Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From universal biomarkers to individualized trajectory-based profiles: charting the multi-omic future of precision medicine in IBD

Dimitrios Kioroglou

dkioroglou@cicbiogune.es

CIC bioGUNE https://orcid.org/0000-0003-3495-4585

Montse Baldán-Martín

mbaldanm@gmail.com

Hospital Universitario de La Princesa

Rubén Gil-Redondo

CIC bioGUNE https://orcid.org/0000-0002-3541-9085

Nieves Embade

CIC bioGUNE https://orcid.org/0000-0001-9878-3290

Mikel Azkargorta

CIC bioGUNE

Monika Gonzalez-Lopez

CIC bioGUNE https://orcid.org/0000-0003-1506-0332

Ana M Aransay

CIC bioGUNE https://orcid.org/0000-0002-8271-612X

Isabel Moreno-Indias

Instituto de Investigación Biomédica de Málaga

Paula J. Martínez

Hospital Universitario de La Princesa

Fabio Suárez-Trujillo

Hospital Universitario de La Princesa

Alfredo J. Lucendo

Instituto de Investigación Sanitaria de Castilla-La Man- cha

Cristina Rodríguez Gutiérrez

nstituto de Investigación Sanitaria de Navarra

Ana Gutiérrez-Casbas

Instituto Investigación Sanitaria y Biomédica de Alicante

Eduardo Martín Arranz

Instituto de Investigación Sanitaria La Paz
Luis Bujanda Fernández de Piérola
Instituto Biodonostia
Iago Rodríguez-Lago
Hospital Universitario de Galdakao
Ruth M de Francisco García
Instituto de Investigación Sanitaria del Principado de Asturias
Eva Iglesias-Flores
Hospital Universitario Reina Sofía
Fernando Bermejo
Hospital Universitario de Fuenlabrada
Manuel Van Domselaar
Hospital Universitario de Torrejón
María José García
Hospital Universitario Marqués de Valdecilla and IDIVAL
Luis Fernández-Salazar
Hospital Clínico Universitario de Valladolid
Francisco J. Garcia-Alonso
Hospital Universitario Río Hortega
Xavier Calvet Calvo
Consorci Corporació Santitària Parc Taulí
Leyanira Torrealba
Hospital Universitario Dr. Josep Trueta
Iria Bastón-Rey
Hospital Clínico Universitario de Santiago
Hector Pallares Manrique
Hospital Juan Ramón Jiménez
Yolanda Ber Nieto
Hospital San Jorge, Huesca, Spain
Eduardo Leo Carnerero
Hospital Universitario Virgen del Rocío
María J. Casanova
Hospital Universitario de La Princesa
Francisco J. Tinahones
Instituto de Investigación Biomédica de Málaga
Felix Elortza
CIC bioGUNE
Oscar Millet
CIC bioGUNE https://orcid.org/0000-0001-8748-4105

Urko M. Marigorta
CIC bioGUNE https://orcid.org/0000-0002-9087-6609
María Chaparro
Hospital Universitario de La Princesa
Javier P. Gisbert
Hospital Universitario de La Princesa

Research Article

Keywords: precision medicine, multi-omics, bioinformatics, IBD, anti-TNF, biomarkers

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10968551/v1

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

Additional Declarations: The authors declare potential competing interests as follows: María Chaparro has served as a speaker or received research or educational funding from MSD, AbbVie, Hospira, Pfizer, Takeda, Janssen, Ferring, Shire Pharmaceuticals, Dr. Falk Pharma, Tillotts Pharma, Biogen, Gilead, and Lilly. Javier P. Gisbert has served as a speaker, consultant, or advisory board member, or received research funding from MSD, AbbVie, Pfizer, Kern Pharma, Biogen, Mylan, Takeda, Janssen, Roche, Sandoz, Celgene/Bristol Myers, Gilead /Galapagos, Lilly, Ferring, Faes Farma, Shire Pharmaceuticals, Dr. Falk Pharma, Tillotts Pharma, Chiesi, Casen Fleet, Gebro Pharma, Otsuka Pharmaceutical, Norgine, and Vifor Pharma. Eva Iglesias-Flores has served as a speaker, consultant, or advisory board member, or received research funding from MSD, AbbVie, Takeda, Janssen, Galapagos/Alfasigma, Lilly, Ferring, Dr. Falk Pharma, and Adacyte.

From universal biomarkers to individualized trajectory-based profiles: charting the multi-omic future of precision medicine in IBD

Dimitrios Kioroglou, PhD [1,*,#], Montse Baldán-Martín, PhD [2,28,*,#], Rubén Gil-Redondo, PhD [3], Nieves Embade, PhD [3], Mikel Azkargorta, PhD [4], Monika Gonzalez-Lopez, PhD [5], Ana M Aransay, PhD [5,28], Isabel Moreno-Indias, PhD [6], Paula J. Martínez, PhD [2,28], Fabio Suárez-Trujillo [2,28], Alfredo J. Lucendo [7,28], Cristina Rodríguez Gutiérrez, MD, PhD [8], Ana Gutiérrez-Casbas [9,28], Eduardo Martín Arranz [10], Luis Bujanda Fernández de Piérola [11], Iago Rodríguez-Lago [12], Ruth M de Francisco García [13], Eva Iglesias-Flores [14], Fernando Bermejo [15], Manuel Van Domselaar [16], María José García [17], Luis Fernández-Salazar [18], Francisco J. Garcia-Alonso [19], Xavier Calvet Calvo [20,28], Leyanira Torrealba [21], Iria Bastón-Rey [22], Hector Pallares Manrique [23], Yolanda Ber Nieto [24], Eduardo Leo Carnerero [25], María J. Casanova [2,28], Francisco J. Tinahones [6,29], Felix Elortza, PhD [4,28], Oscar Millet, PhD [3], Urko M. Marigorta, PhD [26,27,&], María Chaparro, MD, PhD [2,28,&], Javier P. Gisbert, MD, PhD [2,28,&]

[*] corresponding authors: dkioroglou@cicbiogune.es, mbaldanm@gmail.com
[#] authors share first authorship
[&] These authors contributed equally as senior authors.

Affiliations

1. Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Derio, Basque Country, Spain.

2. Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria Princesa (IIS-Princesa), Universidad Autónoma de Madrid (UAM).

3. Precision Medicine and Metabolism Laboratory, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Derio, Basque Country, Spain.

4. Proteomics Platform, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), ProteoRed-ISCIII, Bizkaia Technology Park, Derio, Basque Country, Spain.

5. GAP, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Derio, Basque Country, Spain. Carlos III Health Institute (ISCIII), 28029 Madrid, Spain.

6. IBIMA Plataforma BIONAND, Instituto de Investigación Biomédica de Málaga, 29590 Malaga, Spain; Clinical Management Unit of Endocrinology and Nutrition, University Hospital Virgen de la Victoria, 29010 Malaga, Spain; CIBER Physiopathology of Obesity and Nutrition (CIBERObn), Carlos III Health Institute (ISCIII), 28029 Madrid, Spain.

7. Hospital General de Tomelloso, Instituto de Investigación Sanitaria de Castilla-La Mancha (IDISCAM), Dept. de Gastroenterology, Ciudad Real, Spain.

8. Hospital Universitario de Navarra, Instituto de Investigación Sanitaria de Navarra (IdiSNA), Pamplona, Spain.

9. Hospital General Universitario de Alicante, Instituto Investigación Sanitaria y Biomédica de Alicante, Gastroenterology, Alicante, Spain.

10. Hospital Universitario La Paz and Instituto de Investigación Sanitaria La Paz (IdiPaz), Gastroenterology, Madrid, Spain.

11. Hospital Universitario Donostia and Instituto Biodonostia. Universidad del País Vasco (UPV/EHU), Donostia, Spain.

12. Hospital Universitario de Galdakao, Biocruces Bizkaia HRI, Gastroenterology, Vizcaya, Spain.

13. Hospital Universitario Central de Asturias and Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain.

14. Department of Gastroenterology, Hospital Universitario Reina Sofía, IMIBIC, UCO, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Córdoba, Spain.

15. Hospital Universitario de Fuenlabrada, Gastroenterology, Madrid, Spain.

16. Hospital Universitario de Torrejón, Madrid, Spain.

17. Hospital Universitario Marqués de Valdecilla and IDIVAL, Gastroenterology, Santander, Spain.

18. Department of Gastroenterology, Hospital Clínico Universitario de Valladolid, Instituto de Investigación Biosanitaria de Valladolid (IBioVALL), Department of Medicine, Dermatology and Toxicology, Universidad de Valladolid, Valladolid, Spain.

19. Hospital Universitario Rio Hortega, Valladolid, Spain.

20. Consorci Corporació Santitària Parc Taulí, Unitat de Malalties Digestives, Sabadell, Spain.

21. Hospital Universitario Dr. Josep Trueta., Girona, Spain.

22. Hospital Clínico Universitario de Santiago, Gastroenterology and Hepatology, Santiago de Compostela, Spain.

23. Hospital Juan Ramón Jiménez, Huelva, Spain.

24. Hospital San Jorge, Huesca, Spain.

25. Hospital Universitario Virgen del Rocío, Sevilla, Spain

26. Integrative Genomics Lab, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Derio, Basque Country, Spain.

27. Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Bizkaia, Spain.

28. Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Madrid, Spain.

29. Department of Medicine and Dermatology, Faculty of Medicine, University of Malaga, 29010 Malaga, Spain.

Abstract

Inflammatory bowel disease (IBD) exhibits highly heterogeneous responses to anti-TNF therapy. Previous biomarker discovery relied on single-omic and static timepoint hypothesis testing, yielding limited clinical efficacy and highlighting the need to evaluate dynamic trajectories. While multi-omics offer holistic insights, traditional single-omic paradigms still persist, whereas the challenges of their high dimensionality are overlooked. To address these collective limitations, we profiled six omic layers from tissue, fecal and serum samples of 33 IBD patients before and after anti-TNF treatment. We developed an analytical framework that operates in a biologically interpretable reduced feature space to extract coordinated multi-omic response trajectories. To isolate true therapeutic response from pathological fluctuations, trajectories were anchored to 20 non-IBD individuals. Finally, baseline comparisons against an independent cohort of 38 IBD patients that failed to respond to anti-TNF and received alternative treatments, highlighted the potential of the captured multi-omic profiles to serve as a framework for personalized therapy selection.

Introduction

Inflammatory bowel disease (IBD) is a chronic inflammatory condition of the gastrointestinal tract, primarily encompassing ulcerative colitis (UC) and Crohn's disease (CD). Despite the transformative impact of anti-TNF therapies on disease control and quality of life, clinical outcomes remain heterogeneous. Approximately one-third of patients fail to respond initially (primary non-response), while a substantial proportion lose response over time (secondary loss of response). This variability is an outcome of an intricate interactions between biological background, demographic characteristics and life style that highlights the persistent unmet need to guide therapy selection.

Early efforts to decipher anti-TNF treatment response predominantly relied on single-omics modalities. In this setting, response inference was typically based on univariate hypothesis testing frameworks, where individual molecular features were tested independently for differential abundance between responders and non-responders IBD patients. While this approach has been instrumental in identifying candidate biomarkers, it inherently assumes feature independence and often fails to capture coordinated biological processes, limiting its predictive and mechanistic utility [1]. Moreover, significance thresholds implemented in hypothesis testing frameworks impose over-reliance on binary decision-making from a continuous measurement [2]. This practice can exaggerate the importance of threshold-crossing findings while underemphasizing cohort-dependent variability, thereby limiting reproducibility and interpretability across IBD studies.

Furthermore, relying solely on baseline levels to identify features that differentiate future responders from non-responders is a common practice [3, 4, 5, 6]. However, this approach may capture natural fluctuations in baseline molecular levels, without establishing whether the selected features interact with the treatment. Consequently, features that appear significantly associated with treatment response in one cohort may fail to replicate, or even reverse directionality in another [5, 6]. This instability is further exacerbated by small sample sizes and heterogeneity in disease severity, treatment history, and environmental exposures. These challenges have led to the realization that baseline biomarkers are often insufficient, highlighting a critical need to focus on dynamic response trajectories rather than single timepoints [7].

Advancements in high-throughput technologies have enabled the generation of deeply characterized multi-omic datasets, offering a holistic view of the molecular landscape regulating treatment response in IBD [8]. However, traditional workflows frequently analyze these multi-omic layers in isolation [9, 10], thereby limiting our capacity to fully capture the complex cross-talk dynamics that occur between them. Moreover, the high-dimensional feature space inherent to multi-omics data, combined with limited sample sizes can produce unstable estimates even when regularized approaches are implemented [11]. Consequently, there is a need to develop methodological strategies capable of working with compressed signals derived from high-dimensional feature spaces, capturing coordinated multi-omic dynamics rather than isolated signals.

In the current study, we generated a comprehensive multimodal dataset spanning six omic layers from a cohort of 33 IBD patients, sampled longitudinally at baseline and 14-weeks after anti-TNF treatment. Our analytical framework was specifically designed to overcome the computational and statistical hurdles of multi-omics data and address the previous described methodological limitations. Thus, we implemented dimensionality reduction to handle the high-dimensional feature space, and extracted latent biological variation associated with coordinated multi-omic response trajectories. To prevent these trajectories from merely capturing natural fluctuations within the disease state, we anchored therapeutic response to an

inflammation-resolved state by cross-sectionally profiling an additional reference group of 20 non-IBD individuals. Moreover, feature selection was driven by effect-size evaluation and Bayesian modeling was implemented to derive credible estimates and robustly quantify statistical uncertainty.

Through this approach, we demonstrate how leveraging compressed multi-omic signals alongside reference-informed trajectories can uncover molecular profiles that capture treatment-response dynamics. Finally, cross-cohort pre-treatment comparisons with IBD patients that failed anti-TNF therapy and received alternative treatments, revealed relationships that underscore the potential of these profiles to serve as a framework for individualized therapy selection.

Methods

M.1 Ethics and study cohort

The cohort involved naive to biologics IBD patients from an observational, prospective, multicenter and non-commercial study involving the use of anti-TNF therapy based on clinical indication (BioIBD cohort). The selection of a specific therapeutic anti-TNF agent was at the discretion of the treating physician. No randomized treatment assignment was performed, and all drugs were used according to their approved indications and prescribing conditions for IBD.

The disease activity was determined by simple endoscopic score (SES-CD) for Crohn's disease and the Mayo endoscopic subscore (MES) in ulcerative colitis, and evaluated before initiation of the treatment (baseline) and at week-14. Endoscopic response in CD was defined as a $> 50\%$ decrease in SES-CD 14-weeks after initiating the treatment. In UC patients, endoscopic response was defined as a ≥ 1 point decrease in MES 14-weeks after initiating treatment. At baseline, serum, fecal and intestinal tissue samples were obtained from all included patients.

We additionally incorporated cross-sectional samples from non-IBD participants, whose intestinal tissue samples were obtained from colonscopy performed due to reasons of colorectal cancer screening, changes in bowel habits, or rectal bleeding. The ileocolonic mucosa of these participants was macroscopically and histologically normal, as confirmed both from a colonoscopic and histopathological evaluation.

The study protocol was registered in both European and United States clinical trial registries (EudraCT number 2018-002925-47 and ClinicalTrials.gov identifier NCT03885713). Prior to sample collection, written informed consent was obtained from all participants. The protocol was reviewed and approved by local institutional review boards. The study was conducted in accordance with the clinical and ethical principles of the Declaration of Helsinki. All experimental procedures were approved by the Ethics Research Committee of the coordinating hospital (Hospital Universitario de La Princesa, Madrid, Spain). All data were coded to protect participant confidentiality.

M.2 Integrated multi-omics and inflammatory metrics

To monitor systemic inflammation, we measured the levels of C-reactive protein (CRP) and quantified with NMR the glycoprotein acetylation A and B (GlycA, GlycB) and the supramolecular phospholipid composite (SPC) from serum samples.

Regarding multi-omics, we profiled six omics using intestinal tissue, fecal, and serum samples. Tissue-level analyses included proteomics alongside coding (mRNA) and non-coding (microRNA) transcriptomics, fecal samples underwent 16S amplicon sequencing (metataxonomics) and serum analysis comprised NMR-based lipoproteomics and metabolomics.

While non-IBD patients were sampled at a single timepoint, IBD patients were profiled longitudinally at two timepoints: baseline (pre-treatment) and week-14 (post-treatment).

M.3 Multi-omics sample preparation, data curation and confounders adjustment

Sample preparation for each omic is provided in the supplementary methods (SM1). Here we report the data curation and confounders adjustment each omic underwent individually prior to multi-omic integration and modeling.

M.3.1 Transcriptomics data curation

After removing genes and microRNAs with fewer than 10 counts in more than 90% of the samples, we corrected for differences in sequencing depth and stabilized count variance using the variance-stabilizing transformation (VST) in pyDESeq2 [12]. The resulting VST-normalized counts were then utilized for all downstream analyses.

M.3.2 Proteomics data curation

Data curation was performed using the Perseus platform [13] where initially we corrected batch effects using the available ComBat plugin. Intensities were then $\log_2$-transformed, filtered to remove proteins with excessive missing values, and missing data were imputed from a normal distribution. Final intensities were normalized using quantile normalization.

M.3.3 Metataxonomics data curation

After filtering low quality sequences, we inferred and quantified 16S amplicon sequence variants (ASVs) using DADA2[14] within the QIIME 2 framework[15], adapting the parameters specifically for Ion Torrent data. Taxonomic assignment was performed using a VSEARCH[16] consensus classifier against the Silva reference database (v138.1)[17], clustered at 99% identity and ASV counts were aggregated at genus level. To account for the compositional nature of the data and stabilize variance, raw ASV counts were transformed using a centered log-ratio (CLR) transformation via the SCIPY [18] Python package prior to downstream analysis.

M.3.4 Lipoproteomics, metabolomics and inflammation biomarkers data curation

Metabolic features were quantified employing specialized Bruker software suites. Specifically, 41 metabolites were derived using B.I.Quant-PS 2.1.0, 112 lipoprotein subfractions were extracted via B.I.LISA 1.1.0 PL-5009-01/002, and 5 distinct inflammatory parameters were evaluated using PhenoRisk PACS RuO 1.1.0. Alongside quantification, automated Bruker BioBank QC reports were generated for each sample to assess analytical and preparative performance. Expert personnel meticulously reviewed these reports, monitoring critical spectral indicators such as linewidth and residual water signals, ensuring that only samples passing all rigorous quality thresholds were included in the final dataset.

M.3.5 Confounder adjustment across curated omic features

For each patient, metadata regarding intestinal tissue location and demographic characteristics, such as gender, age, body mass index (BMI) and smoking, were recorded. To account for the confounding effects of these physiological covariates on the curated levels of the omic profiles, we utilized linear mixed-effects models implemented via the Python package STATSMODELS [19]. For each individual omic feature, a model was fitted using the longitudinal samples from IBD patients and cross-sectional samples from non-IBD controls as reference, with patient identity specified as a random effect. To ensure that the disease-specific signal was not absorbed by the confounders due to correlation, disease status was included in the model as a fixed effect. This variable was stratified into five distinct states: non-IBD (reference group), responder at baseline or week-14, and non-responder at baseline or week-14. For omic features where the patient random-effect variance was estimated as zero (singular fit), confounding effects were instead removed using ordinary least-squares (OLS) regression.

M.4 Dimensionality reduction of omic profiles and creation of modules

M.4.1 Dimensionality reduction of proteomics

To reduce the dimensionality of the proteomics data, we utilized the 50 distinct biological pathways defined by the Molecular Signatures Database (MSigDB) [20] Hallmark gene sets. For each pathway, a protein expression matrix was constructed containing the curated, confounders-adjusted levels of its constituent proteins (columns) across all longitudinal IBD samples and cross-sectional non-IBD controls (rows). Each matrix was then scaled to unit variance and principal component analysis (PCA) was subsequently performed to extract the first principal component (PC1) that included the corresponding patients scores (eigenscores). Each score represents a linear combination of the proteins abundances of a given pathway for a patient at a given timepoint. Finally, we compiled these 50 PC1 vectors into a new summary matrix. We refer to the columns of this matrix as modules in the manuscript.

M.4.2 Dimensionality reduction of transcriptomics

While the dimensionality reduction approach for mRNA transcriptomics mirrored that of the proteomics pipeline, miRNA transcriptomics utilized a target-centric mapping strategy. To map miRNAs to functional Hallmark pathways, we utilized the MSigDB C3 target gene module to identify the downstream gene targets for each miRNA. We then cross-referenced these targets with the 50 Hallmark gene sets and performed an over-representation analysis using the hypergeometric test in SCIPY. To ensure robustness, following FDR-correction, we only retained pathways that were significantly enriched (padj ≤ 0.05) and contained more than one mapping miRNA. Finally, for each selected pathway, a matrix of the constituent miRNA curated levels was subjected to PCA after scaling to unit variance, and the PC1 scores (eigenscores) were extracted to construct a new pathway-level activity summary matrix. We refer to the columns of this matrix as modules in the manuscript.

M.4.3 Dimensionality reduction of metataxonomics, lipoproteomics and metabolomics

For these omics, we stratified features into biologically and functionally interpretable feature subsets. Specifically, fecal metataxonomic data were grouped at the family level, while lipoproteomic and metabolomic features were grouped into subsets reflecting distinct biological axes (supplementary file 1). Each resulting sub-matrix was then scaled to unit variance and subjected to PCA. Finally, the PC1 scores (eigenscores) were extracted for each patient to construct an omic-specific summary matrix. We refer to the columns of this matrix as modules in the manuscript.

M.5 Multi-omic integration and identification of latent factors that capture biological axes of variation

To integrate the available omics in an unsupervised manner, we employed the Python framework for multi-omics factor analysis (MOFA) [21]. MOFA is a statistical framework designed to decompose the sources of heterogeneity in multi-omics datasets by inferring an interpretable, low-dimensional representation of the data via latent factors, usually denoted as Z. It represents a generalization of the classical PCA, meaning that the latent factors (Z) capture major sources of variation across multiple omics instead of a single omic.

To proceed with the integration, we joint the available curated and dimensionally reduced into modules multi-omic profiles from non-IBD and IBD patients, including omics from both

baseline and week-14 timepoints. To ensure a completely unbiased, data-driven discovery of the underlying biological variation, indication of disease status, response status and longitudinal timepoints was intentionally omitted from the model specification. Moreover, the model was configured with omic-level scaling to prevent feature-rich omics from dominating the analysis, and it was initialized to identify a maximum of 10 latent factors using Gaussian likelihoods to model the data distributions. Finally, to ensure robustness and focus subsequent downstream analyses on the most informative axes of variation, an automated factor-filtering criterion was applied, instructing MOFA to drop any factor with total captured variance $R^2 < 0.1\%$.

M.6 Latent factors polarity alignment and standardization

To ensure biological interpretability and comparability across the latent factors (Z) identified by MOFA, a post-processing normalization step was applied to the factor scores and their corresponding weights (W).

This step is required due to the sign-flipping invariance inherent in MOFA's matrix factorization, which renders the initial orientation of the latent factors arbitrary. Therefore, to establish a biologically intuitive directionality, we evaluated the baseline state of each factor Z_k. Specifically, if the mean factor score of the IBD patients at baseline was lower than the mean of the non-IBD individuals, the signs of both the factor scores (Z_k) and their corresponding feature weights (W_k) were inverted. This directional anchoring provides a uniform orientation across all factors, allowing downstream analysis to be intuitively interpreted along an axis of disease severity versus non-IBD resolution.

Furthermore, the raw factor scores exhibit arbitrary scales that depend on the input dimensions and variance of the underlying omics. To contextualize these latent axes within a clinically meaningful reference frame, the factor scores were standardized via z-score transformation using the empirical mean and standard deviation of the non-IBD individuals.

M.7 Hierarchical Bayesian modeling of latent factors

To model the longitudinal trajectories and treatment-response dynamics of the standardized latent factors, we implemented a linear mixed-effects framework within a hierarchical Bayesian architecture using PyMC [22].

Each standardized factor score ($Z_{i,k}$) for patient i and factor k was assumed to follow a normal distribution centered around a patient-specific expected mean ($\mu_{i,k}$) with observation-specific residual variance (σ_k^2):

$$Z_{i,k} \sim \mathcal{N}(\mu_{i,k}, \sigma_k) \tag{1}$$

To capture the precise interaction of disease, time, and treatment response, the expected mean ($\mu_{i,k}$) for each observation was modeled using the following linear predictor:

$$\begin{aligned} \mu_{i,k} = \beta_{0,k} + \\ \beta_{1,k}\text{IBD}_i + \\ \beta_{2,k}(\text{IBD}_i \times \text{time}_i) + \\ \beta_{3,k}(\text{IBD}_i \times \text{response}_i) + \\ \beta_{4,k}(\text{IBD}_i \times \text{time}_i \times \text{response}_i) + \\ u_i \end{aligned} \tag{2}$$

In the above linear predictor of $\mu_{i,k}$, we parameterized the model to account the asymmetric design of the cohort, where the IBD population was sampled longitudinally whereas the non-

IBD cross-sectionally. To do so, we explicitly nested the time, response, and interaction terms within a binary IBD indicator variable ($\text{IBD}_i = 1$ for IBD, 0 for non-IBD). This way the model successfully isolates temporal and clinical variance strictly to the patient populations in which they were actually observed.

To account for the dependency structure of repeated measures (baseline and week-14) within individual subjects, a patient-specific random intercept (u_i) was introduced to the linear model. Moreover, given our relatively small sample size ($N = 53$), standard centered parameterizations of random effects can suffer from poor geometry (the "Neals funnel" problem), leading to divergent transitions and inefficient Markov Chain Monte Carlo (MCMC) sampling. To circumvent this, we implemented a non-centered parameterization for the random intercepts. The patient-specific offsets were modeled deterministically by scaling a standard normal vector by a global group standard deviation:

$$\begin{aligned} u_{\text{raw},i} &\sim \mathcal{N}(0,1), \quad \text{for subject } i \in \{1,\ldots,53\} \\ \sigma_{u,k} &\sim \text{Half-Normal}(0.5) \\ u_i &= u_{\text{raw},i} \times \sigma_{u,k} \end{aligned} \tag{3}$$

Weakly informative, regularizing priors were assigned to all parameters to stabilize posterior estimation while allowing the experimental data to drive inference. The exact prior structure for each factor k was defined as:

$$\begin{aligned} \beta_{0,k} &\sim \mathcal{N}(\mu = 0, \sigma = 0.5) \\ \beta_{1,k}, \beta_{2,k}, \beta_{3,k}, \beta_{4,k} &\sim \mathcal{N}(\mu = 0, \sigma = 1.0) \\ \sigma_k &\sim \text{Half-Normal}(\sigma = 0.5) \end{aligned} \tag{4}$$

Moreover, the fixed-effect regression coefficients (β) can be interpreted as follows:

- $\beta_{0,k}$: Represents the intercept that establishes the baseline reference state, capturing the mean of the non-IBD individuals. Because the latent factors were previously standardized using this group, a narrower prior ($\sigma = 0.5$) was applied to reflect the empirical knowledge that this anchor should closely approximate a non-IBD physiological baseline.

- $\beta_{1,k}$: Quantifies the baseline disease effect. This term statistically controls for baseline variations between IBD and non-IBD patients, ensuring that the longitudinal effects estimated by the model reflect true clinical trajectories rather than initial differences in disease status.

- $\beta_{2,k}$: Measures the primary longitudinal trajectory over time within the overall IBD population, capturing the background temporal effect independent of treatment success.

- $\beta_{3,k}$: Controls for pre-existing biological variations between future responders and non-responders at baseline.

- $\beta_{4,k}$: This term isolates the specific longitudinal trajectory of treatment responders over time relative to non-responders, directly testing the primary hypothesis of whether successful therapy drives the latent factor profile back toward the non-IBD reference space defined by the intercept ($\beta_{0,k}$).

Finally, posterior distributions were sampled using the No-U-Turn Sampler (NUTS), an advanced Hamiltonian Monte Carlo (HMC) algorithm extension implemented in PyMC. Convergence of the independent MCMC chains was rigorously assessed using the Gelman-Rubin

diagnostic, visual inspection of trace plots and effective sample sizes (ESS) to confirm robust parameter estimation and the complete elimination of divergent transitions.

M.8 Reconstruction-based identification and selection of high-trajectory modules

To isolate the specific biomolecular modules driving distinct clinical trajectories between treatment responders and non-responders , we performed a multi-staged module selection workflow on the reconstructed module-based omic profiles. This process prioritized modules exhibiting a robust, treatment-associated shift over time while screening out baseline noise and uninformative biological variations.

Initially, using the specific latent factors that demonstrated clinically credible longitudinal dynamics (Z_1,Z_3 and Z_5) along with their corresponding feature loading weights, we reconstructed the modules for all omics that met a variance threshold of cumulative $R^2 > 10\%$. This step effectively filtered out background biological variations unlinked to the therapeutic response.

Following reconstruction, we performed a robust trajectory quantification. More specifically, within each IBD-subtype (UC and CD) and for each module, group medians were calculated using a robust leave-one-subject-out (LOSO) resampling framework to prevent individual patient outliers from biasing the cohort trends. Using these stable metrics, we calculated two parameters that are referred to as module homeostatic displacement (MHD) and standardized response-associated trajectory gap (SRG). The MHD metric refers to the net biological distance of a module from a given IBD patient to the non-IBD state, and was calculated using the Euclidean distance between the corresponding baseline and week-14 vectors. The SRG metric represents the z-score standardized absolute difference between the median MHDs of responders versus non-responders.

$$\text{MHD}_{i,m} = \left\| \vec{v}_{i,m}^{\,\text{w14}} - \vec{v}_{i,m}^{\,\text{baseline}} \right\|_2, \quad \forall i \in \{1, \ldots, N_{\text{IBD}}\}, \forall m \in \{1, \ldots, N_{\text{omic-modules}}\} \tag{5}$$

$$\text{RG}_m = \left| \text{median}(\text{MHD})_m^{\text{Rs}} - \text{median}(\text{MHD})_m^{\text{NRs}} \right|, \quad \forall m \in \{1, \ldots, N_{\text{omic-modules}}\} \tag{6}$$

$$\text{SRG}_m = \frac{\text{RG}_m - \mu_{\text{RG}}}{\sigma_{\text{RG}}}, \quad \forall m \in \{1, \ldots, N_{\text{omic-modules}}\} \tag{7}$$

Finally, we discarded modules where the median values of both IBD groups at both timepoints fell entirely within the interquartile range (IQR) of the non-IBD individuals, as well as those where the variance captured by PC1 fell below the random-noise threshold defined by the Marchenko-Pastur distribution limit. From the remaining modules, we selected those with SRG ≥ 1.

M.9 Associative Modeling of treatment response via baseline distance to non-IBD state

To evaluate whether baseline latent factors Z_1,Z_3 and Z_5 divergence from non-IBD state is associated with subsequent therapeutic response, we developed a two-stage computational framework combining distance-based metrics with Bayesian logistic regression.

Initially, we constructed a vector $\vec{z}_i$ for each IBD patient i at baseline using the corresponding values of the latent factors Z_1,Z_3 and Z_5.

$$\vec{z}_i^{\text{baseline}} = \begin{bmatrix} z_{1,i} \\ z_{3,i} \\ z_{5,i} \end{bmatrix} \tag{8}$$

We then performed the same for every j-th non-IBD individual.

$$\vec{z}_j = \begin{bmatrix} z_{1,j} \\ z_{3,j} \\ z_{5,j} \end{bmatrix} \tag{9}$$

Subsequently, we calculated a disease divergence metric by computing the Euclidean distance d_{ij} between the baseline latent factor vector $\vec{z}_{i,baseline}$ of every i-th IBD patient and the vector $\vec{z}_j$ of every j-th non-IBD individual.

$$d_{ij} = \left\| \vec{z}_i^{\text{baseline}} - \vec{z}_j \right\|_2, \quad \forall i \in \{1, \ldots, N_{\text{IBD}}\}, \forall j \in \{1, \ldots, N_{\text{non-IBD}}\} \tag{10}$$

To minimize the impact of individual outlier reference profiles and obtain a robust estimate of a patient's baseline divergence from the non-IBD state, we calculated the net baseline Euclidean distance (D_i) for each IBD patient as the median of these pairwise distances:

$$D_i = \text{median}\big(\{d_{ij} \mid j = 1, \ldots, N_{\text{non-IBD}}\}\big), \quad \forall i \in \{1, \ldots, N_{\text{IBD}}\} \tag{11}$$

To evaluate the association between this baseline multi-omic latent divergence and the eventual clinical outcome at week-14, we constructed a Bayesian logistic regression model with PyMC. Due to sample size constraints within IBD subtypes, explicitly partitioning UC and CD into a hierarchical structure led to severe MCMC convergence issues and parameter divergences. Thus, UC and CD cohorts were pooled into a unified IBD group to guarantee stable posterior estimation.

The probability of treatment response (p_i) for IBD patient i was modeled as a function of their baseline net distance (D_i) alongside clinical confounders after applying scaling to the continuous covariates (D_i, age and BMI).

$$\text{logit}(p_i) = \alpha + \beta_1 D_i + \beta_2 \text{Age}_i + \beta_3 \text{BMI}_i + \beta_4 \text{Sex}_i + \beta_5 \text{Smoking}_i \tag{12}$$

$$\text{Response}_i \sim \text{Bernoulli}(p_i) \tag{13}$$

Weakly informative priors were assigned to the intercept (α) and all standardized regression coefficients to regularize the parameter space and prevent overfitting:

$$\begin{aligned} \alpha &\sim \text{Normal}(\mu = 0, \sigma = 1.5) \\ \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 &\sim \text{Normal}(\mu = 0, \sigma = 1) \end{aligned} \tag{14}$$

Posterior distributions were sampled using the NUTS sampler and convergence of the independent MCMC chains was assessed using the Gelman-Rubin diagnostic, visual inspection of trace plots and effective sample sizes (ESS) to confirm robust parameter estimation.

M.10 Dimensionality reduction and visualization of high dimensional data

To visualize high-dimensional data, we employed the uniform manifold approximation and projection algorithm (UMAP), a non-linear dimensionality reduction approach chosen for its ability to preserve both the local and global geometric structure of the data. To ensure strict

visual comparability across different datasets, UMAP was performed using an identical parameter set for all analyses. The number of nearest neighbors ($n_neighbors$) was set to 50 to prioritize the preservation of global data structure and broader cluster relationships over fine-grained and localized noise. The minimum distance (min_dist) between embedded points was set to 0.5 to allow for a clearer, more interpretable visualization of distinct cluster densities. Additionally, the cosine distance metric was utilized to compute topological spaces based on directional feature correlation rather than magnitude. Prior to dimensionality reduction, all datasets were scaled to unit variance and projections were then generated using the umap-learn [23] Python package.

M.11 RLR-transformation of selected modules features

For each omic, we created the union of the features of those modules that displayed SRG ≥ 1 and were shared between the UC and CD groups. Then, we normalized each feature using a reference-group log2-ratio (RLR) transformation, calculating the log2-ratio of the unconfounded levels of each IBD patient relative to the geometric mean of the non-IBD group.

$$G_k = \sqrt[N_{\text{NIBD}}]{\prod_{j=1}^{N_{\text{NIBD}}} x_{jk}}, \quad \forall k \in \{1, \ldots, N_{\text{features}}\} \tag{15}$$

$$\text{rlr}(x_{ik}) = \log_2\left(\frac{x_{ik}}{G_k}\right), \quad \forall i \in \{1, \ldots, N_{\text{IBD}}\}, \forall k \in \{1, \ldots, N_{\text{features}}\} \tag{16}$$

M.12 Consolidation of supplementary cohorts with anti-TNF cohort and processing

We incorporated the same multi-omic datasets from three additional patients groups, two UC and one CD, that we had generated before treatment and with documented endoscopic response status at week-14. After merging the supplementary cohorts with the anti-TNF cohort, we applied the same data curation and confounders adjustment pipeline as described in the Methods M.3. Note that confounders impact was adjusted using only OLS regression. Next, for each omic, we considered the union of module features that we selected with the anti-TNF cohort and applied RLR-transformation to the unconfounded feature levels.

M.13 Patients classification for dyslipoproteinemias

We implemented the diagnostic algorithm provided by Sniderman et al. [24] to identify participants with potential underlying dyslipoproteinemias. The algorithm resembles a decision-tree that based on the levels of total cholesterol, triglycerides and apolipoprotein B (ApoB) classifies the individuals into six major dyslipoproteinemias: chylomicrons, chylomicrons+VLDL, chylomicrons+VLDL remnants, VLDL, LDL and VLDL+LDL.

Results

The study cohort comprised 53 participants, categorized into non-IBD individuals (n=20), UC (n=19) and CD (n=14) patients. The IBD patients were treated with anti-TNF and classified as responders (Rs) and non-responders (NRs) after 14-weeks from the initiation of the treatment. Figure 1 shows an overview of the experimental design we followed in this study and Figure 2 an overview of the cohort's characteristics.

For each participant, the demographic characteristics sex, age, BMI and smoking status were available (Figures 2B-2E). To assess cohort homogeneity and evaluate the influence of confounding factors, a factor analysis of mixed data (FAMD) was performed using the PRINCE [25] Python library (Figure 2F). Examination of the first principal component (PC1) revealed partial cohort separation, driven primarily by high loadings from sex and age (Figure 2H). While the influence of sex was anticipated due to its binary nature, the separation by age was largely driven by a small subset of UC patients under 30 years of age. Additionally, the absence of smokers within the UC subgroup contributed to cohort variation along the second principal component (PC2). Nevertheless, both components collectively accounted for approximately 10% of the total observed variance, indicating small impact on the overall structure of the cohort. Consequently, while these demographic factors warrant consideration as potential confounders, the cohort remains sufficiently homogeneous to investigate the multi-omic characteristics.

NMR-based serum inflammation biomarkers showed a post-treatment convergence of IBD patients towards median levels of non-IBD

To evaluate baseline systemic inflammation, we initially compared CRP levels between cohorts. Overall, mean baseline CRP levels were comparable across the IBD groups ($\mu_{UC} = 2.1$, $\mu_{CD} = 1.39$), though trends differed by treatment response (supplementary Table ST1). Specifically, UC responders had higher baseline CRP levels than non-responders ($\mu_{Rs} = 3.25$, $\mu_{NRs} = 0.52$), whereas the opposite pattern was observed in the CD group ($\mu_{Rs} = 0.78$, $\mu_{NRs} = 1.85$). In the UC group, the CRP levels correlated weakly with the disease index, with responders showing slightly higher mean MES scores than non-responders ($\mu_{Rs} = 2.63$, $\mu_{NRs} = 2.25$). Conversely, in the CD group the SES-CD scores did not reflect the pattern observed with the CRP, as responders demonstrated higher mean SES-CD scores than non-responders ($\mu_{Rs} = 9.0$, $\mu_{NRs} = 7.5$). Regarding the non-IBD group, we had available only three CRP observations ($\mu = 0.09$) that did not allow us to draw robust comparisons with the IBD patients.

On the other hand, serum-based inflammatory biomarkers allowed to evaluate the trajectory of the systemic inflammation of the IBD patients compared to non-IBD. More specifically, we NMR-quantified GlycA, GlycB and SPC to map the dynamic cumulative inflammatory burden and monitor changes in mucosal and systemic disease activity over time. GlycA and GlycB serve as integrated measures of the hepatic acute-phase glycoprotein response, whereas SPC reflects inflammation-driven alterations in lipoprotein metabolism. Following confounders adjustment (see Methods M.3.5), we observed a wide distribution among non-IBD individuals and substantial overlap with the IBD groups for SPC at baseline, whereas they displayed smaller variability and overlap to IBD groups for baseline levels of GlycA and GlycB (Figure 3). Interestingly, following treatment intervention (week-14), the median levels of all three biomarkers in IBD patients shifted toward the median values of the non-IBD group, with responders showing smaller variability compared to non-responders.

These findings indicate that while baseline systemic metabolic profiles vary widely even in the absence of IBD, the acute-phase glycoprotein signals (GlycA and GlycB) serve as tighter indicators of baseline disease status but still not very different than non-IBD individuals. Nevertheless, the post-treatment convergence of the IBD groups toward non-IBD medians suggests that the later represent a systemic normalization group that mirrors treatment response.

The non-IBD multi-omic profile defines a molecular reference state for therapeutic response

The cohort's multi-omic dataset included three omics from intestinal tissue (mRNA, miRNA and proteomics), one 16S amplicon-based omic from fecal samples (metataxonomics) and two NMR-quantified omics from serum samples (lipoproteomics and metabolomics). After data curation (see Methods M.3.1-M.3.4), mRNA dataset included 22,529, miRNA 445, proteomics 3586, metataxonomics 292, lipoproteomics 112 and metabolomics 41 features. Following feature-level confounding adjustment (Section M.3.5), we evaluated the longitudinal trajectory of the IBD cohort in relation to the non-IBD group using UMAP (see Methods M.10). This analysis was applied to each omic data layer individually as well as to the integrated dataset, embedding the non-IBD group and both timepoints of the IBD patients.

Our single-layer analyses revealed a greater separation between non-IBD and IBD patients at baseline within the tissue-based omics compared to fecal and serum-based layers (Figure 4). This pronounced separation is biologically expected, as tissue-derived data capture a localized molecular view directly at the site of pathology, whereas the rest of the omics incorporate distal or systemic signals that inherently introduce greater background noise.

Regarding trajectory, treatment responders demonstrated a clear tendency to cluster closely with the non-IBD group at week-14, a trend that was amplified following multi-omic integration. Nevertheless, UMAP topology is greatly influenced by data layers with a disproportionately large number of features, thus tissue omics had great impact on the observed clustering.

Taken together, these findings complement the previous observations of systemic inflammation and indicate that the non-IBD group effectively represents a molecular normalization state even at molecular level, capturing a multi-omic configuration that therapeutic response actively mirrors. Consequently, the next logical step in our analytical approach was to extract a targeted multi-omic signal capable of isolating this longitudinal trajectory of responders as they transition toward this non-inflamed reference state.

Multi-omics integration identified three latent factors that capture response trajectory toward non-IBD state

Prior to extracting a trajectory-based multi-omic signal, we addressed the challenge of highly disparate feature dimensions across omic layers. To do so, we implemented an independent dimensionality reduction workflow for each omic layer. Specifically, PCA was applied to biologically defined feature subsets, hereafter referred to as omic modules, and PC1 was extracted to serve as a module score for each participant at baseline and week 14 (see Methods M.4). This dimensionality reduction strategy effectively compressed the feature space into 50 modules for both mRNA and proteomics, 21 for miRNA, 48 for metataxonomics, 15 for lipoproteomics, and 17 for metabolomics (supplementary file 1).

Next, we consolidated the module scores across all available omic layers, pooling data from both non-IBD and IBD patients and across both timepoints. We subsequenty applied multi-omics factor analysis, without indicating clinical annotations such as disease status, therapeutic response, and timepoint (see Methods M.5). This data-driven approach identified nine latent

factors, with the first five factors capturing high amount of the observed variance in five omics (Figure 5A).

Since the latent factors were derived in a fully unsupervised manner, we next sought to identify which factors specifically captured the underlying dynamics driving the trajectory of IBD responders toward the non-IBD reference state. To achieve this, we first contextualized the factor scores within the non-IBD frame of reference (see Methods M.6) and subsequently applied a separate hierarchical Bayesian model to each factor (see Methods M.7). After verifying that all models achieved reliable sampling and stable estimates for all parameters ($\hat{R}$=1, rank-normalized ESS > 400, supplementary Tables ST2-ST10) [26], we observed that factors 1 and 3 captured the longitudinal convergence of responders toward the non-IBD state in both UC and CD patients, while factor 5 specifically captured this normalized trajectory predominantly within CD patients (Figures 5C,5E, supplementary Figure SF1).

Using these three latent factors, we reconstructed the modules of four omics (mRNA, miRNA, proteomics and lipoproteomics) where the factors captured a cumulative observed variance of $R^2 > 10\%$ (see Methods M.8). After merging the reconstructed omics and performed UMAP, we observed a clear trajectory where responders IBD patients shifted toward the non-IBD group at week-14 (Figure 5B).

To evaluate whether the patients baseline latent molecular state constrains or facilitates this capacity to normalize after treatment, we implemented a Bayesian logistic regression model (see Methods M.9). The objective of this model was to evaluate how this low-dimensional embedding of the baseline latent factors captures the structural relationship between initial multi-omic positioning and the ultimate probability of achieving a non-IBD state.

After verifying that the model achieved reliable sampling and stable estimates for all parameters ($\hat{R}$=1, rank-normalized ESS > 400, supplementary Table ST11), we observed a credible association between the baseline Euclidean distance of the latent factors embedding and the probability of treatment response (Figure 5D). Specifically, the closer the distance to the non-IBD group, the higher was the probability of response assigned to an IBD patient.

Expected response probabilities were higher overall in the UC responders compared to the CD group, though in both IBD groups only responders passed the threshold of 85% expected probability (Figure 5F). Among the two patients that crossed this threshold, one was a UC responder presenting with high MES (3.0) and the highest CRP levels (14.4 mg/L) in the cohort, and the other was as CD responder with mild SES-CD (6.0) and low CRP levels (0.1 mg/L). Regarding these parameters, we observed very weak correlation between the assigned response probabilities and baseline CRP levels or disease index in both IBD groups (supplementary Figure SF2). Finally, we observed a high degree of uncertainty in the assigned probabilities that can be primarily attributed to the modest sample size of our cohort and on the fact that the Euclidean distance was based on a compact three-dimensional embedding.

Taken together, this exercise showed that a three-dimensional latent space isolates a multi-omic signal that highlights the response trajectory toward normalization. Nevertheless, the latent space applies a considerable compression to the multi-omic information, limiting the ability of a classification model to generate highly precise probability estimates. However, it provides a biological view independent of CRP levels and disease index, suggesting the latent factors are unlikely to be merely reflecting global inflammation burden. They may instead represent molecular programs related to disease organization or resolution states, with treatment-specific information potentially embedded in the latent structure.

Response trajectory is characterized by a synchronized multi-omic axis of inflammatory, metabolic and structural adaptation

To isolate the molecular programs that are associated with the observed trajectory, we focused on the omic modules within the reconstructed space. In this space, an omic module value approaching zero signifies proximity to the non-IBD reference state. Leveraging this framework, we calculated and tracked the median module values for responders and non-responders both before and after treatment (Figure 6A). To isolate the modules exhibiting the most pronounced divergence in longitudinal convergence between the two clinical groups, we introduced the standardized response-associated trajectory gap (SRG) (see Methods M.8). This metric quantifies the distance between responders and non-responders in units of standard deviations relative to the non-IBD group, effectively pinpointing the omic modules that are highly associated with the therapeutic response. After considering modules with SRG ≥ 1, we isolated nine mRNA modules, two miRNA modules, and one lipoproteomics module that were consistently altered across both UC and CD. For the proteomic layer, responders exhibited difference for seven modules in UC and eight for CD, with a subset of four modules shared between the two IBD groups (supplementary file 2).

To facilitate interpretability and remove potential impact of sample size differences between the IBD groups, we mapped the selected omic modules back to the original unconfounded data space, and considered those that were shared between the IBD groups. After forming the union of the modules features, we created a multi-omic profile of 1,589 features (mRNA=1,383, miRNA=67, proteomics=127, lipoproteomics=12, supplementary file 2) and applied RLR-transformation to each feature (see Methods M.11). This enabled us to evaluate the treatment's interaction with this multi-omic profile, quantified as log2-fold-change relative to the geometric mean of the non-IBD group.

Overall, all IBD patients displayed fluctuations of less than 1.5-fold change relative to the average of the non-IBD group, indicating that in this multi-omic profile the lack of treatment response is not governed by massive, isolated molecular shifts, but rather by subtle, coordinated multi-omic dysregulations (Figure 6B). This is aligned with our previous observations where latent gene networks that did not pass differential expression thresholds could influence the performance of a response classifier to anti-TNF [27]. Furthermore, while the divergent paths between responders and non-responders revealed an overall trend in treatment dynamics, these trajectories are not uniform and present overlapping between the clinical groups. This indicates that treatment response is driven by highly heterogeneous molecular sub-phenotypes, suggesting that patients may arrive at the same clinical outcome through small variations of the captured multi-omic profile.

Among the most prominent modules, T-cell mediated inflammation (allograft rejection) demonstrated a more pronounced normalization toward the non-IBD state in responders than non-responders, and presented a consistent trend in both mRNA and proteomics layers. This is aligned with *in vitro* observations where an anti-TNF agent inhibited a growth factor in T cell populations of intestinal cultures [28]. While we cannot entirely exclude the possibility that non-responders simply require an extended timeline to achieve molecular normalization, the fact that IL-2/STAT5 signaling not only failed to resolve but intensified in a subset of non-responders strongly suggests the persistence of active T-cell survival and proliferative programs.

Moreover, epithelial-mesenchymal transition (EMT), inflammatory response, and TNF-α signaling via NF-κB were among the modules that showed the least normalization in non-responders. Persistent TNF-α/NF-κB signaling suggests ongoing inflammatory activity that may be sustained by TNF-independent mechanisms or incomplete suppression of inflamma-

tory signaling by anti-TNF. Likewise, persistent EMT activation may reflect continued tissue remodeling, stromal activation, and extracellular matrix remodeling [29]. This pattern is compatible with progressive fibrotic changes and impaired epithelial repair, potentially involving mediators such as TGF-β and matrix-remodeling enzymes [30], which may become less dependent on acute TNF signaling and therefore less responsive to anti-TNF therapy.

Further supporting these observations, microRNAs targeting the protein secretion module exhibited an attenuated normalization trajectory specifically in non-responders. A notable example is miR-192-5p, which maintained persistently suppressed expression levels in non-responders following anti-TNF administration (supplementary Figure SF3). This specific microRNA has been implicated in regulating autophagy and preserving intestinal epithelial barrier function, and its reduced levels have been correlated with compromised gut integrity during active inflammation [31]. Additionally, similar patterns were also observed for miR-145-5p whose suppressed levels have also been associated with increased inflammation [32].

Finally, the VLDL lipid export signal module identified a distinct subgroup of non-responders who exhibited an increase in mean lipid levels following anti-TNF treatment. A closer inspection in the features that were included in this module showed that this change was associated mainly with a shift in the internal lipid composition of the largest VLDL subfractions, specifically marked by an increase in V1TG (VLDL-1 triglycerides) in both IBD groups (supplementary Figure SF4).

Taken together, the above observations demonstrated that the latent response trajectory is governed by a multi-omic axis, characterized not only by classic inflammatory cascades, but also by highly synchronized systemic lipid transport mechanisms and microRNA-mediated cellular pathways. Although we could identify characteristic trajectories for certain non-responders, overall we observed intricate relationships between the omic features, highlighting the heterogeneous nature of non-response.

Baseline multi-omic profile provides insights into patient heterogeneity and therapy selection

To evaluate the utility of the captured multi-omic profile on external patients, we leveraged independent, pre-treatment multi-omic datasets that we had generated from three additional patient groups, two UC and one CD, with documented endoscopic response status at week-14. The UC cohorts included patients treated with either tofacitinib (n=11, Rs=5, NRs=6) or vedolizumab (n=12, Rs=8, NRs=4), whereas the CD cohort comprised patients treated with ustekinumab (n=15, Rs=4, NRs=11). The demographic characteristics of these cohorts are given in the supplementary Table ST12. After merging the supplementary cohorts with the anti-TNF cohort (see Methods M.12), we RLR-transformed the levels of the multi-omic profile of 1,589 features that we captured with the anti-TNF cohort. Moreover, since the multi-omic profile included a signal from serum lipoproteomics, we classified the patients based on the Sniderman et al [24] diagnostic algorithm for dyslipoproteinemias (see Methods M.13). This exercise identified one patient with VLDL+LDL dyslipoproteinemia, six with VLDL and ten with chylomicrons+VLDL remnants. Given that the supplementary cohorts comprised patients who had already failed anti-TNF treatment, our objective was to assess whether the multi-omic profile could identify relationships bridging the primary anti-TNF cohort and the supplementary cohorts. To do so, we applied UMAP dimensionality reduction using the 1,589 multi-omic features. Since the UMAP was based on the cosine distance, spatial proximity of the patients in the resulting embedding reflects shared structural and co-regulatory profiles across the transcriptomic, proteomic, and lipoproteomic layers.

The UMAP embedding demonstrated varying degrees of resolution between clinical phe-

notypes (Figure 7A). Overall, we observed a more pronounced spatial separation between treatment responders and non-responders within the UC cohort compared to the CD cohort, indicating that baseline systemic profiles in UC may possess clearer predictive topology for therapeutic efficacy. Of note, we observed cases where responders and non-responders to anti-TNF clustered in close proximity (cases 1 and 2 supplementary Figure SF5). Closer inspection revealed that, although these patients exhibited an homoiomic profile based on the mean RLR-transformed features at the module level, analysis at the individual omic features uncovered similar mRNA profiles but subtle, patient-specific differences, particularly within the proteomic layer (supplementary Figures SF6 and SF7). This suggests that close proximity is greatly influenced by similar profiles regarding the mRNA layer, whose high dimensionality in the multi-omic profile results in high contribution to the embedding, whereas different clinical outcomes appear to be associated with case-specific variations in features levels across the omic layers.

Special focus was directed toward patients from the supplementary cohorts who projected into immediate proximity with individuals from the anti-TNF cohort. Interestingly, within the UC cohort, we identified specific cases where patients who responded to either vedolizumab or tofacitinib clustered in very close proximity to anti-TNF non-responders (cases 3 and 4 supplementary Figure SF5). This shared baseline proximity suggests that these anti-TNF non-responders possess a multi-omic profile that could potentially rely on alternative inflammatory mechanisms, such as the JAK-STAT pathway or leukocyte trafficking, rather than a purely TNF-α-mediated drive. In a similar manner, we also identified responders to ustekinumab clustering close to anti-TNF non-responders in the CD cohort (case 5 supplementary Figure SF5). Of note, we observed a subset of ustekinumab non-responders that exhibited the most aggregated multi-omic profile among both IBD groups that matched the profile of CD non-responders to anti-TNF at week-14 (Figure 7B, case 6 supplementary Figure SF5). This suggests an intense systemic activation that either highlights the need for higher dosing strategies or combination biologic approaches.

On the other hand, we observed complex patterns of shared proximity in certain cases. Specifically, we observed certain responders to anti-TNF localized near tofacitinib non-responders (case 7 supplementary Figure SF5). This proximity could suggest that the latter group are characterized by an inflammatory endotype that effectively bypasses both the TNF-α and JAK-STAT pathways, making them potential candidates for alternative treatments such as vedolizumab. Moreover, we observed certain patients who responded to vedolizumab and tofacitinib clustering near anti-TNF responders (case 8 supplementary Figure SF5). Since co-localization reflects similar multi-omic profiles, this proximity suggests that historical failure on anti-TNF therapy likely represents a pharmacokinetic mismatch for these patients.

Furthermore, we observed an association between dyslipoproteinemia and distinct multi-omic profiles that can be linked to treatment outcomes. Within the non-IBD group, the only individual that was classified with VLDL dyslipoproteinemia, co-localized at a distance from the main non-IBD cluster. This finding aligns with our previous observations in a independent study of an ostensibly healthy cohort, where individuals with this type of dyslipoproteinemia defined a distinct subpopulation based on a multi-omic profile, whereas those characterized by chylomicrons and VLDL remnants did not provide informative profiles [33]. Moreover, in the CD cohort we observed a case where an anti-TNF non-responder clustered in immediate proximity to the non-IBD group (case 9 supplementary Figure SF5). Although the shared topology creates the initial expectation of a highly manageable disease profile, the lack of treatment response may be explained by an underlying metabolic dysregulation, since this individual was classified with VLDL dyslipoproteinemia. Additionally, a notable impact of lipid

phenotypes was also observed among patients treated with ustekinumab. Specifically, a ustekinumab patient presenting with VLDL dyslipoproteinemia clustered closely with anti-TNF non-responder, yet exhibited different treatment outcome to nearby ustekinumab responder who lacked lipid dysregulation (case 10 supplementary Figure SF5). Finally, we observed a case of dual therapeutic non-response in the UC cohort that initially appeared to contradict our previous observations, which suggested that anti-TNF non-responders could benefit from switching to vedolizumab due to their baseline co-localization. Interestingly, the lack of response in the vedolizumab patient was accompanied by mixed VLDL and LDL dyslipoproteinemia (case 11 supplementary Figure SF5).

Collectively, these observations demonstrate that the captured multi-omic profile incorporates underlying biological information that can assist the interpretation of the treatment outcome and potentially guide therapy selection. Furthermore, the recurrent co-occurrence of dysregulated lipid profiles across non-responders to diverse treatements, suggests that the baseline dyslipoproteinemia represents a confounding metabolic endotype associated with broader lack of response to treatment.

Discussion

Biomarker discovery associated with anti-TNF treatment response has evolved from early reliance on single-omic layers and cross-sectional cohorts toward sophisticated multi-omic integration and longitudinal profiling, enabling a more dynamic and holistic understanding of therapeutic resistance and efficacy over time. Despite these advancements, challenges remain in developing methodologies capable of managing the high dimensionality inherent to multi-omic data and evaluate cross-omic temporal dynamics across the response trajectory. Our study addresses this gap by integrating six omic layers, using a dimension-reduction strategy designed to capture a unified multi-omic profile across the entire treatment trajectory. Furthermore, to account for biological variations and potential noise from fluctuations within a pathological range, we anchored these trajectories against an inflammation-resolved state derived from a non-IBD group.

The compression of the omic features into PCA-derived and biologically interpretable modules, represents a method that conceptually resembles the feature extraction processes employed by neural networks. However, a critical distinction lies in data efficiency. Neural networks generally require large sample sizes to effectively optimize and identify meaningful feature representations, whereas our approach leverages highly informative and biologically interpretable omic subsets, making it particularly suitable for the limited sample sizes that characterize IBD cohorts. Nevertheless, further efforts are needed to refine and identify more informative disease-specific subsets to improve both the interpretability and potential of this framework.

The deconstruction of the multi-omic response trajectory into latent factors enabled the identification of biological variation across multiple omic layers, capturing coordinated molecular programs that would likely be difficult to detect by studying each omic layer in isolation. This highlighted the convergence of the responders IBD patients toward the non-IBD state. Interestingly, despite the well-established role of the microbiome in IBD and the prevailing assumption that connects microbial abundances to treatment response, the latent factor 4, that accounted for substantial variance within the metataxonomics layer, was not associated with the response trajectory. Instead, it exhibited only baseline differences across the clinical groups that equalized after treatment. In relation to the non-IBD group, previous studies have reported conflicting findings regarding differences in microbial abundances between IBD patients and non-IBD individuals, with one study suggesting that the absence of significant differences may be attributable to the inclusion of patients with inactive disease [34, 35]. On the other hand, comparisons between IBD patients have showed significant differences in microbial abundances between responders and non-responders after anti-TNF treatment [36, 37], albeit with substantial observed overlap between the clinical groups. This suggests that either latent factor 4 captured variation within the metataxonomics layer that is independent of treatment response, or that response-associated microbial differences are of relatively small magnitude, making them sensitive to study-specific variability and therefore more likely to be identified in feature-level differential abundance analyses than as a dominant latent factor associated with response.

The latent factors associated with the response trajectory defined a set of 16 omic modules comprising a multi-omic profile of 1,589 molecular features. At the module level, this profile effectively distinguished the markedly dysregulated molecular state of certain non-responders but lacked the resolution to capture the subtle overall differences separating responders from non-responders. In contrast, examination of the individual features provided the granularity needed to identify these differences, while also revealing that they were largely patient-specific

rather than shared across individuals, limiting their utility as universal biomarkers. Although an independent validation cohort for anti-TNF treatment was not available, the inclusion of additional cohorts of IBD patients who failed to respond to anti-TNF therapy allowed us to investigate the relationship between shared multi-omic profiles and clinical outcomes across treatment groups. This proved particularly valuable, as the close proximity of anti-TNF non-responders to patients who responded to alternative therapies suggests that this multi-omic profile may provide a foundation for guiding personalized treatment selection.

Regarding the observed association between dyslipoproteinemia and lack of treatment response, a previous study of anti-TNF therapy reported no significant short-term changes in lipid profiles but observed increases in total cholesterol and triglyceride levels after few years of follow-up [38]. Although these findings do not establish a causal relationship between anti-TNF therapy and lipid dysregulation, the consistent presence of dyslipoproteinemia in our supplementary cohorts, particularly among non-responders, suggests that it may represent a clinically relevant component of the molecular profile associated with treatment response. These observations support the inclusion of dyslipoproteinemia in the comprehensive evaluation of patients when assessing the likelihood of therapeutic response.

Regarding study limitations, the external validation of our results is hindered by a scarcity of cohorts possessing technical characteristics compatible with our multi-omic dataset generation, a challenge noted as a broader obstacle in multi-omic studies [39]. Furthermore, conventional cross-validation strategies are difficult to apply to latent factor models since the inferred latent space depends on stochastic optimization. Consequently, different training subsets may yield latent factors that differ in orientation and composition, complicating direct comparisons and the assessment of factor reproducibility across folds. This sensitivity also suggests that variations in cohort size may lead to molecular profiles that diverge from our findings. Therefore, rather than relying solely on increasing sample size, future efforts should prioritize the development of frameworks designed to integrate multi-omic profiles across heterogeneous cohorts.

Conclusion

In summary, our study introduces a methodological approach that deconvolutes the complex dynamics of anti-TNF treatment response into trajectory-informed molecular profiles, while overcoming the computational challenges inherent to multi-omic data. By highlighting the highly patient-specific nature of these molecular responses, our findings underscore a necessary paradigm shift in disease management. Specifically, the shift from universal biomarkers toward the use of individualized multi-omic profiles. This profile-driven strategy provides a foundation to advance precision medicine in IBD by matching patients with the most effective therapeutic intervention based on inter-patient molecular similarities.

References

[1] Jiratchaya Nuanpirom and Varodom Charoensawan. Multimodal bioinformatic analyses of genome-scale expression beyond gene-centric differential expression. *Briefings in Bioinformatics*, 27(2):bbag152, 2026.

[2] Ronald L Wasserstein and Nicole A Lazar. The asa statement on p-values: context, process, and purpose, 2016.

[3] Youlian Zhou, Zhenjiang Zech Xu, Yan He, et al. Gut microbiota offers universal biomarkers across ethnicity in inflammatory bowel disease diagnosis and infliximab response prediction. *MSystems*, 3(1):10–1128, 2018.

[4] Yael Haberman, Rebekah Karns, Phillip J Dexheimer, et al. Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nature communications*, 10(1):1–13, 2019.

[5] Bram Verstockt, Sare Verstockt, Jonas Dehairs, et al. Low trem1 expression in whole blood predicts anti-tnf response in inflammatory bowel disease. *EBioMedicine*, 40:733–742, 2019.

[6] Renaud Gaujoux, Elina Starosvetsky, Naama Maimon, et al. Cell-centred meta-analysis reveals baseline predictors of anti-tnfα non-response in biopsy and blood of patients with ibd. *Gut*, 68(4):604–614, 2019.

[7] Stefan Schreiber, Konrad Aden, Florian Tran, and Philip Rosenstiel. Rise of precision medicine: can it deliver on its promise in ibd? *Gut*, 75(1):176–188, 2026.

[8] Liru Chen, Chuhan Zhang, Ruixuan Niu, et al. Multiomics biomarkers for predicting efficacy of biologic and smallmolecule therapies in adults with inflammatory bowel disease: A systematic review. *United European Gastroenterology Journal*, 13(4):517530, December 2024. ISSN 2050-6414. doi: 10.1002/ueg2.12720. URL `http://dx.doi.org/10.1002/ueg2.12720`.

[9] Jonathan Wei Jie Lee, Damian Plichta, Larson Hogstrom, et al. Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. *Cell Host & Microbe*, 29(8):12941304.e4, August 2021. ISSN 1931-3128. doi: 10.1016/j.chom.2021.06.019. URL `http://dx.doi.org/10.1016/j.chom.2021.06.019`.

[10] Neha Mishra, Konrad Aden, Johanna I. Blase, et al. Longitudinal multi-omics analysis identifies early blood-based predictors of anti-tnf therapy response in inflammatory bowel disease. *Genome Medicine*, 14(1), 2022. ISSN 1756-994X. doi: 10.1186/s13073-022-01112-z. URL `http://dx.doi.org/10.1186/s13073-022-01112-z`.

[11] Richard D Riley, Joie Ensor, Kym IE Snell, et al. Importance of sample size on the quality and utility of ai-based prediction models for healthcare. *The Lancet Digital Health*, 7(6), 2025.

[12] Maximilien Colange, Guillaume Appé, Léa Meunier, Solène Weill, Akpéli Nordor, and Abdelkader Behdenna. Differential expression analysis with inmoose, the integrated multi-omic open-source environment in python. *BMC bioinformatics*, 26(1):160, 2025.

[13] Stefka Tyanova, Tikira Temu, Pavel Sinitcyn, et al. The perseus computational platform for comprehensive analysis of (prote) omics data. *Nature methods*, 13(9):731–740, 2016.

[14] Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. Dada2: High-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7):581–583, 2016.

[15] Evan Bolyen, Jai Ram Rideout, Matthew R Dillon, et al. Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature biotechnology*, 37(8): 852–857, 2019.

[16] Torbjørn Rognes, Tomáš Flouri, Ben Nichols, Christopher Quince, and Frédéric Mahé. Vsearch: a versatile open source tool for metagenomics. *PeerJ*, 4:e2584, 2016.

[17] Christian Quast, Elmar Pruesse, Pelin Yilmaz, et al. The silva ribosomal rna gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(D1): D590–D596, 2012.

[18] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

[19] Skipper Seabold, Josef Perktold, et al. Statsmodels: econometric and statistical modeling with python. *scipy*, 7(1):92–96, 2010.

[20] Arthur Liberzon, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P Mesirov. Molecular signatures database (msigdb) 3.0. *Bioinformatics*, 27(12):1739–1740, 2011.

[21] Ricard Argelaguet, Britta Velten, Damien Arnol, et al. Multi-omics factor analysisa framework for unsupervised integration of multi-omics data sets. *Molecular systems biology*, 14(6):MSB178124, 2018.

[22] Oriol Abril-Pla, Virgile Andreani, Colin Carroll, et al. Pymc: a modern, and comprehensive probabilistic programming framework in python. *PeerJ Computer Science*, 9:e1516, 2023.

[23] Tim Sainburg, Leland McInnes, and Timothy Q Gentner. Parametric umap embeddings for representation and semisupervised learning. *Neural Computation*, 33(11):2881–2907, 2021.

[24] Allan Sniderman, Patrick Couture, and Jacqueline De Graaf. Diagnosis and treatment of apolipoprotein b dyslipoproteinemias. *Nature Reviews Endocrinology*, 6(6):335–346, 2010.

[25] Max Halford. Prince. URL `https://github.com/MaxHalford/prince`.

[26] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved r for assessing convergence of mcmc (with discussion). *Bayesian analysis*, 16(2):667–718, 2021.

[27] Dimitrios Kioroglou, Ainize Peña-Cearra, Ana M Corraliza, et al. Mitochondrial dysfunction: unraveling the elusive biology behind anti-tnf response during ulcerative colitis. *Inflammatory bowel diseases*, 31(5):1366–1379, 2025.

[28] Ulrike Billmeier, Walburga Dieterich, Markus F Neurath, and Raja Atreya. Molecular mechanism of action of anti-tumor necrosis factor antibodies in inflammatory bowel diseases. *World journal of gastroenterology*, 22(42):9300, 2016.

[29] Sara Lovisa, Giannicola Genovese, and Silvio Danese. Role of epithelial-to-mesenchymal transition in inflammatory bowel disease. *Journal of Crohn's and Colitis*, 13(5):659–668, 2019.

[30] Dulce C Macias-Ceja, M Teresa Mendoza-Ballesteros, María Ortega-Albiach, M Dolores Barrachina, and Dolores Ortiz-Masià. Role of the epithelial barrier in intestinal fibrosis associated with inflammatory bowel disease: relevance of the epithelial-to mesenchymal transition. *Frontiers in Cell and Developmental Biology*, 11:1258843, 2023.

[31] Peishan Qiu, Kezhi Zhou, Youwei Wang, et al. Revitalizing gut barrier integrity: role of mir-192-5p in enhancing autophagy via rictor in enteritis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 327(3):G317–G332, 2024.

[32] Rana Al-Sadi, Jessica Engers, and Raz Abdulqadir. Talk about micromanaging! role of micrornas in intestinal barrier function. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 319(2):G170–G174, 2020.

[33] Dimitrios Kioroglou, Rubén Gil-Redondo, Nieves Embade, et al. Multi-omic integration sets the path for early prevention strategies on healthy individuals. *NPJ Genomic Medicine*, 10(1):35, 2025.

[34] Jason Lloyd-Price, Cesar Arze, Ashwin N. Ananthakrishnan, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, 569(7758):655662, May 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1237-9. URL `http://dx.doi.org/10.1038/s41586-019-1237-9`.

[35] Xochitl C Morgan, Timothy L Tickle, Harry Sokol, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology*, 13(9):R79, 2012.

[36] Laura Sanchis-Artero, Juan Francisco Martínez-Blanch, Sergio Manresa-Vera, et al. Evaluation of changes in intestinal microbiota in crohns disease patients after anti-tnf alpha treatment. *Scientific Reports*, 11(1):10016, 2021.

[37] Yong Eun Park, Hye Su Moon, Dongeun Yong, et al. Microbial changes in stool, saliva, serum, and urine before and after anti-tnf-α therapy in patients with inflammatory bowel diseases. *Scientific reports*, 12(1):6359, 2022.

[38] José Miranda-Bautista, Celia de Gracia-Fernández, María López-Ibáñez, et al. Lipid profile in inflammatory bowel disease patients on anti-tnfα therapy. *Digestive diseases and sciences*, 60(7):2130–2135, 2015.

[39] Arno R Bourgonje, Harry van Goor, Klaas Nico Faber, and Gerard Dijkstra. Clinical value of multiomics-based biomarker signatures in inflammatory bowel diseases: challenges and opportunities. *Clinical and translational gastroenterology*, 14(7):e00579, 2023.

Declarations

Abbreviations

BMI = body mass index
CD = Crohn's disease
CRP = C-reactive protein
ESS = effective sample size GlycA = glycoprotein acetylation A
GlycB = glycoprotein acetylation B
IBD = inflammatory bowel disease
LOSO = leave-one-subject-out
MES = Mayo endoscopic subscore
MHD = module homeostatic displacement
MOFA = multi-omics factor analysis
NMR = nuclear magnetic resonance
NRs = non-responders IBD patients
PC = principal component
PCA = principal component analysis
Rs = responders IBD patients
SES-CD = simple endoscopic score for Crohn's disease
SPC = supramolecular phospholipid composite
SRG = standardized response-associated trajectory gap
UC = ulcerative colitis
UMAP = uniform manifold approximation and projection
VST = variance-stabilizing transformation

Authors contributions

D.K: Conceptualization (computational framework), Methodology (bioinformatic implementation), Formal Analysis, Visualization, Writing - Original Draft. M.B.M: Resources (patient recruitment), Project administration (management of patients samples and cross-lab coordination). R.G.R, N.E, M.A, M.G: Resources (performed laboratory experiments), Data Curation (data preprocessing). A.M.A, F.E, O.M: Supervision (supervised laboratory experiments). I.M.I, P.J.M., F.S.T., A.J.L., C.R.G., A.G.C., E.M.A., L.B.F.P., I.R.L, R.M.F.G., E.I.F., F.B., M.V.D., M.J.G., L.F.S., F.J.G.A., X.C.C., L.T., I.B.R., H.P.M., Y.B.N., E.L.C., M.J.C., F.J.T.: Resources (patient recruitment and clinical samples). U.M.M.: Conceptualization (project initiation and scientific context), Project administration (establishment of the collaboration and cross-lab coordination). M.C, J.P.G.: Conceptualization (clinical study design), Resources (patient recruitment and clinical samples), Methodology (patient enrollment criteria and treatment protocols), Funding acquisition. All authors have reviewed and approved the manuscript.

Funding

This research has been funded by grants from Pfizer Inc. Aspire program (Grant ID: 53589311), CIBEREHD (Intramural Project 2021) and the Instituto de Salud Carlos III (PI18/00622). This grant is supported by the Plan Estatal de Investigación Científica y Técnica y de Innovación 20172020 and the ISCIII-Subdirección General de Evaluación y Fomento de la Investigación, and is co-financed by the European Regional Development Fund, Programa Operativo

Pluriregional de España 2021-2027 (Policy Objective 1: A more competitive and intelligent Europe). The grant will be governed in accordance with Regulations (EU) 2021/1060 and 2021/1058 of the European Parliament and of the Council of 24 June 2021. D.K received financial support by the Severo Ochoa Centre of Excellence Program and Units of Excellence Maria Maeztu under the State Subprogram for Institutional Strengthening of the State Research Agency and the Ministry of Science and Innovation (Grant CEX2021-01136-S funded by MICINN/AEI/https://doi.org/10.13039/501100011033). CIC bioGUNE is supported by the Basque Department of Industry, Tourism and Trade (Etortek, Elkartek and Emaitek Programs) and the Innovation Technology Department of Bizkaia County.

Conflict of interest statement

María Chaparro has served as a speaker or received research or educational funding from MSD, AbbVie, Hospira, Pfizer, Takeda, Janssen, Ferring, Shire Pharmaceuticals, Dr. Falk Pharma, Tillotts Pharma, Biogen, Gilead, and Lilly. Javier P. Gisbert has served as a speaker, consultant, or advisory board member, or received research funding from MSD, AbbVie, Pfizer, Kern Pharma, Biogen, Mylan, Takeda, Janssen, Roche, Sandoz, Celgene/Bristol Myers, Gilead /Galapagos, Lilly, Ferring, Faes Farma, Shire Pharmaceuticals, Dr. Falk Pharma, Tillotts Pharma, Chiesi, Casen Fleet, Gebro Pharma, Otsuka Pharmaceutical, Norgine, and Vifor Pharma. Eva Iglesias-Flores has served as a speaker, consultant, or advisory board member, or received research funding from MSD, AbbVie, Takeda, Janssen, Galapagos/Alfasigma, Lilly, Ferring, Dr. Falk Pharma, and Adacyte.

Data availability

The curated multi-omic datasets used in this study, without confounders-adjustment (see Methods M.3), have been deposited in Zenodo under the accession code: https://doi.org/10.5281/zenodo.21471655. Code for accessing the available data is given in the supplementary Methods (SM2).

Acknowledgements

We would like to acknowledge the contribution of Ricardo Conde (Precision Medicine and Metabolism Lab, CIC bioGUNE), for his assistance in defining the metabolomics and lipoproteomics subsets of features.

Figures

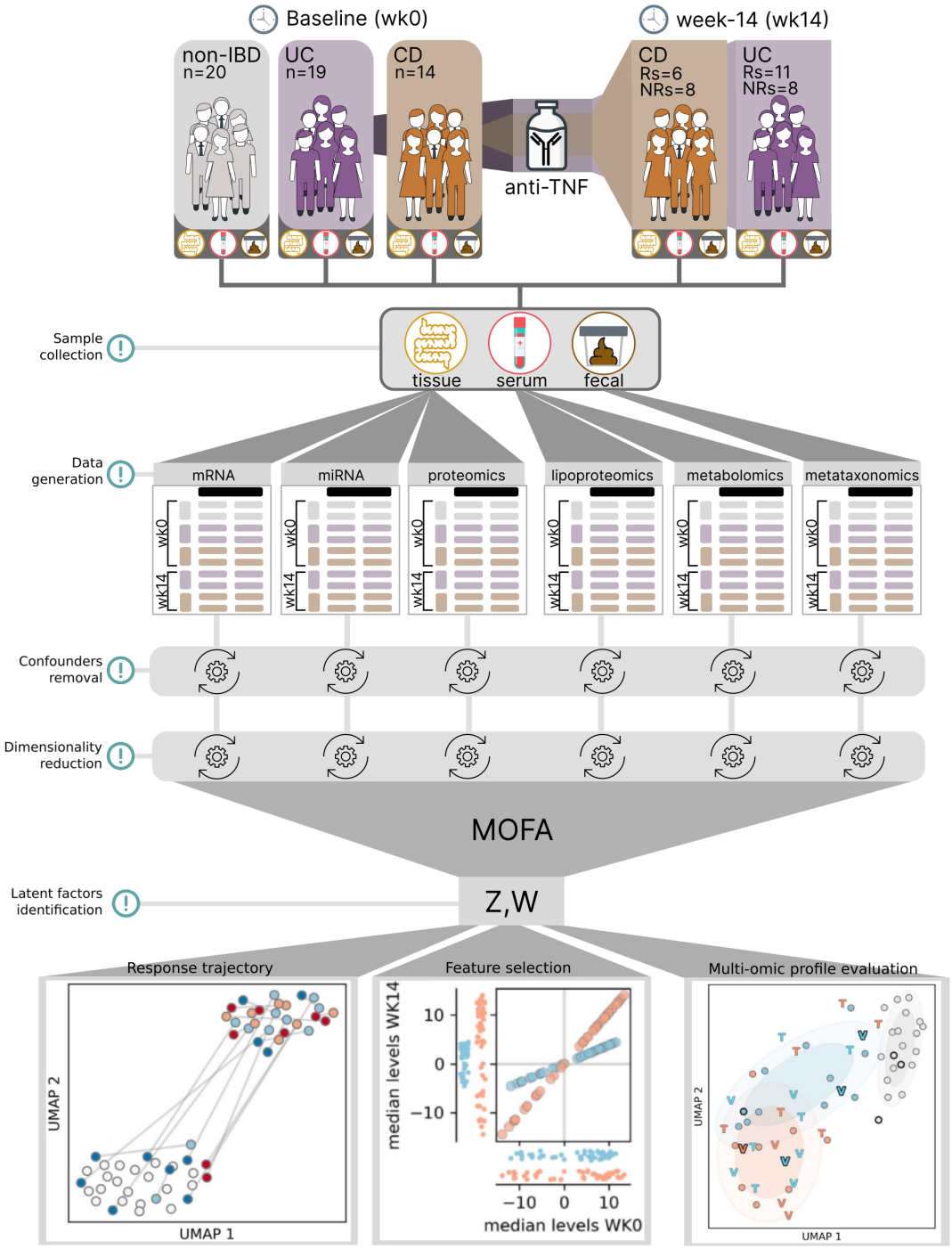


Figure 1: Experimental design. Tissue, serum, and fecal samples were collected from a cohort comprising non-IBD individuals, UC and CD patients. Non-IBD individuals provided cross-sectional samples, while IBD patients provided longitudinal samples before and 14-weeks after anti-TNF treatment where they were classified as responders (Rs) or non-responders (NRs). Six omics datasets were generated from the collected samples, followed by confounder removal and dimensionality reduction for each layer. After multi-omic integration, multi-omics factor analysis (MOFA) was performed to identify latent factors. These factors were used to capture treatment response trajectories, identify associated features, and construct multi-omic profiles to evaluate inter-patient relationships.

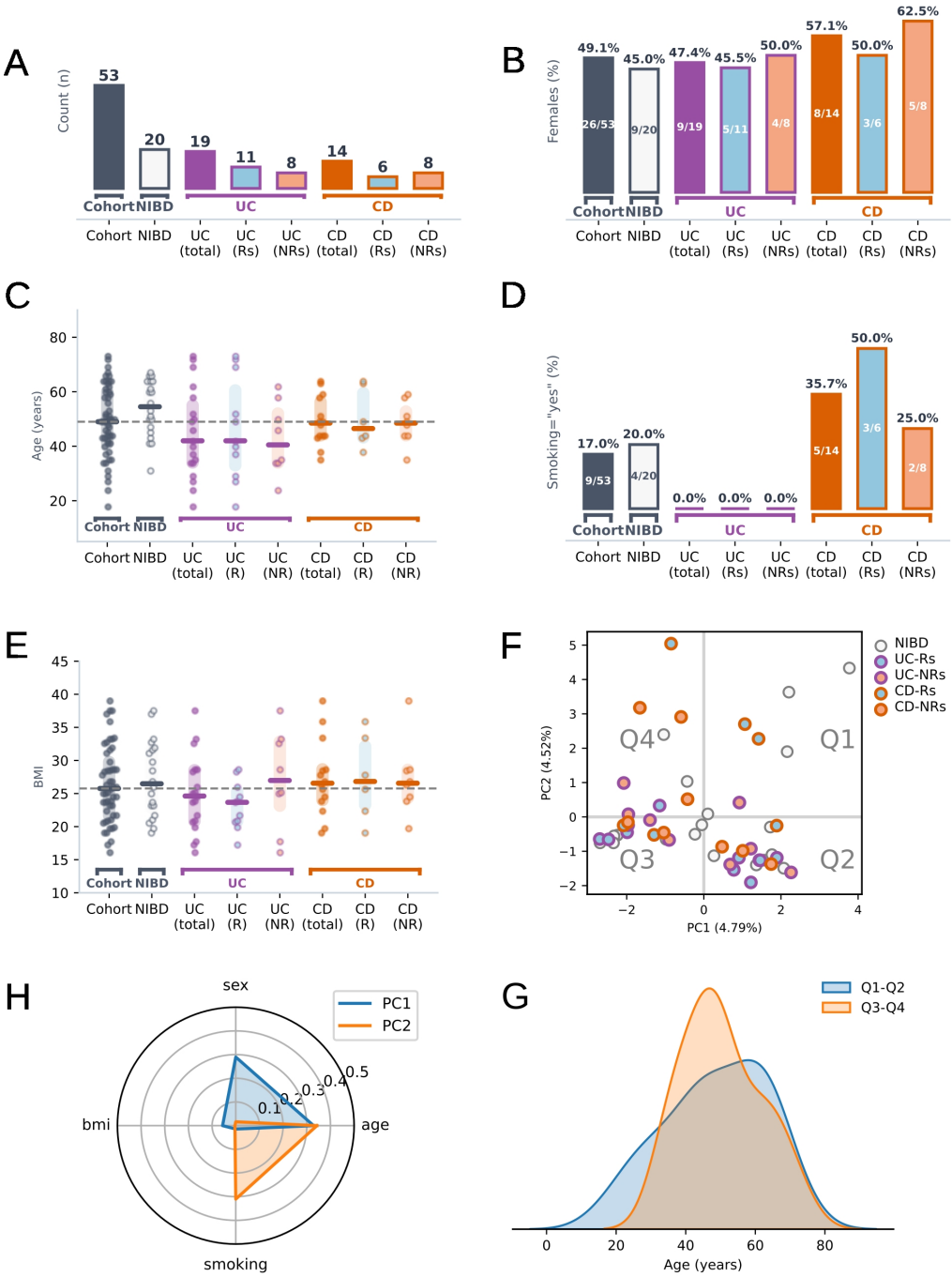


Figure 2: Characterization and demographic profiling of the study cohort. **Panel A**: distribution of number of participants in the cohort and in the groups non-IBD (NIBD) UC and CD. The IBD patients were further subdivided into responders (Rs) and non-responders (NRs). **Panels B-E**: distributions of the demographic characteristics sex, age, smoking status and BMI. **Panel F**: variation of the cohort across the first two principal components based on a factor analysis of mixed data using demographic characteristics. **Panel H**: loadings of the demographic features on the principal components. **Panel G**: age distributions of individuals that the were collocated in the quadrants Q1-Q2 and Q3-Q4 of the panel F.

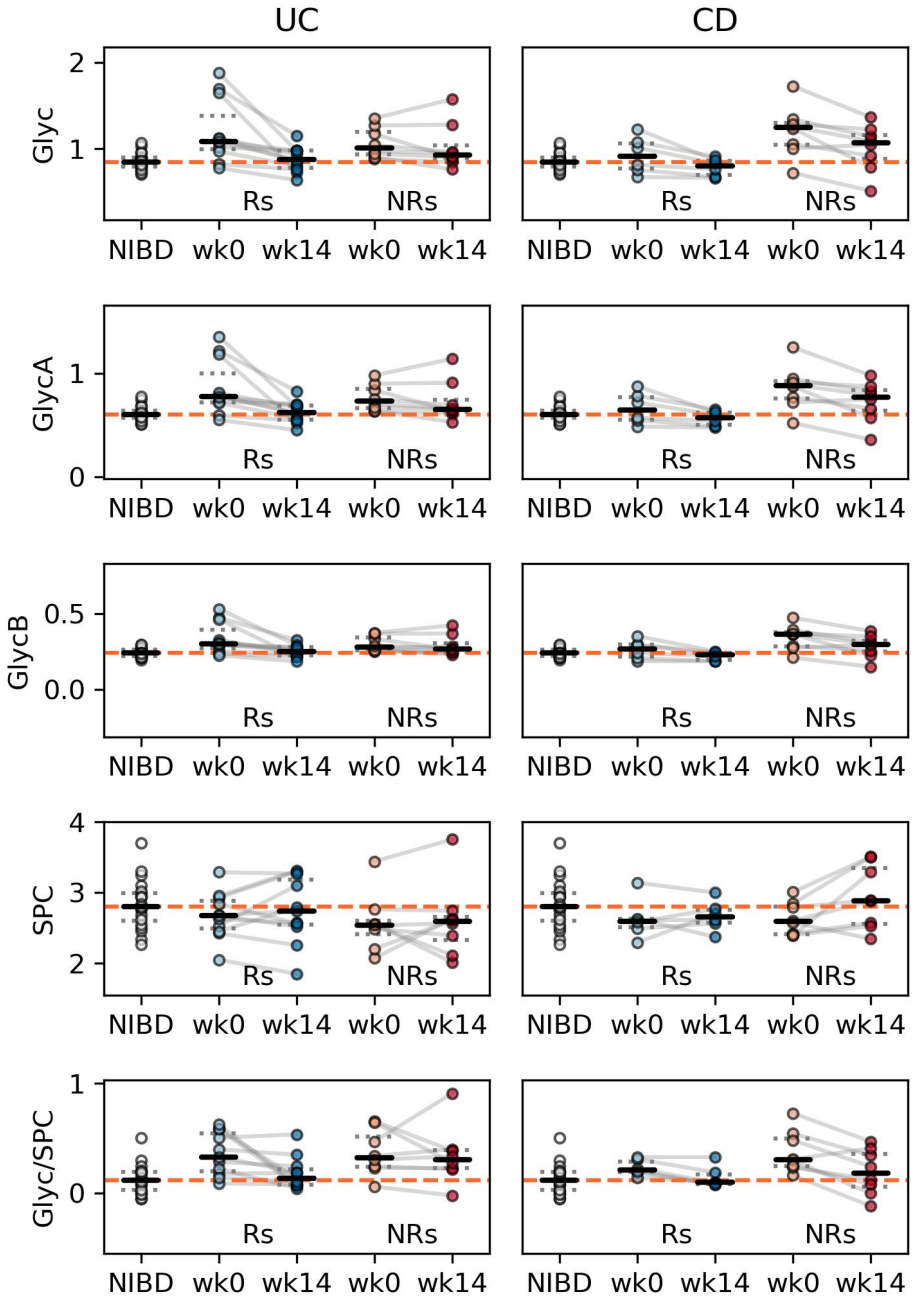


Figure 3: Serum inflammatory biomarker levels across patient groups and treatment time-points. Confounder-adjusted, NMR-quantified levels of GlycA, GlycB, and SPC in UC and CD responders (Rs) and non-responders (NRs) at baseline (wk0) and post-anti-TNF treatment (wk14), compared to non-IBD individuals (NIBD). Combined glycoprotein levels Glyc (GlycA + GlycB) and the Glyc/SPC ratio are also shown. For each clinical group, solid lines represent the median and dotted lines indicate the interquartile range (IQR). The horizontal colored dashed line marks the median value of the non-IBD group.

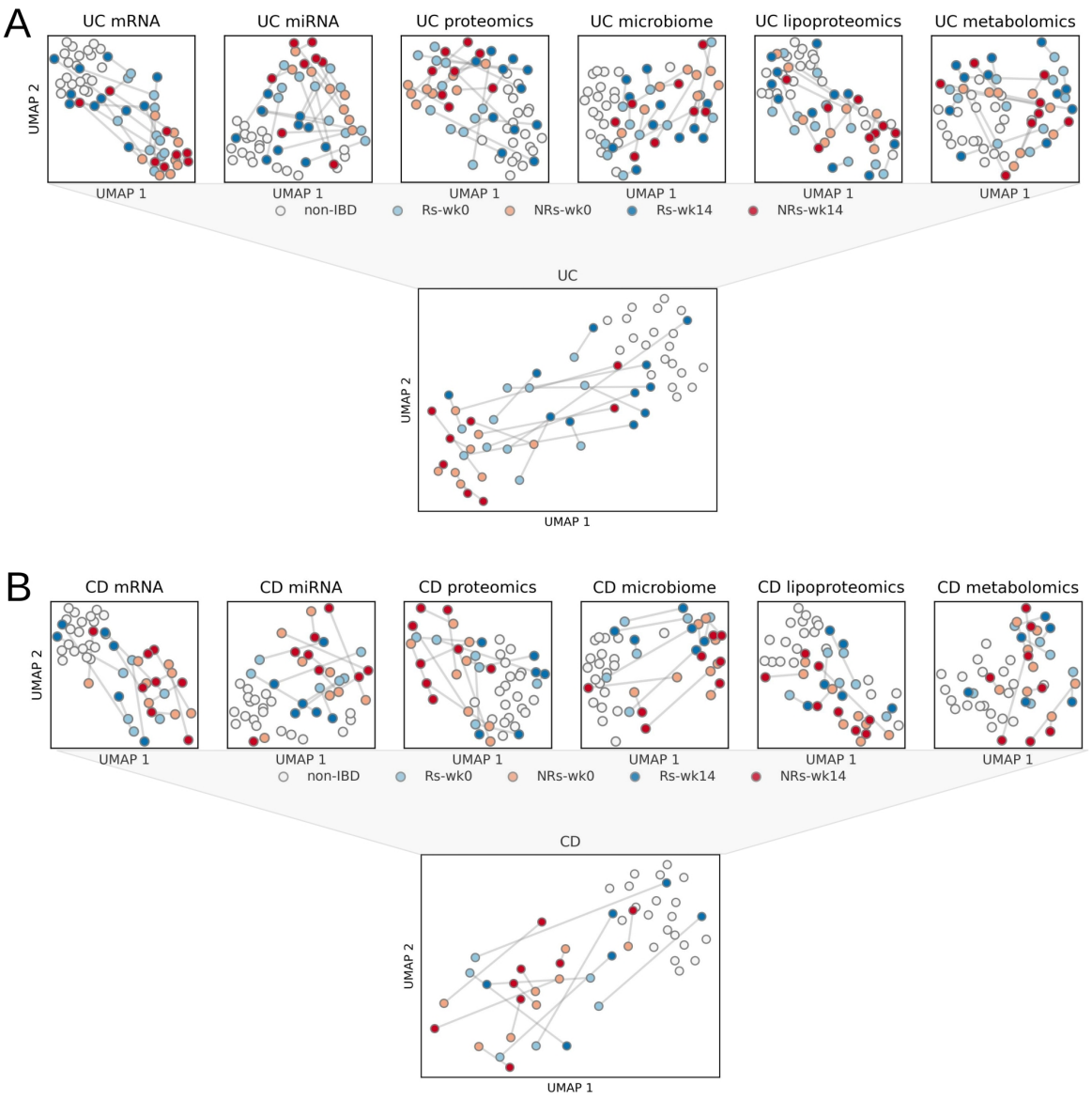


Figure 4: UMAP visualization of single and merged omics datasets. **Panel A**: UMAPs of non-IBD and UC patients for each available omic separately and after merging all omics. **Panel B**: UMAPs of non-IBD and CD patients for each available omic separately and after merging all omics.

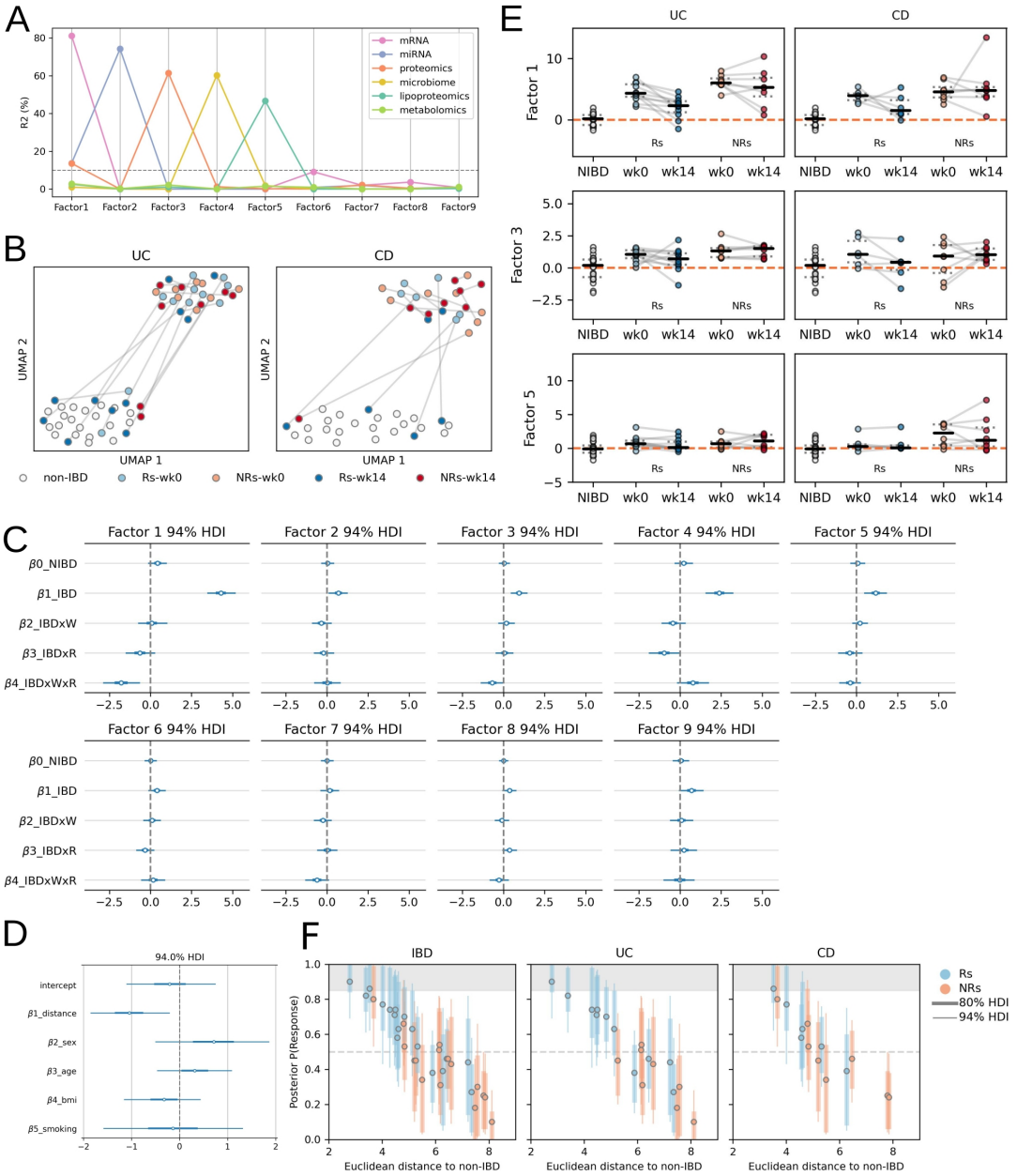


Figure 5: Multi-omics factor analysis. **Panel A**: Variance explained (%) by each latent factor across omic layers. **Panel B**: UMAP of reconstructed omics using selected latent factors, showing clustering non-IBD and IBD responders (Rs) and non-responders (NRs) at baseline (wk0) and after anti-TNF treatment (wk14). **Panels C-D**: Forest plots showing estimated posterior effects of clinical covariates on a given latent factor (C), and Euclidean distance to the non-IBD group on treatment response (D). Points denote posterior medians and inner and outer horizontal bars represent 50% and 94% highest density intervals (HDIs), respectively. **Panel E**: Latent factors scores of UC and CD responders (Rs) and non-responders (NRs) at baseline (wk0) and post-anti-TNF treatment (wk14), compared to non-IBD individuals (NIBD). For each clinical group, solid lines represent the median and dotted lines indicate the interquartile range (IQR). The horizontal colored dashed line marks the median value of the non-IBD group. **Panel F**: Logistic regression posterior response probabilities based on Euclidean distance to the non-IBD group. Points indicate expected probabilities, vertical bars represent 80% (inner) and 94% (outer) HDIs. The horizontal dashed line marks the 50% response probability threshold.

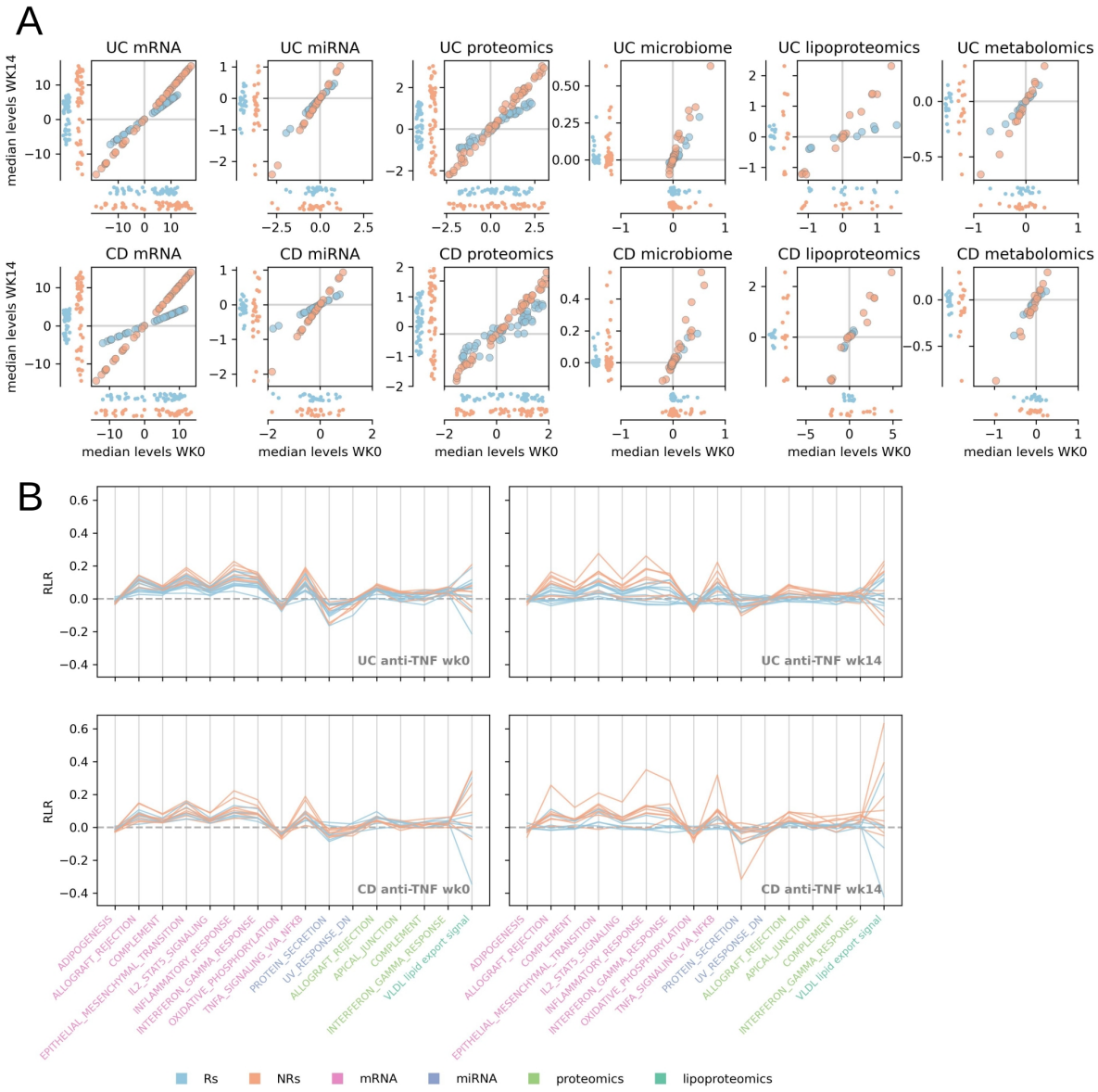


Figure 6: **Panel A**: Median levels of omic modules from reconstructed omic layers based on selected latent factors for responders (Rs) and non-responders (NRs) IBD patients at baseline (wk0) and post-anti-TNF treatment (wk14). **Panel B**: Confounder-adjusted and rlr-transformed mean levels of the features included in the omic modules that captured response trajectory. Each line represents a responder (Rs) or non-responder (NRs) IBD patient before (wk0) and after (wk14) treatment.

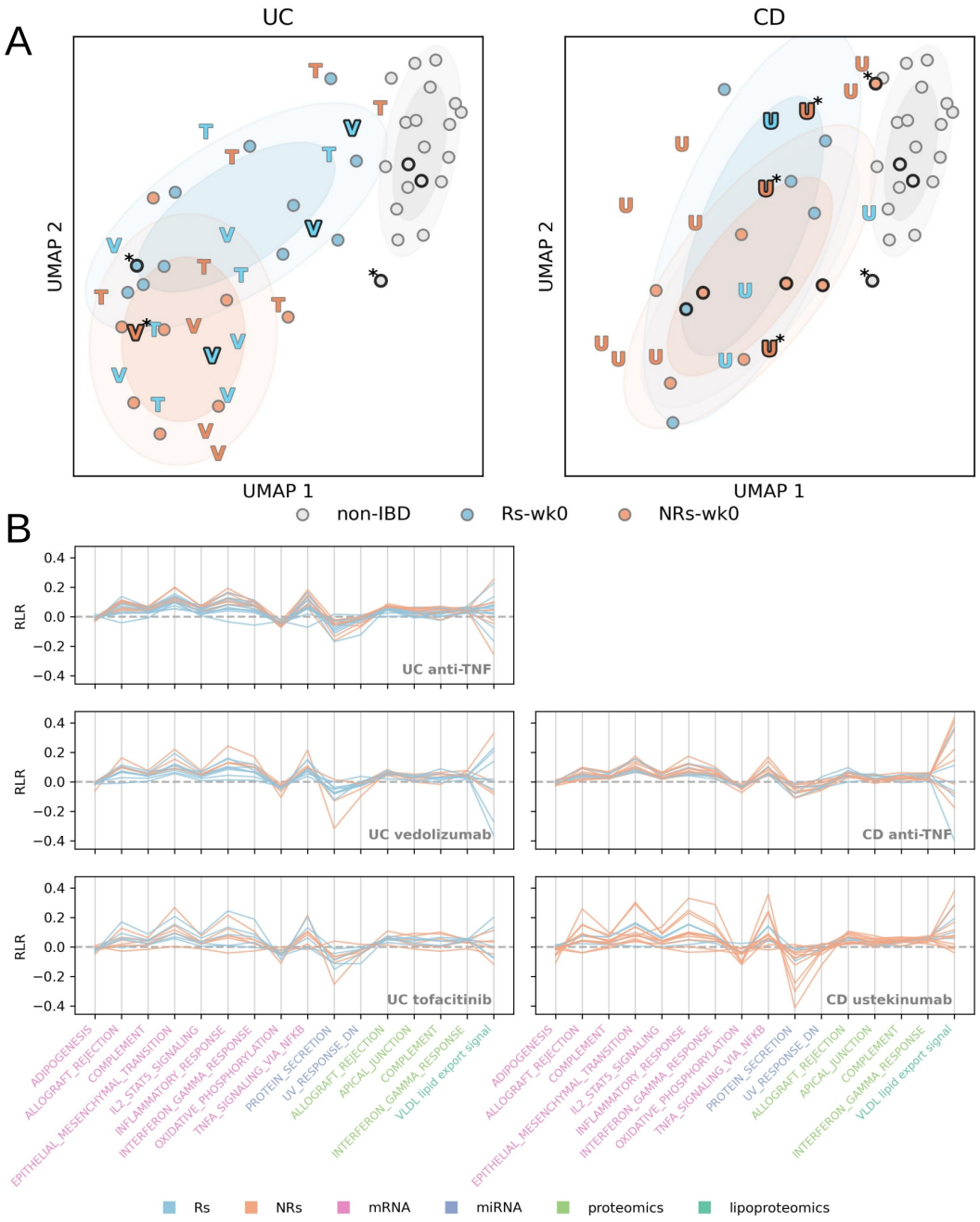


Figure 7: **Panel A**: UMAP before initializing treatment (wk0) based on the multi-omic profile that captured response trajectory in the anti-TNF cohort. Color denotes non-IBD individual or responder (Rs) and non-responder (NRs) IBD patient. Patients that were treated with anti-TNF are denoted with circle, whereas patients that were treated with vedolizumab, tofacitinib or ustekinumab are indicated with the first letter of treatment. Patients that were classified with dyslipoproteinemia are denoted with thick black lines, and asterisks mark those patients that were specifically classified with VLDL or VLDL+LDL dyslipoproteinemia. The overlaid contours indicate the 50% (inner) and 80% (outer) probability coverage of a fitted Gaussian Mixture Model, highlighting the highest density regions of the non-IBD group, and the responders or non-responders patients to anti-TNF. **Panel B**: Confounder-adjusted and rlr-transformed mean levels of the features included in the omic modules that captured response trajectory in the anti-TNF cohort. Each line represents a responder (Rs) or non-responder (NRs) IBD patient before initializing anti-TNF or alternative treatment.

Supplementary Files

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

- supplementarymethods.pdf
- supplementaryfile1.xlsx
- supplementaryfile2.xlsx