

TRANS-OMIC KNOWLEDGE TRANSFER MODELING INFERS GUT MICROBIOME BIOMARKERS OF ANTI-TNF RESISTANCE IN ULCERATIVE COLITIS

Alan Trinh^{1,4} Ran Ran^{2,4}, Douglas K Brubaker^{2,3*}

¹Indiana University School of Medicine, Indianapolis IN, USA. ²Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA. ³Regenstrief Center for Healthcare Engineering, Purdue University, West Lafayette, IN, USA. ⁴These Authors Contributed Equally.

*Email: dkbrubak@purdue.edu

A critical challenge in analyzing multi-omics data from clinical cohorts is the re-use of these valuable datasets to answer biological questions beyond the scope of the original study. Transfer Learning and Knowledge Transfer approaches are machine learning methods that leverage knowledge gained in one domain to solve a problem in another. Here, we address the challenge of developing Knowledge Transfer approaches to map trans-omic information from a multi-omic clinical cohort to another cohort in which a novel phenotype is measured. Our test case is that of predicting gut microbiome and gut metabolite biomarkers of resistance to anti-TNF therapy in Ulcerative Colitis patients. Three approaches are proposed for Trans-omic Knowledge Transfer, and the resulting performance and downstream inferred biomarkers are compared to identify efficacious methods. We find that multiple approaches reveal similar metabolite and microbial biomarkers of anti-TNF resistance and that these commonly implicated biomarkers can be validated in literature analysis. Overall, we demonstrate a promising approach to maximize the value of the investment in large clinical multi-omics studies by re-using these data to answer biological and clinical questions not posed in the original study.

Keywords: Trans-omic. Multi-omic. Transfer Learning. Knowledge Transfer. Microbiome. Colitis.

1. Introduction

The generation of matched multi-omics datasets from large clinical cohorts has resulted in identification of novel biomarkers of disease progression and therapeutic response in cancers [1-3], inflammatory [4, 5], and other complex human diseases [6, 7]. The Integrative Human Microbiome Project (iHMP) is a recent effort to understand the complex host and microbiome drivers of inflammatory bowel disease (IBD) [8], type 2 diabetes (T2D) [7], and preterm birth (PTB) [9] through the integration of human multi-omics. The development of computational tools to integrate molecular data across scales [10] and relate signatures to human phenotypes [11] has been a critical parallel and synergistic effort to experimental advances that have expanded the scope of molecular profiling. Because sequencing remains one of the highest costs to scaling clinical multi-omics, patients are recruited with defined criteria to ensure sufficient statistical power to answer the primary study questions. Therefore, though the molecular data in these cohorts are rich in detail and scope, the clinical and phenotypic variables are often sparse and limited in scope.

Transfer Learning, the use of information gained solving one problem to inform the solution of a different one, is a machine learning area suited to maximize the value of the financial, research, and patient efforts required to generate clinical datasets. The profiling of different molecular data types in multi-omic cohorts encode *trans-omic* information that enables correlation of signals across scales. If one of these scales is present in single-omic cohort matched to new phenotypic variables, the trans-omic relationships in the multi-omics study could reveal molecular associations with the

phenotypic variables in the single-omic cohort. We term this *Trans-omic Knowledge Transfer* and suggest that this approach represents a largely untapped reservoir of opportunity to reuse data from clinical cohorts to answer questions beyond those posed in the original study.

Here, we examine potential strategies for Trans-omic Knowledge Transfer to associate multi-omic signatures from one cohort of patients to a drug-resistance phenotype in another (Figure 1). Our objective is to identify gut microbial taxa and metabolites in the IBD Multi-omics Database (IBDMDB)[8] predictive of anti-tumor necrosis factor alpha (anti-TNF) therapeutic response in Ulcerative Colitis (UC) patients, a phenotype not in the original IBDMDB data.

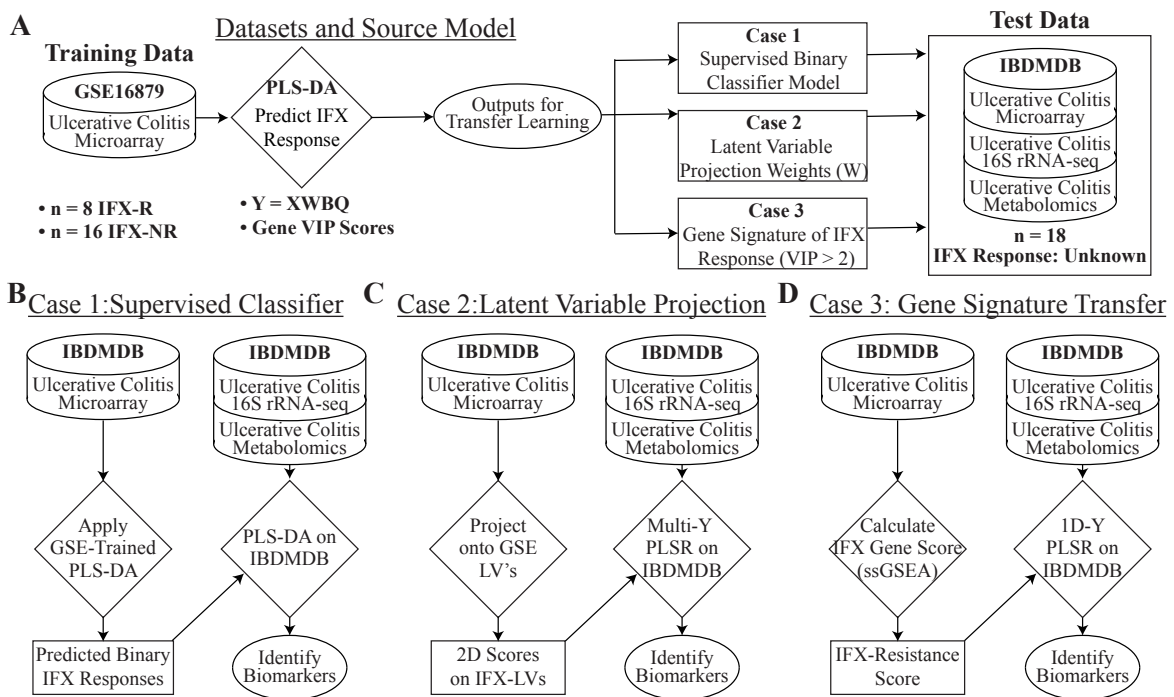


Figure 1. Study Overview. (A) Trans-omic Knowledge Transfer to predict biomarkers of response to an anti-TNF drug (Infliximab: IFX) training on UC patient gene expression data (GSE16879, “GSE-Data”) and predicting on a test set (IBDMDB). (B) Case 1: Supervised Classifier Transfer. The PLS-DA model constructed on the training data is applied to IBDMDB gene expression data to predict IFX response. New PLS-DA models are constructed to associate microbial taxa and metabolites to the predicted IFX response. (C) Case 2: Relative Separation Transfer. The weights matrix W is extracted from the GSE-trained PLS-DA model and the IBDMDB gene expression data are projected onto the GSE latent variables (LV). PLS-R models are trained to associate microbial taxa and metabolites to the positions of the IBDMDB samples on the IFX response, GSE-trained LVs. (D) Case 3: Signature Transfer. Genes predictive of IFX response in the GSE-trained model are extracted via Variable Importance of Projection (VIP) analysis to define an IFX-response Gene Set. Single Sample Gene Set Enrichment Analysis (ssGSEA) constructs a resistance score for this signature in IBDMDB data. PLS-R models are built to associate taxa and metabolites to the resistance score.

UC is a chronic inflammatory condition of the digestive tract that impacts the large intestine and results in progressively worsening inflammation and intestinal damage [12]. Patients typically progress through a sequence of therapies including antibiotics and general immunosuppressive drugs, to more targeted biologic agents, the most common of which are anti-TNF agents [13]. However, 10-40% [14] of patients will exhibit primary resistance, and up to 50% [15] of initial responders will eventually acquire resistance, depending on the disease type and experiment design.

Therefore, anti-TNF resistance represents a major clinical problem in UC and the identification of microbial and metabolite biomarkers of response could aid in the development of probiotic and prebiotic approaches to enhance response and overcome resistance.

Since information about anti-TNF response is not available in the meta-data for the IBDMDB, we developed three transfer learning approaches to leverage UC patient gene expression data matched to anti-TNF response in another cohort to infer an anti-TNF response gradient and biomarkers in IBDMDB. To compare these approaches, we held the initial model constant, a Partial Least Squares Discriminant Analysis (PLS-DA), and compared three strategies for knowledge transfer we term (1) Supervised Model, (2) Relative Separation, and (3) Signature Transfer. We show how each of these methods enables discovery of cross-cohort biomarkers, assess consistency of different approaches, and offer recommendations on how to generalize these approaches to other classes of machine learning models for trans-omic, cross-cohort biomarker discovery.

2. Methods

2.1 Datasets – Download and Processing

Multi-omics data was obtained from the integrated Human Microbiome Project (iHMP) IBD Multi-omics Database [16, 17]. Large intestine samples from Ulcerative colitis and non-IBD control patients were selected if each unique patient had all three of the following data types: gut metabolomics data, 16S rRNA seq data, and colorectal transcriptomics. The cohort consisted of 18 UC patients with all three sets of data. 16S rRNA-seq and gut metabolomic data were log₂ normalized, and the transcriptomic data was z-scored normalized. Gene expression data for UC patients matched to Infliximab response information were obtained from Gene Expression Omnibus (GEO) from dataset GSE16879 (N = 24) [18, 19]. Data were log₂ RMA normalized [20] and the top 33% of most variable genes were selected for analysis. The IBDMDB gene expression dataset was filtered for just these top 33% of most variable genes from GSE16879 to ensure comparability.

2.2 Partial Least Squares and Variable Importance of Projection Analysis

Partial Least Squares Discriminant Analysis (PLS-DA) and Regression (PLS-R) models were trained in MATLAB_R2022a using the ‘plsregress’ function. For training the initial model with GSE16879 gene expression data, models with 1 to 8 latent variables (LV) were assessed using 6-fold cross-validation. Percent variance explained in Y (infliximab response) and minimized mean squared error (MSE) were examined to select the optimal number of LVs for Knowledge Transfer to the test set. Test set models using metabolomics or microbial taxa information were trained examining 1 to 8 LVs using 6-fold cross-validation and the optimal number of LVs were selected based on percent variance explained and minimized MSE. For gene expression, metabolomics, and 16S rRNA-seq data, predictive features were identified in the PLS models using variable importance of projection (VIP) analysis. A VIP score assesses the weighted variance captured by a feature in a PLS model relative to the total variance captured in the model. A feature with VIP score greater than 1 is considered significantly predictive and higher VIP scores indicate more predictive features.

While other methods for predictive modeling do exist (e.g. random forest, neural networks) that we could examine here, the strength of PLS-DA and PLS-R is the ease of interpretation of the loading coefficients on the latent variables, enabling us to use the inferred biological signatures as validation lists. This is necessary since the nature of Trans-omic Knowledge Transfer involves prediction on a test set for which we cannot know the ground-truth biological signatures.

2.3 Case 1: Supervised Classifier Transfer

Following PLS-DA model training on the GSE16879 gene expression data predicting Infliximab response, the model regression coefficients matrix β , was extracted. We applied β to the z-score normalized IBDMDB gene expression data, filtered for overlapping genes with the top 33% most variable genes, to predict an Infliximab response for the IBDMDB samples. Predicted values greater than 0 were marked as “sensitive” or “1” and less than 0 were marked as “resistant” or “-1”. We then used these labels to construct PLS-DA models for the IBDMDB metabolomics and 16S rRNA-seq data predicting the Infliximab response variable. Models were trained and predictive metabolites and microbial taxa were extracted via the procedures described in 2.2.

2.4 Case 2: Relative Separation Transfer

Following PLS-DA model training on the GSE16879 gene expression data predicting Infliximab response, the model weights matrix W was extracted. We multiplied W by the z-score normalized IBDMDB gene expression data, filtered for overlapping genes with the top 33% most variable genes, to predict an Infliximab response for the IBDMDB samples, to infer the scores of IBDMDB samples on GSE16879-trained latent variables. Using this continuous Y matrix, we constructed PLS-R models for the IBDMDB gut metabolomics and 16S rRNA-seq data to predict separation of IBDMDB samples on Infliximab response-associated latent variables. Following model training as described in 2.2 for IBDMDB metabolomics and 16S rRNA-seq data, predictive metabolites and microbial taxa were extracted via VIP analysis.

2.5 Case 3: Signature Transfer

Following PLS-DA model training on the GSE16879 gene expression data predicting Infliximab response, significantly predictive genes were extracted via VIP analysis at a threshold of $VIP > 2$. These genes were used to define a gene set for analysis via single sample Gene Set Enrichment Analysis (ssGSEA). We analyzed the IBDMDB gene expression data in R (v4.1.1) using the package ssGSEA2.0 to infer patient-specific Infliximab resistance pathway scores. After running ssGSEA2.0, sample-specific Infliximab-resistance scores for the IBDMDB patients were extracted for downstream analysis. We trained PLS-R models with the IBDMDB metabolomics and 16S rRNA-seq data to predict the Infliximab resistance scores. Models were trained and significantly predictive metabolites and microbial taxa were extracted via VIP analysis as described in 2.2.

2.6 Data Code Availability

All data required to reproduce the findings in the manuscript is publicly available through GEO [18] or the IBDMDB [8] portals. All code required to reproduce the findings in this manuscript is available at: <https://github.com/WeldonSchool-BrubakerLab/psb2023.git>

3. Results

3.1 Training the source PLS-DA model on Infliximab Response-Matched Transcriptomics

We trained a PLS-DA model predicting Infliximab response from large intestine pinch biopsy gene expression data in GSE16879 as the foundational model we held constant in comparing Trans-omic Knowledge Transfer modeling approaches. Processed gene expression data were filtered for the top 33% most variable genes (5,779 genes) as our X-block and a PLS-DA model was trained using a

binary Y variable for Infliximab response (Sensitive 1, Resistant -1) (Figure 2A). UC samples primarily separated by Infliximab response on LV1 and a two latent variable model minimized prediction error across 6-fold cross-validation. Using Variable Importance of Projection (VIP) analysis, we identified genes in the model predictive of Infliximab response (2,266 genes $VIP > 1$) and extracted a 70 gene Infliximab resistance signature ($VIP > 1$) to define an Infliximab resistance gene set for trans-omic models (Figure 2B-2C). Of those genes, 55 were up-regulated in Infliximab resistant patients relative to sensitive patients.

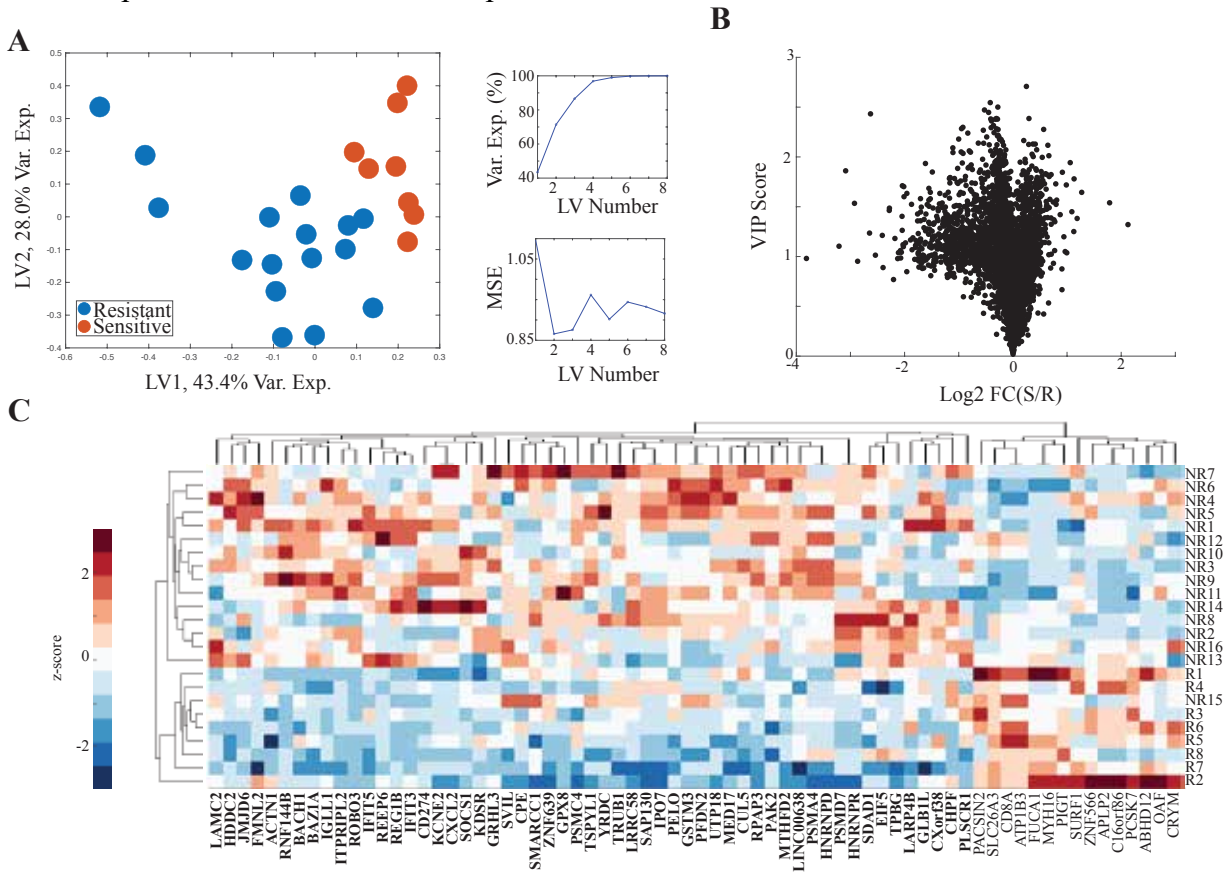


Figure 2. Training the underlying PLS-DA model. (A) Scores of GSE16879 ulcerative colitis samples in PLS-DA latent variables (LV) predicting Infliximab sensitivity or resistance. Two LV were selected explaining 71.4% variance and minimizing MSE. **(B)** Volcano plot of genes by VIP score and log2 fold change between R and NR patients. **(C)** Heatmap of Infliximab response-associated genes at $VIP > 2$ used to construct the gene set for Case 3: Signature Transfer Modeling. R-Responder, NR- Non-responder or resistant. Bolded genes are up-regulated in Infliximab resistant patients.

3.2 Inferring Gut Microbial Taxa Predictive of Infliximab Response

Having trained the initial PLS-DA model predicting Infliximab response from gene expression data in UC patients from GSE16879, we examined three approaches for Trans-omic Knowledge Transfer to identify gut microbial taxa predictive of Infliximab response using gene expression and 16S rRNA-seq data from IBDMDB (Figure 3). For our first case, Supervised Classifier Transfer, we applied the PLS-DA model trained on GSE16879 gene expression data to the IBDMDB gene expression data to predict a binary Infliximab response variable for these patients. Having made this prediction, we trained a new PLS-DA model (2 LV – 73.2% variance explained) predicting

IBDMDB Infliximab response labels from IBDMDB 16S rRNA-seq data (Figure 3A). The microbial taxa information was able to stratify the predicted labels primarily along LV1.

For the second case, Relative Separation Transfer, we applied the weights matrix W from the GSE16879-trained PLS-DA model to the gene expression data from IBDMDB to calculate the scores of IBDMDB samples on GSE16879 latent variables. We then trained a PLS-R model (2 LV – 59.9% variance explained) predicting IBDMDB scores on GSE16879 LV1 and LV2 using IBDMDB 16S rRNA-seq data (Figure 3B). Positive scores on GSE16879 LV1 and LV2 were associated with Infliximab resistance (Figure 2). We observed that these scores were not well stratified by the PLS-R model, but some separation could be observed on 16S rRNA-seq LV1.

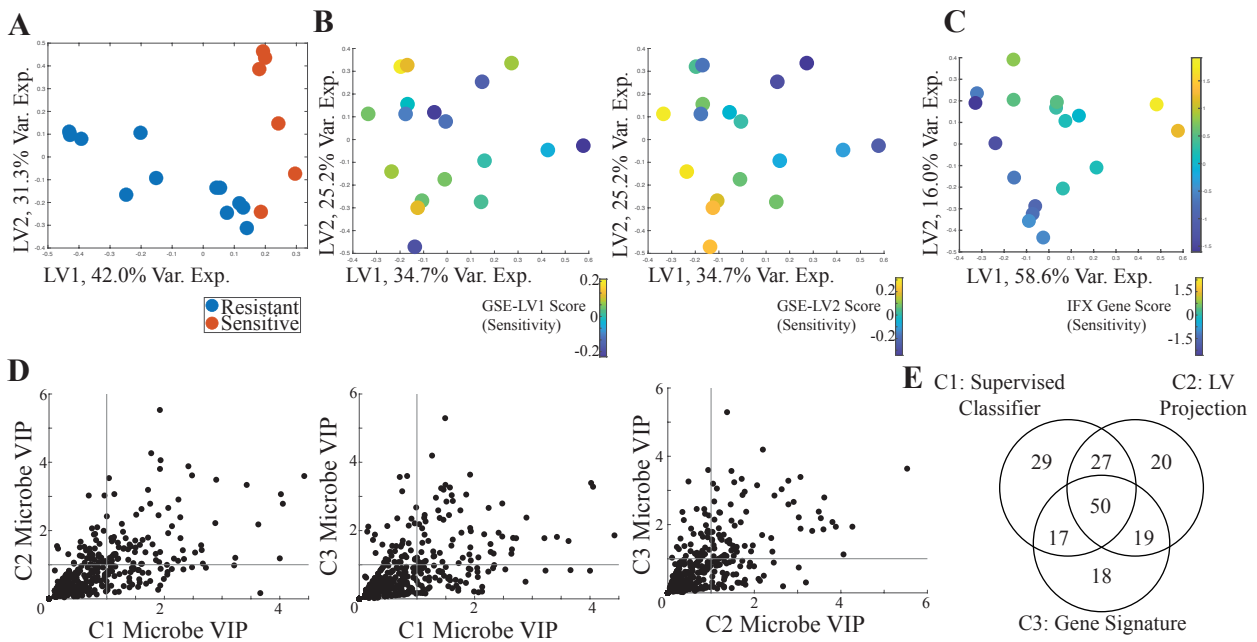


Figure 3. Gut Microbial Predictors of Infliximab Response. (A) Case 1 Supervised Classifier Transfer. Scores plot for a PLS-DA model predicting IBDMDB inferred Infliximab response classes from 16S rRNA-seq data. Infliximab response classes were predicted by applying the GSE16879 trained PLS-DA model to the IBDMDB gene expression data. (B) Case 2 Relative Separation Transfer. Scores plot for a PLS-R model trained to predict separation of IBDMDB samples on GSE16879 PLS-DA latent variables from 16S rRNA-seq data. Plots are colored by IBDMDB sample scores on GSE16879 latent variables 1 and 2 inferred using IBDMDB gene expression data. (C) Case 3 Signature Transfer: PLS-R model predicting IBDMDB Infliximab resistance gene score from 16S rRNA-seq data. (D) Comparison of microbial taxa VIP scores from Case 1 (C1), Case 2 (C2), and Case 3 (C3) PLS Knowledge Transfer models. (E) Venn diagram of the number of Infliximab response-associated taxa (VIP > 1) across models.

For the third case, Signature Transfer, we used the 70 genes with VIP scores greater than 2 from the GSE16879 PLS-DA model to define an Infliximab resistance gene set for single sample Gene Set Enrichment Analysis (ssGSEA) of the IBDMDB gene expression data. In brief, ssGSEA calculates an enrichment score for a pathway or gene set, for each sample in a dataset based on the cumulative expression of genes within that sample gene set. Here, we used ssGSEA to calculate an Infliximab resistance score for each sample in the IBDMDB gene expression data and then trained a PLS-R model predicting that score using the IBDMDB 16S rRNA-seq data (Figure 3C). We observed very strong separation of IBDMDB samples by Infliximab resistance scores along the

microbial taxa latent variables. Compared to the Supervised Classifier Transfer and Relative Separation Transfer approaches, the model trained using Signature Transfer generated latent variables capturing the greater proportion of variance between samples.

We performed VIP analysis of the microbial taxa in each Knowledge Transfer model to extract Infliximab response-predictive microbial taxa from each approach. When we compared the extracted features across models by VIP score, we observed that there was relatively little consistency between the biomarkers identified by each approach (Figure 3D). This suggests that while all approaches shared the same base-model, a PLS-DA model trained on GSE16879 gene expression data, the specific procedures of trans-omic knowledge transfer strongly influence the downstream-inferred biomarkers. Despite these differences, we were able to identify a core set of 50 microbial taxa associated with Infliximab response across all approaches (Figure 3E). Of these, 18 have been reported to be associated with anti-TNF- α response in clinical studies (Table 1).

Table 1. Bacteria abundance in response to anti-TNF treatment in IBD patients

SILVA Genus	Effect
Subdoligranulum	responder baseline \uparrow [21] responder post-therapy \uparrow [22]
Blautia	responder baseline \uparrow [23] responder after therapy \downarrow [21, 22]
Butyricicoccus	responder after therapy \uparrow [24]
Fusicatenibacter	responder after therapy \uparrow [22]
Roseburia	responder baseline \uparrow [24] responder after therapy \uparrow [25]
Clostridium sensu stricto 1	responder baseline \uparrow [24]
Faecalibacterium	responder baseline \uparrow [26] responder baseline \uparrow (F. prausnitzii) [21] responder after therapy \uparrow [22, 25] non-responder after therapy \downarrow [22]
Eubacterium hallii group	responder after therapy \uparrow [22] CD responder after therapy \uparrow [24]
Ruminococcacea NK4A214_group	responder baseline \uparrow [24]
Lachnospiraceae NK4A136 group	responder baseline \uparrow [24]
Eubacterium coprostanoligenes group	responder baseline \uparrow [24]
Dialister	non-responder baseline \uparrow responder after therapy \downarrow [21]
Ruminococcacea NK4A214 group	responder baseline \uparrow [24]
Coprococcus	responder after therapy \downarrow [21]
Ruminococcus gnavus group	responder after therapy \uparrow [24] responder baseline \uparrow [23]
Dorea	responder baseline \uparrow [23] responder after therapy \uparrow [22]
Bacteroides	relapser baseline \downarrow [27]
Eubacterium rectale	responder baseline \uparrow [28]

Faecalibacterium is one of the most abundant bacterial genera in the human intestine, and *Faecalibacterium prausnitzii* is the only known species in this genus [29]. Its abundance is reduced in both CD and UC [30-34]. Recent literature shows that a higher pre- and post-treatment level of *Faecalibacterium* correlates to a better anti-TNF- α response [21, 22, 25-27], which may result from the anti-inflammatory effect of a high amount of butyrate produced by *Faecalibacterium prausnitzii* [33]. Furthermore, the pre- and post-treatment level of *Blautia*, *Roseburia*, *Dorea*, and *Ruminococcus gnavus* group is also found to differentiate between anti-TNF- α responders and non-responders [21-25, 28, 35], inferring that they may be useful biomarkers for IBD prognostics.

3.3 Inferring Gut Metabolites Predictive of Infliximab Response

Similarly, to the microbial taxa Knowledge Transfer models, we used the PLS-DA model trained on UC patients from GSE16879 to examine Trans-omic Knowledge Transfer approaches to identify gut metabolites predictive of Infliximab response using gene expression and stool metabolomics data from IBDMDB (Figure 4). For Supervised Classifier Transfer, we used the same Infliximab response labels inferred for analysis of the 16S rRNA-seq data to train a new PLS-DA model using the IBDMDB stool metabolomics data to predict the inferred IBDMDB Infliximab response labels (Figure 4A). A two LV model (83.4% variance explained) strongly separated IBDMDB samples by predicted Infliximab response and appeared to capture more total variance explained in these samples than the 16S rRNA-seq PLS model.

For Relative Separation Transfer, we used the same projections of IBDMDB samples onto GSE16879 gene expression latent variables used in for the 16S rRNA-seq models in Figure 3. We trained a PLS-R model (2 LV – 69.9% variance explained) using gut metabolomics to predict IBDMDB scores on GSE16879 latent variables and observed that the metabolomics data produced clearer separation of projection scores and captured more sample-to-sample variance than the 16S rRNA-seq data (Figure 4B). For Signature Transfer, we trained a new PLS-R model predicting the IBDMDB Infliximab resistance gene score from the IBDMDB metabolomics data (2 LV - 86.0% variance explained) and observed strong separation of IBDMDB samples by resistance score on the inferred metabolomics latent variables (Figure 4C). Like the microbial taxa data, the clearest separation between samples and the largest variance explained was attributable to the Signature Transfer methodology. The models using the metabolomics data produced clearer separation between IBDMDB samples than the microbial taxa data in all matched cases, potentially due to the greater percent variance captured in the Y-block by the metabolomics data.

We performed VIP analysis of the metabolites in each Knowledge Transfer model to extract Infliximab response-predictive metabolites from each approach. Just like with the 16S rRNA-seq models, when we compared the metabolites across models by VIP score, we observed that there was relatively little consistency between the biomarkers identified by each approach (Figure 4D). This strengthens our observation that while all approaches shared the same base GSE16879 trained model, the trans-omic knowledge transfer approach strongly influence the downstream-inferred biomarkers. Despite these differences, we identified core set of 44 microbial taxa associated with Infliximab response across all approaches (Figure 4E).

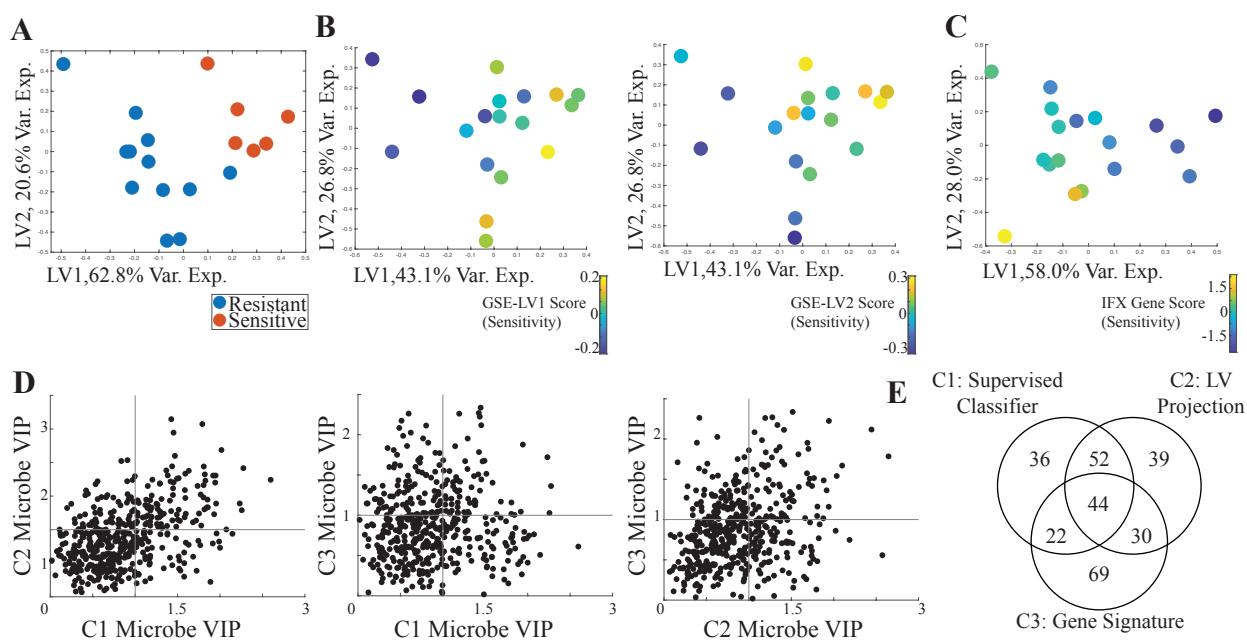


Figure 4. Gut Metabolite Predictors of Infliximab Response. (A) Case 1 Supervised Classifier Transfer. Scores plot for a PLS-DA model predicting IBDMDB inferred Infliximab response classes from gut metabolomics. Infliximab response classes were predicted by applying the GSE16879 trained PLS-DA model to the IBDMDB gene expression data. (B) Case 2 Relative Separation Transfer. Scores plot for a PLS-R model trained to predict separation of IBDMDB samples on GSE16879 PLS-DA latent variables from gut metabolomics. Plots are colored by IBDMDB sample scores on GSE16879 latent variables 1 and 2 inferred using IBDMDB gene expression data. (C) Case 3 Signature Transfer: PLS-R model predicting IBDMDB Infliximab resistance gene score from gut metabolomics data. (D) Comparison of microbial taxa VIP scores from Case 1 (C1), Case 2 (C2), and Case 3 (C3) PLS Knowledge Transfer models. (E) Venn diagram of number of Infliximab response-associated taxa (VIP > 1) across models.

Sphingomyelin (d18:1/16:0), a sphingolipid abundant on the apical side of the gastrointestinal epithelial cell membrane and in the myelin sheath of nerve cells [36], significantly increased in the UC mice [37, 38] and ileum of CD human [39]. It was reported that the sphingomyelin level was elevated in anti-TNF- α non-responding IBD patients' serum [40]. This accumulation potentially results from the downregulation of alkaline sphingomyelinase—one of the sphingomyelin digesting enzymes that exhibit anti-inflammatory properties in colitis mice—in IBD [41–43], inferring that the increase of sphingomyelin may correlate to the aggravation of the inflammation, which manifests as the diminished effect of anti-TNF- α . Our model also identified glycine as a core metabolite. It is an amino acid that has been reported to increase in the feces of adult and pediatric IBD patients [44, 45]. The metabolome profile of pediatric Crohn's Disease patients shows a decrease in pediatric CD patients after anti-TNF- α treatment [46]. Given glycine inhibits the TNF- α activity [47], such a decrease could result from the remission. In the same study, sebacic acid, a breakdown product of fatty acids that is normal in urine, was reported to be more abundant in the non-responder. Furthermore, metabolites like leucine, phosphatidylcholine, and arginine are closely related to TNF- α and associated pathways in IBD [48–51]. Their potential as metabolomic predictors of anti-TNF- α response warrants future studies.

4. Discussion

We show that Trans-omic Knowledge Transfer provides a framework for inferring multi-omic biomarkers of phenotypes across cohorts. The approaches we examined, Supervised Classifier, Relative Separation, and Signature Transfer, have methodological and interpretability differences with advantages and disadvantages. Supervised Classifier Transfer is direct application of a supervised model on a test set. New phenotypic labels are inferred in the test set using one data type, and secondary models are built to infer biomarkers in other data types in the test set. The challenge with this approach generally is that the validity of the inferred phenotypic labels cannot be directly assessed and for a binary phenotype, drug resistant or sensitive, the classification threshold at which the phenotypes are defined in the test set may influence the resulting downstream biomarkers.

Relative Separation Transfer partially addresses issues of classification threshold and by using a projection onto latent variables to define a continuum of anti-TNF resistance states in the test set. This allows for continuous modeling of relative differences in samples in the test set along one or more latent variables defined in the training data. However, the projection procedure appears to be the noisiest of the approaches we tested here based on the lack of clear separation by LV scores and interpreting positions on latent variables, rather than a binary phenotype, is challenging.

Signature Transfer is perhaps the most interpretable Knowledge Transfer approach we examined here. Once the gene signature is extracted from the training set, no other features of the model from the training data are retained, all inference of biomarkers is performed in the test set modeling the signature as a dependent variable. The final model thus only aims to characterize the trans-omic relationships in the test set and relative signature score associated with a phenotype. Separation between samples was clearer in this model compared to the Relative Separation approach and the internal consistency of the model mitigates some concerns of predicted class validity in the Supervised Classifier Transfer approach. Though the association of gene signature activity with anti-TNF resistance is still uncertain in this approach, we recommend Signature Transfer as the most rigorous and interpretable Trans-omic Knowledge Transfer approach among those tested here.

Despite the methodological differences in the three approaches, we find that a common set of microbial and metabolite biomarkers of anti-TNF response can be identified. Validation against literature suggests that consensus biomarkers inferred across approaches have potential clinical benefit. A fourth approach to Trans-omic Knowledge Transfer may be to construct multiple models using the same single-omic training and multi-omic test sets and extract the commonly identified trans-omic features for future biological studies. This *Ensemble* approach to Knowledge Transfer may be further augmented by testing multiple classes of prediction models, such as support vector machines, random forests, or neural networks, and extracting the resulting consensus biomarkers.

Our study has some limitations which may be addressed in future studies to extend the approach. While we present important feasibility and proof of concept data here, a disseminatable software toolbox would increase the impact and applicability of our approaches to other problems. Part of this should include additional benchmarking using pairs of multi-omics datasets, varying gene inclusion threshold percent, and withholding select data types to enable more quantitative validation metrics. While not examined here, in principle our frameworks could be expanded to other -omic data types, including proteomics, scRNA-seq, and metagenomics data, provided data types are encoded in latent variables reflective of data-specific distributional properties.

In conclusion, we demonstrate that Trans-omic Knowledge Transfer modeling is a potentially powerful approach for integrating multi-omics and single-omics data across clinical

cohorts to discover biomarkers of conditions and phenotypes measured in one or the other cohort. To our knowledge, there are no comparable approaches widely implemented for us to compare our approaches and results to, making this an important first feasibility study of the utility of Trans-omic Knowledge Transfer. The paucity of other methods in this space is likely due to the challenge of validating approaches with quantitative metrics, a limitation we acknowledge and propose a solution to for future methodological studies.

In future work, extensions of this approach could account for cohort-specific covariates in biomarker discover to enhance the robustness of the inferred associations. The ability to re-use clinical multi-omics data to answer novel biological questions adds an important tool to preclinical studies of drug resistance and disease biology. Such methods increase the value of the initial investment to generate the cohort by allowing basic and translational scientists to test new hypotheses through computational models of existing data and to potentially advance new therapies.

5. References

1. Chaudhary, K., et al., *Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer*. Clin Cancer Res, 2018. **24**(6): p. 1248-1259.
2. Xiao, Y., et al., *Multi-Omics Profiling Reveals Distinct Microenvironment Characterization and Suggests Immune Escape Mechanisms of Triple-Negative Breast Cancer*. Clin Cancer Res, 2019. **25**(16): p. 5002-5014.
3. Linskrog, S.V., et al., *An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer*. Nat Commun, 2021. **12**(1): p. 2301.
4. Sacco, K., et al., *Immunopathological signatures in multisystem inflammatory syndrome in children and pediatric COVID-19*. Nat Med, 2022. **28**(5): p. 1050-1062.
5. El Saie, A., et al., *Metabolome and microbiome multi-omics integration from a murine lung inflammation model of bronchopulmonary dysplasia*. Pediatr Res, 2022.
6. Bai, B., et al., *Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression*. Neuron, 2020. **105**(6): p. 975-991 e7.
7. Zhou, W., et al., *Longitudinal multi-omics of host-microbe dynamics in prediabetes*. Nature, 2019. **569**(7758): p. 663-671.
8. Schirmer, M., et al., *Dynamics of metatranscription in the inflammatory bowel disease gut microbiome*. Nat Microbiol, 2018. **3**(3): p. 337-346.
9. Fettweis, J.M., et al., *The vaginal microbiome and preterm birth*. Nat Med, 2019. **25**(6): p. 1012-1021.
10. Picard, M., et al., *Integration strategies of multi-omics data for machine learning analysis*. Comput Struct Biotechnol J, 2021. **19**: p. 3735-3746.
11. Chakraborty, S., et al., *Onco-Multi-OMICs Approach: A New Frontier in Cancer Research*. Biomed Res Int, 2018. **2018**: p. 9836256.
12. Head, K.A. and J.S. Jurenka, *Inflammatory bowel disease Part 1: ulcerative colitis--pathophysiology and conventional and alternative treatment options*. Altern Med Rev, 2003. **8**(3): p. 247-83.
13. Rutgeerts, P., S. Vermeire, and G. Van Assche, *Biological therapies for inflammatory bowel diseases*. Gastroenterology, 2009. **136**(4): p. 1182-97.
14. Ben-Horin, S., U. Kopylov, and Y. Chowers, *Optimizing anti-TNF treatments in inflammatory bowel disease*. Autoimmun Rev, 2014. **13**(1): p. 24-30.
15. Papamichael, K. and A.S. Cheifetz, *Therapeutic drug monitoring in inflammatory bowel disease: for every patient and every drug?* Curr Opin Gastroenterol, 2019. **35**(4): p. 302-310.
16. Integrative, H.M.P.R.N.C., *The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease*. Cell Host Microbe, 2014. **16**(3): p. 276-89.
17. Integrative, H.M.P.R.N.C., *The Integrative Human Microbiome Project*. Nature, 2019. **569**(7758): p. 641-648.
18. Arijs, I., et al., *Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment*. PLoS One, 2009. **4**(11): p. e7984.
19. Arijs, I., et al., *Effect of vedolizumab (anti-alpha4beta7-integrin) therapy on histological healing and mucosal gene expression in patients with UC*. Gut, 2018. **67**(1): p. 43-52.
20. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data*. Biostatistics, 2003. **4**(2): p. 249-64.
21. Ventin-Holmberg, R., et al., *Bacterial and Fungal Profiles as Markers of Infliximab Drug Response in Inflammatory Bowel Disease*. Journal of Crohn's & colitis, 2021. **15**(6): p. 1019-1031.
22. Sanchis-Artero, L., et al., *Evaluation of changes in intestinal microbiota in Crohn's disease patients after anti-TNF alpha treatment*. Scientific Reports, 2021. **11**(1).

23. Park, Y.E., 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, 2022. **12**(1).
24. Dovrolis, N., et al., *The interplay between mucosal microbiota composition and host gene-expression is linked with infliximab response in inflammatory bowel diseases*. Microorganisms, 2020. **8**(3).
25. Wang, Y., et al., *Characteristics of faecal microbiota in paediatric Crohn's disease and their dynamic changes during infliximab therapy*. Journal of Crohn's and Colitis, 2018. **12**(3): p. 337-346.
26. Magnusson, M.K., et al., *Anti-TNF therapy response in patients with ulcerative colitis is associated with colonic antimicrobial peptide expression and microbiota composition*. Journal of Crohn's and Colitis, 2016. **10**(8): p. 943-952.
27. Rajca, S., et al., *Alterations in the intestinal microbiome (Dysbiosis) as a predictor of relapse after infliximab withdrawal in Crohn's disease*. Inflammatory Bowel Diseases, 2014. **20**(6): p. 978-986.
28. Yilmaz, B., et al., *Microbial network disturbances in relapsing refractory Crohn's disease*. Nat. Med, 2019. **25**(2): p.323-336.
29. Lopez-Siles, M., et al., *Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics*. The ISME Journal 2017 **11**:4, 2017. **11**(4): p. 841-852.
30. Martinez-Medina, M., et al., *Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis*. Inflamm Bowel Dis, 2006. **12**(12): p. 1136-1145.
31. Joossens, M., et al., *Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives*. Gut, 2011. **60**(5): p. 631-637.
32. Machiels, K., et al., *A decrease of the butyrate-producing species roseburia hominis and faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis*. Gut, 2014. **63**(8): p. 1275-1283.
33. Sokol, H., et al., *Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(43): p. 16731-16736.
34. Swidsinski, A., et al., *Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora*. Inflammatory Bowel Diseases, 2008. **14**(2): p. 147-161.
35. Sakurai, T., et al., *Mucosal microbiota and gene expression are associated with long-term remission after discontinuation of adalimumab in ulcerative colitis*. Scientific Reports, 2020. **10**(1).
36. Duan, R.D. and Å. Nilsson, *Metabolism of sphingolipids in the gut and its relation to inflammation and cancer development*. Progress in Lipid Research, 2009. **48**(1): p. 62-72.
37. Qi, Y., et al., *PPAR α -dependent exacerbation of experimental colitis by the hypolipidemic drug fenofibrate*. American Journal of Physiology - Gastrointestinal and Liver Physiology, 2014. **307**(5): p. 564-573.
38. Fischbeck, A., et al., *Sphingomyelin induces cathepsin D-mediated apoptosis in intestinal epithelial cells and increases inflammation in DSS colitis*. Gut, 2011. **60**(1): p. 55-65.
39. Braun, A., et al., *Alterations of phospholipid concentration and species composition of the intestinal mucus barrier in ulcerative colitis: A clue to pathogenesis*. Inflammatory Bowel Diseases, 2009. **15**(11): p. 1705-1720.
40. Ding, N.S., et al., *Metabonomics and the Gut Microbiome Associated With Primary Response to Anti-TNF Therapy in Crohn's Disease*. Journal of Crohn's & colitis, 2020. **14**(8): p. 1090-1102.
41. Sjöqvist, U., et al., *Chronic Colitis Is Associated With a Reduction of Mucosal Alkaline Sphingomyelinase Activity*. Inflammatory Bowel Diseases, 2002. **8**(4): p. 258-263.
42. Duan, R.D., et al., *Distribution of alkaline sphingomyelinase activity in human beings and animals. Tissue and species differences*. Digestive diseases and sciences, 1996. **41**(9): p. 1801-1806.
43. Andersson, D., et al., *Expression of alkaline sphingomyelinase in yeast cells and anti-inflammatory effects of the expressed enzyme in a rat colitis model*. Digestive Diseases and Sciences, 2009. **54**(7): p. 1440-1448.
44. Bjerrum, J.T., et al., *Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals*. Metabolomics, 2015. **11**(1): p. 122-133.
45. Kolho, K.L., et al., *Faecal and Serum Metabonomics in Paediatric Inflammatory Bowel Disease*. Journal of Crohn's and Colitis, 2017. **11**(3): p. 321-334.
46. Wang, Y., et al., *Microbial and metabolic features associated with outcome of infliximab therapy in pediatric Crohn's disease*. Gut Microbes, 2021. **13**(1): p. 1-18.
47. Hasegawa, S., et al., *Cysteine, histidine and glycine exhibit anti-inflammatory effects in human coronary arterial endothelial cells*. Clinical and experimental immunology, 2012. **167**(2): p. 269-274.
48. Laplante, M. and D.M. Sabatini, *mTOR signaling in growth control and disease*. Cell, 2012. **149**(2): p. 274-293.
49. Treede, I., et al., *TNF-alpha-induced up-regulation of pro-inflammatory cytokines is reduced by phosphatidylcholine in intestinal epithelial cells*. BMC gastroenterology, 2009. **9**.
50. Treede, I., et al., *Anti-inflammatory effects of phosphatidylcholine*. The Journal of biological chemistry, 2007. **282**(37): p. 27155-27164.
51. Krzystek-Korpacka, M., et al., *Transcriptional and Metabolomic Analysis of L-Arginine/Nitric Oxide Pathway in Inflammatory Bowel Disease and Its Association with Local Inflammatory and Angiogenic Response: Preliminary Findings*. International journal of molecular sciences, 2020. **21**(5).