# Peripheral blood DNA methylation signatures predict response to vedolizumab and ustekinumab in adult patients with Crohn's disease: The EPIC-CD study

Vincent W. Joustra[1*], Andrew Y.F. Li Yim[2,3*], Peter Henneman[3*], Ishtu Hageman[1,2], Tristan de Waard[4], Evgeni Levin[4], Alexandra J. Noble[5], Thomas P. Chapman[5,6], Femke Mol[2], Sarah van Zon[1], Donghyeok Lee[4], Colleen G.C. McGregor[5], Alex T. Adams[5], Jack J. Satsangi[5**], Wouter J. de Jonge[2,7**], Geert R. D'Haens[1**]; EPIC-CD Consortium.

**EPIC-CD Consortium:** Hans Paulich, Inge van Welsen, Rishand Ramkisoen, Yara Abeling, Marloes Zwart, Nicole Berkers, Caroline Verseijden, Anje te Velde, Nicola Radcliffe, Aisling Quinn, Beatriz Camesella Perez, Tom Thomas, Rahul Ravindran, James Chivenga, Simon Travis.

[1]Amsterdam UMC - University of Amsterdam, Department of Gastroenterology & Hepatology, Amsterdam, The Netherlands.
[2]Amsterdam UMC - University of Amsterdam, Tytgat Institute for Liver and Intestinal Research, Amsterdam, The Netherlands.
[3]Amsterdam UMC - University of Amsterdam, Genome Diagnostics Laboratory, Department of Human Genetics, Amsterdam, The Netherlands.
[4]Horaizon BV, Delft, The Netherlands.
[5]Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Translational Gastroenterology Unit & NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom.
[6]Department of Gastroenterology, St Richard's and Worthing Hospitals, University Hospitals Sussex NHS Foundation Trust, West Sussex, UK
[7]University of Bonn, Department of Surgery, Germany.

* These authors share first authorship
** The 3 last authors are shared senior authors

*Address for correspondence:* Prof. Geert D'Haens PhD, Amsterdam University Medical Centers, Department of Gastroenterology and Hepatology, Room C2-208, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
*Email:* g.dhaens@amsterdamumc.nl

*Funding:* This project was supported by a grant from the Helmsley Foundation [2019-1179]

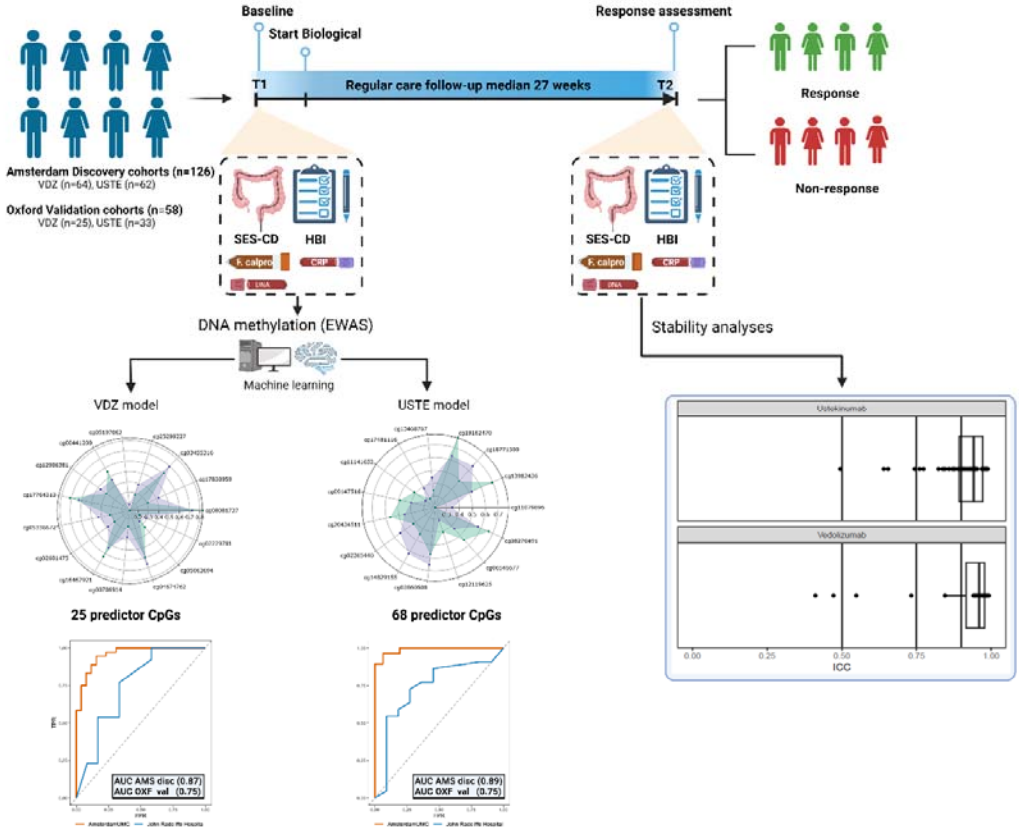
Short title: DNA methylation predicts response to biologicals in CD

Acknowledgement: Floris de Voogd for his assistance in evaluation of ultrasound images.

Conflicts of interest: The AmsterdamUMC has a patent pending for the vedolizumab and ustekinumab response prediction models presented in this manuscript. ALY received honoraria from Janssen, Johnson & Johnson, DeciBio and was employed by GSK. WJ received honoraria from Janssen, Johnson & Johnson and is a cofounder of AIBiomics BV. GD received speaker fees from Janssen, Johnson & Johnson. EL is a cofounder of

38 Horaizon BV and AIBiomics BV. The remaining authors disclose no conflicts. TC received honoraria from Janssen
39 and Takeda. The remaining authors have no conflicts of interest to declare.

# Graphical abstract



# Abstract

Biological therapeutics are now widely used in Crohn's disease (CD), with evidence of efficacy from randomized trials and real-world experience. Primary non-response is a common, poorly understood problem. We assessed blood methylation as a predictor of response to vedolizumab (VDZ, anti-a4b7 integrin) or ustekinumab (USTE, anti-IL-12/23p40). We report a two-center, prospective cohort study in which we profiled the peripheral blood DNA methylome of 184 adult male and female CD patients prior to and during treatment with VDZ or USTE in a discovery (n=126) and an external validation cohort (n=58). We defined epigenetic biomarkers that were stable over time and associated with combined clinical and endoscopic response to VDZ or USTE with an area under curve (AUC) of 0.87 and 0.89, respectively. We validated these models in an external cohort yielding an AUC of 0.75 for both VDZ and USTE. These data will now be prospectively tested in a multicenter randomized clinical trial.

# Introduction

Crohn's disease (CD) is an incurable, chronic, relapsing inflammatory bowel disease (IBD) caused by a complex interplay between the environment, gut microbiome, and a dysregulated immune system in genetically susceptible patients[1,2]. Accessibility to high-throughput "omics" technology has enhanced our understanding of the underlying molecular pathogenesis of CD leading to the development of several monoclonal antibodies or biologicals that target specific inflammatory pathways in an effort to suppress the inflammation and to induce and or maintain a state of clinical and endoscopic remission[3]. Currently, the repertoire of approved biologicals in CD includes anti-TNF antibodies (infliximab (IFX) and adalimumab (ADA)), the anti-$\alpha_4\beta_7$ integrin antibody vedolizumab (VDZ) and the anti-IL12/23p40 antibody ustekinumab (USTE). More recently, specific IL23p19 antibodies and JAK inhibitors have also been approved for clinical use[4-6]. Despite the established efficacy of these biological treatments to induce corticosteroid-free *clinical remission* in up to 65% of CD patients, sustained *endoscopic remission* is observed in not more than a third of patients after 1 year of treatment[7-9]. This creates a clinical challenge since therapeutic guidelines suggest to use endoscopic remission as a target. An increasingly common clinical scenario is the choice between VDZ or USTE as second-line treatment for patients who have not responded to anti-TNF therapies.

To date, treatment selection has been based on a trial-and-error approach. Given the limited efficacy, many patients are therefore treated with insufficiently effective treatment, which is associated with an increased risk of complications (stenosis, fistula, abscesses, nutritional deficiencies) and surgery. The development of strategies that allow selection of treatment based on the likelihood of response is an important unmet need. Although various efforts using clinical[10], transcriptomic[11-13], proteomic[14] or microbial[15,16] technologies have been investigated and reported, no predictive biomarkers have made their way to clinical application[17-19].

DNA methylation is one of the most studied epigenetic features characterized by the covalent binding of methyl groups to nucleotides, most often a cytosine in a cytosine-phosphate-guanine (CpG) sequence in humans[20]. DNA methylation is believed to play an essential role in the regulation of gene expression, thereby determining cellular phenotype and behavior without altering the DNA sequence itself[21,22]. Within the context of IBD, DNA methylation has gained particular interest due to its dynamic interaction with the environment and suggested epigenetic-microbial crosstalk[23-25]. A number of recent studies demonstrated differential DNA methylation profiles associated with the presence of CD and/or specific CD-phenotypes[26,27] in

93 peripheral blood leukocytes (PBL)[28-30], intestinal mucosa[31] and specific cell-types thereof[32,33].
94 Most studies proposed a potential role of the DNA methylome in diagnostics and prediction
95 of treatment response.

96 In the current study we performed an epigenome-wide association study (EWAS) in which
97 we identified and validated prognostic DNA methylation signatures in peripheral blood of
98 adult CD patients associated with objective therapeutic response to VDZ and USTE. We
99 subsequently interrogated whether these signatures were stable over time and assessed the
100 influence of common confounding variables. We further performed additional validation
101 analyses of the model against patients with previous non-response to one or both drugs.
102 Finally, we investigated the association of the identified DNA methylation biomarkers with
103 gene expression through transcriptomic analyses.

104

# Results

## Study population

We prospectively recruited a discovery cohort of 126 adult patients at the IBD Center of Amsterdam University Medical Centers (Amsterdam UMC), Amsterdam, Netherlands. All patients had active symptomatic and endoscopic CD and were scheduled to start VDZ (N = 64) or USTE (N = 62). Evidence for active disease was documented with a validated clinical score (Harvey Bradshaw Index (HBI), median 8 (interquartile range (IQR) 4-12)), biochemical tests (serum C-reactive protein (CRP), median 6.1 mg/L (IQR 2.2-14.8) and fecal calprotectin (FCP), median 903 µg/g (IQR 278-1816)) and also endoscopic signs of inflammation measured with the simple endoscopic score for CD (SES-CD), median 9 (IQR 6-15). In addition, an external validation cohort of 58 adult CD patients starting VDZ (N = 25) or USTE (N = 33) biological therapy were recruited at the John Radcliffe Hospital, Oxford, United Kingdom. Peripheral blood leukocyte (PBL) samples were obtained prior to treatment initiation and upon response assessment at a median of 27 (IQR 20-33) weeks into treatment, when patients were classified as responder (R) or non-responder (NR). A detailed overview of all the clinical characteristics across the different cohorts and treatments can be found in the methods section as well as in **Table 1**.

## Blood DNA methylation profiling predicts response to vedolizumab and ustekinumab

To identify prognostic biomarkers of VDZ and USTE response, we performed supervised machine learning through stability selected gradient boosting[34,35] on blood samples obtained shortly before the start of treatment (**Fig. 1a**). The model was trained on the discovery cohort acquired in Amsterdam and subsequently validated in the validation cohort acquired in Oxford. We were able to generate response-predicting models with an area under the curve (AUC) of 0.87 with a standard deviation of 0.06 and 0.89 with a standard deviation of 0.08 for VDZ and USTE, respectively, when testing against the discovery cohort (**Fig. 1b**). The models comprised 25 and 68 differentially methylated CpGs for VDZ and USTE, respectively (**Fig. 1c, Supplementary Fig. 1-2** and **Supplementary Tables 1-2**). Validating our models against the independent validation cohort yielded an AUC of 0.75 for both VDZ and USTE (**Fig. 1b**), indicating reproducible response-associated differences in DNA methylation prior to the start of treatment.

While the discrepancy between discovery and validation performances is expected as part of machine learning, we hypothesized that the differences were partially due to differences in

response assessments. During the COVID-19 pandemic, endoscopic assessments were scheduled less frequently resulting in the validation cohort consisting of patient samples in whom response was defined on both strict (combined clinical-, biochemical- and endoscopic evaluation) or modified criteria (combined clinical- and biochemical evaluation). We therefore investigated whether a difference in performance could be observed between both methods of evaluation. This stratification indicated that accuracy was optimized by using the strictly defined combination of clinical- and endoscopic endpoints for both VDZ ($AUC_{strict}$ = 0.83 versus $AUC_{modified}$ = 0.66) and USTE ($AUC_{strict}$ = 0.83 versus $AUC_{modified}$ = 0.72) (**Supplementary Fig. 3a**). Application of the strict criteria yielded performances similar to the discovery dataset thereby confirming that our model is likely more capable at predicting response as defined using combined clinical-, biochemical- and endoscopic evaluations.

In addition, we were interested whether the performance of our models was affected by prior exposure to anti-TNF medication. We observed a better performance of our models among anti-TNF naïve compared to anti-TNF exposed patients for both VDZ ($AUC_{non\text{-}exposed}$ = 0.85 versus $AUC_{exposed}$ = 0.66) and USTE ($AUC_{non\text{-}exposed}$ = 0.97 versus $AUC_{exposed}$ = 0.63) (**Supplementary Fig. 3b**).

Focusing on the practical implications in a clinical setting, we calculated a sensitivity of 0.769 and a specificity of 0.67 for VDZ and both a sensitivity and specificity of 0.73 for USTE (**Table 2**). Next, we computed the likelihood ratio of response and the post-test probability of response to aid clinicians in accurately predicting response following a positive test outcome. From Lowenberg *et al.* [7], we note that 50 of the 110 VDZ-treated CD patients presented with endoscopic response at week 52 indicating a pre-test probability of response of 0.45. At the calculated sensitivity of 0.77 and a specificity of 0.67, the likelihood ratio of response is 2.31, thereby making the post-test probability of response 0.65 for VDZ. Similarly, for USTE, prior research showed that 75 of the 179 USTE-treated CD patients presented with endoscopic response at week 52[8], indicating a pre-test probability of response of 0.42. At the calculated sensitivity and specificity of 0.73 the likelihood ratio is 2.67 and the post-test probability of response is 0.66. Taken together, the probability that a patient would actually respond to VDZ or USTE when classified as responder is 0.20 and 0.24 higher than the current standard of care.

## Methylation status of predictor CpGs remains stable over time

As baseline samples were acquired pre-treatment, we sought to understand whether the initiation of treatment affected the methylation status of the predictor CpGs. To this end, we compared samples obtained at the time of response assessment (T2) with samples obtained

pre-treatment (T1). We could not identify any statistically significant differences in DNA methylation levels for any of the predictor CpGs (**Fig. 2a**). Indeed, comparing the differences over time suggested that the mean difference between R and NR was similar both pre-treatment and at response assessment (**Fig. 2b**). Furthermore, a two-way, mixed, consistency intra-class correlation (ICC) analysis indicated highly-stable DNA methylation over time with 24 out of 25 VDZ and 62 out of 68 USTE predictor CpGs presenting ICC values ≥0.75 (**Fig. 2c**, **Supplementary Fig. 1** and **2**). This observation was corroborated by interrogating our previous longitudinal consistency analysis of peripheral blood DNA methylation from 46 adult IBD patients collected at 2 time points with a median of 7 years (range, 2-9 years) in between[36]. Here, we observed that the majority (16 out of 25 VDZ and 52 out of 68 USTE) of the predictor CpGs presented "good" ($0.75 \leq ICC < 0.9$) to "excellent" ($0.9 \geq ICC$) stability[37] over a median span of 7 years (**Fig. 2c**). As a final validation, we utilized the prognostic model to predict response to therapy of the samples obtained at response assessment, where we obtained better performances compared to the pre-treatment samples ($AUC_{VDZ} = 0.97$; $AUC_{USTE} = 1.00$) (**Fig. 2d**). Altogether, our observations suggest that response-associated differences in DNA methylation detected prior to treatment remain stable during treatment.

## Patients who previously experienced treatment failure to multiple biological treatments are classified correctly

Having established that the prognostic response prediction models for both drugs perform well both pre- and into treatment, we explored the performance of our model in a further independent cohort of 33 adult CD patients from whom blood had been collected after sequential treatment with anti-TNF, VDZ, and/or USTE. Treatment failure was determined with endoscopy in 28 (82%), MRI in 2 (6%), or the need for surgical resection in 4 (12%), combined with biochemical biomarkers such as CRP in 12 (35%) and/or FCP in 16 (47%) and clinical parameters in 32 (94%) patients. In this cohort, a relatively large proportion of patients had extensive disease location (67.6%), perianal disease (52.9%), and/or IBD-related surgery (70.6%) in the past (**Supplementary Table 3**).

Both VDZ and USTE response prediction models accurately predicted NR, with 24 out of 27 (88.9%) VDZ-NR patients and 24 out of 26 (92.3%) USTE-NR patients correctly identified. Focusing specifically on the patients with previous non-response to both VDZ and USTE the models correctly classified 16 out of 19 (84%) as NR. Notably, 12 of out of these 19 patients (75%) were anti-TNF experienced.

## Assessment of potential confounding variables

Through multiple linear regression analyses we observed that 22 of the 25 (88%) VDZ response-associated CpGs and 38 of the 68 (55%) USTE response-associated CpGs (**Fig. 3a** and **Supplementary Fig. 1-2**) presented statistically significant differences. The large discrepancy between the linear regression analyses and the USTE response-associated CpGs indicates that a more complex non-linear relationship exists among the response-associated predictor CpGs and underscores that statistical p-values may not equal biologic or functional relevance and/or importance at an individual level. As it has been established that the peripheral blood DNA methylome is associated with certain phenotypic characteristics such as sex, age, smoking status, as well as the underlying cellular composition[38-41], we investigated whether any of these variables confounded our results. In doing so with linear regression analysis, we observed that 12 out of 22 (55%) and 30 out of 38 (79%) markers remained significantly associated with response for VDZ and USTE, respectively (**Fig. 3a**). In terms of effect size, the mean percentage methylation difference between R and NR of the predictor CpGs on average decreased by 15% and increased by 5% in VDZ and USTE, respectively (**Fig. 3b**). To understand whether the confounding variables are capable of predicting response, we constructed a prediction model solely based on the confounding variables using the discovery cohort and tested this on the validation cohort. The confounder model yielded an AUC of 0.57 and 0.65 for VDZ and USTE, respectively, against the validation cohort indicating worse performance than the prediction model based on CpGs only (**Fig. 3c**). Our results indicate that while there appears to be a significant association with sex, age, smoker status and blood cell distribution, the CpG-only prediction model outperforms a confounder-only model.

We next investigated whether the predictor CpGs were significantly associated with severity of systemic and intestinal inflammation at baseline measured using CRP and FCP, as was previously reported by Somineni *et al.*[42]. For VDZ, 5 predictor CpGs significantly associated with CRP whereas only a single CpG was associated with FCP (**Supplementary Fig. 4**). For USTE, we observed 9 predictor CpGs associated with CRP and 2 predictor CpGs with FCP (**Supplementary Fig. 4**). In all cases, mean differences in methylation were smaller than 0.5%. These observations suggest that the majority of the identified predictor CpGs are independent of inflammatory status at onset of therapy.

## RFPL2 presents concordant response-associated differential methylation and expression

To understand the functional relevance of the predictor CpGs, we annotated the them to their respective genes based on whether they were located in either a gene promoter or enhancer

resulting in the annotation of 20 VDZ response-associated predictor CpGs to 16 unique genes (**Supplementary Table 1**) and 43 USTE response-associated predictor CpGs to 46 genes (**Supplementary Table 2**). We performed transcriptomic analyses on a subset of samples from the discovery cohort (VDZ; $N_{T1}=10_R10_{NR}$, $N_{T2}=10_R8_{NR}$ and USTE; $N_{T1}=14_R11_{NR}$, $N_{T2}=10_R9_{NR}$). Comparing R with NR identified pretreatment differences for predictor CpG-associated genes *TULP4* (p-value$_{T1}$ = 4.38E-02) and *RFPL2* (p-value$_{T1}$ = 4.72E-02) for VDZ (**Fig. 4a, b** and **Supplementary Table 4**), with *RFPL2* presenting a significant inverse correlation between DNA methylation and gene expression (Pearson r = -0.65; p-value = 1.39E-03) (**Fig. 4c**). For USTE, we observed significant differences in the expression of predictor CpG-associated genes *MRC1* (p-value$_{T1}$ = 4.25E-04) and *TMEM191B* (p-value$_{T1}$ = 7.31E-03) (**Fig. 4a** and **d** and **Supplementary Table 5**) with *TMEM191B* presenting a significant positive correlation between DNA methylation and gene expression (Pearson r = 0.57; p-value = 3.11E-03) (**Fig. 4e**). We next investigated whether predictor CpG-associated genes presented differential expression after the start of treatment by comparing R with NR at response assessment. VDZ predictor CpG-associated genes *MCM2* (p-value$_{T2}$ = 2.40E-03) and *RFPL2* (p-value$_{T2}$ = 3.88E-03) were differentially expressed at response assessment (**Fig. 4a, f** and **Supplementary Table 4**) with *RFPL2*, once again, presenting a significant inverse correlation between DNA methylation and expression (Pearson r = -0.55; p-value = 0.017) (**Fig. 4g**). For USTE, predictor CpG-associated genes *POTEF* (p-value$_{T2}$ = 1.47E-02), *HDAC4* (p-value$_{T2}$ = 2.18E-02), *PARP4* (p-value$_{T2}$ = 3.49E-02) and *MARK3* (p-value$_{T2}$ = 2.87E-02) presented differential expression at response assessment (**Fig. 4a, h** and **Supplementary Table 5**), but did not show any significant correlation with DNA methylation. Taken together, our results show that some of the response predictor CpG-associated genes present differential expression either pretreatment and/or during response assessment, with VDZ- and USTE- response-associated genes *RFPL2* and *TMEM191B*, respectively, presenting a significant correlation between DNA methylation and gene expression. Nonetheless, we acknowledge that most predictor CpG-associated genes present no response-associated differential expression.

# Discussion

Although the introduction of biologicals has transformed the care of patients with CD, the current clinical practice of treatment selection remains suboptimal. Nonetheless, a large body of real-world evidence studies suggests there could be a place for a more individualized approach[43,44]. Therefore, in parallel with ongoing drug development, predictive biomarkers that allow for selection of successful medical therapy would represent a major step forward in clinical care.

Here, we conducted a longitudinal case-control study where we identified methylation signatures composed of 25 and 68 markers associated with combined endoscopic, biochemical and clinical response to VDZ and USTE, respectively, in a cohort of adult CD patients in whom treatment response was assessed with stringent criteria. We were able to build models with significant predictive performance at an AUC > 0.85 for both models. The models demonstrated similar performance in an independent, external validation with an AUC of 0.75 for both models.

Recent real-world data and (post-hoc) findings from both the GEMINI and UNITI trials indicate superior response to both VDZ and USTE in anti-TNF naïve patients[45-49]. In our discovery cohorts, 77% of VDZ and 98% of the USTE-treated patients were previously exposed to anti-TNF medication. Stratifying the patients in the validation cohort by previous anti-TNF exposure showed that both models performed noticeably better in anti-TNF naïve rather than exposed patients. However, the number of patients included in both subset comparisons are relatively small and further exploration using larger groups patients are needed. Furthermore, we demonstrate the ability of our models to effectively identify patients with previous non-response to both VDZ and USTE treatment. This holds true for both anti-TNF naïve and experienced patients, providing significant importance for clinical practice, as the precise prediction of non-response to both drugs offers clinicians the opportunity to make informed decisions. For anti-TNF experienced patients, this could mean a direct switch towards newer modes-of-action (i.e. JAK-inhibitors).

While this study is the first to demonstrate the utility of DNA methylation profiling in whole blood for objective response to VDZ and USTE in CD patients, previous analyses from two separate studies explored its application to predict anti-TNF response in both CD and UC patients[50,51]. In the first study, the authors sought to identify an anti-TNF response-associated profile combining integrated methylation and gene expression data from samples taken before and 2 weeks into treatment using a primary endpoint of clinical remission at week 14[50]. The observations were made using a mixed cohort of 37 IBD (18 CD, 19 UC) patients

and validated against a publicly available gene expression data of 20 CD patients. In the second study, the authors did not show replication of these observations using methylation data of 385 patients, as part of the previously published PANTS study[51,52]. In this study, bio-naïve patients with active CD started treatment with adalimumab or infliximab. Interestingly, the authors report 323 differentially methylated positions annotated to 210 genes identified at baseline that were significantly associated with serum drug concentrations at week 14. This profile could potentially be of interest to identify patients in need of intensified anti-TNF drug monitoring or dosing. The methodological differences and lack of endoscopic data in the previous experiments preclude direct comparison with our study in which outcome assessments were more stringent.

Through two separate stability analyses, we demonstrated both short- and long-term hyper stability of the majority of our identified CpG markers indicating their independence of treatment as well as the resultant difference in inflammation. The latter is further evidenced by the lack of correlation between the methylation status of the predictor CpGs and both baseline CRP and FCP**,** therapy switch or even CD-related surgery[53] suggesting that the CpGs are response-predictors but do not directly mediate inflammation. Nonetheless, confounder analysis indicated potential confounding for some of the predictor CpGs for age, sex, smoking status and the blood cell distribution. We observed however that prediction modeling using the confounders only performed significantly worse than the CpG model, indicating that confounders do not contribute substantially to the predictive performance.

We note that the predictor CpGs annotate to genes associated with MHC class I (*HLA-C*) and cell migration (*TSPEAR, NID2*)[54,55] for VDZ and macrophage function (*RHOJ*, *MARK3* and *PCGF3*)[56-61] and polarization (*PKNOX1* and *MRC1*)[62-64], histone remodeling and Th17-differentiation (*HDAC4*)[65-69] and TGF-β signaling (*SMAD1* and *EBF3*)[70-77] for USTE. Integrating DNA methylation with their transcriptomic data indicates that VDZ-response-associated *RFPL2*, which encodes an E3 ubiquitin ligase[78,79], and USTE-response-associated *TMEM191B*, which encodes a transmembrane protein, present concordant differential methylation and expression. However, their exact role in either IBD and/or response to therapy remains unknown to date. Accordingly, beyond the utility of the predictor CpGs in classifying response to therapy, identifying their role in the pathogenesis and etiology of non-response remains challenging at present and hence a subject for future studies.

Our study derives its main strength from the sampling and endpoint assessment strategy. In addition, patients with anti-drug antibodies or without a measurable serum drug concentration and those that stopped treatment due to adverse events were excluded prior to

the selection of this cohort. Non-responders therefore reflect a more homogenous group of true biological (i.e. pharmacodynamic) non-response rather than failure due to pharmacokinetics or intolerance. Second, the relatively small sample size was justified by performing a sample size estimation based on prior pilot experiments in combination with previous studies on statistical power in EWAS[80]. Lastly, besides stability of the observed methylation differences during induction- and maintenance treatment, our markers demonstrated stable differences between R and NR over time, which was further corroborated when interrogating our 7-year longitudinal DNA methylation survey. This time-independent behavior of the predictor CpGs indicates that exposure over time and change in inflammatory parameters does not affect the response-associated behavior of the CpGs, suggesting that the CpGs are very stable, which in turn increases its utility in clinical practice.

There are however some limitations to this study. First, a post-hoc power analysis of the mean effect size difference between responders and non-responders indicated a mean absolute difference of 9.9% and 7.6% for VDZ and USTE, respectively. At the collected samples sizes, this would translate to a statistical power of approximately 98.4% and 64.3% for VDZ and USTE, respectively[80]. Despite the lower power for USTE, we note that the model remained performant in predicting response in the external cohort. Second, in the external cohort response assessment was less stringent in approximately 70% of the patients due to the COVID-19 pandemic during which we encountered a notable reduction in the access to non-essential endoscopies, particularly in the UK[81]. Nonetheless, the modified response criteria are a reflection of clinical parameters that physicians commonly use in daily practice. This pragmatic approach enhances the overall generalizability of our results to a much larger IBD population. Notably, the analysis where we included the available endoscopic outcomes in the subset of UK patients for whom these data were available, enhanced the performance to an AUC of 0.83 for both the VDZ and USTE models, reinforcing the validity of both models in identifying objective responders to these biological therapies. While we acknowledge this limitation, we believe that our comprehensive approach provides valuable insights into the predictive capabilities of the models under diverse clinical scenarios. Third, while we purposely used PBL samples as these are minimally invasive and easily obtained during daily clinical practice, PBL represents a mixed cellular population. Therefore, the specific cell types responsible for the observed predictive signal remain unidentified[82]. It should be noted however that after correcting for the blood cell distribution, several predictor CpGs remained statistically significant, indicating independence of cellular composition. Third, the majority of the predictor CpG loci identified are situated within gene introns, complicating the biological interpretation of our findings. Although the identified predictor CpGs collectively serve as a strong predictor of response, we acknowledge that the underlying biology behind this

observation is more complicated than merely the inverse correlation between DNA methylation and gene expression. Lastly, while strict removal of most catalogued and predicted genetic variant-binding probes, we acknowledge that there is still a possibility that underlying genetic differences could have influenced our outcome[83]. Nonetheless both models performed effectively in both the discovery- (Dutch) and the validation (UK) cohorts.

In summary, our findings pave the way towards personalized medicine for CD. Several US and European studies report a significant reduction in pharmaco-economic burden of CD if clinical and endoscopic remission can be achieved with adequate treatment compared to the cost of treatment of IBD patients on suboptimal medication[84-88]. In the absence of a predictive biomarker panel, current endoscopic response rates at week 52 have been reported around 45% in VDZ and 42% in USTE treated CD patients[7,8]. Taken together, our biomarker panels could potentially increase these proportions by approximately 20% for VDZ and 24% for USTE, thereby significantly impacting healthcare costs and disease burden in these patients. We acknowledge that clinical validation of our findings in a randomized prospective trial, comparing our method of pre-treatment selection with current clinical practice, is needed to firmly demonstrate both clinical and economic benefit[89]. To this end, the Omicrohn trial as part of the ongoing Horizon Europe funded METHYLOMIC project has been launched and is currently underway[90].

# Online Methods

## Study population and design

We prospectively recruited adult male and female CD patients that presented with a combination of clinical, biochemical and endoscopic disease activity at ileo-colonoscopy and were scheduled to start VDZ or USTE treatment within 1 year after the last endoscopy at the Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands (discovery cohort) and the John Radcliffe Hospital, Oxford, United Kingdom (validation cohort). All patients were naïve to the biological of interest.

Patients were treated according to standard-of-care protocols, which for VDZ meant that patients were given 300 mg infusions at week 0, 2 and 6 followed by infusions at an 8 week interval. For USTE, standard-of-care involved patients receiving a single intravenous infusion (6 mg/kg rounded to 260 mg, 390 mg or 520 mg) at week 0 and subsequent 90 mg subcutaneous injections at an 8 week interval. For both VDZ and USTE, interval intensification every 6 or 4 weeks with an additional week 10 infusion for VDZ or extra intravenous boost infusion for USTE at the treating physicians' discretion. To ensure assessment of mechanistic and not pharmacokinetic failures of each biological treatment, only patients with measurable serum concentrations without anti-drug antibodies at response assessment were used for methylation analyses.

The project was approved by the medical ethics committee of the Academic Medical Hospital (METC NL57944.018.16 and NL53989.018.15) and written informed consent was obtained from all subjects prior to sampling, as well as by the National Health Service Research Ethics committee. (REC reference: 21/PR/0010 Protocol number: 14833 IRAS project ID: 266041). UK patients were recruited and consented under the ethics of the Translational Gastrointestinal Unit biobank IBD cohort ethics (09/H1204/30) and GI cohort ethics (16/YH/0247 and 21/YH/0206). The quality of the collected data and study procedures were assessed by an independent monitor.

## Vedolizumab discovery and validation cohort characteristics

The VDZ discovery cohort consisted of 64 patients ($N_R$ = 36, $N_{NR}$ = 28) of which 49 (77%) had previously been exposed to anti-TNF treatment and 9 (14%) to USTE. Response was defined on endoscopic as well as clinical/biochemical criteria in this group (see section "Definitions of response"). In 62 (97%) patients, a follow-up endoscopy was performed. The remaining 2 patients were classified as non-responders as a result of urgent surgical

intervention due to worsening of disease without CD-associated complications, such as stenosis or perforating disease. R and NR presented overall comparable clinical characteristics, with no significant differences in age, sex and smoking behavior. Importantly, serum VDZ concentrations at T2 were not significantly different between R and NR (median 15 (IQR 7.7-20.5) versus median 14 (IQR 3.6-27.5), p-value = 0.77) although more NR patients received an additional VDZ infusion at week 10 (35.7% vs 13.9%, p-value = 0.04). Notably, patients in the NR group had more frequently been exposed to anti-TNF treatment (89.3% vs 66.7%, p-value = 0.03) and/or USTE (25% vs 5.6%, p-value = 0.03) compared to R group. Thirteen of the 15 R patients received VDZ as first-line biologic, had a shorter disease duration, lower rates of previous surgery and perianal disease as well as a higher percentage of B1 phenotype compared to anti-TNF experienced patients.

The VDZ validation cohort consisted of 25 additional patients ($N_R$ = 14 and $N_{NR}$ = 11) from Oxford with a median 9 (IQR 3-15) year disease duration. Seventeen (68%) did not undergo a follow-up colonoscopy to assess endoscopic response due to restrictions during the COVID-19 pandemic and therefore were assessed using the modified definition of response, which was defined as a combination of clinical and biochemical parameters (see “Definitions of response” below). Out of these 25 patients, 13 (52%) were biological naïve, 12 (48%) were anti-TNF-experienced and 7 (28%) were previously treated with USTE, which was enriched among the NR group (54.5% vs 7.1%, p-value = 0.01). All other clinical characteristics were comparable in R and NR, including age, sex, and smoking behavior. As in the VDZ discovery cohort, the majority (69%) of biological naïve patients were responders to VDZ.

## Ustekinumab discovery and validation cohort characteristics

The USTE discovery cohort consisted of 62 patients ($N_R$=30, $N_{NR}$=32) of which 60 (98%) were previously exposed to anti-TNF and 26 (42%) to VDZ. A follow-up endoscopy was performed in 58 (94%) patients. The remaining 4 were evaluated with serial intestinal ultrasound using validated criteria assessed by an expert IBD ultrasonographist. Clinical characteristics did not significantly differ between R and NR, although the R population consisted of more female patients (R = 80%, NR = 56.3%, p-value = 0.05). No significant differences in treatment intensification (21.9% vs 6.7%, p-value = 0.08), extra intravenous boost infusions (p-value = 0.26) or serum USTE concentrations at T2 between R and NR were observed (median 3.0 (IQR 1.8-5.5) versus median 4.8 (IQR 2.1-8.6), p-value = 0.31).

The USTE validation cohort consisted of 33 patients ($N_R$=22 and $N_{NR}$=11), with a median disease duration of 11 (IQR 3-20) years. Twenty-five (76%) did not undergo follow-up

colonoscopy as a result of pandemic-related restrictions on non-essential endoscopy and therefore were assessed using the modified definition of response. Of the 33 included patients, 9 were biological naïve (52%), 21 (63.6%) were anti-TNF-experienced and 5 (15.2%) were previously treated with VDZ. Responders presented a significantly longer disease duration compared to non-responders (median 15 vs 5 years, p=0.05). No significant differences in age, sex, and smoking behavior were observed.

## Sample collection and storage protocols

In all patients, whole peripheral blood leukocyte (PBL) samples were collected for measurement of epigenome-wide DNA methylation prior to the start of VDZ/USTE, before the baseline endoscopy or the first infusion (time point 1) and after an interval of 6-9 months into treatment (time point 2) using 4.0-6.0mL BD ethylenediaminetetraacetic acid (EDTA) vacutainer tubes. For the Amsterdam discovery cohort, samples were subsequently aliquoted into 1.10mL micronic tubes before storing at -80 ºC until further handling. The Oxford validation samples were directly frozen at -80 ºC in preparation for later extraction. At both time points, additional PBL samples were stored for the purpose of targeted gene expression analyses from a subset of the Amsterdam discovery patients using 2.0-mL PAXgene Blood RNA tubes and were frozen at –20 ºC for 24 h before storing at –80 ºC until further handling.

## Definitions of response

At response assessment, patients were classified as responders (R) or non-responders (NR) based on a strict combination of endoscopic, biochemical and clinical criteria: ≥50% reduction in the endoscopic SES-CD score, corticosteroid-free clinical remission (≥3 point drop[91] in HBI or HBI ≤4 and no systemic steroids) and/or biochemical response (CRP reduction ≥50% or CRP≤5 mg/L and FCP reduction ≥50% or FCP ≤250 µg/g). Modified response was defined as a combination of corticosteroid-free clinical- (HBI ≤4) and biochemical (CRP ≤5 mg/L and/or FCP ≤250 µg/g) remission between week 26-52 without treatment change through week 52.

## DNA isolation and in vitro DNA methylation analysis

For the Amsterdam discovery cohorts, genomic DNA was extracted using the QIAsymphony. We next assessed the quantity of DNA using the FLUOstar OMEGA and quality of the high-molecular weight DNA on a 0.8% agarose gel. Following these steps, 750ng of DNA per sample was randomized per plate, to limit batch effects, after which genomic DNA was

bisulfite converted using the Zymo EZ DNA Methylation kit and analyzed on the Illumina HumanMethylation EPIC BeadChip array. Aforementioned work was performed at the Core Facility Genomics, Amsterdam UMC, Amsterdam, the Netherlands.

For the Oxford validation cohorts, genomic DNA was extracted using the Qiagen Puregene Blood Core Kit C at the Oxford Translational Gastroenterology Unit, University of Oxford. DNA samples were assessed for quality using the NanoDrop spectrometer (NanoDrop 1000, Thermo Scientific). A total of 750ng of DNA per sample was randomized per plate, to limit batch effects, after which genomic DNA was bisulfite converted using the Zymo EZ DNA Methylation kit and analyzed on the Illumina HumanMethylation EPIC BeadChip array at UCL Genomics, University College London, London, United Kingdom.

## Raw methylation data pre-processing

DNA methylation data processing was orchestrated in Snakemake (v7.14.1)[92]. Raw methylation data was imported into the R statistical environment (v4.3.1) using the Bioconductor minfi[93,94] package (v1.44). Quality control was performed using shinyMethyl (v1.38.0)[95] for probe level quality control, ewastools (v1.7.2)[96] to ensure paired samples were correctly labeled, and the Horvath clock[97] to ensure a proper match with the metadata. One patient sample from the VDZ and one patient sample from the USTE discovery cohort were removed due to a discrepancy in the predicted sex with the annotated sex. Raw signals were normalized using functional normalization[98]. Probes were annotated to their gene of interest using the provided Illumina annotations, which were further enhanced with enhancer data as obtained using the publicly accessible promoter-capture Hi-C data[99]. We subsequently calculated the methylation signal in the form of percentage methylation. Technical artefacts as a result of batch, plate and plate position were removed using ComBat (v0.9.7)[100] as implemented in the sva (v3.50.0)[101] package, a tool for removing known batch effects in microarray data using the parametric empirical Bayes framework, using the default parameters. Probes hybridizing to allosomes were removed to identify sex-independent differences. Moreover, probes hybridizing to known and tentative genetic variants were removed to identify true methylation signals. Known genetic variants were identified based on their presence in dbSNP (v137), whereas tentative genetic variants were based on the methylation signal displaying a tri- or bimodal distribution, a hallmark of genetic variants underlying DNA methylation[102], as determined using gaphunter[103] set to a threshold of 0.25. The eventual number of CpGs used for training the prediction models were 806,308 and 808,815 for VDZ and USTE, respectively. For the validation of the prediction models, raw methylation data from the validation cohort was pre-processed together with the discovery cohort using functional normalization and ComBat to mitigate batch effects introduced by the

different experimental setup. From the combined dataset, the predictor CpGs were extracted and the prediction model was recalibrated against the discovery dataset where after predictions were made on the validation dataset.

## Machine learning models: stability selected gradient boosting analyses

The machine learning modeling was divided into two steps, namely feature selection and validation (**Fig. 1a**). Feature selection was performed on the discovery cohort, collected at the AmsterdamUMC, whereas validation was performed on the validation cohort, collected at the John Radcliffe Hospital. During the feature selection procedure, the model was at no point exposed to the validation cohort. To identify baseline epigenetic markers associated with response/non-response to treatment, we implemented a rigorous supervised machine learning approach using stability selected gradient boosting[34,35,104] combined with covered information disentanglement (CID)[105]. Gradient boosting is an algorithm for supervised learning, which operates through stepwise improvement of weak learners thereby minimizing the overall prediction error against the observed data. We opted for gradient boosting as it captures linear, non-linear and interaction effects better than traditional linear regression approaches[104]. CID, in conjunction with stability selected gradient boosting, represents an approach for feature selection that assigns permutation-based feature importance that, unlike other methods for assigning feature importance, is unbiased by multicollinearity[105].

We first removed CpGs with low variance and then applied univariate feature selection using a F-test. Stability selection was subsequently employed to identify reliable biomarkers by splitting randomly the discovery data into an 80% training data and a 20% test data using stratified shuffle split and repeating this process 100 times to mitigate overfitting[106]. During each split, we computed the CID for each CpG by randomly permuting it 100 times and calculating the mean feature importance and assessing the average effect of permutation on the model's performance (i.e., predicted versus true outcome)[105]. After all 100 iterations, the mean feature importance per iteration was averaged and compared against a randomly generated noise variable that was included throughout the entire modeling process where CpGs with an aggregated feature importance ranked above a random variable were termed predictor CpGs and retained for future analyses.

Having determined the predictor CpGs, we subsequently validated the predictive performance internally and externally. Internal validation was performed by extracting the predictor CpGs, training an ensemble of 100 gradient boost models on the 80% discovery training data and utilizing all models to predict against the withheld 20% discovery test data. External validation was performed using a similar approach where the discovery and

validation cohort were merged, the predictor CpGs extracted, and an ensemble of 100 models trained on the discovery cohort. The resultant models were then requested to predict the response in the external validation cohort. In both cases, the output of each model yielded a prediction score on a scale of 0 to 1 per sample. The prediction scores were aggregated by calculating the mean prediction score per sample, representing the final, ensembled, output of the model. This final prediction score was used to calculate the receiver operator characteristic (ROC) to assess performance. The resultant prediction scores were subsequently converted into classes by freezing the model at the Youden index, thereby balancing the true positive rate relative to the false positive rate. Analyses were not conducted separately for males and females as the cohorts would become too small. Aforementioned analyses were performed in Python (v3.10) (https://www.python.org), with packages scikit-learn and xgboost for the model development.

## In silico DNA methylation analysis

Differential methylation analyses were performed in R using limma[107] (v3.46) and eBayes[108]. Separate analyses were run either regressing against the slide and slide position only, or in combination with common confounding variables: age, sex, smoking behavior and the estimated blood cell distribution[109,110]. The blood cell distribution was estimated using the method described by Houseman *et al.* [109] and implemented by Salas *et al*.[110] for the Illumina HumanMethylation EPIC BeadChip array, where methylation profiles of the current data are compared against a reference methylation profiles of cell-sorted neutrophils, B cells, monocytes, NK cells, CD4+ T cells and CD8+ T cells, enabling the inference of the cellular composition. Intra-class correlation analyses were performed by conducting a two-way, mixed, consistency analysis comparing samples obtained during response assessment with samples obtained pretreatment using irr (v0.84.1). Statistical significance was defined as a false discovery rate-adjusted p-value < 0.05. Visualizations were generated using ggplot2[111] (v3.3.5).

## RNA expression and data processing

Transcriptomic analyses was conducted through RNA sequencing, wherein mRNA was extracted utilizing the QIAsymphony system, converted into cDNA and sequenced in a paired-end format on the Illumina NovaSeq6000 at the Amsterdam UMC Core Facility Genomics, generating a dataset comprising 40 million 150 bp-reads. Following that, *in silico* data processing was orchestrated in Snakemake where read quality was assessed using FastQC (v0.11.8) and summarized using MultiQC (v1.0)[92,112]. Raw reads were aligned to the human genome (GRCh38) using the STAR aligner (v2.7.0), with annotations provided by the Ensembl v95 annotation. Post-alignment processing was done in SAMtools (v1.9), followed

by read counting using the featureCounts function from the Subread package (v1.6.3)[113-115]. Differential expression analysis (DE) analysis, was carried out within the R statistical environment (v4.3) using the Bioconductor package DESeq2 (v1.38.3). We specifically focused on genes associated with the predictor CpG loci based on the latter's location in either promoter or enhancer regions. Differentially expressed genes (DEGs) were identified based on their significant differences, defined as those with a Benjamini–Hochberg-adjusted p-value <0.05. Visualization of the results was accomplished using ggplot2 (v3.4.0)[116].

## Sample size estimation

We based our sample size on an initial pilot experiment and the calculations performed by Tsai and Bell[80] where we maintained a nominal p-value threshold of 0.05. A pilot experiment using a subset of VDZ treated patients (7 R and 5 NR) indicated on average that the most differentially methylated CpGs presented a mean difference in percentage methylation of 10% when comparing R with NR. We subsequently consulted the power calculations reported by Tsai and Bell[80], who conducted a case-control epigenome wide simulation study. Assuming an approximately equal number of cases and controls and a mean difference in percentage methylation of at least 10% at a nominal p-value threshold of 0.05, a statistical power of at least 80% would be achieved if we included 40 patients (20 R and 20 NR) per drug. To further eliminate the possibility of being underpowered, we aimed to collect at least 60 patients for the discovery cohort per drug, which would be supplemented by at least 20 patients for the validation cohort.

## Statistical analysis of clinical variables

Baseline characteristics of all included patients were summarized using descriptive statistics. Categorical variables are presented as percentages and continuous variables as median annotated with the interquartile range (IQR). Differences in distribution between responders, non-responders and the different cohorts were assessed using a chi-square test (categorical variables) or Mann-Whitney U (continuous variables). Two-tailed probabilities were used with a p-value ≤0.05 were considered statistically significant. Analyses of clinical data were performed in IBM SPSS statistics (v26). To estimate the probability of a patient responding to either VDZ or USTE after being predicted to be a responder, we calculated the post-test probability using the following formulas:

(1) $$pre\ test\ odds = \frac{pre\ test\ probability}{1 - pre\ test\ probability}$$

(2) $$likelihood\ ratio = \frac{sensitivity}{1 - specificity}$$

627 (3) $post\ test\ odds = likelihood\ ratio * pre\ test\ odds$

628 (4) $post\ test\ probability = \frac{post\ test\ odds}{post\ test\ odds+1}$

629 The pre-test probabilities were obtained from the largest VDZ[7], and USTE[8] treatment-
630 response studies to date. The sensitivity and specificity were calculated by determining the
631 number of true positives, true negatives, false positives and false negatives when predicting
632 response in the validation cohort.

633 # Data availability

634 The raw DNA methylation- (.idat) and gene expression (.fastq.gz) data alongside the de-
635 identified patient metadata as reported on in this study have been published under controlled
636 access for research purposes at the European Genome-phenome Archive (EGA). The DNA
637 methylation data for the VDZ discovery cohort can be found under accession ID:
638 EGAD00010002651. The DNA methylation data for the VDZ validation cohort can be found
639 under accession ID: EGAD00010002652. The DNA methylation data for the USTE discovery
640 cohort can be found under accession ID: EGAD00010002649. The DNA methylation data for
641 the USTE validation cohort can be found under accession ID: EGAD00010002650. The
642 RNA-sequencing data for the VDZ discovery cohort can be found under accession ID:
643 EGAD50000000385. The RNA-sequencing data for the USTE discovery cohort can be found
644 under accession ID: EGAD50000000386.

645 # Code availability

646 All Snakemake, bash calls and R scripts have been made available on GitHub and can be
647 found at https://github.com/ND91/HGPRJ0000008_EPICCD_multi_drug.git. The machine
648 learning modeling was performed using proprietary algorithms founded on the same
649 statistical principles as those of gradient boosting, permutation importance, and covered
650 information disentanglement. The code for these techniques is openly available at
651 https://xgboost.ai and https://github.com/JBPereira/CID or https://scikit-learn.org/stable.

# Tables

**Table 1: Baseline characteristics discovery- and validation cohorts.** Values in bold are significant, percentages shown are valid percentages. ADA: adalimumab. VDZ: vedolizumab. USTE: ustekinumab. R: responder. NR: non-responder, SD: standard deviation. IQR: interquartile range, CRP: C-reactive protein. FCP: Fecal calprotectin. HBI: Harvey Bradshaw Index. SES-CD: simple endoscopic disease activity score, Immunomodulator: azathioprine, mercaptopurine, thioguanine, methotrexate. Anti-TNF: infliximab, adalimumab or golimumab.

| | Amsterdam (discovery) | | | Oxford (validation) | | | Amsterdam (discovery) | | | Oxford (validation) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VDZ R (n=36) | VDZ NR (n=28) | P-val | VDZ R (n=14) | VDZ NR (n=11) | P-val | USTE R (n=30) | USTE NR (n=32) | P-val | USTE R (n=22) | USTE NR (n=11) | P-val |
| **Sex, n (%), female** | 17 (47.2) | 18 (64.3) | 0.17 | 6 (42.9) | 3 (27.3) | 0.42 | 24 (80) | 18 (56.3) | 0.05 | 13 (59.1) | 5 (45.5) | 0.63 |
| **Age, years, median (IQR)** | 36 (25-52) | 28 (32-56) | 0.98 | 41 (21-65) | 42 (24-70) | 0.56 | 38 (29-56) | 35 (23-45) | 0.80 | 43 (35-55) | 30 (24-52) | 0.25 |
| **Disease duration, years, median (IQR)** | 12 (3-19) | 9 (4-20) | 0.80 | 12 (2-15) | 8 (3-17) | 0.41 | 12 (7-23) | 10 (5-21) | 0.81 | 15 (6-24) | 5 (2-12) | 0.05 |
| **Ethnic background, n (%), White European** | 29 (80.6) | 19 (67.9) | 0.24 | 13 (92.9) | 9 (81.8) | 0.80 | 25 (83.3) | 22 (68.8) | 0.18 | 21 (95.5) | 7 (87.5) | 0.10 |
| **CRP, mg/L, median (IQR)[a]** | 3.9 (2.0-11.7) | 6.2 (3.2-19.2) | 0.26 | 5.2 (1.4-11.6) | 5.2 (1.3-23.4) | 0.97 | 5.7 (1.8-19.2) | 7.8 (3.4-19.1) | 0.45 | 6.1 (1.6-18.0) | 4.8 (2.3-10.9) | 0.30 |
| **FCP, ug/g, median (IQR)[b]** | 1003 (412-1802) | 1322 (389-2654) | 0.79 | - | - | - | 809 (180-1599) | 729 (221-2347) | 0.69 | - | - | - |
| **Total baseline HBI, mean (±SD) or median (IQR)[c]** | 7.4 (±4.2) | 9.3 (±4.9) | 0.11 | 3 (1-8) | 4 (1-8) | 0.53 | 8.4 (±5.3) | 8.2 (±5.9) | 0.90 | 6 (4-9) | 4 (2-8) | 0.69 |
| **Total baseline SES-CD, median (IQR)** | 8 (6-12) | 10 (5-13) | 0.27 | 7 (6-12) | 7 (6-22) | 0.92 | 10 (6-17) | 9 (6-14) | 0.62 | 8 (5-16) | 8 (3-19) | 0.73 |
| **Endoscopic evaluation at follow-up** | 36 (100) | 26 (93) | 0.19 | 4 (29) | 4 (36) | 1.00 | 29 (97) | 29 (91) | 0.61 | 6 (27) | 2 (18) | 0.69 |
| **Disease location, n (%)** | | | 0.50 | | | 0.21 | | | 0.78 | | | 0.43 |
| **- Ileal disease (L1)** | 13 (36.1) | 8 (28.6) | | 3 (21) | 2 (18) | | 6 (20.0) | 6 (18.8) | | 7 (32) | 5 (46) | |
| **- Colonic disease (L2)** | 9 (25.0) | 5 (17.9) | | 2 (14) | 5 (45) | | 7 (23.3) | 10 (31.3) | | 8 (36) | 2 (18) | |
| **- Ileocolonic disease (L3)** | 14 (38.9) | 15 (53.6) | | 9 (64) | 4 (36) | | 17 (56.7) | 16 (50.0) | | 6 (27) | 3 (27) | |
| **- Upper GI involvement (L4)** | 1 (2.8) | 2 (7.1) | 0.41 | - | - | - | - | - | - | - | 1 (9) | 0.13 |
| **Disease behavior, n (%)** | | | 0.60 | | | 0.53 | | | 0.39 | | | 0.06 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **- Non stricturing non-penetrating (B1)** | 16 (44.4) | 9 (32.1) | | 9 (82) | 10 (91) | | 10 (33.3) | 12 (37.5) | | 15 (68) | 5 (46) | |
| **- Stricturing (B2)** | 10 (27.8) | 10 (35.7) | | 2 (18) | 1 (9) | | 11 (36.7) | 15 (46.9) | | 5 (23) | 2 (18) | |
| **- Penetrating (B3)** | 10 (27.8) | 9 (32.1) | | - | - | | 9 (30.0) | 5 (15.6) | | - | 3 (27) | |
| **- Perianal disease (p)** | 8 (22.2) | 11 (39.3) | 0.14 | 1 (7) | 2 (18) | 0.48 | 10 (33.3) | 12 (37.5) | 0.73 | 6 (27) | 3 (27) | 0.88 |
| **Previous IBD related surgery, n (%)** | 14 (38.9) | 16 (57.1) | 0.15 | 4 (28.6) | 4 (36.4) | 0.88 | 23 (76.7) | 21 (65.6) | 0.34 | 11 (50) | 7 (63) | 0.29 |
| **Concomitant medication, n (%)** | | | | | | | | | | | | |
| **- Immunomodulators** | 2 (5.6) | 2 (7.1) | 0.80 | - | 2 | - | 1 (3.3) | - | 0.23 | 1 | - | - |
| **- Prednisone taper scheme** | - | - | | - | - | - | 7 (23.3) | 4 (12.5) | 0.26 | - | - | - |
| **Previous treatment exposure, n (%)** | | | | | | | | | | | | |
| **- Immunomodulators** | 28 (77.8) | 24 (85.7) | 0.42 | 13 (92.9) | 8 (72.7) | 0.17 | 29 (96.7) | 32 (100) | 0.23 | 17 (77.3) | 5 (45.5) | 0.07 |
| **- Anti-TNFs** | 24 (66.7) | 25 (89.3) | 0.03 | 5 (35.7) | 7 (63.9) | 0.25 | 30 (100) | 31 (96.9) | 0.25 | 14 (63.6) | 7 (63.3) | 1.00 |
| **- Vedolizumab** | - | - | - | - | - | - | 12 (40.0) | 14 (43.8) | 0.77 | 2 (9.1) | 3 (27.3) | 0.15 |
| **- Ustekinumab** | 2 (5.6) | 7 (25.0) | 0.03 | 1 (7.1) | 6 (54.5) | 0.01 | - | - | - | - | - | - |
| **Smoking, n (%)[d]** | | | | | | | | | | | | |
| **- Active** | 6 (17.1) | 2 (7.1) | 0.22 | 3 (21.4) | 1 (9.1) | 0.30 | 6 (20) | 5 (15.6) | 0.65 | 5 (22.7) | 1 (9.1) | 0.24 |

**Table 2.** Predictive performance metrics of the prognostic biomarkers on the discovery cohort acquired at the AmsterdamUMC (n=126) and the validation cohort acquired at the John Radcliffe Hospital (n=58). TP = True positives. TN = True negatives. FP = False positives. FN = False negatives. AUC = Area under the receiver operator characteristic curve. VDZ: vedolizumab. USTE: ustekinumab.

| | **VDZ discovery** | **VDZ validation** | **USTE discovery** | **USTE validation** |
|---|---|---|---|---|
| **TP** | 33 | 10 | 27 | 16 |
| **TN** | 22 | 8 | 29 | 8 |
| **FP** | 4 | 4 | 2 | 3 |
| **FN** | 3 | 3 | 2 | 6 |
| **AUC** | 0.87 | 0.75 | 0.89 | 0.75 |
| **Recall** | 0.92 | 0.77 | 0.93 | 0.73 |
| **Precision** | 0.89 | 0.71 | 0.93 | 0.84 |
| **F1-score** | 0.90 | 0.74 | 0.93 | 0.78 |

# Legends

## Figures

**Fig. 1: Predictive model using stability selected gradient boosting for response to therapy**. **a)** An overview of the feature selection and supervised machine learning approach for predicting response to VDZ and USTE as used in the current study. Int. train: Internal training set. Int. test: Internal test set. ROC: Receiver operator characteristic. **b)** Receiver operating characteristics plots showing the mean area under the curve (AUC) performance of the discovery (n=126) and validation (n=58) cohorts. **c)** Left: radar plots presenting the standardized difference in methylation between R (purple) and NR (green) for the top 15 predictor CpGs. Right: Aggregated feature importance of the top 15 predictor CpGs.

**Fig. 2: Longitudinal stability analyses. a)** Volcano plot representing the differential methylation analyses when comparing into treatment (T2) with pretreatment (T1) where grey dots represent CpG loci located on the Illumina HumanMethylation EPIC BeadChip array and black dots represent response-associated predictor CpGs. X-axis represents mean difference in percentage methylation, Y-axis represents the statistical significance as depicted in $-\log_{10}$(p-value) with a dashed line at p-value = 0.05. **b)** Scatterplot showing the correlation of differential DNA methylation between R and NR pretreatment (T1) and into treatment (T2). Grey dots represent CpG loci located on the Illumina HumanMethylation EPIC BeadChip array and black dots represent response-associated predictor CpGs. **c)** Boxplot of the two-way, consistency, intra-class correlation (ICC) coefficients of the predictor CpGs calculated when comparing pretreatment and into treatment as well as the ICC coefficients of the predictor CpGs obtained from a previous study on long-term stability of DNA methylation in IBD patients[53]. The vertical dashed grey lines represent classification boundaries introduced by Koo and Li[37], with blocks representing poor (ICC < 0.5), moderate (0.5 ≤ ICC < 0.75), good (0.75 ≤ ICC < 0.9), and excellent (0.9 ≥ ICC). **d)** Receiver operating characteristic plots representing the predictive performance into treatment (T2; black) and pre-treatment (T1; grey) as reference.

**Fig. 3: Analyses of potential confounding variables. a)** Volcano plot representing the change in response-associated differential methylation when correcting for the potential confounding variables age, sex, and estimated cellular composition (black) or not (grey). X-axis represents mean difference in percentage methylation, Y-axis represents the statistical significance as depicted in $-\log_{10}$(p-value). The dotted line represents a threshold set at p = 0.05. **b)** Boxplot of the change in effect size after correcting for the confounding variables as calculated by $(\beta_{corrected} - \beta_{uncorrected})/\beta_{uncorrected}$. **c)** Receiver operator characteristic curve comparing the response-prediction model for the CpG model (grey) with the confounder model (blue).

**Fig. 4: Integrative analyses of predictor CpG-associated genes. a)** Scatterplot showing the effect size (Wald statistic) of the response-associated difference in pretreatment (T1; X-axis) and into treatment (T2; Y-axis). Grey dots represent all genes measured, black dots represent non-differentially expressed predictor CpG-associated genes, red, blue and purple dots represent predictor associated genes that are differentially expressed at T1, T2, or T1 and T2, respectively. **b)** Boxplot of the T1 $\log_2$(expression) for VDZ response predictor CpG-associated genes *TULP4* and *RFPL2* stratified by response. **c)** Scatterplot of the T1 $\log_2$(expression) of *RFPL2* (X-axis) relative to the percentage DNA methylation (Y-axis) for the predictor CpG cg12906381 annotated with the Pearson correlation coefficient and the associated p-value. **d)** Boxplot of the T1 $\log_2$(expression) for USTE response predictor CpG-associated genes *MRC1* and *TMEM191B* stratified by response. **e)** Scatterplot of the T1 $\log_2$(expression) of *TMEM191B* (X-axis) relative to the percentage DNA methylation (Y-axis) for the predictor

CpG cg13982436 annotated with the Pearson correlation coefficient and the associated p-value. **f)** Boxplot of the T2 $\log_2$(expression) for VDZ response predictor CpG-associated genes *MCM2* and *RFPL2* stratified by response. **g)** Scatterplot of the T2 $\log_2$(expression) of *RFPL2* (X-axis) relative to the percentage DNA methylation (Y-axis) for the predictor CpG cg12906381 annotated with the Pearson correlation coefficient and the associated p-value. **h)** Boxplot of the T2 $\log_2$(expression) for USTE response predictor CpG-associated genes *POTEF, HDAC4, PARP4* and *MARK3* stratified by response.

## Supplementary Tables

**Supplementary Table 1**: Differential methylation summary statistics of the VDZ-response predictor CpGs. Columns represent the Illumina CpG identifier, the associated HGNC gene symbol, the chromosome, the location on the chromosome (build: hg19), the mean difference in percentage methylation between responders and non-responders, the nominal p-value associated with the mean difference, the intra-class correlation coefficient between pretreatment and during response assessment.

**Supplementary Table 2**: Differential methylation summary statistics of the USTE-response predictor CpGs. Columns represent the Illumina CpG identifier, the mean difference in percentage methylation between responders and non-responders, the nominal p-value associated with the mean difference, the genomic coordinates on the human genome (build: hg19), the annotated gene name.

**Supplementary Table 3**: Clinical characteristics multi-biological failure cohort.

**Supplementary Table 4**: Differential expression summary statistics of the VDZ-response predictor CpGs-associated genes. Columns represent the Ensembl gene ID, the HGNC gene name, the mean difference in log2(fold change) between responders and non-responders and the nominal p-value associated with the mean difference.

**Supplementary Table 5**: Differential expression summary statistics of the USTE-response predictor CpGs-associated genes. Columns represent the Ensembl gene ID, the HGNC gene name, the mean difference in log2(fold change) between responders and non-responders and the nominal p-value associated with the mean difference.

## Supplementary Figures

**Supplementary Fig. 1**: Longitudinal differential methylation of vedolizumab response-associated predictor CpGs. Individual boxplots and temporal stability of the 25 VDZ-associated predictor CpGs over time. Top: jitterplot visualization with samples obtained from the same donor connected by a dotted line annotated with the two-way consistency intra-class correlation coefficient and the 95% confidence intervals. Bottom: boxplot visualization grouped by response and stratified by timepoint. Annotations include nominal p-value as determined through response-associated differential methylation analyses for samples obtained pretreatment (T1) or during response assessment (T2). Green and red represent responders (R) and non-responders (NR), respectively.

**Supplementary Fig. 2**: Longitudinal differential methylation of ustekinumab response-associated predictor CpGs. Individual boxplots and temporal stability of the 68 USTE-associated predictor CpGs over time. Top: jitterplot visualization with samples obtained from the same donor connected by a dotted line annotated with the two-way consistency intra-class correlation coefficient and the 95% confidence intervals. Bottom: boxplot visualization grouped by response and stratified by timepoint. Annotations include nominal p-value as determined through response-associated differential methylation analyses for samples obtained pretreatment (T1) or during response assessment (T2). Green and red represent responders (R) and non-responders (NR), respectively.

**Supplementary Fig. 3**: Stratified prediction analyses. Receiver operating characteristics plots showing the mean area under the curve (AUC) performance of the validation cohort patients that were a) assessed using either the strict (NVDZ = 8, NUSTE = 8; blue) or modified (NVDZ =17, NUSTE = 25; pink) response criteria, and b) either anti-TNF exposed (NVDZ = 12, NUSTE = 21; black) or non-exposed (NVDZ = 13, NUSTE = 12; yellow) for VDZ (left) and USTE (right).

**Supplementary Fig. 4**: Analyses of markers of IBD-associated inflammation. Volcano plot representing the differential methylation analyses regressing against a) C-reactive protein (CRP) and b) fecal calprotectin (FCP) pretreatment (T1) for VDZ (left) and USTE (right). Grey dots represent all CpG loci located on the Illumina HumanMethylation EPIC BeadChip array and black dots represent response-associated predictor CpGs. X-axis represents mean difference in percentage methylation, Y-axis represents the statistical significance as depicted in –log10(p-value). CpGs with a nominal p-value < 0.05 are annotated.

# References

1. de Souza, H.S. & Fiocchi, C. Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol* **13**, 13-27 (2016).
2. Liu, J.Z.*, et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986 (2015).
3. Alsoud, D., Vermeire, S. & Verstockt, B. Biomarker discovery for personalized therapy selection in inflammatory bowel diseases: Challenges and promises. *Curr Res Pharmacol Drug Discov* **3**, 100089 (2022).
4. Juillerat, P., Grueber, M.M., Ruetsch, R., Santi, G., Vuillemoz, M. & Michetti, P. Positioning biologics in the treatment of IBD: A practical guide - Which mechanism of action for whom? *Curr Res Pharmacol Drug Discov* **3**, 100104 (2022).
5. Loftus, E.V., Jr.*, et al.* Upadacitinib Induction and Maintenance Therapy for Crohn's Disease. *N Engl J Med* **388**, 1966-1980 (2023).
6. D'Haens, G.*, et al.* Risankizumab as induction therapy for Crohn's disease: results from the phase 3 ADVANCE and MOTIVATE induction trials. *Lancet* **399**, 2015-2030 (2022).
7. Lowenberg, M.*, et al.* Vedolizumab Induces Endoscopic and Histologic Remission in Patients With Crohn's Disease. *Gastroenterology* **157**, 997-1006 e1006 (2019).
8. Sands, B.E.*, et al.* Ustekinumab versus adalimumab for induction and maintenance therapy in biologic-naive patients with moderately to severely active Crohn's disease: a multicentre, randomised, double-blind, parallel-group, phase 3b trial. *Lancet* **399**, 2200-2211 (2022).
9. Colombel, J.F.*, et al.* Infliximab, azathioprine, or combination therapy for Crohn's disease. *N Engl J Med* **362**, 1383-1395 (2010).
10. Dulai, P.S.*, et al.* Development and Validation of Clinical Scoring Tool to Predict Outcomes of Treatment With Vedolizumab in Patients With Ulcerative Colitis. *Clin Gastroenterol Hepatol* **18**, 2952-2961 e2958 (2020).
11. Verstockt, B.*, et al.* Expression Levels of 4 Genes in Colon Tissue Might Be Used to Predict Which Patients Will Enter Endoscopic Remission After Vedolizumab Therapy for Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol* **18**, 1142-1151 e1110 (2020).
12. Verstockt, B.*, et al.* Low TREM1 expression in whole blood predicts anti-TNF response in inflammatory bowel disease. *EBioMedicine* **40**, 733-742 (2019).
13. Verstockt, B.*, et al.* DOP70 An integrated multi-omics biomarker predicting endoscopic response in ustekinumab treated patients with Crohn's disease. *Journal of Crohn's and Colitis* **13**, S072-S073 (2019).
14. Soendergaard, C., Seidelin, J.B., Steenholdt, C. & Nielsen, O.H. Putative biomarkers of vedolizumab resistance and underlying inflammatory pathways involved in IBD. *BMJ Open Gastroenterol* **5**, e000208 (2018).
15. Ananthakrishnan, A.N.*, et al.* Gut Microbiome Function Predicts Response to Anti-integrin Biologic Therapy in Inflammatory Bowel Diseases. *Cell Host Microbe* **21**, 603-610 e603 (2017).
16. Lee, J.W.J.*, et al.* Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. *Cell Host Microbe* **29**, 1294-1304 e1294 (2021).
17. Stevens, T.W.*, et al.* Systematic review: predictive biomarkers of therapeutic response in inflammatory bowel disease-personalised medicine in its infancy. *Aliment Pharmacol Ther* **48**, 1213-1231 (2018).
18. Gisbert, J.P. & Chaparro, M. Predictors of Primary Response to Biologic Treatment [Anti-TNF, Vedolizumab, and Ustekinumab] in Patients With Inflammatory Bowel Disease: From Basic Science to Clinical Practice. *J Crohns Colitis* **14**, 694-709 (2020).

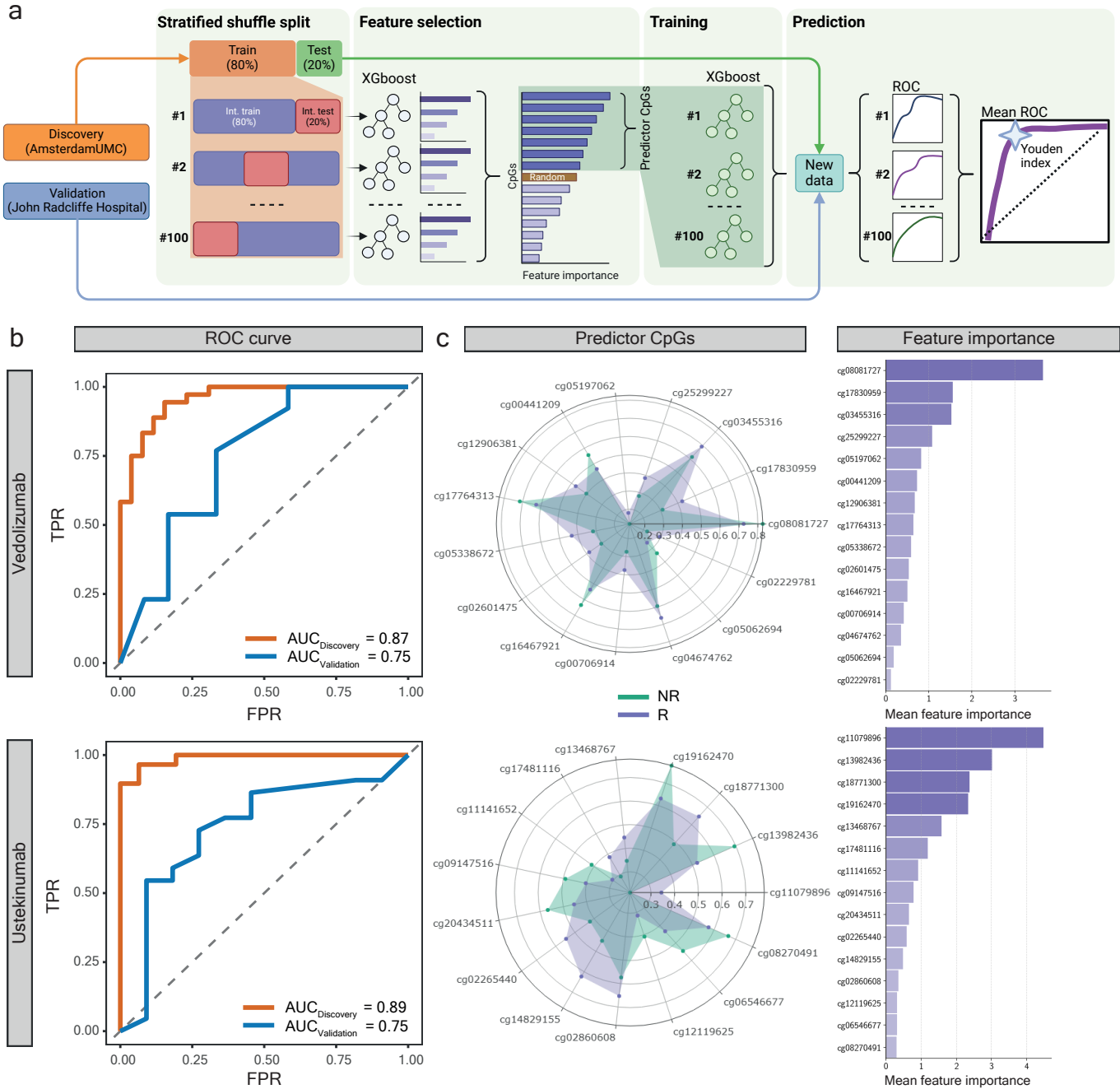
19. Zilbauer, M. & Heuschkel, R. Disease Prognostic Biomarkers in Inflammatory Bowel Diseases-A Reality Check. *J Crohns Colitis* **16**, 162-165 (2022).
20. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).
21. Kaluscha, S.*, et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat Genet* **54**, 1895-1906 (2022).
22. de Mendoza, A.*, et al.* Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. *Genome Biol* **23**, 163 (2022).
23. Hornschuh, M., Wirthgen, E., Wolfien, M., Singh, K.P., Wolkenhauer, O. & Dabritz, J. The role of epigenetic modifications for the pathogenesis of Crohn's disease. *Clin Epigenetics* **13**, 108 (2021).
24. Peery, R.C., Pammi, M., Claud, E. & Shen, L. Epigenome - A mediator for host-microbiome crosstalk. *Semin Perinatol* **45**, 151455 (2021).
25. Qin, Y. & Wade, P.A. Crosstalk between the microbiome and epigenome: messages from bugs. *J Biochem* **163**, 105-112 (2018).
26. Li, Y.*, et al.* Intestinal mucosa-derived DNA methylation signatures in the penetrating intestinal mucosal lesions of Crohn's disease. *Sci Rep* **11**, 9771 (2021).
27. Sadler, T.*, et al.* Genome-wide analysis of DNA methylation and gene expression defines molecular characteristics of Crohn's disease-associated fibrosis. *Clin Epigenetics* **8**, 30 (2016).
28. Adams, A.T.*, et al.* Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn□s Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory Bowel Diseases* **20**, 1784--1793 (2014).
29. Ventham, N.T.*, et al.* Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nature Communications* **7**, 13507 (2016).
30. Li Yim, A.Y.F.*, et al.* Peripheral blood methylation profiling of female Crohn's disease patients. *Clin Epigenetics* **8**, 65 (2016).
31. Howell, K.J.*, et al.* DNA Methylation and Transcription Patterns in Intestinal Epithelial Cells From Pediatric Patients With Inflammatory Bowel Diseases Differentiate Disease Subtypes and Associate With Outcome. *Gastroenterology* **154**, 585--598 (2018).
32. Gasparetto, M.*, et al.* Transcription and DNA Methylation Patterns of Blood-Derived CD8(+) T Cells Are Associated With Age and Inflammatory Bowel Disease But Do Not Predict Prognosis. *Gastroenterology* (2020).
33. Li Yim, A.Y.F.*, et al.* Whole-Genome DNA Methylation Profiling of CD14+ Monocytes Reveals Disease Status and Activity Differences in Crohn's Disease Patients. *J Clin Med* **9**(2020).
34. Meinshausen, N. & Bühlmann, P. Stability Selection *Journal of the Royal Statistical Society Series B: Statistical Methodology* **72**, 417/473 (2010).
35. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. . in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (New York, 2016).
36. Joustra V., L.Y.A. Long-term temporal stability of peripheral blood DNA methylation alterations in patients with inflammatory bowel disease. (BioRxiv, 2022).
37. Koo, T.K. & Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* **15**, 155-163 (2016).
38. Tsai, P.C.*, et al.* Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin Epigenetics* **10**, 126 (2018).
39. Houseman, E.A., Kelsey, K.T., Wiencke, J.K. & Marsit, C.J. Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. *BMC Bioinformatics* **16**, 95 (2015).

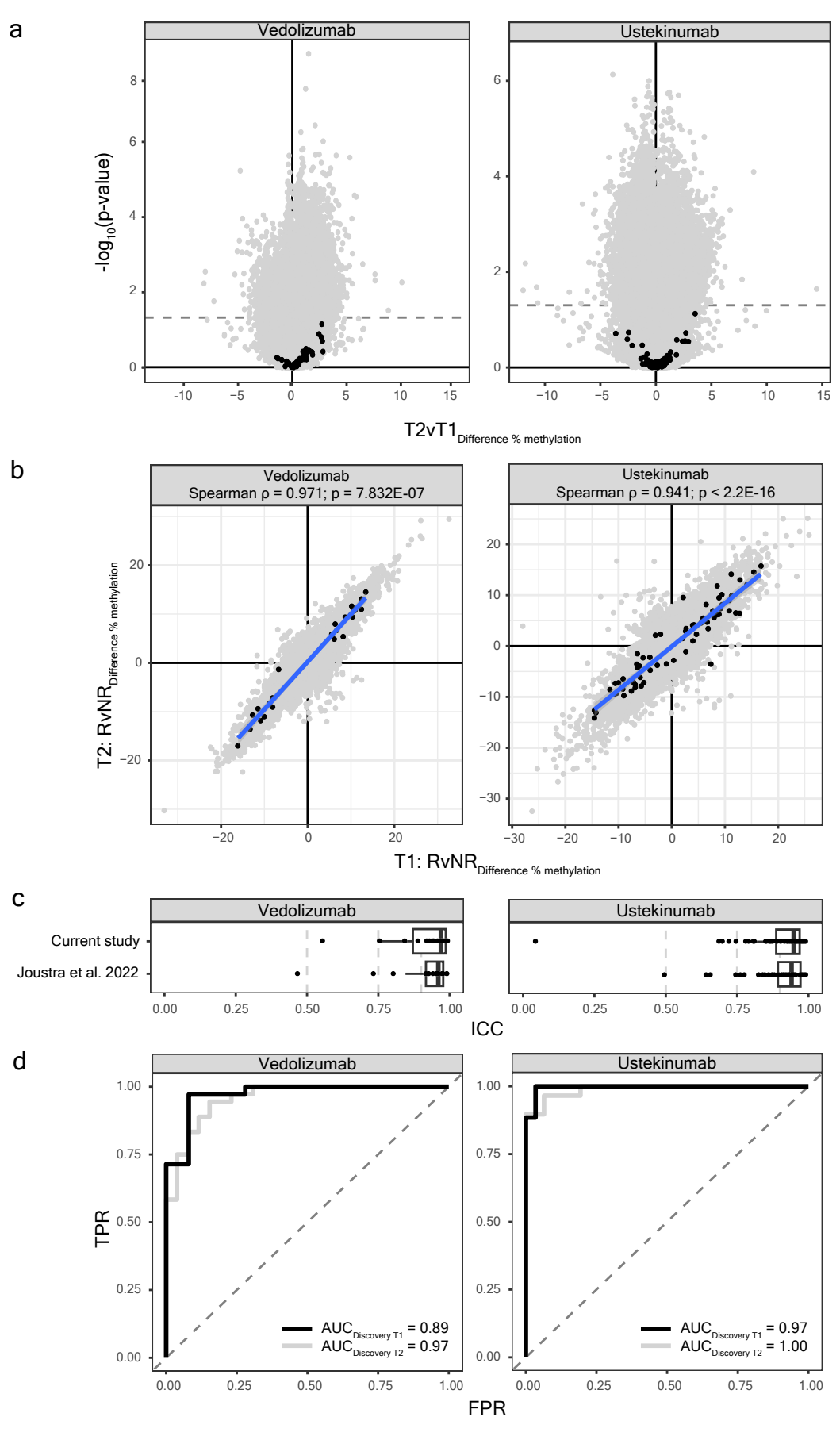
40. Solomon, O.*, et al.* Meta-analysis of epigenome-wide association studies in newborns and children show widespread sex differences in blood DNA methylation. *Mutat Res-Rev Mutat* **789**(2022).
41. Dobbs, K.R.*, et al.* Age-related differences in monocyte DNA methylation and immune function in healthy Kenyan adults and children. *Immun Ageing* **18**, 11 (2021).
42. Somineni, H.K.*, et al.* Blood-Derived DNA Methylation Signatures of Crohn Disease and Severity of Intestinal Inflammation. *Gastroenterology* (2019).
43. Kappelman, M.D.*, et al.* Real-World Evidence Comparing Vedolizumab and Ustekinumab in Antitumor Necrosis Factor-Experienced Patients With Crohn's Disease. *Am J Gastroenterol* **118**, 674-684 (2023).
44. Torres, J.*, et al.* ECCO Guidelines on Therapeutics in Crohn's Disease: Medical Treatment. *J Crohns Colitis* **14**, 4-22 (2020).
45. Siegel, C.*, et al.* P714 Mucosal healing in CD with vedolizumab versus other biologics: endoscopic outcomes during long-term routine care in a multinational observational study. *Journal of Crohn's and Colitis* **17**, i844-i844 (2023).
46. Danese, S.*, et al.* Treat to target versus standard of care for patients with Crohn's disease treated with ustekinumab (STARDUST): an open-label, multicentre, randomised phase 3b trial. *Lancet Gastroenterol Hepatol* **7**, 294-306 (2022).
47. Feagan, B.G.*, et al.* Rapid Response to Vedolizumab Therapy in Biologic-Naive Patients With Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol* **17**, 130-138 e137 (2019).
48. Sandborn, W.J.*, et al.* Long-term efficacy and safety of ustekinumab for Crohn's disease through the second year of therapy. *Aliment Pharmacol Ther* **48**, 65-77 (2018).
49. Feagan, B.G.*, et al.* Ustekinumab as Induction and Maintenance Therapy for Crohn's Disease. *N Engl J Med* **375**, 1946-1960 (2016).
50. Mishra, N.*, et al.* Longitudinal multi-omics analysis identifies early blood-based predictors of anti-TNF therapy response in inflammatory bowel disease. *Genome Med* **14**, 110 (2022).
51. Lin, S.*, et al.* Whole blood DNA methylation changes are associated with anti-TNF drug concentration in patients with Crohn's disease. *J Crohns Colitis* (2023).
52. Kennedy, N.A.*, et al.* Predictors of anti-TNF treatment failure in anti-TNF-naive patients with active luminal Crohn's disease: a prospective, multicentre, cohort study. *Lancet Gastroenterol Hepatol* **4**, 341-353 (2019).
53. Joustra, V.*, et al.* Long-term Temporal Stability of Peripheral Blood DNA Methylation Profiles in Patients With Inflammatory Bowel Disease. *Cell Mol Gastroenterol Hepatol* **15**, 869-885 (2023).
54. Ido, H.*, et al.* Molecular dissection of the alpha-dystroglycan- and integrin-binding sites within the globular domain of human laminin-10. *J Biol Chem* **279**, 10946-10954 (2004).
55. Ulazzi, L.*, et al.* Nidogen 1 and 2 gene promoters are aberrantly methylated in human gastrointestinal cancer. *Mol Cancer* **6**, 17 (2007).
56. Yuan, L.*, et al.* RhoJ is an endothelial cell-restricted Rho GTPase that mediates vascular morphogenesis and is regulated by the transcription factor ERG. *Blood* **118**, 1145-1153 (2011).
57. Park, S.Y.*, et al.* RhoA/ROCK-dependent pathway is required for TLR2-mediated IL-23 production in human synovial macrophages: Suppression by cilostazol. *Biochem Pharmacol* **86**, 1320-1327 (2013).
58. Zeng, R.J., Zhuo, Z.W., Luo, Y.J., Sha, W.H. & Chen, H. Rho GTPase signaling in rheumatic diseases. *Iscience* **25**(2022).
59. Sandi, M.J.*, et al.* MARK3-mediated phosphorylation of ARHGEF2 couples microtubules to the actin cytoskeleton to establish cell polarity. *Sci Signal* **10**(2017).
60. Hu, Y.J.*, et al.* PCGF3 promotes the proliferation and migration of non-small cell lung cancer cells via the PI3K/AKT signaling pathway. *Exp Cell Res* **400**(2021).

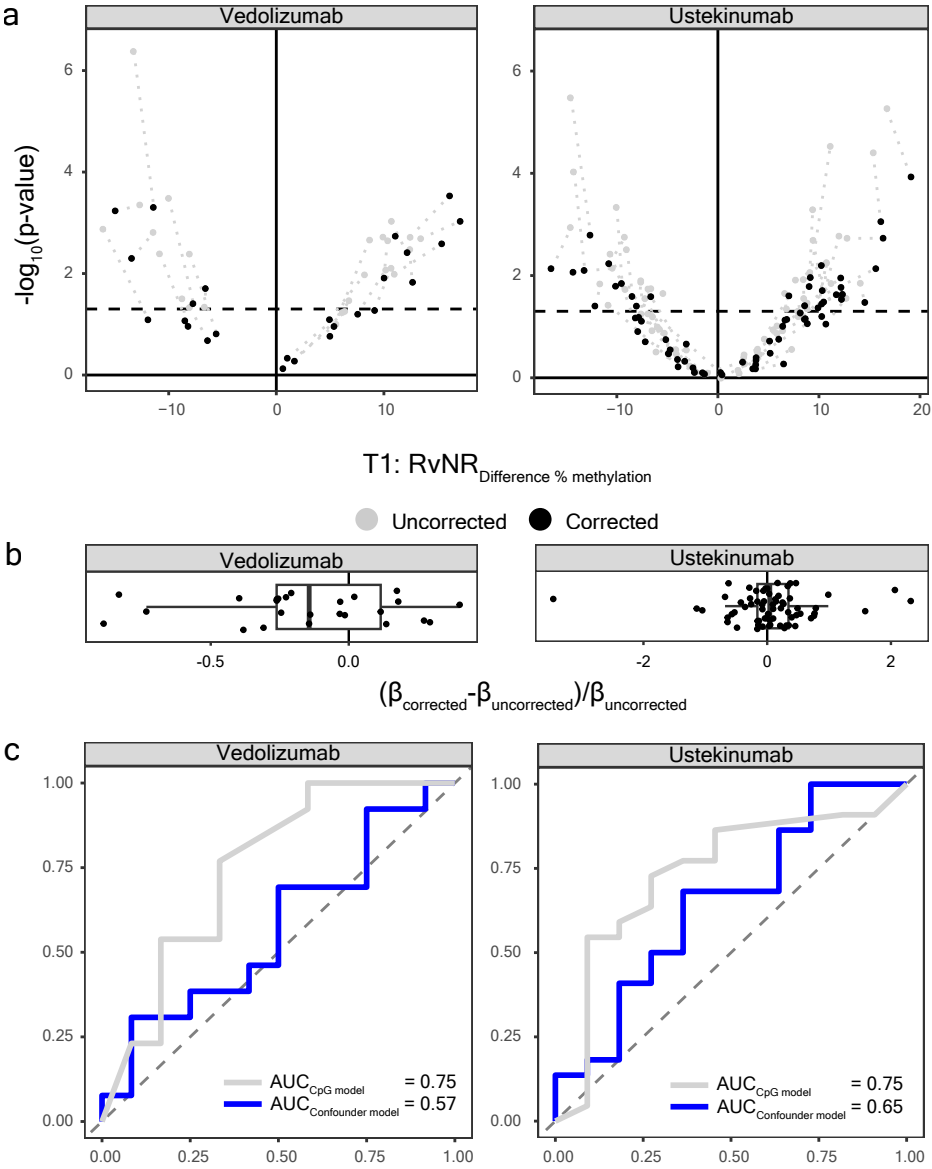
61. Zhuang, G.Q., *et al.* A Novel Regulator of Macrophage Activation miR-223 in Obesity-Associated Adipose Tissue Inflammation. *Circulation* **125**, 2892-+ (2012).
62. Jiao, P., *et al.* miR-223: An Effective Regulator of Immune Cell Differentiation and Inflammation. *Int J Biol Sci* **17**, 2308-2322 (2021).
63. Wright, P.B., *et al.* The mannose receptor (CD206) identifies a population of colonic macrophages in health and inflammatory bowel disease. *Sci Rep* **11**, 19616 (2021).
64. Koelink, P.J., *et al.* Anti-TNF therapy in IBD exerts its therapeutic effect through macrophage IL-10 signalling. *Gut* **69**, 1053-1063 (2020).
65. Lu, W., *et al.* The microRNA miR-22 inhibits the histone deacetylase HDAC4 to promote T(H)17 cell-dependent emphysema. *Nat Immunol* **16**, 1185-1194 (2015).
66. Glauben, R., Sonnenberg, E., Wetzel, M., Mascagni, P. & Siegmund, B. Histone Deacetylase Inhibitors Modulate Interleukin 6-dependent CD4(+) T Cell Polarization in Vitro and in Vivo. *J Biol Chem* **289**, 6142-6151 (2014).
67. Brusselle, G.G. & Bracke, K.R. MicroRNA miR-22 drives T(H)17 responses in emphysema. *Nat Immunol* **16**, 1109-1110 (2015).
68. Dou, B., Ma, F.Z., Jiang, Z.Y. & Zhao, L. Blood HDAC4 Variation Links With Disease Activity and Response to Tumor Necrosis Factor Inhibitor and Regulates CD4+T Cell Differentiation in Ankylosing Spondylitis. *Front Med-Lausanne* **9**(2022).
69. Brusselle, G.G. & Bracke, K.R. MicroRNA miR-22 drives T(H)17 responses in emphysema. *Nat Immunol* **16**, 1109-1110 (2015).
70. Browning, L.M., *et al.* TGF-beta-mediated enhancement of TH17 cell generation is inhibited by bone morphogenetic protein receptor 1alpha signaling. *Sci Signal* **11**(2018).
71. Kuczma, M. & Kraj, P. Bone Morphogenetic Protein Signaling Regulates Development and Activation of CD4(+) T Cells. *Vitam Horm* **99**, 171-193 (2015).
72. Lu, L., *et al.* Synergistic effect of TGF-beta superfamily members on the induction of Foxp3(+) Treg. *Eur J Immunol* **40**, 142-152 (2010).
73. Park, S.H., *et al.* Hypermethylation of EBF3 and IRX1 genes in synovial fibroblasts of patients with rheumatoid arthritis. *Mol Cells* **35**, 298-304 (2013).
74. Stockinger, B. & Veldhoen, M. Differentiation and function of Th17 T cells. *Curr Opin Immunol* **19**, 281-286 (2007).
75. Lyakh, L., Trinchieri, G., Provezza, L., Carra, G. & Gerosa, F. Regulation of interleukin-12/interleukin-23 production and the T-helper 17 response in humans. *Immunol Rev* **226**, 112-131 (2008).
76. Zhao, L., Ghetie, D., Jiang, Z. & Chu, C.-Q. How Can We Manipulate the IL-23/IL-17 Axis? *Current Treatment Options in Rheumatology* **1**, 182-196 (2015).
77. Morishima, N., Mizoguchi, I., Takeda, K., Mizuguchi, J. & Yoshimoto, T. TGF-beta is necessary for induction of IL-23R and Th17 differentiation by IL-6 and IL-23. *Biochem Bioph Res Co* **386**, 105-110 (2009).
78. Rajkovic, A., Lee, J.H., Yan, C. & Matzuk, M.M. The ret finger protein-like 4 gene, Rfpl4, encodes a putative E3 ubiquitin-protein ligase expressed in adult germ cells. *Mech Dev* **112**, 173-177 (2002).
79. Huang, C. Roles of E3 ubiquitin ligases in cell adhesion and migration. *Cell Adh Migr* **4**, 10-18 (2010).
80. Tsai, P.C. & Bell, J.T. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* **44**, 1429-1441 (2015).
81. Rees, C. & Penman, I. Has the COVID-19 pandemic changed endoscopy in the UK forever? *Lancet Gastroenterol Hepatol* **8**, 6-8 (2023).
82. Lappalainen, T. & Greally, J.M. Associating cellular epigenetic models with human phenotypes. *Nat Rev Genet* **18**, 441-451 (2017).
83. Villicana, S., *et al.* Genetic impacts on DNA methylation help elucidate regulatory genomic processes. *Genome Biol* **24**, 176 (2023).

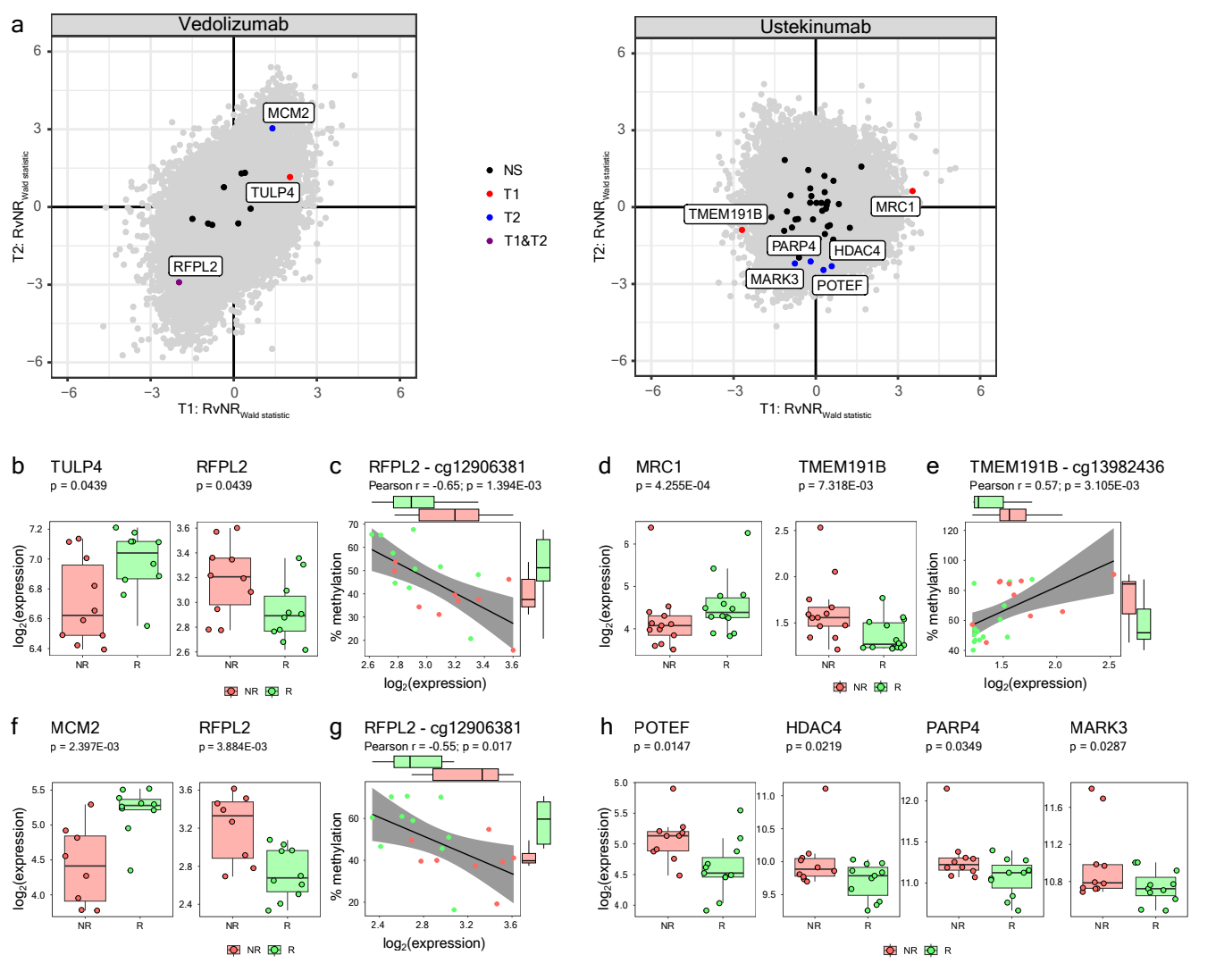
84. Yu, A.P., Cabanilla, L.A., Wu, E.Q., Mulani, P.M. & Chao, J. The costs of Crohn's disease in the United States and other Western countries: a systematic review. *Curr Med Res Opin* **24**, 319-328 (2008).
85. Mehta, F. Report: economic implications of inflammatory bowel disease and its management. *Am J Manag Care* **22**, s51-60 (2016).
86. Jackson, B., Con, D., Ma, R., Gorelik, A., Liew, D. & De Cruz, P. Health care costs associated with Australian tertiary inflammatory bowel disease care. *Scand J Gastroenterol* **52**, 851-856 (2017).
87. Ghosh, N. & Premchand, P. A UK cost of care model for inflammatory bowel disease. *Frontline Gastroenterol* **6**, 169-174 (2015).
88. Rubin, D.T., Mody, R., Davis, K.L. & Wang, C.C. Real-world assessment of therapy changes, suboptimal treatment and associated costs in patients with ulcerative colitis or Crohn's disease. *Aliment Pharmacol Ther* **39**, 1143-1155 (2014).
89. Noor, N.M.*, et al.* A biomarker-stratified comparison of top-down versus accelerated step-up treatment strategies for patients with newly diagnosed Crohn's disease (PROFILE): a multicentre, open-label randomised controlled trial. *Lancet Gastroenterol Hepatol* **9**, 415-427 (2024).
90. DNA methylation markers to predict treatment success of biologicals in Crohn’s disease.
91. Vermeire, S., Schreiber, S., Sandborn, W.J., Dubois, C. & Rutgeerts, P. Correlation between the Crohn's disease activity and Harvey-Bradshaw indices in assessing Crohn's disease severity. *Clin Gastroenterol Hepatol* **8**, 357-363 (2010).
92. Molder, F.*, et al.* Sustainable data analysis with Snakemake. *F1000Res* **10**, 33 (2021).
93. Aryee, M.J.*, et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369 (2014).
94. Fortin, J.P., Triche, T.J., Jr. & Hansen, K.D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558-560 (2017).
95. Fortin, J.P., Fertig, E. & Hansen, K. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. *F1000Res* **3**, 175 (2014).
96. Heiss, J.A. & Just, A.C. Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. *Clin Epigenetics* **10**, 73 (2018).
97. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013).
98. Fortin, J.P.*, et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* **15**, 503 (2014).
99. Javierre, B.M.*, et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384 e1319 (2016).
100. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
101. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883 (2012).
102. Daca-Roszak, P.*, et al.* Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: implications for comparative population studies. *BMC Genomics* **16**, 1003 (2015).
103. Andrews, S.V., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D. & Fallin, M.D. "Gap hunting" to characterize clustered probe signals in Illumina methylation array data. *Epigenetics Chromatin* **9**, 56 (2016).
104. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **29**, 1189-1232 (2001).

105. J., P., E., S., A., Z. & E., L. Covered Information Disentanglement: Model Transparency via Unbiased Permutation Importance. in *The Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)* (2022).
106. Demsar, J. & Zupan, B. Hands-on training about overfitting. *PLoS Comput Biol* **17**, e1008671 (2021).
107. Ritchie, M.E.*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
108. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
109. Houseman, E.A.*, et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
110. Salas, L.A.*, et al.* An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* **19**, 64 (2018).
111. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, (Springer-Verlag New York, 2016).
112. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016).
113. Dobin, A.*, et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
114. Li, H.*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
115. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
116. H, W. ggplot2: Elegant Graphics for Data Analysis [Internet]. *New York, NY: Springer-Verlag New York* (2009).


a
Stratified shuffle split
Feature selection
Training
Prediction
Train (80%)
Test (20%)
Int. train (80%)
Int. test (20%)
#1
#2
#100
XGboost
CpGs
Random
Feature importance
Predictor CpGs
XGboost
New data
ROC
Mean ROC
Youden index
Discovery (AmsterdamUMC)
Validation (John Radcliffe Hospital)
b
ROC curve
Vedolizumab
TPR
FPR
AUC Discovery = 0.87
AUC Validation = 0.75
Ustekinumab
AUC Discovery = 0.89
AUC Validation = 0.75
c
Predictor CpGs
NR
R
Feature importance
Mean feature importance


a
Vedolizumab
Ustekinumab
-log10(p-value)
T2vT1Difference % methylation
b
Vedolizumab
Spearman ρ = 0.971; p = 7.832E-07
Ustekinumab
Spearman ρ = 0.941; p < 2.2E-16
T2: RvNRDifference % methylation
T1: RvNRDifference % methylation
c
Vedolizumab
Ustekinumab
Current study
Joustra et al. 2022
ICC
d
Vedolizumab
Ustekinumab
TPR
FPR
AUCDiscovery T1 = 0.89
AUCDiscovery T2 = 0.97
AUCDiscovery T1 = 0.97
AUCDiscovery T2 = 1.00


a
Vedolizumab
Ustekinumab
-log10(p-value)
T1: RvNR Difference % methylation
Uncorrected
Corrected
b
Vedolizumab
Ustekinumab
(βcorrected-βuncorrected)/βuncorrected
c
Vedolizumab
Ustekinumab
AUCCpG model = 0.75
AUCConfounder model = 0.57
AUCCpG model = 0.75
AUCConfounder model = 0.65


a
Vedolizumab
Ustekinumab
MCM2
TULP4
RFPL2
TMEM191B
MRC1
PARP4
HDAC4
MARK3
POTEF
NS
T1
T2
T1&T2
T2: RvNR Wald statistic
T1: RvNR Wald statistic
b
TULP4
p = 0.0439
RFPL2
p = 0.0439
log2(expression)
NR
R
c
RFPL2 - cg12906381
Pearson r = -0.65; p = 1.394E-03
% methylation
log2(expression)
d
MRC1
p = 4.255E-04
TMEM191B
p = 7.318E-03
e
TMEM191B - cg13982436
Pearson r = 0.57; p = 3.105E-03
f
MCM2
p = 2.397E-03
RFPL2
p = 3.884E-03
g
RFPL2 - cg12906381
Pearson r = -0.55; p = 0.017
h
POTEF
p = 0.0147
HDAC4
p = 0.0219
PARP4
p = 0.0349
MARK3
p = 0.0287