CRP Predicts Biochemical but Not Clinical Remission in Ustekinumab-Treated Crohn's Disease: A Remission-Specific Whole-Blood Transcriptomic Signature

Simbarashe G. Magwenzi
NYNOSK LLP, 71-75 Shelton Street, Covent Garden, London, United Kingdom, WC2H 9JQ
Corresponding author: simbarashe.magwenzi@nynosk.com

Running title: Transcriptomic biomarkers for ustekinumab response in Crohn's disease
Keywords: Crohn's disease; ustekinumab; transcriptomics; machine learning; biomarkers; C-reactive protein; remission; Random Forest; SHAP

Abstract

Background and Aims
Clinical outcome prediction in ustekinumab-treated Crohn's disease remains a significant unmet need. C-reactive protein (CRP) is widely used as a biochemical marker of inflammation, yet its utility as a predictor of clinical remission alongside blood transcriptomics has not been formally evaluated.

Methods
Predictive model training used GSE207465 (n=353, ustekinumab-treated, UNITI-2 trial; whole-blood RNA, HuGene 2.1 ST). Co-primary endpoints were CRP normalisation (serum CRP <5 mg/L at week 8) and clinical remission (CDAI score <150, week 8). Baseline CRP served as a reference comparator via logistic regression. A 60-gene baseline expression panel was selected from 1,165 remission-specific differentially expressed genes; a Random Forest classifier was trained with five-fold stratified cross-validation and SHAP feature attribution. External validation used GSE186963 (n=23, infliximab-treated; Clariom S) via a 25-gene cross-platform subset and performance was quantified by AUC with 2,000-iteration bootstrap 95% confidence intervals.

Results
Baseline CRP strongly predicted CRP normalisation (AUC 0.840, 95% CI 0.756-0.911) but was near chance for clinical remission (AUC 0.564, 95% CI 0.454-0.675). The 60-gene panel achieved a higher point-estimate AUC for remission than CRP alone (0.641, 95% CI 0.525-0.751), though confidence intervals overlapped substantially. The top SHAP feature, *PDE2A-AS1* (phosphodiesterase 2A antisense lncRNA), was non-significant in the overall differential expression analysis (logFC 0.013, Padj 0.808), but it was a significant remission-specific transcriptional marker (logFC 0.288, -log10P 3.376). *BANK1* showed opposing associations with the two endpoints. External validation was negative (AUC 0.517, 95% CI 0.265-0.768), consistent with biologic-specific transcriptomic prediction.

Conclusions
Baseline CRP and whole-blood gene expression predict discordant endpoints in ustekinumab-treated Crohn's disease. A remission-specific signature including *PDE2A-AS1* adds incremental predictive value beyond CRP alone. Prospective validation in adequately powered ustekinumab cohorts is required before clinical utility can be established.

Introduction

Crohn's disease is a chronic, relapsing inflammatory bowel condition associated with considerable morbidity and healthcare utilisation. Ustekinumab, a monoclonal antibody directed against the shared p40 subunit of interleukin-12 and interleukin-23, was approved for the treatment of moderate-to-severely active Crohn's disease following demonstration of efficacy in the UNITI-1 and UNITI-2 phase 3 randomised controlled trials.[1] Clinical remission (CDAI score <150) at week 8 was achieved in 15.9% and 20.9% of ustekinumab-treated patients in UNITI-1 (130 mg and 6 mg/kg intravenous, respectively) and in 30.6% and 40.2% in UNITI-2, with substantially higher rates in the latter reflecting enrolment of patients with prior conventional but not necessarily prior biologic therapy failure.[1,2]

The identification of baseline biomarkers that reliably predict clinical outcomes in ustekinumab-treated Crohn's disease represents an important and unmet clinical need. Current treat-to-target guidelines recommend objective endpoint monitoring, with clinical remission and normalisation of inflammatory markers as the preferred treatment targets.[3,4] Serum CRP is widely employed as an accessible biochemical surrogate of intestinal inflammation in Crohn's disease.[5] However, the relationship between CRP normalisation and clinical remission is imperfect, and the predictive value of baseline CRP for each endpoint may differ substantially. Based on published evidence available at the time of writing, no study has formally evaluated the degree to which these two endpoints share or diverge in their molecular predictors using genome-wide baseline transcriptomic data.

Transcriptomic approaches to biologic response prediction have been pursued using both mucosal biopsies and peripheral blood. Baseline mucosal gene expression patterns have been associated with anti-tumour necrosis factor (anti-TNF) response in cohort studies of inflammatory bowel disease.[6] In the blood-based domain, a personalised network framework applied to peripheral blood from infliximab-treated Crohn's disease patients identified a monocyte-derived RAC1-PAK1 signalling axis predictive of response, validated across IBD and rheumatoid arthritis cohorts.[7] Whole-blood transcriptomics offers the practical advantage of accessibility in routine clinical care and may capture systemic immune dynamics not reflected by a single acute phase protein.

Machine learning (ML) methods, particularly tree-based ensemble approaches such as Random Forest and XGBoost, are well-suited to high-dimensional transcriptomic feature spaces. Coupled with SHapley Additive exPlanations (SHAP) analysis, which quantifies how much each baseline gene expression value raises or lowers an individual patient's predicted probability of achieving a clinical outcome, these models produce biologically interpretable feature attributions.[8]

The present study utilised publicly available whole-blood RNA microarray data from the UNITI-2 trial[1] (GEO accession GSE207465[9]; n=353) to examine two co-primary endpoints: CRP normalisation (<5 mg/L at week 8) and clinical remission (CDAI, week 8). The aims were: (i) to quantify the dissociation between baseline CRP's predictive performance for each endpoint; (ii) to develop and evaluate a whole-blood transcriptomic prediction model for clinical remission; (iii) to

characterise the molecular features driving model predictions using SHAP analysis; and (iv) to assess the directional consistency of candidate biomarkers in an independent external cohort.

Methods

Data sources and study design

This study used two publicly available whole-blood RNA microarray datasets retrieved from the NCBI Gene Expression Omnibus (GEO).[10] The training cohort comprised GSE207465,[9] derived from the UNITI-2 phase 3 induction trial of ustekinumab in Crohn's disease.[1] From the series of 1,733 samples, 353 patients were included based on receiving ustekinumab (intravenous induction), having complete paired baseline (week 0) as well as week 8 samples, and having available CRP data. Whole-blood RNA was profiled on the Affymetrix HuGene 2.1 ST microarray. Data were accessed in normalised and $\log_2$-transformed form as deposited. All analyses were performed on publicly available, fully anonymised data; no additional ethical approval was required.

For external validation, GSE186963[11] was identified as the sole publicly available longitudinal whole-blood RNA dataset from biologic-treated Crohn's disease patients with paired baseline and post-treatment samples as well as CRP data. Indeed, the study investigators noted that no other such dataset existed at the time of their publication.[7] GSE186963 comprises the primary discovery cohort of Gerassy-Vainberg et al. (Cell Reports Medicine, 2024), containing 23 infliximab-treated Crohn's disease patients with complete paired baseline (week 0) and week 14 samples, profiled on the Affymetrix Clariom S array, with institutional review board approval (0052-17-RMB, Rambam Health Care Campus).[7] Datasets derived from solid tissue biopsies, single-timepoint cross-sectional studies, or those lacking CRP data were not suitable for the present analysis. The use of infliximab rather than ustekinumab and the different post-treatment timepoint (week 14 versus week 8) are acknowledged methodological constraints reflecting the absence of a suitable independent ustekinumab whole-blood dataset in the public domain.

Outcome definitions

Two co-primary endpoints were defined in the training cohort: CRP normalisation (serum CRP <5 mg/L at week 8) and clinical remission (CDAI score <150 at week 8; the CDAI incorporates stool frequency, abdominal pain, general wellbeing, and extraintestinal manifestations into a composite score).[12] Of 353 included patients, 140 (39.7%) achieved clinical remission and 192 (54.4%) achieved CRP normalisation; thus, approximately 42 and 57 patients in the held-out test set (n=106) achieved each endpoint respectively. In the external validation cohort, clinical remission was defined by Gerassy-Vainberg et al. as cessation of diarrhoea and abdominal cramping (or complete fistula closure) at week 14 combined with the treating physician's decision to continue infliximab at the current dose and schedule, with partial responders adjudicated by a hierarchical algorithm incorporating steroid dependency, CRP, calprotectin dynamics, and clinical state at week 26.[7] Of 23 patients, 15 (65.2%) met this criterion. This clinical remission-based endpoint is analogous in principle to the CDAI-based remission criterion used in the training cohort, though the assessment method and post-treatment timepoint differ.

Data preprocessing and feature scaling

Baseline gene expression values from all 353 training cohort patients were standardised using z-score scaling prior to model training. Scaling was applied across the full training dataset before train-test splitting, which represents a minor acknowledged limitation. For cross-platform validation, the 25-gene Clariom S subset was standardised independently within the validation cohort (n=23) to avoid information transfer between training and validation data.

Differential expression analysis

Treatment-responsive genes (baseline versus week 8) were identified separately for patients achieving remission (n=140) and those not achieving remission (n=213) using the limma package[13] in R (version 4.2.2) with Benjamini-Hochberg false discovery rate correction; genes with adjusted $P<0.05$ were considered significant. A Venn diagram analysis identified genes uniquely and significantly changing in remitters, but not in non-remitters. In the overall population analysis (all 353 patients, baseline versus week 8), 5,710 genes were significant (2,341 upregulated, 3,369 downregulated; adjusted $P<0.05$), with $\log_2$ fold-changes in the range -0.14 to +0.22.

Gene panel selection

From the 1,165 genes uniquely and significantly changing in remitters from baseline to week 8 (adjusted $P<0.05$), the top 30 most upregulated and top 30 most downregulated genes by absolute $\log_2$ fold-change (logFC) were selected to form a 60-gene baseline expression panel (Supplementary Table S1). Selection was performed on the hypothesis that transcripts uniquely mobilised during the remission process may exist in a distinct baseline expression state in patients with the biological capacity to achieve remission; their pre-treatment levels were therefore expected to constitute a predictive signature of future treatment outcome. Baseline CRP was excluded from the gene panel and evaluated separately as a clinical reference comparator. Full technical details are provided in Supplementary Methods.

Model development

A 70:30 stratified train-test split (random seed 10) yielded 247 patients for training and 106 for independent testing. Two classification approaches were evaluated: Random Forest (RF)[14] using scikit-learn[15] and XGBoost.[16] Hyperparameters were optimised separately per endpoint and are available from the corresponding author upon reasonable request. Five-fold stratified cross-validation assessed within-training performance. A logistic regression model trained on baseline CRP alone served as the clinical reference comparator. All AUC values were computed from predicted class probabilities. All machine learning was performed in Python 3 (version 3.13).

Model evaluation

The primary performance metric was AUC, representing the probability that the model assigns a higher predicted risk to an outcome-positive patient than an outcome-negative patient (0.5 = chance; 1.0 = perfect discrimination). Bootstrap 95% confidence intervals were computed from 2,000 iterations on the held-out test set (n=106). Overlap of confidence intervals between models was used to assess statistical support for superiority claims.

SHAP feature attribution

SHAP TreeExplainer[17] was applied to the trained RF classifier for each endpoint using the held-out test set. SHAP values quantify each baseline gene's contribution to the predicted probability of achieving the outcome for each patient. Beeswarm plots show the direction and magnitude of feature effects; bar plots show mean absolute SHAP value. Gene symbols were substituted for probe identifiers using a validated probe-to-gene mapping table (Supplementary Table S1), with *PDE2A-AS1* confirmed as the correct symbol for probe 105369379_at (antisense lncRNA to phosphodiesterase 2A, distinct from the PDE2A protein-coding gene).

Cross-platform validation and external cohort

Of the 60 training panel genes (HuGene 2.1 ST probes), 25 were available on the Clariom S platform (Supplementary Table S1); notable absences included *PDE2A-AS1, LINC00900*, and all T-cell receptor segment probes. A 25-gene RF model was trained on the training cohort using the same remission-endpoint hyperparameters and evaluated on the held-out training test set (AUC 0.624), then applied to the independently standardised validation cohort (n=23). Spearman correlations between baseline gene expression and clinical outcomes were computed separately in training and validation cohorts to assess directional consistency.

Statistical analysis

Differential expression analysis was performed in R (version 4.2.2) using the limma package.[13] All machine learning and statistical analyses were performed in Python 3 (version 3.13) using scikit-learn,[15] pandas, numpy, and scipy. Spearman rank correlations were computed for gene-outcome associations. No formal sample size calculation was performed, as this was a retrospective analysis of publicly available data. Analysis code is available from the corresponding author on reasonable request.

Results

Cohort characteristics

The training cohort (GSE207465, UNITI-2) comprised 353 ustekinumab-treated Crohn's disease patients with paired baseline and week 8 whole-blood RNA profiles as well as CRP data. Clinical remission at week 8 was achieved by 140 patients (39.7%) and CRP normalisation by 192 patients (54.4%). The observed remission rate of 39.7% in the training cohort is consistent with the UNITI-2 population from which the dataset was derived.[1] Baseline CRP had a mean of 15.9 mg/L (SD 22.1; range 0.1-137.0 mg/L), consistent with an active inflammatory population and the right-skewed distribution typical of CRP. The external validation cohort (GSE186963, Gerassy-Vainberg et al.[7]) comprised 23 infliximab-treated Crohn's disease patients profiled on the Clariom S platform, with clinical remission at week 14 achieved by 15 patients (65.2%); baseline CRP was 18.7 mg/L (SD 29.7; range 1.0-125.7 mg/L). Full cohort characteristics are summarised in Table 1.

Transcriptional landscape of ustekinumab treatment response

Differential expression analysis across all 353 ustekinumab-treated patients from baseline to week 8 identified 5,710 significantly changed genes: 2,341 upregulated and 3,369 downregulated

(adjusted P<0.05; Figure 2A). Fold changes were modest (log_2 range -0.14 to +0.22), consistent with the biological characteristics of whole-blood microarray data.

Remission-stratified differential expression analysis identified 1,165 genes uniquely and significantly changing in remitters only, 907 genes changing in both remitters and non-remitters, and 865 genes uniquely changing in non-remitters (Figure 2C). The 60-gene panel was selected from the 1,165 remission-specific genes (Figure 2B; Supplementary Table S1). *PDE2A-AS1*, the top SHAP feature for the remission model, was non-significant in the overall analysis (logFC 0.013, adjusted P=0.808) but showed a substantially larger and significant fold change in the remission-stratified analysis (logFC 0.288, -log10P 3.376; Figure 2A versus 2B). This finding illustrates the analytical value of outcome-stratified differential expression: genes with remission-specific signals are diluted to non-significance in an unstratified analysis yet carry genuine discriminatory information at baseline.

CRP dissociation: discordant predictive performance across endpoints

Baseline CRP demonstrated markedly different discriminatory performance across the two co-primary endpoints (Table 2, Figure 3). For CRP normalisation at week 8, logistic regression on baseline CRP alone achieved an AUC of 0.840 (95% CI 0.756-0.911), reflecting the inverse relationship between baseline CRP level and the likelihood of biochemical normalisation: patients with lower baseline CRP face a shorter biochemical distance to the normalisation threshold. For clinical remission, CRP alone was near chance (AUC 0.564, 95% CI 0.454-0.675). An AUC of 0.564 indicates that in only 56.4% of randomly selected patient pairs comprising one remitter and one non-remitter, baseline CRP correctly assigned the higher predicted probability to the remitter, barely above the 50% expected by chance.

The 60-gene transcriptomic panel (RF model with CRP excluded) achieved an AUC of 0.686 (95% CI 0.580-0.784) for CRP normalisation, substantially lower than CRP alone. For clinical remission, the gene panel achieved AUC 0.641 (95% CI 0.525-0.751) compared with CRP alone (0.564, 95% CI 0.454-0.675); confidence intervals overlapped (overlap region 0.525-0.675), precluding definitive conclusions of superiority. XGBoost achieved comparable performance to RF for both endpoints (CRP normalisation AUC 0.682; remission AUC 0.616; Table 2), supporting robustness across tree-based ensemble methods. Five-fold cross-validation produced mean AUCs of 0.663 (SD 0.064) for the CRP normalisation RF and 0.557 (SD 0.071) for the remission RF.

SHAP analysis: molecular determinants of remission prediction

SHAP analysis of the RF remission model identified *PDE2A-AS1* as the feature with the highest mean absolute SHAP value (0.045; Figure 4B), followed by *LINC00900* (0.029), *MIR519D, FBXL13*, *ECHDC3*, and *EIF1AX*. The beeswarm plot (Figure 4A) demonstrated directional coherence: higher baseline *PDE2A-AS1* expression was associated with positive SHAP contributions towards remission prediction, whilst T-cell receptor segment genes and *EIF1AX* showed negative contributions when highly expressed. These features are heterogeneous in biological function, encompassing long non-coding RNAs (*PDE2A-AS1*, *LINC00900*), a microRNA (*MIR519D*), an E3 ubiquitin ligase adaptor (*FBXL13*), a mitochondrial enzyme (*ECHDC3*), and a translation initiation factor (*EIF1AX*).

PDE2A-AS1 encodes a long non-coding antisense RNA overlapping the phosphodiesterase 2A (PDE2A) gene locus. PDE2A degrades the second messengers cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), which regulate inflammatory cytokine production and leucocyte activation. The clinical relevance of *PDE2A-AS1* expression in Crohn's disease has not been previously described. However, the directionally coherent SHAP pattern observed, combined with *PDE2A-AS1* increasing specifically in remitters from baseline to week 8 (logFC +0.288) whilst showing negligible change in the overall population (logFC 0.013), supports the hypothesis that this lncRNA marks a remission-associated transcriptional state at baseline.

BANK1 demonstrates opposing associations with co-primary endpoints

BANK1 (B-cell scaffold protein with ankyrin repeats 1) demonstrated opposing associations with the two co-primary endpoints (Figure 5A). In the training cohort, baseline *BANK1* expression was positively correlated with CRP normalisation (Spearman rs=+0.182, P<0.05) but negatively correlated with clinical remission (rs=-0.186, P<0.05). The negative association with clinical remission was directionally consistent in the external validation cohort (rs=-0.330), providing preliminary cross-cohort support for this finding.

BANK1 is a B-cell-specific scaffold protein implicated in B-cell receptor signalling and has been identified as a susceptibility gene in systemic lupus erythematosus.[18] High baseline *BANK1* expression in peripheral blood may reflect elevated B-cell activity. Notably, Gerassy-Vainberg et al. demonstrated that monocytes are the primary cellular source of infliximab-response-associated transcriptional signal in the same validation cohort,[7] raising the hypothesis that a B-cell-dominant baseline immune phenotype may inversely associate with the monocyte-driven response dynamics relevant to both ustekinumab and infliximab. This finding warrants prospective validation and mechanistic investigation.

Cross-cohort consistency of remission-associated gene signatures

Spearman correlations were computed for the 25 genes present in both platforms (Figure 5B). *FBXL13* showed the strongest cross-cohort consistency (rs=+0.145 training; rs=+0.550 validation). *ECHDC3* and *GRINA* also showed positive directional consistency (*ECHDC3*: rs=+0.181 training, +0.138 validation; *GRINA*: rs=+0.145 training, +0.303 validation). Negative directional consistency was observed for *ZNF430* (rs=-0.150 training, -0.261 validation) and *CXCL8* (rs=-0.135 training, -0.124 validation). These associations should be treated as hypothesis-generating given the small external validation sample (n=23), the different biologic agent and endpoint definition, as well as the absence of formal multiple testing correction.

External validation

The 25-gene RF model, which achieved an AUC of 0.624 on the training cohort held-out test set, was applied to the independently standardised external validation cohort. The AUC for clinical remission prediction was 0.517 (95% CI 0.265-0.768, 2,000 bootstrap iterations), not meaningfully above chance; the wide confidence interval reflects the limited statistical power of the 23-patient cohort and precludes conclusions in either direction (Supplementary Figure S1). The validation cohort used a clinical remission-based endpoint (symptom cessation criteria at week 14) analogous in principle to the training endpoint (CDAI <150 at week 8), reducing the endpoint mismatch concern. The primary factors explaining the negative result are the different biologic mechanism

(infliximab targeting TNF-alpha versus ustekinumab targeting IL-12/23) and the different post-treatment timepoint. Notably, Gerassy-Vainberg et al. demonstrated that an infliximab-specific transcriptomic signature (RAC1-PAK1 axis) successfully predicted response in the same cohort (AUC 0.89),[7] confirming that it can discriminate response groups and that the failure is specific to the ustekinumab-trained panel rather than an inherent property of the dataset. The negative external validation is reported in full accordance with the TRIPOD statement.[19]

Discussion

This study reports three principal findings. First, baseline CRP demonstrates a marked dissociation in discriminatory performance across clinical and biochemical endpoints in ustekinumab-treated Crohn's disease, with strong prediction of CRP normalisation (AUC 0.840), but near chance performance for clinical remission (AUC 0.564). Second, a 60-gene whole-blood baseline transcriptomic panel selected from remission-specific differentially expressed genes achieves modestly higher remission prediction than CRP alone (AUC 0.641), though confidence intervals overlap and superiority cannot be definitively established at this sample size. Third, *PDE2A-AS1*, a remission-specific lncRNA non-significant in the overall population analysis, emerges as the primary SHAP feature for the remission model, illustrating how outcome-stratified transcriptomic analyses can reveal biologically relevant features that remain invisible in unstratified approaches.

The CRP dissociation finding has important clinical implications. CRP is routinely used as a monitoring biomarker in Crohn's disease treat-to-target protocols, and CRP normalisation is an established therapeutic target.[3,4] The present data demonstrate that baseline CRP contributes almost no discriminatory information for clinical remission. This is mechanistically logical: CRP normalisation is inherently influenced by baseline CRP level, as patients with lower baseline CRP face a shorter biochemical distance to the <5 mg/L threshold. Clinical remission, captured by CDAI, incorporates abdominal pain, stool frequency, and general wellbeing, and may depend on immune and neuroenteric processes not captured by a single acute phase protein. These findings argue that CRP normalisation and clinical remission should not be treated as interchangeable endpoints when evaluating predictive biomarker performance. Studies using CRP normalisation as a surrogate for biomarker discovery may not identify markers relevant to clinical remission.

PDE2A-AS1 represents a biologically plausible but unexplored candidate in Crohn's disease. *PDE2A*, the sense-strand gene, encodes phosphodiesterase 2A, which degrades both cAMP and cGMP in a cGMP-stimulated manner. Cyclic nucleotide signalling plays roles in regulating inflammatory cytokine production and leucocyte activation. *PDE2A-AS1* is a long non-coding antisense RNA that may regulate PDE2A expression in cis, a mechanism described for several lncRNA-sense gene pairs. The finding that *PDE2A-AS1* increases specifically in remitters from baseline to week 8 (logFC +0.288) but shows negligible change in the full patient population (logFC 0.013) suggests a remission-associated transcriptional programme at this locus. Whether *PDE2A-AS1* represents a cause, consequence, or correlate of remission-associated immune reprogramming cannot be determined from this retrospective observational analysis, and experimental validation is required.

The opposing associations of *BANK1* with CRP normalisation (positive) and clinical remission (negative) add biological complexity to these findings. BANK1 is a B-cell-specific adaptor protein previously identified as a susceptibility locus in systemic lupus erythematosus.[18] High baseline *BANK1* expression may reflect elevated B-cell activity, consistent with attenuation of T-helper-17-driven intestinal inflammation with preserved B-cell activation that the anti-IL-12/23 mechanism of ustekinumab does not fully suppress. This dual-role pattern is consistent with the existence of immunological patient subtypes with discordant response characteristics in inflammatory bowel disease.

The cross-cohort directional consistency of *FBXL13, ECHDC3, GRINA, ZNF430*, and *CXCL8* provides preliminary evidence that some remission-associated blood transcriptomic signals may generalise across biologic agents. *CXCL8*, encoding the neutrophil-attracting chemokine interleukin-8, shows consistent negative correlation with response in both cohorts, consistent with its role as a marker of active intestinal inflammation. These directional consistencies should be treated as hypothesis-generating given the small validation sample and multiple testing considerations.

The negative external validation result is mechanistically coherent when considered alongside Gerassy-Vainberg et al., who applied a distinct analytical framework to the same external cohort (GSE186963) and identified a RAC1-PAK1 signalling axis in peripheral monocytes as a predictive biomarker for infliximab response, achieving an AUC of 0.89 in the discovery cohort.[7] The RAC1-PAK1 axis is a downstream effector of TNF-alpha signalling in innate immune cells, directly implicating the pharmacological target of infliximab. By contrast, the transcriptomic panel in the present study was selected from genes changing specifically in patients achieving remission with ustekinumab, an agent acting primarily on T-helper cell differentiation via IL-12/23 blockade. The failure of ustekinumab-trained features to predict infliximab remission in the same cohort where IFX-specific features succeed supports the concept that blood transcriptomic predictors are biologic-specific rather than generalisable as pan-IBD tools. This has practical implications for biomarker discovery study design: predictive transcriptomic panels may need to be developed within the target biologic's mechanism of action.

Several limitations must be acknowledged. The analysis was retrospective and used data not originally collected for biomarker discovery. The training cohort comprised 353 patients with 106 in the held-out test set, yielding wide bootstrap confidence intervals. It is noted that the mean cross-validation AUC for the remission model (0.557) was lower than the test-set estimate (0.641), suggesting some degree of optimism in the held-out estimate; the cross-validation value provides a more conservative guide to expected generalisation. The external validation cohort was small (n=23) and lacked 35 of 60 panel genes including *PDE2A-AS1*. The negative result therefore reflects both these constraints and the fundamental difference in biologic mechanism. Scaling was applied before train-test splitting, representing a mild information leak expected to have negligible impact on AUC estimation. Whole-blood transcriptomics reflects a cellular mixture and may attenuate signals from specific leucocyte subsets; sorted-cell or single-cell approaches may provide greater resolution in future studies.

Taken together, the findings demonstrate that baseline CRP and whole-blood gene expression capture distinct biological processes relevant to different clinical outcomes in ustekinumab-treated Crohn's disease. The endpoint-specific nature of transcriptomic prediction, and the biologic-specific

failure of cross-agent generalisation, have direct implications for how future biomarker discovery programmes in this field should be designed and validated. Prospective studies collecting paired whole-blood RNA and clinical outcome data in ustekinumab-treated Crohn's disease, with pre-specified model thresholds and sufficient power for remission as the primary endpoint, are needed to determine whether a validated transcriptomic panel could contribute to clinical decision-making in precision medicine for inflammatory bowel disease.

Author Contributions

SGM conceived the study, designed and performed the experiments, conducted experimental analysis, interpreted results, developed data visualization tools and prepared the manuscript.

Acknowledgements

Artificial intelligence assistance was used in the preparation of this manuscript, including data analysis, figure generation, and text drafting. All AI-generated content was reviewed, verified, and edited by the author. The use of AI assistance was in accordance with the policies of the target journal and established scientific publishing norms.

Gratitude is extended to the UNITI-2 study investigators, Gerassy-Vainberg et al. and to the NCBI Gene Expression Omnibus for making the GSE207465 and GSE186963 datasets publicly available.

Conflicts of Interest

None.

Funding

None.

Data Availability Statement

All gene expression data are publicly available through the NCBI Gene Expression Omnibus under accession numbers GSE207465 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207465) and GSE186963 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186963).

Analysis code is available from the corresponding author upon reasonable request.

References

1. Feagan BG, Sandborn WJ, Gasink C, et al. Ustekinumab as Induction and Maintenance Therapy for Crohn's Disease. *New England Journal of Medicine*. 2016;375(20):1946-1960. doi:10.1056/nejmoa1602773

2. Sandborn WJ, Gasink C, Gao LL, et al. Ustekinumab Induction and Maintenance Therapy in Refractory Crohn's Disease. *New England Journal of Medicine*. 2012;367(16):1519-1528. doi:10.1056/NEJMoa1203572

3. Torres J, Bonovas S, Doherty G, et al. ECCO guidelines on therapeutics in Crohn's disease: Medical treatment. *J Crohns Colitis*. 2020;14(1):4-22. doi:10.1093/ecco-jcc/jjz180

4. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: An Update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) Initiative of the International Organization for the Study of IBD (IOIBD): Determining Therapeutic Goals for Treat-to-Target strategies in IBD. *Gastroenterology*. 2021;160(5):1570-1583. doi:10.1053/j.gastro.2020.12.031

5. Vermeire S, Assche G Van, Rutgeerts P. C-Reactive Protein as a Marker for Inflammatory Bowel Disease. *Inflamm Bowel Dis*. 2004;10(5):661-665. doi:10.1097/00054725-200409000-00026.

6. Arijs I, Li K, Toedter G, et al. Mucosal gene signatures to predict response to infliximab in patients with ulcerative colitis. *Gut*. 2009;58(12):1612-1619. doi:10.1136/gut.2009.178665

7. Gerassy-Vainberg S, Starosvetsky E, Gaujoux R, et al. A personalized network framework reveals predictive axis of anti-TNF response across diseases. *Cell Rep Med*. 2024;5(1). doi:10.1016/j.xcrm.2023.101300

8. Lundberg SM, Allen PG, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. *Advances in Neural Information Processing Systems* 30. 2017. p. 4765-4774.

9. National Center for Biotechnology Information. GSE207465. *Gene Expression Omnibus*. November 8, 2022. Accessed June 5, 2026. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207465

10. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-210. doi:10.1093/nar/30.1.207

11. National Center for Biotechnology Information. GSE186963. *Gene Expression Omnibus*. November 1, 2024. Accessed June 5, 2026. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186963

12. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *The Lancet*. 1980;315(8167):514. doi:10.1016/s0140-6736(80)92767-1

13. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007

14. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.

15. Pedregosa F, Michel V, Grisel O, Blondel M, Prettenhofer P, Weiss R, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-30. Available from: http://scikit-learn.sourceforge.net

16. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery; 2016. p. 785-794. doi:10.1145/2939672.2939785.

17. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67. doi:10.1038/s42256-019-0138-9

18. Kozyrev S V., Abelson AK, Wojcik J, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat Genet*. 2008;40(2):211-216. doi:10.1038/ng.79

19. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med*. 2015;13(1). doi:10.1186/s12916-014-0241-z

Tables

Table 1. Cohort characteristics

Characteristic	Training cohort GSE207465 (ustekinumab)	Validation cohort GSE186963 (infliximab)
Array platform	HuGene 2.1 ST (GPL17586)	Clariom S (GPL23159)
Source publication	UNITI-2 trial (NCT01369329) Feagan et al. 2016[1]	Gerassy-Vainberg et al. 2024[7]
Total patients	353	23
Treatment	Ustekinumab (IV induction)	Infliximab (IV induction)
Post-treatment timepoint	Week 8	Week 14
Clinical outcome	Remission (CDAI <150)	Remission (symptom-based, week 14)[7]
Outcome achieved	140/353 (39.7%)	15/23 (65.2%)
CRP <5 mg/L at post-treatment	192/353 (54.4%)	13/23 (56.5%)
Baseline CRP, mean (SD) mg/L	15.9 (22.1)	18.7 (29.7)
Baseline CRP range (mg/L)	0.1-137.0	1.0-125.7

CRP: C-reactive protein; CDAI: Crohn's Disease Activity Index; IV: intravenous; SD: standard deviation. The validation cohort clinical outcome was defined by symptom-based criteria (cessation of diarrhea/cramping or fistula closure plus physician decision to continue therapy; partial responders adjudicated by algorithm incorporating CRP and calprotectin dynamics).[7]

Table 2. Model performance on the held-out test set (n=106) and external validation (n=23)

Model	CRP norm. AUC (95% CI)	Remission AUC (95% CI)	CRP norm. CV AUC (mean +/- SD)	Remission CV AUC (mean +/- SD)
Baseline CRP alone (logistic regression)	0.840 (0.756-0.911)	0.564 (0.454-0.675)	Not evaluated	Not evaluated
60-gene panel, RF (CRP excluded)	0.686 (0.580-0.784)	0.641 (0.525-0.751)	0.663 +/- 0.064	0.557 +/- 0.071
60-gene panel, XGBoost (CRP excluded)	0.682	0.616	0.635 +/- 0.089	0.531 +/- 0.074
25-gene cross-platform panel, RF (training test set)	0.628	0.624	0.640 +/- 0.095	0.545 +/- 0.102
25-gene cross-platform panel, RF (external validation)	Not evaluated*	0.517 (0.265-0.768)	--	--

*AUC: area under the receiver operating characteristic curve (0.5=chance; 1.0=perfect discrimination). Bootstrap 95% CIs (2,000 iterations) for all models. CV: five-fold stratified cross-validation on training set (n=247). RF: Random Forest. CRP: C-reactive protein. All gene panel models excluded baseline CRP as a feature. ROC curve for external validation shown in Supplementary Figure S1. *CRP normalisation external validation not performed as a pre-specified analysis.*

Figures

Figure 1

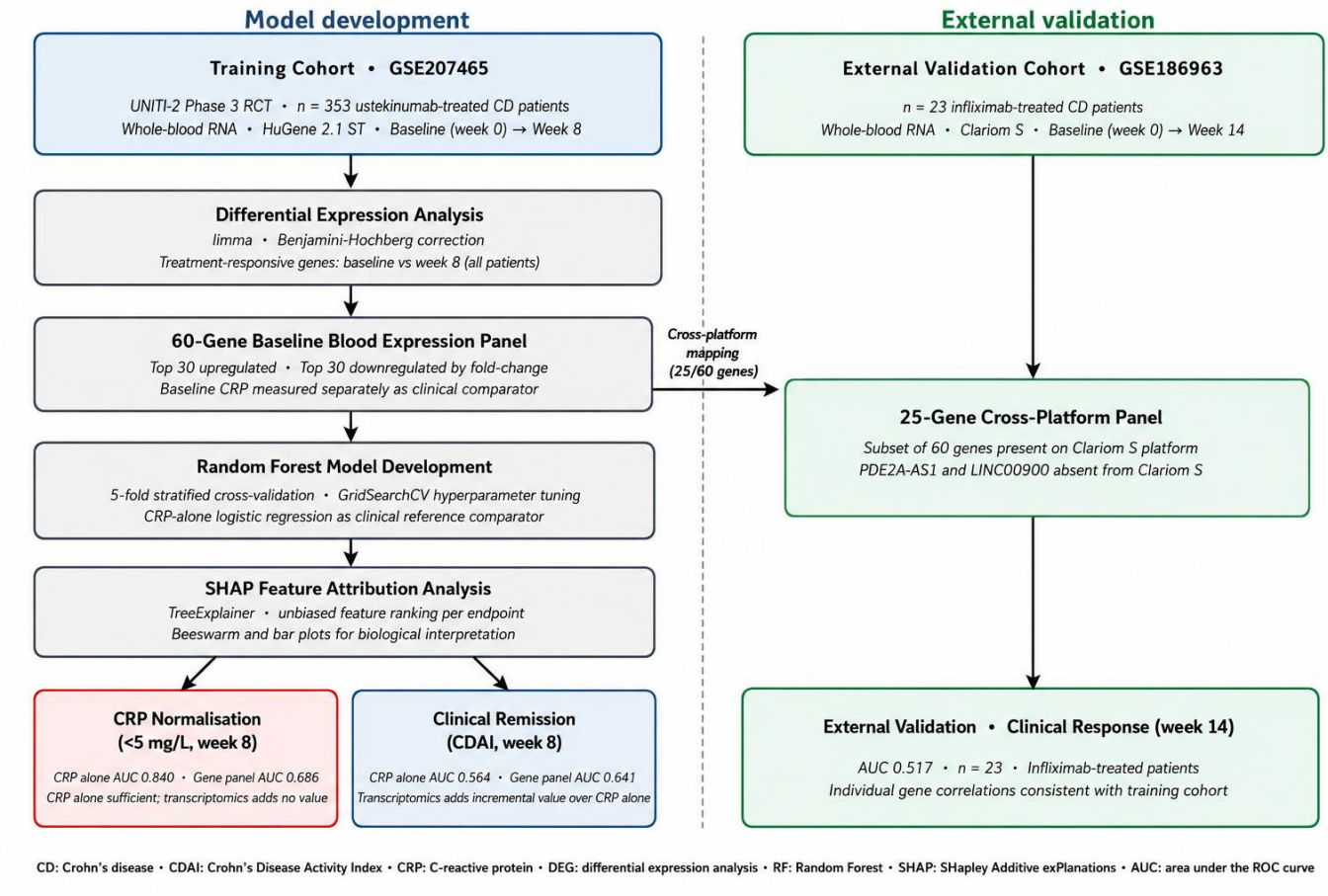


Figure 1. Study design schematic

Schematic overview of the study design. The training pipeline (left) comprised 353 ustekinumab-treated Crohn's disease patients from the UNITI-2 phase 3 RCT[1] (GSE207465[9]; whole-blood RNA, HuGene 2.1 ST). Remission-stratified differential expression analysis identified remission-specific transcriptional changes, from which a 60-gene baseline expression panel was selected. Random Forest models were trained for two co-primary endpoints with baseline CRP as a reference comparator; SHAP analysis identified the primary molecular determinants per endpoint. Cross-platform mapping identified 25 genes available on the Clariom S array for external validation (right) in 23 infliximab-treated patients (GSE186963[11], Gerassy-Vainberg et al.[7]). CD: Crohn's disease; CDAI: Crohn's Disease Activity Index; CRP: C-reactive protein; RF: Random Forest; SHAP: SHapley Additive exPlanations.

Figure 2

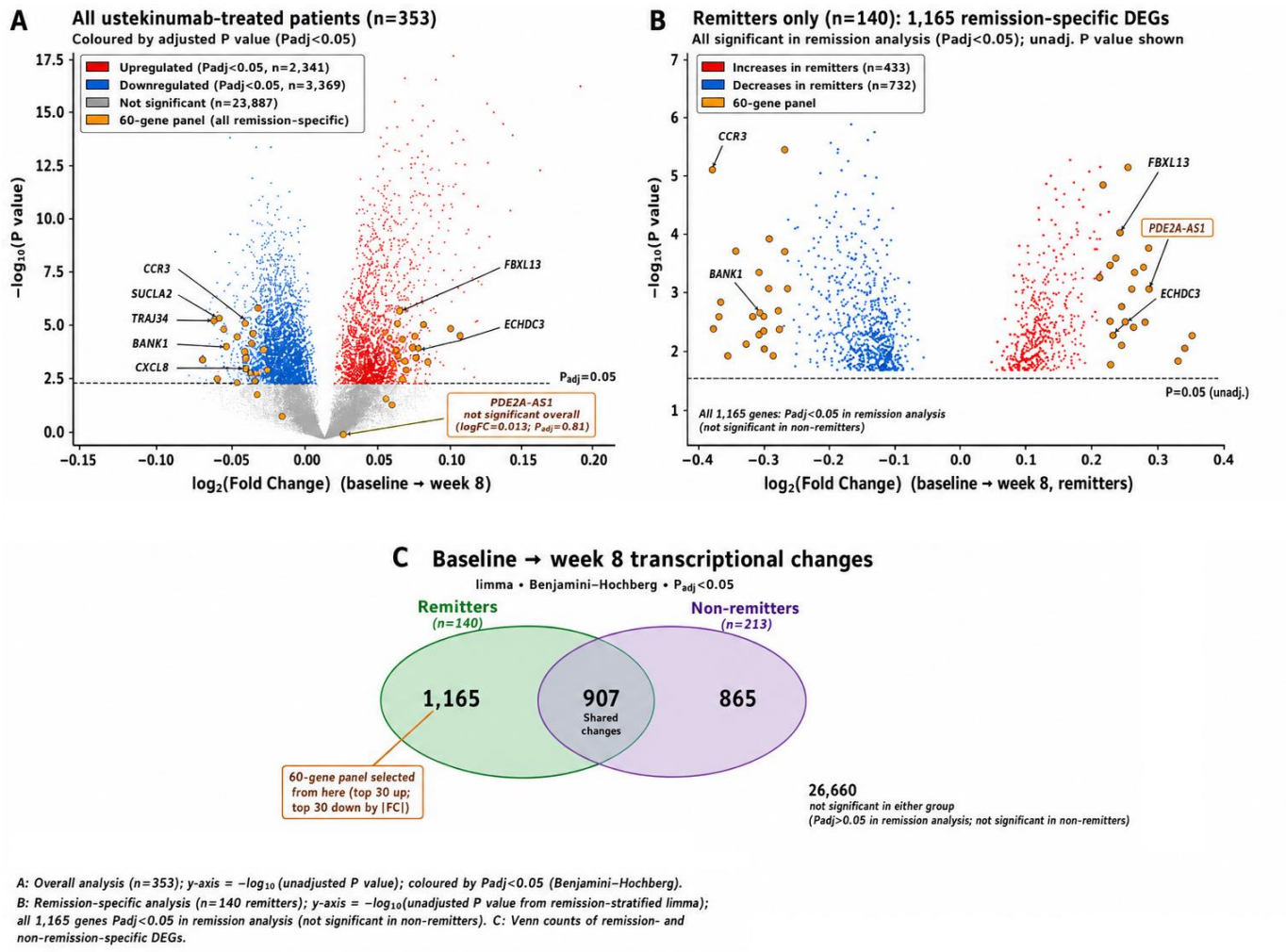
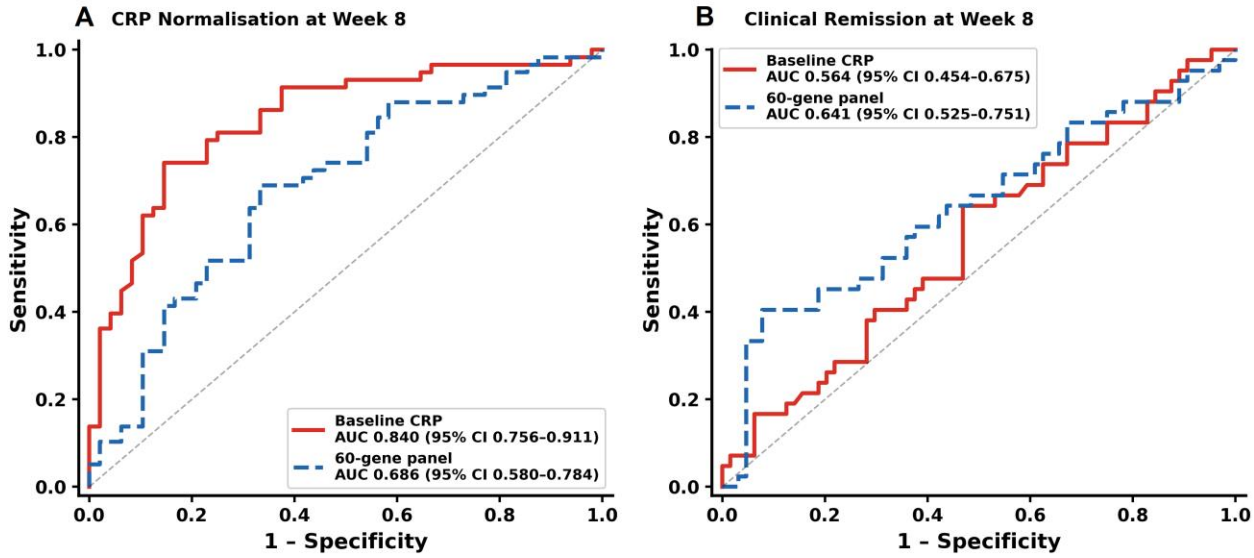


Figure 2. Transcriptional landscape of ustekinumab treatment response

Panel A: Volcano plot of differential expression (baseline versus week 8) across all 353 ustekinumab-treated patients. The y-axis shows -$\log_{10}$(unadjusted P value); the dashed threshold line indicates the unadjusted P value corresponding to the Benjamini-Hochberg adjusted P=0.05 boundary for this analysis. Red: upregulated (adjusted P<0.05, n=2,341); blue: downregulated (adjusted P<0.05, n=3,369); grey: not significant (n=23,887); orange: 60-gene panel. *PDE2A-AS1* (callout) is non-significant in the overall analysis (logFC: 0.013, adjusted P=0.808) despite being the top SHAP feature for the remission model. Panel B: Volcano plot of differential expression in remitters only (n=140; baseline versus week 8). All 1,165 genes shown are uniquely and significantly changed in remitters (adjusted P<0.05 in remission-stratified analysis; not significant in non-remitters). The y-axis shows -$\log_{10}$(unadjusted P value from the remission-stratified analysis); threshold line at unadjusted P=0.05 (all genes are above this threshold). Orange: 60-gene panel; *PDE2A-AS1* is significant here (logFC: +0.288). Note: axes differ between Panels A and B as they represent different analyses. Panel C: Venn diagram of differentially expressed genes (adjusted

P<0.05) from baseline to week 8 in remitters (n=140) and non-remitters (n=213). The 60-gene panel was selected from the 1,165 remission-specific genes.

Figure 3

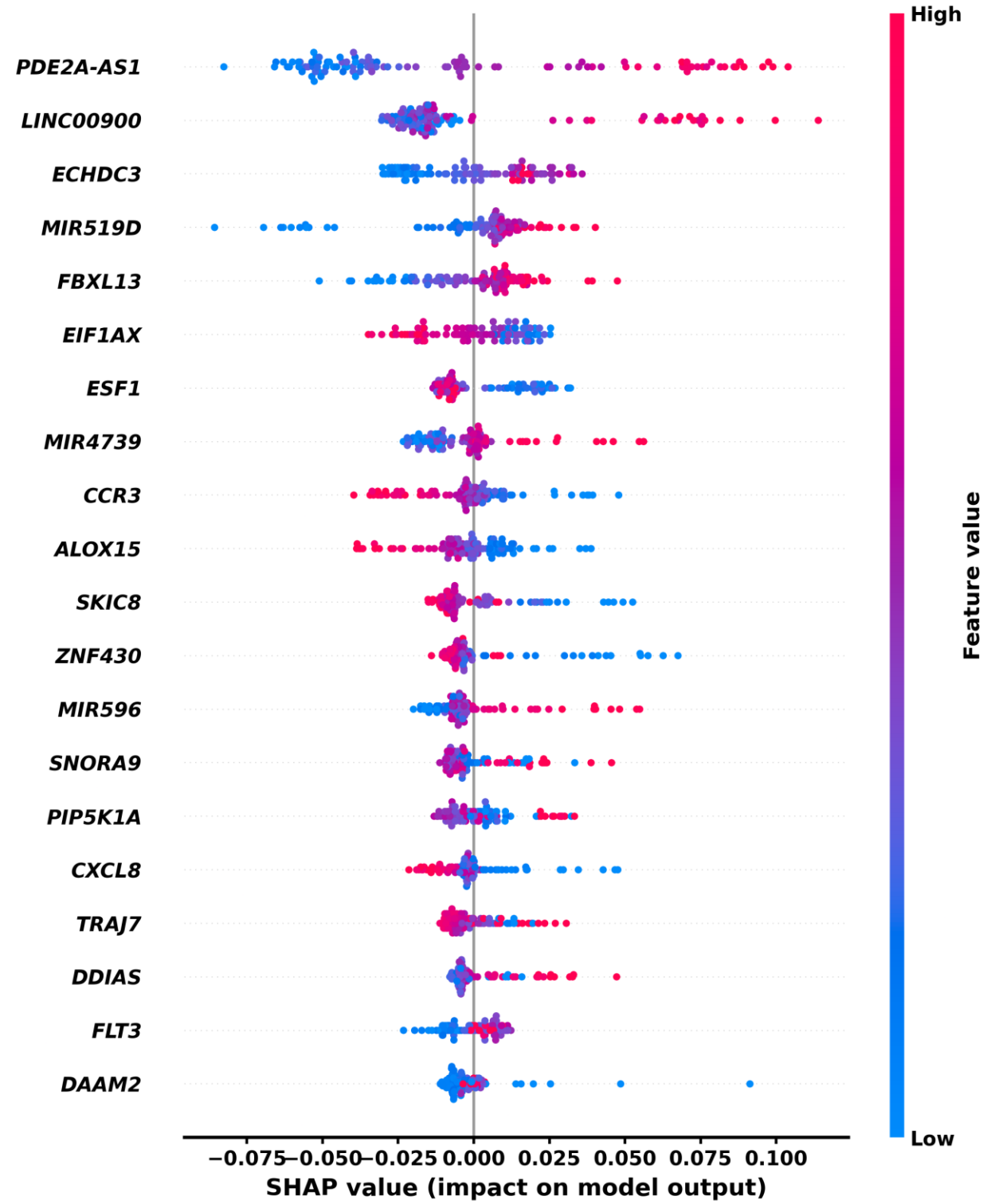


Gene panel: 60 baseline whole-blood transcriptomic features (baseline CRP excluded). AUC: area under the receiver operating characteristic curve. 95% CI: 2,000 bootstrap iterations, held-out test set (n=106).

Figure 3. Dissociation of CRP and transcriptomic predictive performance across endpoints
Receiver operating characteristic (ROC) curves comparing baseline CRP alone (logistic regression; red solid) with the 60-gene whole-blood transcriptomic panel (Random Forest, CRP excluded; blue dashed) for two co-primary endpoints. Panel A: CRP normalisation at week 8. Baseline CRP strongly predicts biochemical normalisation (AUC 0.840, 95% CI 0.756-0.911) whilst the gene panel performs substantially lower (AUC 0.686, 95% CI 0.580-0.784). Panel B: Clinical remission at week 8. CRP alone is near chance (AUC 0.564, 95% CI 0.454-0.675) whilst the gene panel achieves a higher point estimate (AUC 0.641, 95% CI 0.525-0.751) with overlapping confidence intervals, precluding definitive superiority claims. AUC: area under the ROC curve; 95% CI: 2,000 bootstrap iterations, held-out test set (n=106).

Figure 4

A


PDE2A-AS1
LINC00900
ECHDC3
MIR519D
FBXL13
EIF1AX
ESF1
MIR4739
CCR3
ALOX15
SKIC8
ZNF430
MIR596
SNORA9
PIP5K1A
CXCL8
TRAJ7
DDIAS
FLT3
DAAM2
−0.075 −0.050 −0.025 0.000 0.025 0.050 0.075 0.100
SHAP value (impact on model output)
High
Feature value
Low

B

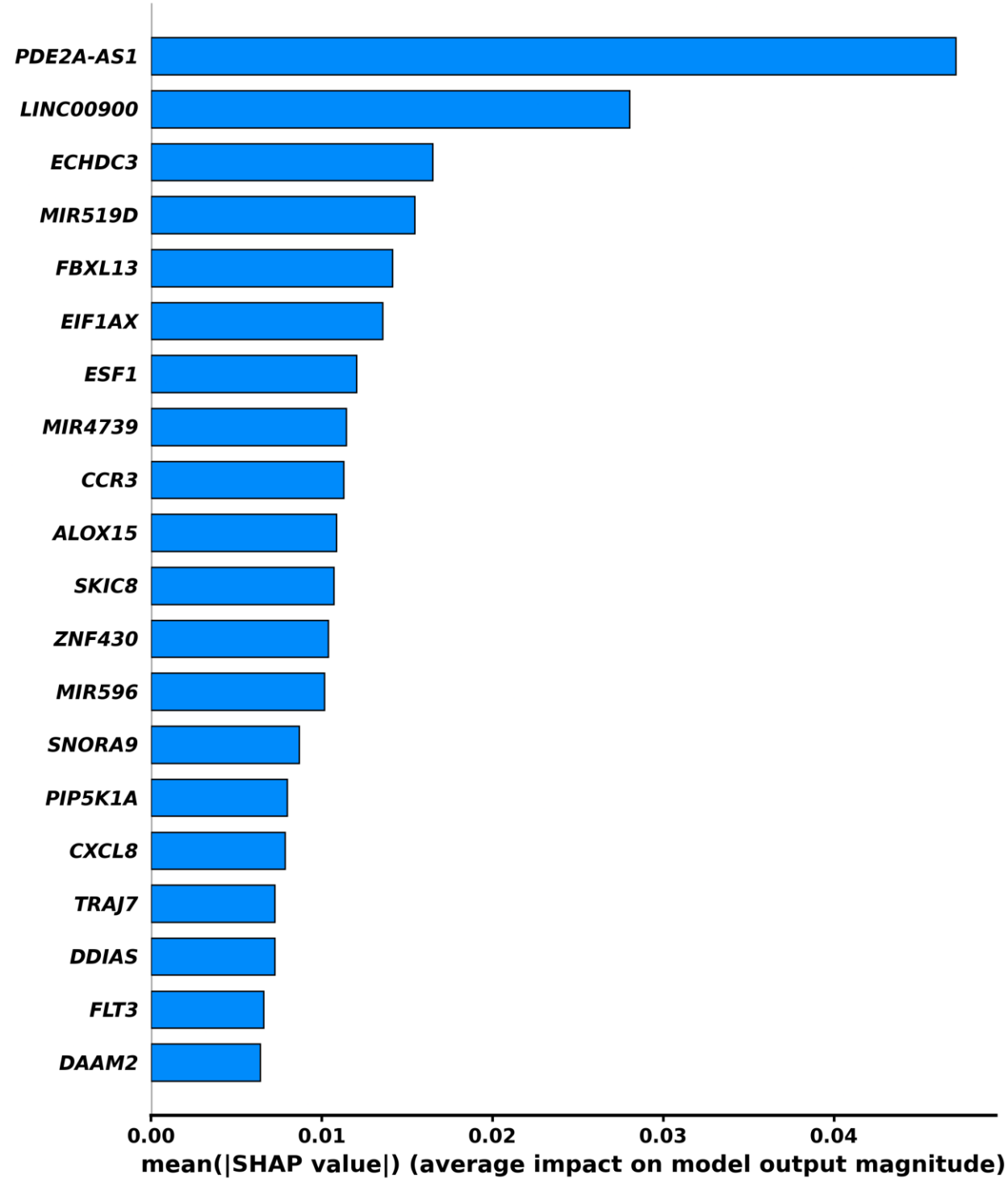
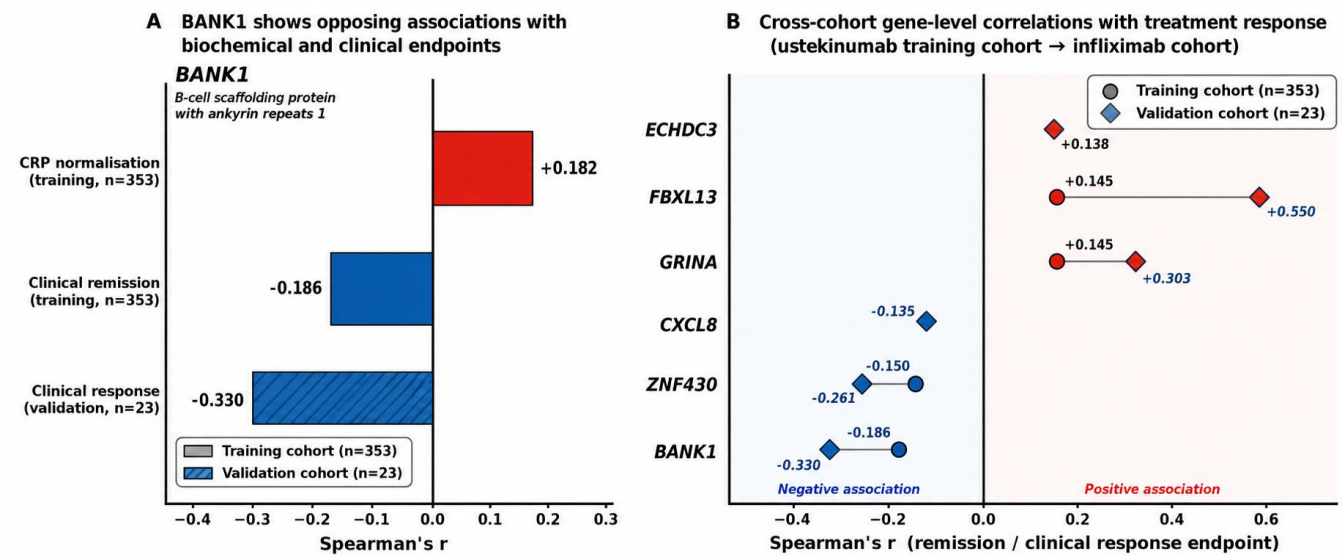


Figure 4. SHAP feature attribution for the Random Forest remission model

Panel A: Beeswarm plot of SHAP (SHapley Additive exPlanations) values for the top 20 features of the RF remission classifier, applied to the held-out test set (n=106). Each point represents one

patient; horizontal position = SHAP value (positive: pushes prediction towards remission; negative: away from remission). Point colour = baseline expression level (red: high, blue: low). Gene names are in italic. *PDE2A-AS1* is the top-ranked feature (mean absolute SHAP value 0.045); higher baseline *PDE2A-AS1* expression is associated with positive contributions towards remission prediction. Panel B: Bar chart of mean absolute SHAP value for the top 20 features, summarising overall feature contribution across all test-set patients.

Figure 5



Values are Spearman rank correlation coefficients. Correlations computed for the remission endpoint in the training cohort and the clinical response endpoint in the validation cohort. BANK1 appears in both panels; note opposing associations in Panel A.

Figure 5. BANK1 dual associations and cross-cohort consistency

Panel A: Horizontal bar chart showing Spearman rank correlation coefficients for *BANK1* (B-cell scaffold protein with ankyrin repeats 1) with CRP normalisation (training, rs=+0.182), clinical remission (training, rs=-0.186), and clinical response (validation cohort, n=23, rs=-0.330; hatched bar). *BANK1* demonstrates opposing associations with the two training-cohort endpoints, with directional consistency of the negative remission association in the independent validation cohort (Gerassy-Vainberg et al., GSE186963).[7] Panel B: Connected dot plot showing Spearman correlations with the remission or response endpoint for six genes with directional consistency across both cohorts. Filled circles: training cohort (n=353); filled diamonds: validation cohort (n=23). Positive associations (*ECHDC3, FBXL13, GRINA*) in red; negative associations (*CXCL8, ZNF430, BANK1*) in blue. Validation cohort used infliximab and a different response definition; associations should be treated as hypothesis-generating.