, a substantial decrease (>50%) in biomarker dynamics (fecal calprotectin if available and serum CRP when calprotectin was not available), as an indicator of response to treatment. For subjects who were not steroid dependent and exhibited no substantial biomarker dynamics, response was defined according to the clinical state at week 26.

IFX levels and antibodies to IFX (ATI) measurements[11,12] were available for 28 of the patients. From these, we excluded responders for which 2 subsequent IFX level measurements below 3 (μg/ml) were observed prior to week 26, assuming their response status was unlikely to be IFX related, and non-responders with measurements of ATI level above 15 (μg/mL), assuming they had a secondary loss of response, unrelated TNFα blockade.

In total, applying this decision tree criteria left 29 responders and 23 non-responders from the two centers.

Immuno-histochemistry markers

We examined plasma cell frequencies by CD138+ IHC staining. For inflammatory macrophages, our *in-silico* deconvolution analysis relied on a gene expression signature of monocyte derived macrophages bearing typical macrophage morophology and phagocytic activity [5]. Given the disease context, this suggested a bias towards inflammatory macrophage phenotype (M1), as such, the expert pathologist performing the IHC, assessed the co-expression of the CD68 and CD86 as well as cell morphology, as these markers are co-expressed by monocytes and CD86 also in other cell subsets (e.g B and T). Specifically, to account for morphological differences between

macrophages and monocytes, mononuclear cells showing broad cytoplasm and oval nucleus were considered as "inflammatory macrophages", while CD68 and CD86-positive monocytes were ignored.

Supplementary Figure and Table Legends

Supplementary figure 1. Predictive gene signatures from previous gene expression differential analysis. Heatmaps showing the gene expression of the 6 previously reported predictive gene signatures in their respective discovery cohort(s). The top colored strip annotates responders and non-responder samples in blue and red respectively. Panel title and subtitle indicate the corresponding gene signature and cohort names respectively. The color scale goes from dark blue (high in responder) to dark red (high in non-responder). The row annotation indicates the group in which each gene was up-regulated. All but 4 genes (all from IRRAT) were up-regulated in non-responders.

Supplementary figure 2. Computational deconvolution of cell subset proportions identifies higher proportions of inflammatory macrophages and plasma cells in non-responders. Estimated immune cell subset proportions in the UC cohorts. Boxplot shows the baseline estimated proportions of each immune cell subset in each cohort for responders (blue) and non-responders (red) to anti-TNFα therapy. Only cell type with at least 75% non-zero values are shown.

Supplementary figure 3. Multi-cohort analysis of estimated cell proportions identifies consistent baseline differences in inflammatory macrophages and plasma cells. Each panel shows estimated group proportion differences (pseudo

median) and 95% confidence interval for a given cell subset, across all discovery cohorts. Missing data is due to cell type/cohort pairs not included in the meta-analysis due to too many zero estimated proportions. The x-axis represents the log2 proportion fold change (i.e. log2(Responders/Non-Responders)). Statistical significance was calculated using Wilcoxon rank sum test (nominal p-value ≤ 0.05), and is shown in red and blue for significantly higher proportions in non-responders and responders respectively. Violet indicates non-significant differences.

Supplementary figure 4. **Differences in cellular biomarkers increase between response groups following treatment.** Deconvolution derived estimates of plasma cells and inflammatory macrophages proportions in responders and non-responder groups in the UC-A and CDc cohorts. Both cell subsets are significantly lower in responders in both cohorts p-values<0.01 BH-adjusted). Strikingly, inflammatory macrophages were undetectable (zero values) in all responders, except for two patients from the CDc cohort.

Supplementary figure 5. Adjusting gene expression for abundances of inflammatory macrophages and plasma cells significantly breaks the association between gene signatures and treatment response status. Heatmaps similar to those in Supplementary figure 1, showing the association between the signatures scores UC-A (left), UC-B (middle) and CDc (right) in their respective cohorts, generated from the original gene expression data (top row) and after adjustment for estimated abundances of inflammatory macrophages and plasma cells (bottom row).

Supplementary figure 6. Differences in cell subset is the driving force of the reported gene signatures for predicting anti-TNFα non-response from baseline. (A) Predictive power of reported gene signatures drops after correction for variations in cellular biomarkers. ROC curves for each original gene signature score in its respective discovery cohort UC-A (mustard), UC-B (blue), and CDc (kaki) on the original data (left) and the data following adjustment for variations in the abundances of both cellular biomarkers (right). Respective AUCs are reduced from 93%, 97% and 100% on the original data, to 57%, 68% and 79% on the adjusted data. (B) The drop in AUC due to cell subset proportions differences is significant. AUC was calculated for each gene signature score in their respective discovery cohort. Densities represent AUC null distributions obtained by adjusting gene expression data with random proportions, in each cohort (panels). Solid line shows achieved AUC in original (unadjusted) data, dashed line indicates AUC obtained after adjustment for estimated proportions of inflammatory and plasma cells.

Supplementary figure 7. Anti-TNFα non-responders exhibit consistent increased numbers of plasma cells in the 2 cohorts, whereas the inflammation severity is broadly distributed. Responders (blue) and non-responders (red) in the 2 IBD patient cohorts: (A) Boxplots of plasma cell quantitative score in the preliminary cohort (left panel, n=20) and main cohort (right panel, n=52). P values < 0.001 for both cohorts (by Student T-test). (B,C) Distribution of responding and non-responding patients in the preliminary cohort (top panels) and main cohort (bottom panels) as shown by patient proportions and patient counts in each score. (B) Histograms of pathologist plasma cell score, (C) Histograms of inflammation severity scores as determined by an expert pathologist.

Supplementary figure 8. **Overlap between cohorts of differentially expressed pathways and genes performed on cellular biomarker adjusted data.** (A) Venn diagrams showing overlap of cellular biomarker-adjusted differentially expressed pathways across cohorts identified by GSEA. The fifteen pathways differentially expressed in all three cohorts were all upregulated in non-responders. (B) Venn diagrams showing overlap of differentially expressed genes across cohorts. Upregulated genes (left), downregulated (right). (C) GSEA enrichment score are driven by consistent leading-edge genes across cohorts. Curves show GSEA enrichment scores for the 3 pathways (rows) that have the most consistent leading-edge genes across cohorts (>25% of their genes in leading-edge in all cohorts).

Supplementary figure 9. Biopsy gene expression decreased after adjustment to inflammatory macrophage and plasma cell proportions. Box plots showing the adjusted expression of TNFα and its receptors, TNFR1 and TNFR2 in the three biopsy cohorts.

Supplementary figure 10. **No difference in the hypothesized model genes in blood between responders and non-responders.** Boxplots showing CCL7, CCR2, TNFR1 mRNA expression as measured in whole blood of 22 responding (blue) and non-responding CD patients, prior to initiation of infliximab therapy.

Table S1. Previously reported gene signatures predictive of response to anti-TNFα. List of the gene signatures used in the analysis shown in Figure 2. For each signature, we provide its member genes and their respective annotations.

Table S2. Compendium of sorted cell expression profiles. Description of the GEO data used to compute the immune contribution to previously reported signature genes

Table S3. Immune contribution to signature gene expression. Results of the preliminary analysis that derived overall cellular origin of all previously reported signature genes.

Table S4. Basis matrix used for the deconvolution of all discovery cohorts.

Excel spreadsheet containing the data of the 359 Affymetrix probeset IDs (HG-U133plusV2), information about their corresponding gene if available (ENTREZ gene ID, SYMBOL and description), and their gene expression levels across 17 immune cell subsets.

Table S5. Estimated immune cell subset proportions

Estimated proportions in all cohorts at baseline and after treatment.

Table S6. Results of the cell type proportion meta-analysis

Table S7. Results of the post-treatment cell type proportion analysis

Table S8. Demographic and clinical information for the validation cohort

Table S9. List of differentially expressed genes when adjusting for inflammatory macrophage and plasma cell proportions.

Table S10-12. Results from IPA analysis: common regulators, enriched biofunctions and networks

References

1 Arijs I, Li K, Toedter G, *et al.* Mucosal gene signatures to predict response to infliximab in patients with ulcerative colitis. *Gut* 2009;**58**:1612–9. doi:10.1136/gut.2009.178665

2 Katherine Li FB. Biomarkers for anti-tnf treatment in ulcerative colitis and related disorders. 2009.https://www.google.com/patents/WO2010044952A3?cl=en (accessed 11 Sep 2017).

3 Arijs I, Quintens R, Van Lommel L, *et al.* Predictive value of epithelial gene expression profiles for response to infliximab in Crohn's disease. *Inflamm Bowel Dis* 2010;**16**:2090–8. doi:10.1002/ibd.21301

4 Mueller TF, Einecke G, Reeve J, *et al.* Microarray analysis of rejection in human kidney transplants using pathogenesis-based transcript sets. *Am J Transplant* 2007;**7**:2712–22. doi:10.1111/j.1600-6143.2007.02005.x

5 Abbas AR, Baldwin D, Ma Y, *et al.* Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun* 2005;**6**:319–31. doi:10.1038/sj.gene.6364173

6 McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). *Biostatistics* 2010;**11**:242–53. doi:10.1093/biostatistics/kxp059

7 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;**8**:118–27. doi:10.1093/biostatistics/kxj037

8 Barrett T, Troup DB, Wilhite SE, *et al.* NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res* 2011;**39**:D1005-10. doi:10.1093/nar/gkq1184

9 Halloran B, Chang J, Shih DQ, *et al.* Molecular patterns in human ulcerative colitis

and correlation with response to infliximab. *Inflamm Bowel Dis* 2014;**20**:2353–63. doi:10.1097/MIB.0000000000000239

10 Khatri P, Roedder S, Kimura N, *et al.* A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *J Exp Med* 2013;**210**:2205–21. doi:10.1084/jem.20122709

11 Vande Casteele N, Gils A, Ballet V, *et al.* Randomised Controlled Trial of Drug Level Versus Clinically Based Dosing of Infliximab Maintenance Therapy in IBD: Final Results of the TAXIT Study. 2013;**1**:A1.https://lirias.kuleuven.be/handle/123456789/425829 (accessed 8 Aug 2017).

12 Ben-Horin S, Yavzori M, Katz L, *et al.* The immunogenic part of infliximab is the F(ab')2, but measuring antibodies to the intact infliximab molecule is more clinically useful. *Gut* 2011;**60**:41–8. doi:10.1136/gut.2009.201533