UiT The Arctic University of Norway

Faculty of Health Science

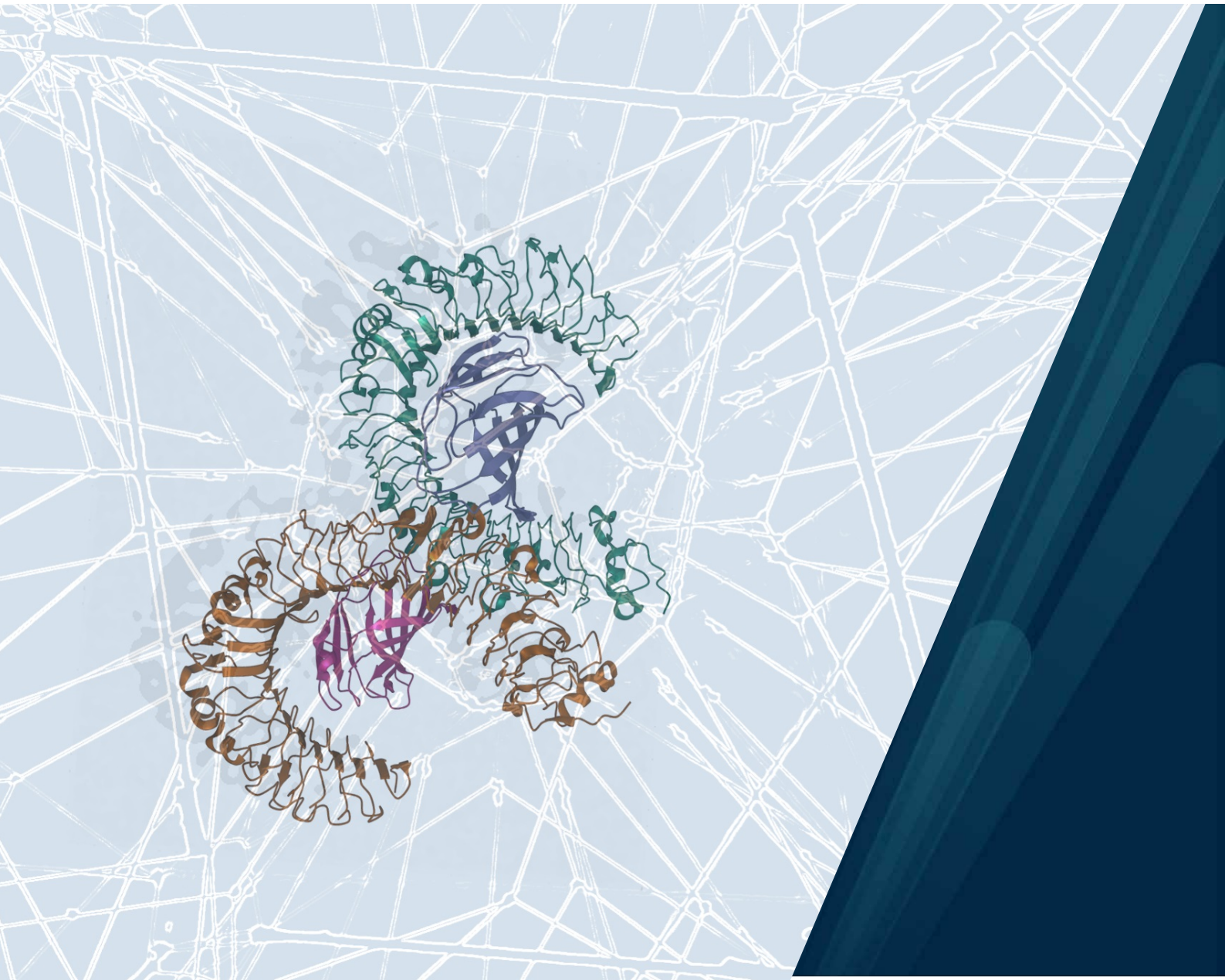
Department of Clinical Medicine

Falsifiable Network Models

A Network-based Approach to Predict Treatment Efficacy in Ulcerative Colitis

Amrinder Singh

A dissertation for the degree of Philosophiae Doctor (Ph.D.) June 2022



A dissertation for the degree of Philosophiae Doctor

Falsifiable Network Models

A Network-based Approach to Predict Treatment Efficacy in Ulcerative Colitis

Amrinder Singh

Tromsø, June 2022

Clinical Bioinformatics Research Group
Department of Clinical Medicine
Faculty of Heath Science
UiT The Arctic University of Norway
Norway

Table of Contents

Acknowledgements

This research was carried out in the Clinical Bioinformatics Group, Department of Clinical Medicine, Faculty of Health Sciences, at University of Tromsø – The Arctic University of Norway from February 2019 to June 2022. My sincere thanks to Helse Nord for financially supporting this research project.

I would like to express my gratitude to everybody that has supported me during my PhD research. This PhD would not have been possible without the guidance of my main supervisor Endre Anderssen. I deeply appreciate your contribution and guidance, which was a major source of motivation throughout my PhD. Without a doubt, it was quite enjoyable learning experience with him. I embraced every scientific and non-scientific conversation with him. I would also like to acknowledge my co-supervisor Ruth Paulssen for sharing her extensive knowledge and valuable opinions. This project was completed in the stipulated time frame due to her timely supervision and guidance. I thank you for being available for my academic and personal matters and for providing me with guidance and motivation. I would like to extend my thanks to Christopher Fenton. I enjoyed your interesting all-around conversation on various topics during lunch break. Your comments and feedbacks on my research articles were valuable. I appreciate the support and supervision you provided briefly while Endre was on his paternity leave. You were often responsive to my doubts which ignited fruitful whiteboard discussions.

I appreciate the friendship of Hagar and her immense support throughout my PhD time. Thank you for our long discussions and your wonderful company. I also like to thank my friends Wei, Mithlesh, Lars, and Line who will always be remembered for our gossips and wonderful memories. I will never forget the memorable pre-Christmas feast and birthday treats. I want to thank Rafi Ahmed for being supportive and most generous during the tough times of PhD. I am looking forward to join you soon in a cricket match in Oslo.

In the last 3 years, I was fortunate to have friends in Tromsø. I want to thank Anuroop, Bhupinder, Sonam, Roopam, Sunny and others who kept my social life alive in the Arctic. I will always remember our late-night parties, delicious weekend dinners and hiking trips. Our trip to Poland will always be one of the golden memories of all time. Thank you, Puneet and Animesh for reviewing my thesis and providing useful suggestions. To Puneet, we can hopefully make our pending trip to Iceland soon. At last, I am overwhelmed by the constant unconditional love and support of my mom Manjeet and sister Gurleen who had always motivated and cared for me. With your support and care, I stood through my worst episodes. I will always try to make you proud in future.

I would like to thank all personnel of the Department of Clinical Medicine for providing required administrative support, especially Arvid Inge Paulsen. I want to thank NORBIS and Digital Life Norway consortiums for providing me the necessary support to attend various courses and conferences. I would like to thank NORBIS for giving me the opportunity to contribute in national bioinformatics consortium as a student board representative.

Undoubtedly, this city "Tromsø" is a paradise on earth with relatively untouched wonderful surrounding nature. A gaze out of a window is the remedy to reduce stress. However, there are flying rats (Seagulls) in the sky which sometimes might be annoying.

Amrinder Singh

Tromsø, June 2022

List of papers

Paper-I

Identifying anti-TNF response biomarkers in ulcerative colitis using a diffusion-based signalling model

Amrinder Singh, Endre Anderssen, Christopher G. Fenton, Ruth H. Paulssen

Bioinformatics Advances, Volume 1, Issue 1, 2021, vbab017, https://doi.org/10.1093/bioadv/vbab017

Paper-II

Identifying predictive signalling networks for Vedolizumab response in ulcerative colitis

Amrinder Singh, Christopher G. Fenton, Endre Anderssen, Ruth H. Paulssen

International Journal of Colorectal Disease (2022)

https://doi.org/10.1007/S00384-022-04176-W

Paper-III

Modelling individual variability of pattern recognition receptor pathway response in IBD

Endre Anderssen, Amrinder Singh, Christopher G. Fenton, Ruth H. Paulssen

Manuscript

Summary

This work is focused on understanding the treatment efficacy of patients with ulcerative colitis (UC) using a network-based approach. UC is one of two forms of inflammatory bowel disease (IBD) along with Crohn's disease. UC is a debilitating condition characterised by chronic inflammation and ulceration of the colon and rectum. UC symptoms occur gradually rather than abruptly, and the degree of symptoms differs across UC patients. Only around 20% of all UC cases can be explained by known genetic variations, implying a more ambiguous aetiology that is yet not fully understood but is thought to involve a complex interplay between genetic and environmental factors.

The available therapy for UC substantially reduces symptoms and achieves long-term remission. However, about one-third of UC patients fail to respond to anti-TNFα therapy and consequently develop long-term side effects due to medication. Non-response to existing antibody-based therapies in subgroups of UC patients is a major challenge and incurs a healthcare burden. Therefore, the disease markers for predicting therapy response to assist individualized therapy decisions are needed. To date, no quantitative computational framework is available to predict treatment response in UC.

We developed a quantitative framework that uses gene expression data and existing biological background information on signalling pathways to quantify network connectivity from receptors to transcription factors (TF) that are involved in UC pathogenesis. Variations in network connectivity in UC patients can be used to identify responders and non-responders to anti-TNFα and anti-Integrin treatment. Our findings allow us to summarize the effect of small gene expression changes on the overall connectivity of a signalling network and estimate the effect this will have on the individual patients' responses. Estimating the network connectivity associated with varied drug responses may provide an understanding of individualized treatment outcomes.

Our model could be used to generate testable hypotheses about how individual genes act together in networks to cause inflammation in UC as well as other immune-inflammatory diseases such as psoriasis, asthma, and rheumatoid arthritis.

Abbreviations

UC: Ulcerative Colitis

IBD: Inflammatory Bowel Disease

CD: Crohn's Disease

TNF: Tumour Necrosis Factor

IFX: Infliximab

VDZ: Vedolizumab

ETR: Etrolizumab

TCZ: Tocilizumab

MTX: Methotrexate

RTX: Rituximab

FDR: False Discovery Rate

FMT: Faecal Microbiota Transplant

HNF4A: Hepatocyte Nuclear Factor 4 Alpha

IBSEN: Inflammatory Bowel disease

JAK 2: Janus Kinase 2

KEGG: Kyoto Encyclopedia of Genes and Genomes

GWAS: Genome-wide Association Studies

HLA: Human Leukocyte Antigen

5-ASA: 5-Aminosalicylic acid

CRP: C-reactive Protein

GEO: Gene Expression Omnibus

IQR: Interquartile Range

ROC: Receiver Operating Characteristic

AUC: Area Under the Curve

PCA: Principal Component Analysis

PLS: Partial Least Square

GO: Gene Ontology

TF: Transcription Factor

MAdCAM-1: Mucosal vascular Addressin Cell Adhesion Molecule 1

VCAM-1: Vascular Cell Adhesion Molecule 1

ICAM-1: Intercellular Adhesion Molecule 1

ITGB4: Integrin Subunit Beta 4

ITGB7: Integrin Subunit Beta 7

lncRNAs: Long non-coding Ribonucleic Acid

PPI: Protein-Protein Interaction

FFAR2: Free Fatty Acid Receptor 2

DSS: Dextran Sodium Sulphate

1 Introduction

1.1 Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a group of ailments with a highly heterogenous disease phenotype and response to therapy. IBD consists of two major clinical entities: Crohn's disease (CD) and ulcerative colitis (UC) which are distinct by their pathophysiological states [1,2]. Both CD and UC share epidemiological, clinical, and therapeutic characteristics. For some, it is difficult to distinguish to diagnosis (10-15%) which is termed indeterminate colitis [3]. IBD manifests in the gastrointestinal tract with chronic, relapsing and remission intervals of mucosal inflammation [4]. Patients with IBD suffer from diarrhoea, rectal bleeding, weight loss, abdominal cramps and extra-intestinal manifestations leading to poor quality of life. IBD is characterised by progressive and destructive disease stages with serious complications and an increased risk of developing colorectal cancer [5]. During the disease course, IBD patients require maintenance treatment to reduce inflammation. As a result of therapy, some IBD patients achieve clinically inactive disease (CID) with less or no inflammation and usually without symptoms [6–8]. IBD is a multifaceted disease triggered by multiple factors including genetic, environmental, abnormal gut microbiota, and dysregulated immune response [9–11]. The disease's aetiology remains elusive despite known factors. IBD is often diagnosed in individuals of diverse age groups. Around 25-35 % of IBD patients are above the age of 60, however, the prevalence of the disease is rapidly growing in the paediatric population [12].

1.1.1 Ulcerative colitis (UC)

UC is a chronic inflammatory illness characterised by persistent, non-transmural (confined to the epithelial lining) inflammation and ulcers (sores) of the large intestine, commencing in the rectum and extending to proximal segments of the colon. UC is typically characterised by continuous, circumferential, and superficial inflammation that is localized to a mucus layer [1,5,13]. Extraintestinal manifestation (EIM) can occur in UC patients including peripheral arthritis, aphthous stomatitis, uveitis, and sclerosing cholangitis [14–16]. Because UC is complex, genetic, environmental, and microbial variables all have a contribution to disrupting gut homeostasis, leading to epithelial barrier impairments and dysregulated immune-inflammatory responses [17]. The disease susceptibility conferred by genetic factors has a lower impact as compared to Crohn's disease. Emerging evidence supports that the relevance of environmental variables in UC aetiology outweighs genetic association, i.e., 5-15 % of UC risk is genetic [18,19].

1.1.2 Epidemiology and Pathophysiology of UC

Epidemiological findings suggest that the incidence and prevalence of UC is rising globally, especially in the white population, affecting 0.3% of the population in North America and northern Europe[20–23]. Walker *et al.* discovered substantial changes in the IBD phenotype and significantly less penetrating disease effects in a South Asian IBD cohort compared to a Northern European IBD cohort. This indicates that IBD is caused by a combination of genetic, environmental, and behavioural factors. A recent systematic analysis that analysed the global disease burden in 195 countries between 1990 and 2017 discovered that there is a huge social, economic, and healthcare burden expected to rise in the future [24]. Burisch *et al*. projected a direct healthcare cost of 4.6-5.6 billion Euros per year [25]. This healthcare burden is projected to increase as the more expensive treatments are developed to address the ever-increasing number of affected IBD patients [26]. Both UC and CD share remarkable epidemiological incidence. However, recent studies have estimated that the incidence and prevalence of UC are twice as much as compared to Crohn's disease. For example, the incidence and prevalence of UC range from 1.2-20 and 7.6-245 cases per 100,000 individuals/year respectively [27–30]. UC can develop at any age, although the incidence age pattern is bimodal, with a significant peak between the ages of 30-40 years and a modest increase between the ages of 50-70 years [31–33].

The growing trend of UC cases in developed countries has reduced over the last decade. According to new epidemiological studies, there is a concerning trend in the incidence rate of UC in the developing world, which was previously low. One rationale for the increasing prevalence of UC in Asia and South America might have been the progressive adoption of Western culture (diet, lifestyle, etc.) [34]. The explanation for the higher prevalence of IBD in the Western world remains unknown; nevertheless, the key variables driving that cause may be an adaptation to a Western lifestyle. The significance of gender in UC prevalence is still being debated, some UC studies have indicated male predominance or a comparable gender balance [28,30,35], while others have shown contradicting results [22]. The reported studies of patients with UC-related mortality ranged between 11-30 %, with the majority of these patients developing colorectal cancer (24 - 44 %), gastrointestinal diseases, postoperative complications (17-100 %), non-alcoholic liver diseases, end-stage liver diseases from primary sclerosing cholangitis, cholangiocarcinoma, etc. [36,37].

1.1.3 Pathogenesis of UC

UC pathophysiology is multifaceted and remains poorly understood. It encompasses a broad range of risk factors, including immunological, genetic, epigenetic (environmental), and microbiological

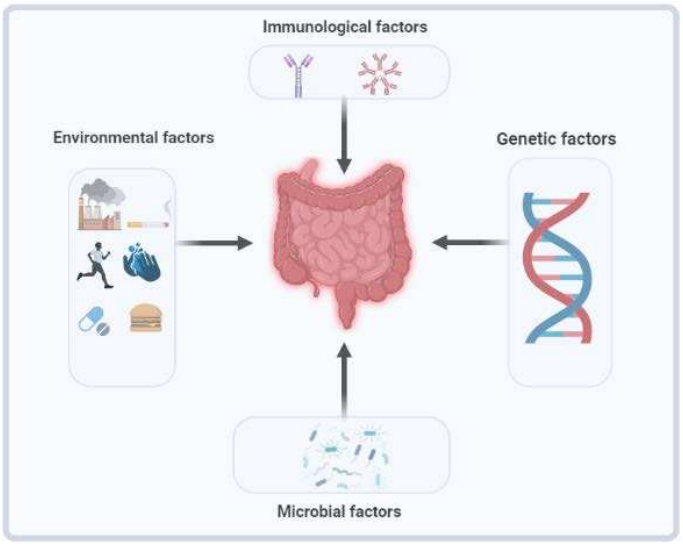


Figure 1. Four key components of IBD pathogenesis

factors. These risk factors primarily contribute to epithelial barrier abnormalities and dysregulated immunological response, which both contribute to disease onset [1,38,39].

Genetic factors

Hereditary factors may play a role in the aetiology of UC, with a family history of UC being the most significant genetic risk factor. Around 8-14 % of UC patients have a family history of IBD, which is more prominent in UC [40]. Danish and Swedish twin studies, for example, demonstrated the involvement of genetic factors for UC, with between 6.3% and 18.8% concordance between monozygotic twins compared to 0-4.5 % in dizygotic twins [41,42].

New high-throughput sequencing technologies, such as next-generation sequencing (NGS), have made it conceivable to study disease-specific genetic factors in UC. Using genome-wide association studies (GWAS), over 200 IBD-specific disease loci were identified, of which about 23 were unique to UC and 47 showed overlap with CD, such as interleukins (IL) 23 and 10, and Janus kinase-2 (JAK-2) pathway genes, etc. [10,43,44]. These genes, which are associated with adaptive immune response, influence the differential immune response in a UC patient as compared to a control [45]. Ntunzwenimanahave *et al.* have identified 145 IBD-associated genetic loci that show endogenous expression in the intestinal epithelial cell and play multiple roles in epithelial structure, function, and microbial defense [46]. Several GWAS studies found that variants in hepatocyte nuclear factor 4-alpha (HNF4A) is positively related to UC. HNF4A is an epithelial-specific gene that functions as a transcriptional regulator of ion transport. The mucosal ion transport function modulates the overall inflammatory condition [47].

A meta-analysis found UC risk genes, revealing that 13% of UC patients had a family history of the disease [19]. These genetic susceptibility loci and genetic risk factors account for about 20-25 % of UC genetic heritability. HLA haplotype DRB1*0103, for instance, shows a substantial association with disease susceptibility and is a high-risk factor for colectomy [48,49]. There are various HLA complexes associated with UC. In the Japanese population, HLA-DR2 has been linked to UC [50]. The discovery of UC-specific risk alleles that translate into a protein plays an important role in epithelial cell adhesion and highlights the significant importance of faulty barrier function in disease development [51]. Other findings revealed that hypermethylation of the protein E-cadherin had a hereditary link to colorectal cancer and ulcerative colitis [52].

Epigenetics is another key aspect of genetics in UC. Epigenetic alterations are both dynamic and reversible genetic changes. DNA methylation, histone modifications, non-coding RNAs, and nucleosome positioning are all examples of epigenetic alterations that are impacted by environmental variables. Chemical adjustments to genomic structure occur rather than changes to genetic sequences [53]. Taman *et al.* found that in UC treatment-naive patients, hypermethylation of genes is associated with homeostasis and microbial defense, while hypomethylation corresponds to immunological response [54]. The epigenetic changes in UC are being explored as potential biomarkers for pharmacological therapy, diagnostic, predictive, and patient-tailored therapeutic tools. DNA methylation and non-coding RNA have been thoroughly examined in IBD, and differently expressed non-coding RNAs between CD and UC have been discovered [55,56].

Environmental factors

Environment factors influence the risk of UC, according to accumulating epidemiological data at both the individual and population levels. Western diet and lifestyle, as well as an increasing trend of urbanization, is one of these factors. Diet is a major environmental risk factor for people with UC because it induces dysbiosis inside the gut microbiota and stimulates proinflammatory substrate production, which damages the mucus barrier and increases intestinal permeability [57]. Consumption of a Western diet high in processed meat, saturated fats, processed sugars, and low consumption of high fiber foods have been linked to gut dysbiosis, which results in worse IBD-related outcomes [58–61]. Rising UC incidence levels have been observed in the Western world, which has a long life expectancy and higher hygienic status, contributing to the "hygiene hypothesis" [62]. According to the hygiene hypothesis, early exposure to enteric and general microbial infection renders the immune system less diverse for antigen recognition in the late phases of life. Smoking enhances the risk of

IBD and is the most recognized environmental factor that modulates immune responses and alters gut flora composition [63,64]. Chronic smoking has been identified as a latent risk factor that influences epithelial mucus profiles, resulting in a dysfunctional mucus barrier with gut inflammation [65]. In contrast to CD, smokers had a decreased likelihood of developing UC and a lower risk of colectomy [66,67]. However, former smoking has been identified as a strong risk factor for UC disease onset while an active smoker has positive effects on reducing the severity of the disease course. Appendectomy has been shown to reduce the risk of developing UC, suggesting that the vermiform appendix has a direct interaction with the gut flora [68]. Exposure to antibiotics also affects the gut microbiota and is implicated in the pathogenesis of IBD [69]. Several studies have found an inverse connection between breastfeeding and the incidence of IBD, with a substantial protective advantage for at least one year of nursing [70]. Patients with UC who are under psychological stress have an increased risk of disease, implicating the brain-gut axis in UC development [71]. Stress may exert its effects through a variety of pathways, including the production of cytokines, the modification of gut flora, etc. [72]. Patients with UC who raised their intensity of regular exercise for 6 months had a 24% reduced chance of developing active UC [73].

Gut microbial flora

The estimated mean size of the gastrointestinal tract (GI) is $32m^2$ with the large intestine measuring around $2m^2$ [74]. The human gut has the most diverse bacterial populations (1012 bacteria/cm^3) in the large intestine, with over 1000 species, however, this number varies by individual.[75]. The human GI tract is the primary location of immune system function due to its diversified and dense bacterial flora population. UC disrupts the complex interplay between the immune system and the intestinal flora, which is triggered by environmental stimuli in genetically susceptible individuals, resulting in inflammation [39]. These environmental factors alter the host defense mucosal barrier, resulting in an exaggerated immune response and a dysregulated equilibrium between beneficial and harmful gut microbial species [76]. The epithelial barrier, which is protected by the mucinous layer, is the primary defense line that provides mucosal immune response against luminal microbes and synthesizes antimicrobial peptides. During UC, there is a decrease in the synthesis of sulphation colonic mucin known as mucin-2 [77]. The epithelial barrier becomes susceptible to damage as a result of alterations in protective substances such as mucin 2 and some defective regulation of tight junctions, resulting in higher permeability and increased inflow of luminal antigens [78]. The intestinal immune system during ulcerative colitis can be manifested by changes in the colonic

epithelial cells and mucosal barrier defects. Patients with mucosal barrier defects have depleted colonic goblet cells as well as a highly permeable mucosal barrier junction [1].

Immunological factors

In UC, the mucosa comprises an immune cell population in which homoeostatic balance between regulatory and effector T-cells such as T-helper (e.g., T_H 1, T_H 2, and T_H 17) cells is disrupted. Traditionally, UC pathogenesis has been identified as a T_H2-like disease as compared to CD which is associated with T_H1 cells. However, some studies have suggested the role of imbalance between T_H17 and regulatory T (T_{reg}) [39,79]. All T_H1, T_H2 and T_H17 cells are effector T helper (T_H) cells and subtypes of CD4+ T cells which secrete various cytokines such as interleukin (IL). For example, IL 4, 5 and 13 are secreted by T_H2, whereas T_H1 secrete TNF-α, IFN-γ and T_H17 secrete IL-17. IL 13 is known for its cytotoxic activity against epithelial cells causing apoptosis and damage to a protein involved in tight junction functioning [78,80]. These T-helper cells play a key protective role for the host against pathogens and prevent the inflow of luminal bacteria, thus their regulation is important for gut homeostasis [81]. Another significant form of immune response is antigen recognition, in which antigens trigger innate immune responses by interacting with macrophages and dendritic cells. Dendritic cells recognize antigens in the lumen by sending dendrites to the epithelium. [82]. There are abundant macrophages and dendritic cells in the lamina propria that use B cells and T cells to activate adaptive and innate immune responses. The population of activated and mature dendritic cells increases dramatically during UC. Dendritic cell's high stimulatory activity plays a significant part in the development of inflammation [83].

Dendritic cells have been shown to express a variety of microbial-recognition receptors, including Toll-like receptors (TLR) and NOD-like receptors. TLR signalling protects against pathogens and epithelial damage, aiding in homoeostasis and epithelial barrier function. TLR3 and 5 are more often expressed by normal epithelial cells than TLR2 and 4, which are either scarce or absent [84]. TLR4 expression was found to be strongly upregulated in patients with UC [85] as well as CD whereas expression of TLR2 and 5 remained unaltered in IBD. Thus, there is a significant difference in TLR expression levels between normal and active UC intestinal epithelium, suggesting that differences in innate immune response may result in UC pathogenesis. TLR polymorphism makes the host more susceptible to intestinal infections or enables adaptive immune responses to be more tolerant of pathogens. Variation in TLR4 (D299G), for example, may be a significant risk factor for UC in Caucasians [86]. TLR pathway activation is critical in the activation of key transcription factors such as

nuclear factor- κB (NF-κB) connected with the inflammatory signalling cascade [87]. The role of NF-κB is complex and cell-type dependent because NF-κB regulates proinflammatory and cell survival functions in macrophages and T cells, [88,89] whereas it has a protective role in epithelial cells [90].

One of the major causes of inflammation in a UC patient is the migration of leukocytes to the mucosal region. During active inflammation, chemoattractants such as CXCL8 are upregulated, contributing to the amplification of the inflammatory response [91]. The increased production of adhesion molecules on the endothelium layer of the mucosa by proinflammatory cytokines such as mucosal addressin cellular adhesion molecule-1 (MadCAM-1) enhances leucocyte adherence, perpetuating an inflammatory cycle. During inflammation, MadCAM-1 interaction with α4β7 integrin directs lymphocyte homing to gut-associated lymphoid tissue [92]. Anti-integrin drugs comprise antibodies to either MadCAM-1 or its ligand α4β7 (e.g., Vedolizumab) and another subunit β7 of heterodimeric integrin (e.g., Etrolizumab) reduces the severe colonic inflammation by preventing lymphocyte recruitment.

1.2 Signal Transduction

Signal transduction is a biochemical process that translates extracellular signals from the environment into specific cellular responses. It occurs in mainly three steps: Reception, Transduction, and Response. During the reception, the cell senses a signalling molecule such as ligands, microbes, or other external stimuli, that adhere to cell surface receptor molecules, e.g., cytokine receptors, tumour necrosis factor receptors (TNFR), etc. The binding of ligand alters the conformation of receptor protein which initiates the activation of the signalling cascade resulting in the cellular response (**Figure 2**). Various signalling such as ligands, receptors, kinases, phosphatases, transcription factors etc, are involved in the signalling cascade, which generates diverse signal transduction pathways. Signal transduction is thought to occur through closely organised networks controlled by a modular domain that drives protein-protein interactions and the reversible construction of signalling complexes. Anomalies in signalling transduction pathways are involved in the pathophysiology of

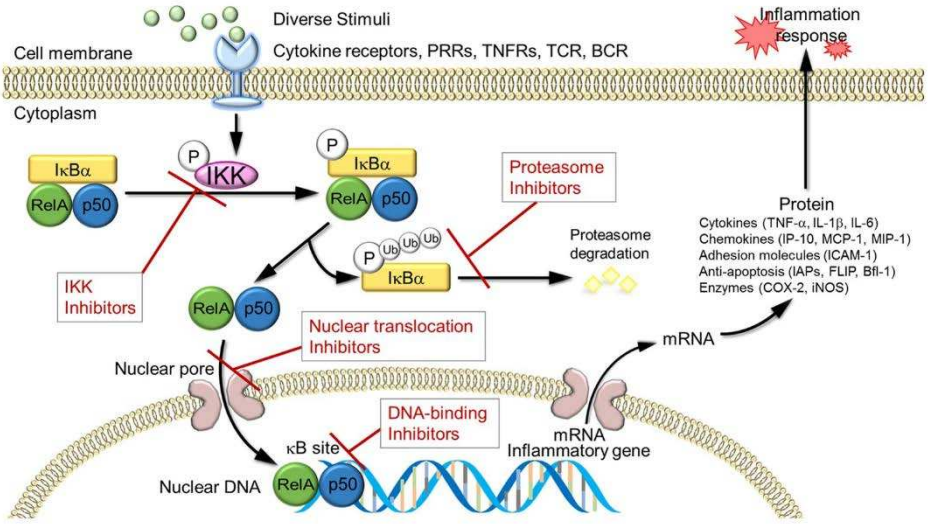


Figure 2. NF-κB-targeted therapeutics in inflammatory diseases [used with permission from [94]]

many diseases. Cancer and diabetes are the most common diseases induced by signalling dysregulation. The mechanism of action of many drugs also involves cell signalling pathways [93].

1.2.1 Inflammatory signalling pathways in UC

Inflammation is the host's defense response to infection; it induces tissue damage and involves the migration of immune cells to the site of injury. It is the primary characteristic of numerous human diseases including UC. Understanding the mechanisms of inflammation is crucial for improving disease outcomes. The intestine is the primary organ affected by inflammation in IBD, including various factors that make disease aetiology immensely complex. Intestinal inflammation is the result of an aberrant, chronic mucosal immune response. It is characterised by the release of pro-inflammatory mediators that disrupt intestinal homeostasis and cause intestinal barrier dysfunction [95]. Multiple inflammatory cytokines such as interleukins (ILs) (e.g., IL-1β, IL-6), interferons, tumour necrosis factor- α (TNF-α), and increased leukocyte recruitment to the site of inflammation characterise intestinal inflammation. The following illustration describes immune cells and cytokines associated with UC aetiology:

These inflammatory cytokines are expressed at relatively higher levels in the intestinal tissues of patients with UC. However, the expression level of these cytokines depends on the disease phase and characteristics of patients [96]. TNF-α is a pleiotropic cytokine that regulates inflammation by inducing IL-1 and IL-6 production, fibroblast proliferation, adhesion molecule expression, and the initiation of an acute immune response [97]. TNF-α triggers gut inflammation by activating and proliferating immune cells, producing cytokines and chemokines, and degrading extracellular matrix [98]. As a result

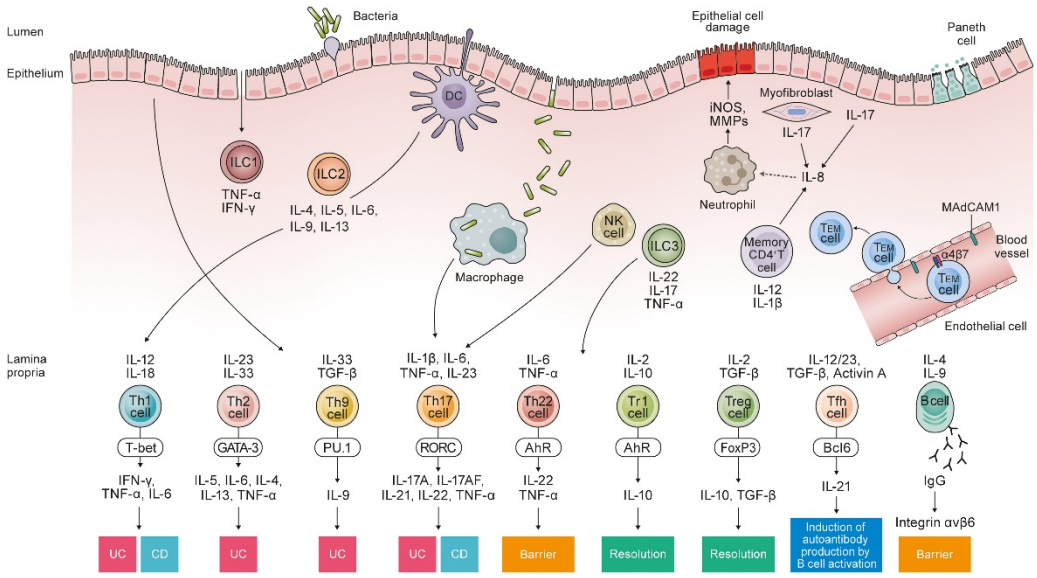


Figure 3. Immune cells and cytokines that contribute to the pathophysiology of UC [used with permission from [96]].

of elevated TNF-α levels, mucosal barrier function diminishes in UC patients with the inflammation [99].

There is currently no comprehensive treatment for UC; however, symptoms of inflammation can be decreased or prevented by achieving remission through therapies. However, the complex aetiology of UC may induce variation in responses to therapy. The key role of TNF in IBD pathogenesis is highlighted by the efficacy of anti-TNF monoclonal antibodies such as Infliximab, Adalimumab, and Golimumab. Anti-TNF therapies have significantly improved outcomes in UC patients [98].

Following antigen invasion in the intestinal mucosa, immune cell trafficking occurs, directing T-cells to the site of inflammation. T-cells adhere to the inflamed site through adhesion molecules such as chemokine receptor 9 (CCR9) and the integrins α4β1 and α4β7. The binding of these adhesion molecules promotes the recruitment and activation of immune cells, resulting in the initiation, propagation, and amplification of gut inflammation [100]. Blocking these adhesion molecules with leucocyte infiltrate is an attractive treatment strategy in IBD. This treatment enables the blocking of the integrin driven cellular signalling using anti-integrin antibodies [101,102].

The pathogenesis of UC is a partial consequence of signal defects driven by the dynamics of cells and cytokines in the signalling pathways. The nature of defects and how it influences disease phenotype varies significantly at patient level [103]. Pathogenic microbes and viruses also contribute, tending to induce a disruption in the signalling events leading to a signal cascade generating a dysregulated immune response. The defect in inflammatory pathways leads to the uninhibited release

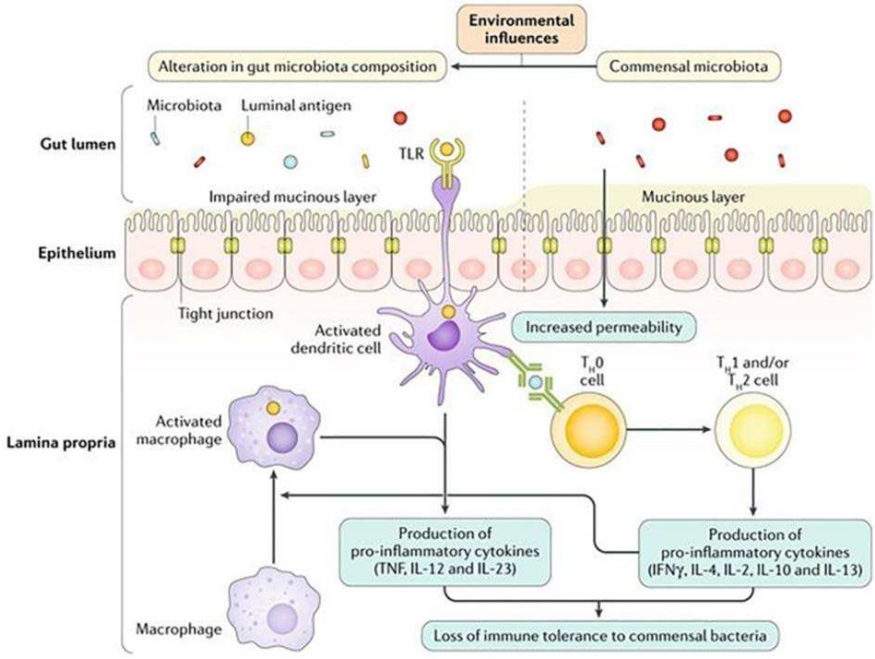


Figure 4. The inflammatory cascade in inflammatory bowel disease [used with permission from [105]]

of effector T cells resulting in a perforated barrier perpetuating inflammation. Major known signalling pathways include P38 MAP kinase (MAPK), Janus Protein Tyrosine Kinase/signal transducer and activator of transcription (JAK/STAT), Fas/FasL, PI3K/Akt, and NF-κB signalling pathways [104]. An overview of the inflammatory cascade in IBD is represented in (**Figure 4**) below:

For homeostasis in the gut, it is essential to have a controlled, tolerance system to prevent an antigen-derived immune response. When this balance of the intestinal immune system is disrupted, it expedites pathogenic attack, increased epithelial permeability of mucosa, and abrupt immune reaction via signalling pathways. Consequently, underlying mucosal tissue gets exposed to the luminal antigens via a leaky intestinal barrier.

Various immuno-inflammatory pathways in UC are associated with epithelial barrier function, antigen recognition, commensal microflora, dysregulated immunological responses, leukocyte recruitment, and genetic factors [103]. NF-κB is one of the major transcription regulators which controls the molecular network for various IBD-associated cellular functions and is also an attractive therapeutic target for IBD. NF-κB pathways are activated via stimuli e.g., bacteria, cytokines, viruses, and growth factors by TLR activation together with MAPK and TNF signal transduction [89,106]. In IBD, TLR/NF-κB signalling has a protective role in wound healing and tissue repair in the intestinal epithelial cell thereby re-establishing gut barrier functionality [107,108].

Moreover, cytokines and their signalling play a significant role in cell communication and perpetuation of UC resulting in mucosal injury, tissue damage, and triggering debilitating immune responses. Immuno-inflammatory effect of cytokine is exerted in cooperation with other related proteins by forming a network of cytokines instead of a single cytokine effect. The cytokine network comprises pro-inflammatory as well as anti-inflammatory cytokines which contribute to the up-and down-regulation of disease progression respectively [109]. The pro-inflammatory cytokines such as IL-1, TNF-α, IL-12, and IL-23 are NF-κB-dependent mediators found to be elevated in patients with IBD [109–111].

Highly conserved recognisable molecules called pathogen-associated molecular patterns (PAMP) are known to initiate the innate immune response when microbes invade intestinal epithelial cells (IEC). Toll-like receptors (TLRs) are the receptors that initiate downstream signalling of inflammatory pathways and are recognized for their critical involvement in mucosal homeostasis. The expression patterns of TLRs in the mucosal epithelial cells vary considerably among individuals [112]. TLR4 is important for the recognition of lipopolysaccharides (LPS) and activates the NF-κB signalling pathway in the regulation of pro-inflammatory cytokines such as ILs, interferons (IFN), and TNF-α [113]. TLR4 has shown an association with UC or/and CD [114,115], TLR2 and TLR4 have been used as treatment targets for IBD [116]. However, TLR3 and TLR5 are constitutively expressed by healthy intestinal cells. Signalling disorders of TLR3 and TLR4 inhibit IRF-3/Type 1 IFN/STAT/ISRE/IRF-1 pathway using different compounds [104].

The abnormalities in signalling pathways that influence disease phenotype can be traced back. Previous studies have shown that the immuno-inflammatory pathway in UC involves tissue damage driven by a complex, dynamic interplay between immune and non-immune cells [117]; cytokines have a pivotal role in this crosstalk. Therefore, simultaneous investigating of multiple pathways rather than single pathways may provide a comprehensive overview of cytokine-directed signalling defects [118] and identify other potential inflammatory mediators contributing to IBD pathogenesis [119,120].

1.3 Diagnosis and clinical presentation of UC

There is no single reference standard for the diagnosis of UC. Diagnosis of UC consists of various alternative procedures such as clinical presentation, endoscopic findings, histology, and exclusion of alternate diagnosis [40]. For the selection of appropriate treatment and prognosis for UC patients, it is crucial to characterise the extent and severity of inflammation. Lab testing mainly involves blood and

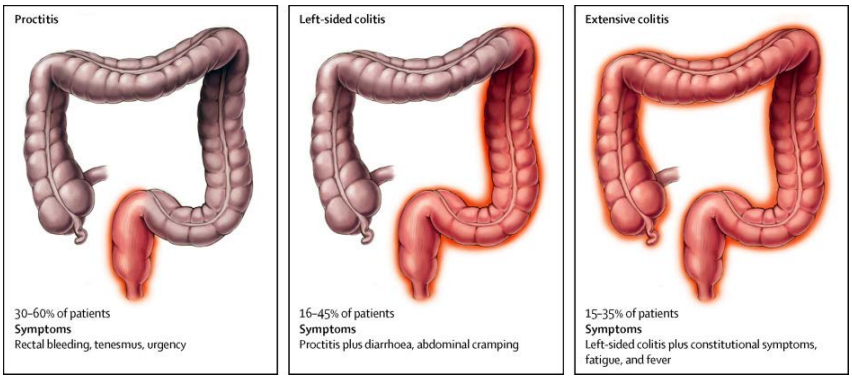


Figure 5. Ulcerative colitis phenotypes by Montreal Classification [used with permission from Mount Sinai health system [1]]

stool testing to confirm the status of inflammatory markers such as elevated erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), low ferritin, anaemia, hypoalbuminemia, faecal calprotectin, and lactoferrin. To examine the status of the inflammation in the colon, endoscopic procedures are used, and biopsies are obtained for further histological analysis. Endoscopic findings mainly include erythema, oedema, spontaneous bleeding, and ulcers [40].

For determining disease activity and clinical outcomes for UC patients, a scoring system has been developed by Baron *et al.*, [121] which divides UC classification into mild, moderate, and severe. Mayo score is a commonly used classification system measuring UC [122], which assesses the disease occurrence and severity. For histological feature assessment, Geboes Index [123] is mostly used. The Montreal classification system [124] classifies patients into four subgroups (**Figure 2**). Based on disease severity, the patient groups are classified into four patient subgroups: S0 (Clinical Remission), S1 (Mild UC), S2 (Moderate UC), and S3 (Severe UC) [40].

For UC activity assessment, there is a scale of invasiveness that consists of clinical, biochemical, endoscopic, and histological activity. UC patients achieve remission (symptom-free state) following treatment. There is no validated definition of remission. Clinical remission is simply defined as full resolution of symptoms which is further classified into endoscopic, histological, and mucosal remission/deep remission. Mucosal remission is an emerging objective in UC treatment [125,126] and is associated with improved long-term clinical remission with reduced risk of colectomy [127].

1.3.1 Treatment interventions

Ulcerative colitis is a chronic inflammation of the colon resulting in life-threatening complications that require lifelong therapy. While UC has no known complete cure and universally effective treatment, treatment strategies can significantly alleviate signs and symptoms of the disease and attain

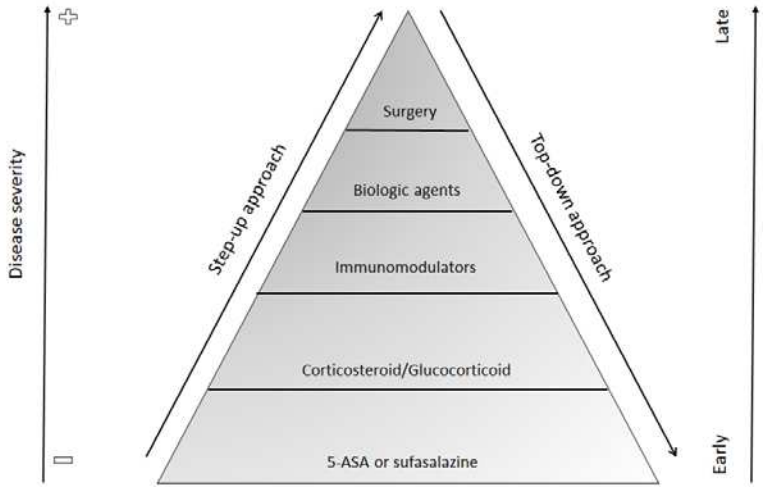


Figure 6. The therapeutic pyramid of inflammatory bowel disease [Adapted from [132]]

long-term remission. However, the definition of remission is highly debatable, and it varies in different contexts based on the resolution of clinical symptoms and mucosal healing [128–130]. The treatment of UC is a stepwise procedure in which therapy decisions are made based on the disease severity and clinical response of previous treatment [118,131]. Following (**Figure 6)** is the typical representation of the therapeutic pyramid of inflammatory bowel disease.

In UC, there are two major classifications of therapy based on their effects: induction and maintenance therapy. Induction therapy is aggressive, effective intervention given to a patient with UC to achieve remission quickly (6-8 weeks) by reducing the inflammation of the gut whereas maintenance therapy is a long-term intervention (>12 weeks) while gut inflammation is under control. Aminosalicylates (aka 5-ASAs, such as Mesalazine and Sulfasalazine), Steroids (such as Prednisolone and Budesonide), biologics (such as Infliximab, Adalimumab, Vedolizumab, Golimumab and Ustekinumab) are used as induction therapy. Maintenance therapy may be switched to a completely new type of drug if there is a relapse. There are various alternatives for maintenance therapy such as Aminosalicylates, Immunomodulators (such as Azathioprine, Mercaptopurine and Methotrexate), and targeted drugs.

There are different medications such as topical, systemic administration of drugs and surgery which are selectively chosen to treat the patient based on the extent of disease symptoms. In cases, when there are severe flare-ups during the disease course, immunomodulators and anti-tumour necrosis factor-alpha (anti-TNFα) inhibitors are used in combination or as monotherapy with a top-down approach (**Figure 6**) to avoid relapses. These targeted drugs are known as biologics and their mechanism of action is based on targeting TNF and an integrin, respectively [1,133]. These TNF

blocking biologics such as Infliximab, Adalimumab, and Golimumab are mainly effective for inducing remission in moderate-to-severe patients with UC [1,134,135]. Among anti-TNF alternatives, Infliximab is a frequently and widely used induction therapy for out-patients with severe UC [136]. The action mechanism of Infliximab is that antibodies bind to TNF and block the interactions with its receptors. Since TNF is a major inflammatory mediator, inhibiting the results of its effects in reduction of the inflammation [137]. However, it is still debatable that these molecules have an additional role in the drug mechanism than merely TNF blocking [134]. TNF has a dual role as a ligand as well as a receptor molecule, while in the soluble form it acts as a ligand whereas its precursor forms transmembrane TNF (mTNF), which plays a role as a receptor molecule that gets activated once in contact with TNF antibody [138].

The anti-integrin biologics are used as maintenance therapy for the moderate-to-severely active UC patients with non-response or loss of response to anti-TNFα therapy [139]. Both Vedolizumab (VDZ) and Etrolizumab (ETZ) are the next-generation, gut-targeted targeted drugs that are humanised monoclonal antibody that selectively blocks β7 subunit of heterodimeric integrin α4β7 [140]. Following (**Figure 7)** shows the action of targeted therapies such as anti-TNFα and anti-integrin respectively on the intestinal immune system in ulcerative colitis.

In the GEMINI-1 cohort study for UC, VDZ was more effective than placebo as induction and maintenance therapy [143]. ETZ has a dual action that selectively targets α4β7 and αEβ7 integrins. α4β7 integrin is a glycoprotein that B and T cells present on the surface of B and T cells [143]. α4β7 interacts with mucosal addressin-cell adhesion molecule 1(MAdCAM-1), which is preferentially expressed in the lamina propria of the gut tract and its associated lymphoid tissues. Anti-integrin therapy mainly targets α4β7-MadCam-1 interaction by blocking α4β7integrin, thereby facilitating both prevention of lymphocyte requirement to intestinal tissue and inhibiting their inflammatory effect on gut mucosa [144,145]. It was found that VDZ therapy has consistent efficacy in durable clinical remission, mucosal healing, and steroid-free remission were all significantly higher in the VDZ groups when compared with placebo regardless of previous anti-TNF therapy [146,147]. As of now, there is no reliable supporting study to guide the choice of individual targeted therapy over others as maintenance therapy in UC. However, a network meta-analysis study on all available targeted drugs such as Infliximab, Adalimumab, Golimumab, and Vedolizumab has found their higher efficacy over the placebo group for maintenance of remission and response [148]. Abrilumab and PF-00547659 are other anti-integrin drugs both administered subcutaneously, and their mode of action is based on blocking α4β7integrin-binding with MAdCAM with high affinity and selectivity [149,150].

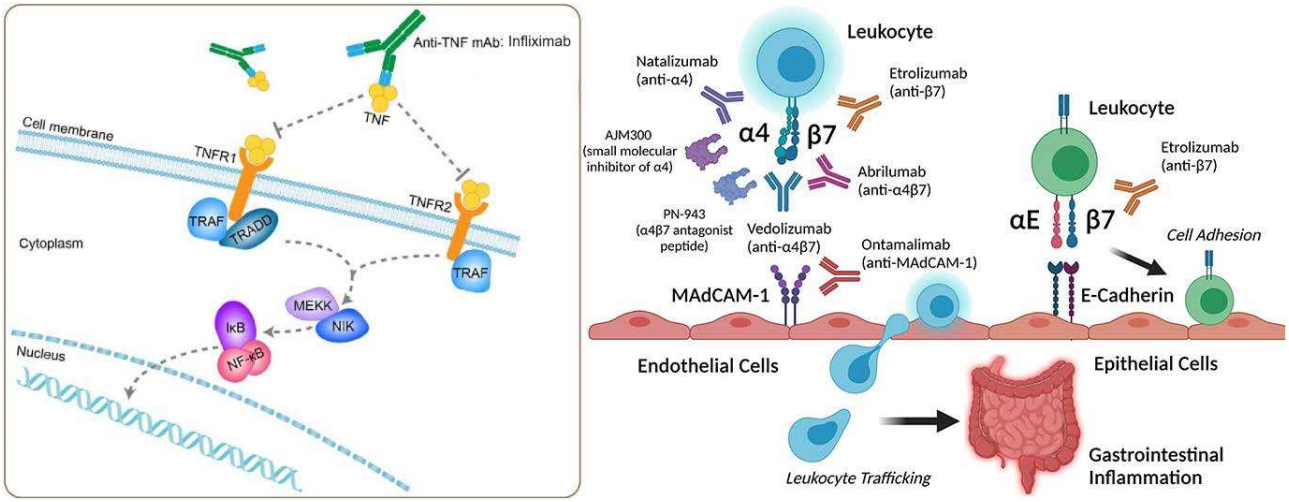


Figure 7. Summary of current and investigational targeted therapies such as anti-TNFα (left) and anti-integrin drug (right) of the intestinal immune system in ulcerative colitis [used with permission from creativebiolabs (left), adopted from [144] (right)]

Alternatives to anti-TNF and anti-integrin therapy are interleukin inhibitor (ILI) targeted drugs which selectively target IL-12 and IL-23, e.g., Ustekinumab, Risankizumab. These are monoclonal antibodies to the p40 (Ustekinumab) and p19 (Risankizumab) subunits of IL-12, IL13 (Ustekinumab) and IL-23 (Risankizumab) respectively. ILI targeted drug prevents binding of IL-12 and IL-23 to IL-12R-β1 receptor chain of IL-12 and IL-23 receptor complexes on the surface of natural killer (NK) and T cells, this inhibits the effect of IL-12 and IL-23 mediated responses to produce cytokines. For those patients with UC who have failed other available targeted therapies, a new class of therapy called small molecule drugs (SMDs) in the league of treatment of patients with moderate-to-severe UC is Tofacitinib [151,152]. Tofacitinib inhibits phosphorylation and JAK activation, which prevents phosphorylation of cytokine receptors. As a result, cytokine receptors cannot dock to initiate transduction and activation of STAT for gene transcription and production of cytokines [40].

The demand for targeted therapy has been dramatically increasing because of its efficacy in attaining favourable outcomes for IBD. A recent study has found that the elevated cost of therapy in the USA is due to the increasing use of targeted drugs in the last decade [153]. As a result, the high cost of targeted drugs imposes a significant burden on healthcare. Nevertheless, despite tremendous advances, about one-third of the patients with IBD are non-responsive to anti-TNFα and 10% lose their response to therapy each year. Moreover, current IBD medications cause infectious, neoplastic side effects (tumour growth) and other unfavourable side effects, therefore usually not prescribed for paediatric IBD patients. Among other emerging alternate IBD therapies, stem-cell therapy has a major advantage for its low immunogenicity profile, and lack of requirement for whole-body chemotherapy

following transplant. It is used to treat damage caused by chronic inflammation in IBD through varied mucosal immune responses [154]. Mesenchymal stem cells (MSCs) are used for their immunomodulatory capability which induces regulatory T-cell production for downregulating mucosal immune response to promote mucosal healing [155]. Another known treatment is called faecal microbial transplant (FMT) where a faecal microbiota sample is collected from individuals with healthy commensal flora and transplanted to the patient with IBD. It is believed that gut microbiota has a pivotal role in the pathogenesis of IBD [95,156]. However, multiple FMT is burdensome on the patient who requires multiple follow-ups for endoscopies and associated healthcare costs.

1.4 Biological networks and network-based approaches

The concept of networks encompasses a wide range of dynamic systems such as natural, social, and biological systems e.g., a cell is a complex network of chemicals connected by various chemical reactions. A cell consists of densely packed diverse macromolecules present in the interior to the extracellular matrix of the cell. The intra- and intermolecular interactions because of molecular crowding are known to cause a crucial role in cellular signalling networks. Traditionally, networks have been modelled as random graphs, however, later it is increasingly recognized that there are underlying robust, organised principles behind network topology and dynamics [157].

Albert-László Barabási[158] has coined the term 'network medicine' which emphasises the use of network-based methods for system-level understanding of disease mechanisms. These techniques seek to deduce network connections and the dynamics of information flow across biological systems to identify putative disease markers [159]. A biological system is a highly complex web-like structure comprising dynamically interacting components such as genes, proteins, and metabolites. A defined biological function is rarely linked to a single molecule; rather, it is the result of an interaction between constituents of the biological system. The interaction of numerous network components at various system-level gives a deep mechanistic knowledge of biological processes. Protein-protein interactions (PPIs) and gene regulatory networks are two examples of network types (GRNs) [160,161].

Unlike random networks, the biological system embraces the complex architecture which is often scale-free [163]. These networks are interchangeably termed as graphs containing nodes connected by links (edges). Nodes with multiple interactions within signalling networks are known as hub nodes, and hub nodes constitute network modules [164]. The notable feature of network hub nodes is their ability to hold the whole network together, although accounting only for 20% of the network nodes. Despite decades of research, the p53 (tumour suppressor) is a prototypical example of a hub protein

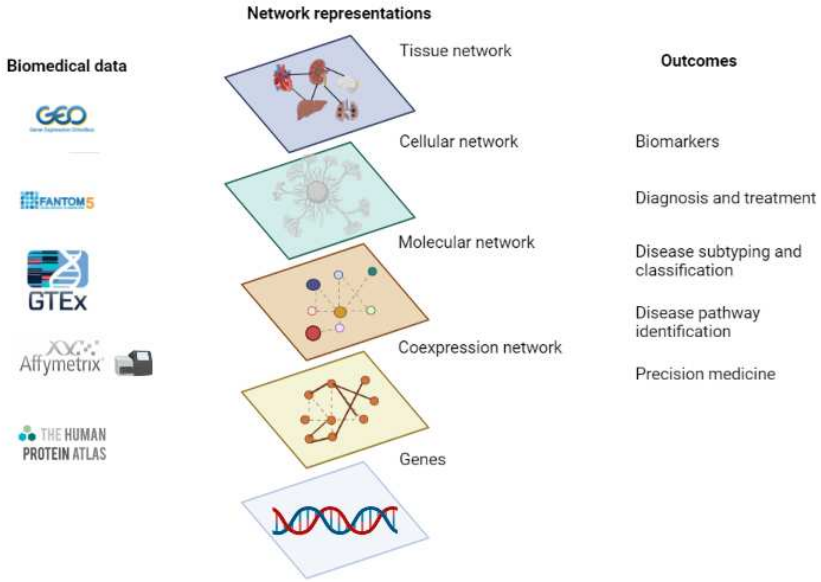


Figure 8. An overview of hierarchical biological network representations with different layers involving various biomedical data types from the genome, transcriptome, and proteome [162]

whose biological functions in the signalling cascade are continuously being discovered [165]. Each of the network modules in the network carries a highly specific biological function. Nodes that connect two or more modules are called bottleneck nodes and their primary function is to exchange information across network modules for carrying out biological subtasks [166]. Bottleneck nodes have high betweenness centrality (a measure of the number of the shortest paths that go through each node). The identification of network modules has been challenging [167] which has inspired researchers for the invention of methods such as random walk-based approaches [168,169].

Networks are a flexible, quantitative tool that provides a conceptual and intuitive framework for visualising and inferring intra- and intercellular molecular interactions. Biological networks are often employed as perceptual inference-making tools to describe, integrate, and analyse data. This framework is used to model the integration of multi-omics data, analysis of the cellular organisation, explore genotype-phenotype association, and the capturing the effect of perturbations on a complex intracellular network (**Figure 5**). The scale-free property of networks allows for the modelling of biological systems at all levels, from molecular to population-scale. [170].

Network-based approaches enable understanding of disease aetiology, identification of putative biomarkers, the discovery of drug targets, identifying the disease pathways, and subtypes of disease states. Network models are heavily dependent on large biological datasets for constructing models. The unprecedented development of high-throughput sequencing and rapid decline in sequencing costs provides impetus to produce a large scale of omics datasets e.g., transcriptomics, proteomics,

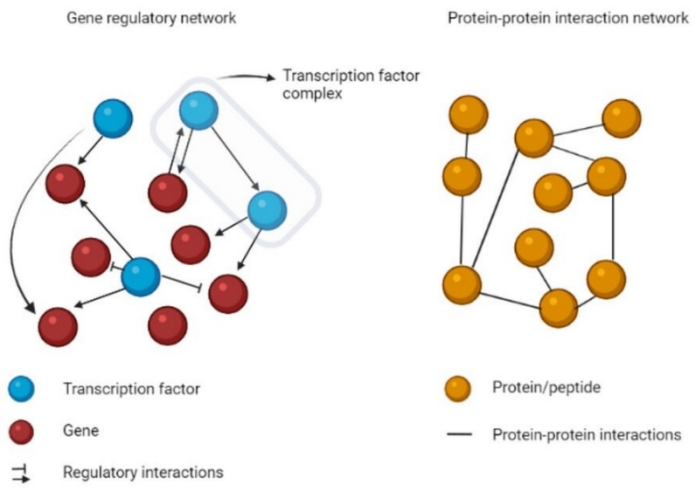


Figure 9. A gene regulatory network vs. a protein-protein interaction network. A gene's regulatory network nodes indicate genes or proteins, and the lines connecting them indicate regulatory interactions [left]. A protein-protein interaction network is composed of nodes that represent genes or proteins and lines that represent regulatory interactions [left]. Nodes in a protein-protein interaction network are always proteins, and the connecting lines indicate physical protein-protein interactions [right] [Adapted from [171]].

lipidomics, microbiomics, etc [172]. These multi-omics datasets facilitate the application of network-based approaches to decipher new insights into the molecular mechanism of diseases. On longitudinal patient data, network-based approaches may be utilised for temporal or dynamic network analysis, in which statistical dependencies of a network at distinct gene expression profiles could be examined to infer a temporal regulatory relationship [173]. To identify disease genes and their associated mechanism, network-based approaches are mainly classified into exploratory and analytical methods [174]. The exploratory method captures biological patterns due to the perturbation whereas analytical methods focus on identifying individual genes and disease-associated pathways.

The perturbation at hubs and bottlenecks across a biological network can be studied using network analysis. The use of network techniques in breast cancer research has seen significant progress over the last decade [175,176]. It has led to the discovery of translatable insights in the study of IBD as well as other complex heterogeneous disorders [177]. Unlike rare diseases caused by a single gene mutation, IBD is driven by complex interactions between different factors such as genetic, environmental, and microbial factors. These interactions span numerous cellular, genomic, epigenetic, and protein levels to manifest the disease phenotype (**Figure 5**). Because of disease heterogeneity among IBD patients, which is primarily influenced by diverse environmental exposures and distinct genetic polymorphisms, multiple pathogenic pathways are activated. Targeting diverse pathways with a one-size-fits-all therapeutic approach precludes effective treatment options toward precision medicine, a key technique for enhancing clinical outcomes with patient-tailored therapy [178–181].

Biological networks may be constructed in several different ways depending on the type of data. Protein-Protein Interaction (PPI) networks, gene regulatory networks, and metabolic networks are notable examples of these biological networks [182]. Using a single biological network has various constraints, such as the gene regulatory network, which requires gene expression measurement and transcriptome data. However, the drawback of adopting network techniques based on correlation measurements is their exclusive dependency on the correlation between genes connected in the network, and correlation does not necessarily imply causation. Some available resources in PPI are quite sparse, resulting in biases in network generation approaches. The following are the commonly used network-based approaches:

1.4.1 Protein-Protein Interaction (PPI) Networks

PPI networks are a mathematical representation of nodes (proteins/peptides) that are connected by edges (connections) in the cell (**Figure 9**). PPI networks provide information about physical interactions between proteins derived from experiments, and computational predictions [182]. The disease context-specific interactions can be incorporated to baseline PPI with the inclusion of an additional layer of a biomedical dataset. Different methodologies may be used to evaluate protein interactions, including experimental methods (yeast two-hybrid tests and affinity purification coupled mass spectrometry) and computational prediction methods (text-mining and machine learning) [183]. Using comprehensive databases like SignaLink3, Reactome, OmniPath, etc., a PPI network with directed PPIs may be utilised to construct intra- and intercellular signalling models. [184].

By integrating gene expression data that represents proteins with their transcripts in the PPI network, PPI network-based approaches may be contextualised to the disease-specific state. These network techniques have been widely used in IBD research to construct IBD-relevant PPI networks to identify novel IBD-relevant gene pathways represented in IBD gene-enriched modules [185]. A recent study used scRNAseq data from healthy, non-inflamed UC, and inflamed UC colonic biopsies to establish cell-type-specific PPI networks of intracellular and intercellular signalling in the colon in healthy and UC states [186].

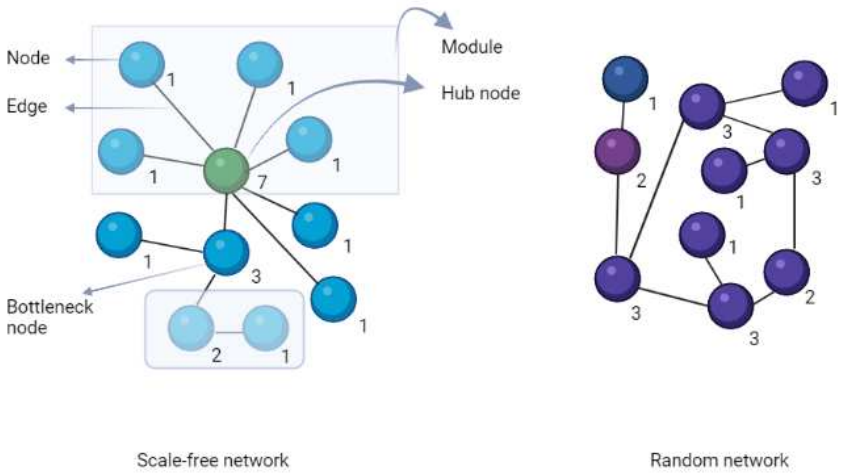


Figure 10. A scale-free network vs. a random network. The number below each node represents the degree (number of interactions) of that node. The majority of nodes in a scale-free network connect with only a few other nodes (blue) [left], whereas only a few interact with many other nodes and act as network hubs (pink). The nodes (blue) in a random network connection with a uniform probability, resulting in most nodes having the same number of connections [right]. [Adapted from [171]].

1.4.2 Gene Regulatory Networks (GRNs)

GRNs are mathematical models that represent the rationale underpinning gene regulatory behaviours. A GRN describes regulatory interaction which involves transcription factors (TFs) and their target genes. A simple GRN model comprises two key events: (1) an active TF binds to gene promoter; (2) binding of TF activates /suppresses the expression level of the target gene. By integrating these two events derived from high-throughput data and existing literature, the GRN model can be reconstructed [187]. GRN reconstruction has been recognized as a reliable technique to describe how cells assimilate biochemical stimuli and respond appropriately, triggering complex cellular programs involving a set of genes [188].

GRNs are a scale-free method for the temporal representation of biological networks. GRN models utilise gene expression data obtained under different experimental conditions and aid in deciphering the behavioural patterns of biological systems in response to perturbations. This allows the identification of important modulators, driver nodes, and subnetwork identification present in the network. GRN captures local network features such as degree and other centrality matrices that impact network neighbourhoods such as network hubs and core nodes. GRN can examine the regulatory pattern of transcription factor (TF) in which TF-target gene information is incorporated using ChIP-seq and/or ATAC-seq data. Because of the increased availability of high-throughput omics technologies, there is unprecedented potential for constructing a new GRN inference approach. (**Figure 11**).

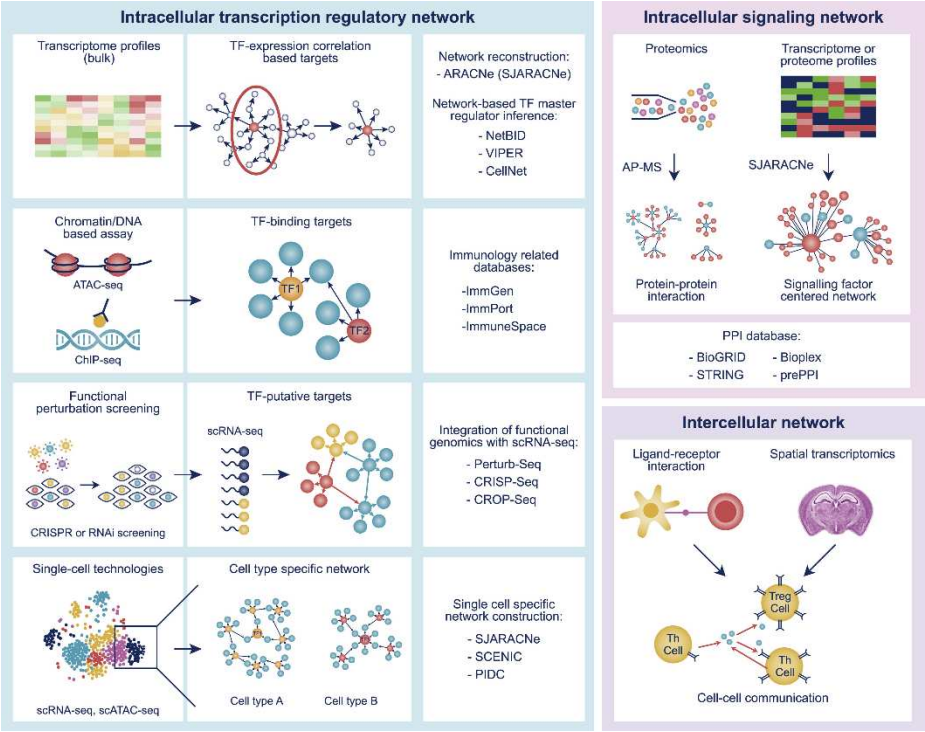


Figure 11. An Overview of Popular Omics Technologies and Network Inference Computational Methodologies [used with permission from [169]*]*

GRN may be used to obtain quantitative estimates of network components such as DNA, mRNA, TFs, and signalling proteins, as well as their interaction patterns. Various gene expression-based approaches for constructing GRNs and Gene Co-expression Networks (GCNs) employing multimodal genomic data have been developed [182]. GCNs are context-specific networks that are constructed by estimating gene expression correlations among related genes. GCNs and GRNs complement each other by sharing a common input requirement: gene expression data. The GRNs and GCNs models can be refined by including biological constraints data, such as potential transcription factor-target gene interactions. These models can make inferences about disease perturbations by integrating disease or tissue-specific information with a baseline prior network. Weighted Gene Co-expression Analysis (WGCNA), a popular GCN technique, estimates a correlation between genes and filters out weak connections to create a scale-free network [189]. ARACNE, C3Net, GENIE3, WGCNA, and other GRN techniques are examples [189–192]. ARACNE is used to reconstruct gene regulatory networks and signalling networks from large-scale gene expression data, as well as to solve network deconvolution challenges [190]. This method identifies direct transcriptional relationships while eliminating the majority of indirect interactions identified by co-expression methodologies. ARACNE and other comparable GRN techniques have a drawback

that is most apparent when assessing inferred GRN. This problem will be discussed in further depth in the discussion chapter.

GRNs can be used to construct condition-specific regulatory models that model transcription factor binding propensities instead of co-expression. This method facilitates defining the involvement of genes or modules in each disease state. These networks can provide useful insights into the highly dynamic nature of gene regulation, as well as the molecular underpinnings of observable phenotypes as a consequence of perturbations. The discovery of driver nodes and network-based biomarkers, such as the differential strength of interactions between TF and target genes in the context of disease, is accomplished by condition-specific GRNs. Passing Attributes between Networks for Data Assimilation (PANDA) [193] algorithm is an example of a condition-specific GRN that allows prior knowledge (PPI, GCN, and a network prior) to be integrated to uncover interaction patterns between transcription factors and target genes for differential targeting analysis. PANDA generates network-derived weighted edge scores that create self-consistent modules which highlight key biological processes. PANDA has been applied in cancer research to identify TFs and genes, as well as prognostic biomarkers and novel core modules of functionally relevant TFs [194,195]. LIONESS (Linear Interpolation to Obtain Network Estimates for Single Samples) [196] techniques have been designed to capture network variability in a population, and are mainly used for sample-specific network inference. LIONESS generates single-sample networks based on correlated gene expression to generate a comprehensive, weighted adjacency matrix.

Network-based biomarkers, typically utilise molecular pathways rather than individual genes, are feasible for multifaceted diseases and can be used for diagnostic, predictive, and prognostic purposes. By combining multiple types of networks, a molecular interaction network is constructed, which has an advantage as a network medicine tool over similarity metrics based solely on expression profiles [197]. This is attributed to the reason that each network type contains information linked to distinct biochemical characteristics for representing an individualized network's phenotype. Jostlin *et al.* discovered 73 unique and 163 IBD-associated genomic loci utilising GCNs and GRNs to evaluate the influence of IBD-associated loci on IBD pathogenesis in a meta-analysis study of 15 IBD GWAS. In this study, 211 co-expression IBD-enriched modules were discovered using WGCNA of gene expression data from diverse types of tissues [10]. Verstockt *et al.* estimated gene dysregulation at multiple phases of CD [198] using WGCNA-derived GCNs. Aschenbrenner *et al.* utilised GCNs [199] to investigate cytokine (IL23) signalling in a recent study on treatment-naïve paediatric CD and non-

inflamed controls. IL23 is a pro-inflammatory cytokine that has been linked to the pathogenesis of IBD.

1.4.3 Multi-layered network approach

Over the last decade, there has been an unprecedented generation of omics data, and it appears that their demand for biological applications is expanding over time. Databases such as OmniPath [200], SignaLink3 [201], and ConsensusPathDB [202] have been developed to construct multi-layered networks to incorporate multi-omics data [203]. These databases provide a comprehensive knowledge base of signalling processes such as interaction, regulation localization, disease, etc. and are valuable resources for tissue or compartment-specific pathway interaction analysis in signalling model construction. The various signalling processes can be integrated into a multi-layered signalling network using network modelling techniques to get valuable insights into disease pathogenesis. These techniques are performed by combining GCNs and GRNs using gene expression data [10]. In contrast, intracellular ligand-receptor signalling networks have recently been studied using single-cell RNAseq data, in which the PPI network and GCNs were integrated [204]. A cell signalling network is an intriguing form of multi-layered network, comprising two key components: upstream and downstream. A directed PPI network with intracellular signalling pathways occurs in the upstream components, whereas a GRN of transcription factor-target gene interactions occurs in the downstream components [166].

Although an example of a cell signalling network is still restricted to IBD, Brooks *et al.* recently created a pipeline "iSNP" that constructs a UC-specific cell signalling network using non-coding SNP data obtained from UC patients [205]. Using iSNP, a patient-specific cell signalling network was created in which patients were categorised into 4 groups, each with its own set of pathogenic pathways. The identification of unique regulatory effects of UC-associated non-coding SNPs in patient subgroups is achieved by an iSNP-derived cell signalling network.

1.4.4 Machine learning-based network approaches

Machine learning (ML) is a subset of artificial intelligence (AI) that allows machines to learn without being explicitly programmed. The learning process of machine learning-based models is automated and refined based on repeated experiences with fresh data provided to the computer. Machine learning-based approaches decode the complexities of highly complicated biological data using mathematical calculations and statistical assumptions. In contrast to other traditional network-based

techniques, the input to ML approaches are features, such as patient characteristics in disease, and these characteristics predict patterns in an unknown volume of data in the future. ML approaches have been extensively used in reconstructing gene regulatory networks [206], PPI networks [207], prognosis prediction [208] and recently used in the analysis of scRNAseq data [209,210]. ‘nnet’, a feed-forward neural network algorithm [211] has recently been used in biomarker discovery [212]. 'nnet' is a generic machine learning approach that uses regularization model parameters to control overfitting. However, deep learning approaches have two important limitations: it requires a significant quantity of data, which is not practical in all clinical trials, and it is prone to low interpretability for model parameters due to their complex model structure. Despite recent advances in network-based techniques for understanding complex UC disease mechanisms, there is still insufficient information to accomplish the goal of precision therapy in IBD [213]. As a result, significant progress in strong novel and integrative network techniques is necessary to broaden the scope of precision medicine in IBD research.

2 Aim of thesis

The patient at risk of developing non-response to the UC treatment is not well understood and approximately one-third of the patient is non-responsive to available targeted therapies. This suggests more complex disease pathogenesis is involved in the development of UC. Identification of patients with a high risk of non-response is a major clinical challenge. The overarching aim of the thesis is to contribute to the understanding of the individualized therapy response. Thus, contributing to the development of precision medicine. To achieve this objective, a quantitative, individualized, computational, framework is developed. Identifying putative complex biomarkers and predictive signalling pathways that can stratify patient subgroups within UC. This may enable patient-tailored therapy selection.

2.1 Hypothesis

The working hypothesis was that modelling patient-specific receptor to transcription factor (TF) network connectivity might provide a tool for more accurate prognosis and precision of treatment in the clinical management of UC.

2.2 Objective of thesis

Objective 1: Developing a model that identifies predictive biomarkers for stratifying anti-TNF treatment responder and non-responder patients with UC.

Objective 2: To identify a predictive signalling network for anti-integrin treatment response in UC.

Objective 3: Predict individual variability in pattern recognition receptor stimulation response.

A complex system that works is invariably found to have evolved from a simple system that works.

- John Gaule

3 Methodology

3.1 Diffusion model

Diffusion has been studied in a variety of scientific fields, including physical, chemical, and biological sciences. The diffusion of salts in water is a well-known phenomenon with analogies in other physical processes such as electrical conduction, heat conduction, and more. In his research [214], Fick described the relationship between the rate of diffusion and concentration gradient. A molar flux due to diffusion is proportional to the concentration gradient.

We opted to represent the biochemical processes that occur during signal transduction using a network that connects cell surface receptors to TFs in the nucleus. In signal transduction, receptor proteins transmit a signal to the adjacent proteins and the signal received by TF controls gene expression. To build a model, we chose a network in which IBD-relevant receptors are connected to TFs. This model represents the molecular events that occur during signal transduction. To quantify these molecular events, we calculate patient-specific network edge weights. The output of the model is the signal received by the TFs over time which is simplified as a time to reach 50% of the maximum signal at the TF (t_{50}). An obtained output contains a matrix of t_{50} score with receptors-TF pair per patient. This matrix can then be used for further statistical analysis or machine learning. We used t_{50} score for the separation of treatment response subgroups. However, in this simplified model the network contains only signal transduction nodes and does not include any negative regulation of transduction or feedback mechanism.

Consider a patient-specific signalling network in which nodes represent proteins such as cytokines, receptors, and kinases that build a signal transduction cascade. If a signal S, which corresponds to a chemical concentration in Fick's law, is applied to a node i the signal flux F along a network edge connecting node i to node j at time t is given by:

$$F(t)_{i \to j} = \left(S_i - S_j\right) * E_i * E_j$$

Where the edge connectivity weight, corresponding to the diffusion constant in Fick's law, is computed using the patient's normalized gene expression levels, E, of the genes coding for the proteins i and j. The signal present at each protein node i connected to J other protein nodes $j \epsilon 1.J$ is then updated at time t + 1 using the sum of all fluxes:

$$S(t+1)_i = S(t)_i + \sum_{j=1}^{J} F(t)_{i \to j}$$

The computation is initialized at time zero by setting all signal levels to zero and then placing one unit of signal on a starting receptor protein. The signal propagates through interconnected proteins throughout the network. To quantify the connectivity, we take the number of time steps to reach 50% of the maximum signal at the TF of interest (t_{50}). This methodology was implemented in R (4.1.0). Simulations were run for 2000 timesteps for all samples in each dataset, generating a new data matrix of t_{50} data with rows for each sample and a column for each receptor-TF pair. All the analysis conducted was implemented with packages in R version 3.6.3 and 4.2.0 (www.r-project.org). Software is available at https://github.com/Amy3100/receptor2tfDiffusion.

In paper III the model is expanded to include genes that can negatively regulate signalling. These nodes function identically but have their signal value (S) reset to zero between each time step. This allows adding nodes that remove signals, acting as signal sinks.

3.2 Network construction

A signalling network was constructed using UC-relevant receptors (IBD GWAS risk genes) and key TFs that describe the connectivity from receptors to key TFs. Gene Ontology (GO) terms were used to restrict the network to relevant signal transduction genes. The final signalling network contains nodes and edges that comprise receptors, signal transducer proteins, and TFs. The network interaction in pattern recognition receptor (PRR), a signalling pathway focusing on negative regulation was used [216]. The network interaction was validated in the STRING database and corresponding gene symbols were updated accordingly. Networks were created individually for TLR2, TLR4, TLR7, TLR9 and NOD2.

3.3 Subnetwork identification

A subnetwork is a modulization of a signalling network into a manageable size which improves the interpretability in the simplified form of the signalling network. It contains multiple TFs that cooperate for regulating target genes in the subnetwork [217]. Subnetworks are constructed by finding the shortest path (using gene t_{50} values) between the receptor and the TFs. It should be noted that a lower t_{50} for each gene suggests a larger accumulation of diffused signals. The AUC for the top 10 receptor-TF pairings is all greater than 0.78. (VDZ response vs non-response). The subnetworks were plotted using the igraph (https://igraph.org/) software. For each subnetwork, the sum of the gene t_{50}

values per branch (receptor to TF) for each sample was calculated. Then, the group branch sums were used to compare the patient sample groups.

3.4 Data sources

3.4.1 Gene Expression

Gene expression datasets were retrieved from a publicly available repository: Gene expression omnibus (GEO). These datasets include Affymetrix microarray e.g., Illumina HT12 arrays and highthroughput RNAseq of patient samples with UC (paper I). For Affymetrix microarray gene expression data, the pre-processing and normalisation (Robust Multi-Array Average) was performed. For RNAseq and Illumina microarray data, quantile normalisation was performed.

Paper-I: The clinical samples were obtained from the public repository Gene expression omnibus (GEO) with accession id: GSE16879. This dataset contains Affymetrix microarray gene expression data of UC patients along with non-IBD controls. This dataset contains the colonic biopsies before the week (W) 0 and after treatment (W6) with an anti-TNF drug (Infliximab). This dataset was used to develop the model. We used three other microarray gene expression datasets for model testing. To evaluate the performance of the model on relevant autoimmune diseases, we retrieved and merged two gene expression datasets of rheumatoid arthritis (RA). This dataset contains gene expression data obtained from paired synovial biopsy samples from the affected knee of RA patients. The samples were collected before (W0) and after treatment (W12) with anti-inflammatory drugs such as Tocilizumab (TCZ), Methotrexate (MTX) or Rituximab (RTX).

Paper-II: In this study, clinical samples of patients with UC were obtained from GEO (GSE73661) [218] containing Affymetrix gene expression data before and after treatment with Vedolizumab (VDZ) or Infliximab (IFX) targeted drugs. This dataset also contains non-IBD controls making a total of 178 samples with 120 colonic mucosal biopsies obtained at week (W) 0 (before treatment) and W6, W12, and W52 (after treatment). Only 112 patient samples treated with VDZ and 12 non-IBD control samples were selected for analysis. The baseline characteristics of patients are summarised in paper-II [218]. For the further testing performance of the model on another anti-integrin targeted drug (Etrolizumab), we retrieved a publicly available high-throughput RNAseq dataset from GEO (GSE72819). This dataset contains baseline colonic biopsies obtained from Etrolizumab (ETZ)-treated patients. A total of 70 samples: 58 non-responder and 12 responder UC patient samples were used.

Paper-III: The clinical samples were obtained from GEO: GSE80325 [219] and GSE137680) [199]. GSE80325 comprises an Illumina HumanHT-12 V4.0 expression array of healthy adult blood stimulated with 15 different stimuli with a total of 291 samples. The baseline characteristics of patients are summarised in paper-III [220]. The stimuli used for induction include agonists for surface TLRs, intracellular TLRs and NOD receptors, recombinant cytokines, and protein kinase C. Following ex-vivo infusion of 15 stimuli in adult healthy blood samples, transcriptional responses were observed after 6 hours. GSE137680 [199] contains *ex vivo* samples of peripheral blood mononuclear cells (PBMCs) from IBD patients stimulated with the PRR ligands LPS and MDP.

3.4.2 Knowledge-based databases

Following prior knowledge, databases were used for building a signalling network.

3.4.2.1 comPPI

For generating PPI for the network-based diffusion model, comPPI database was used. comPPI prevents biologically irrelevant interactions where two proteins share no subcellular localizations and predict compartment-specific biological functions [221]. This PPI database contains 8 subcellular localization datasets with a hierarchical structure of >1600 subcellular localizations, each containing confidence scores based on the likelihood of their interactions. The only interaction with a confidence score > 0.60 was used in constructing the network used in constructing the model.

3.4.2.2 Regulatory circuit

For the true representation of relevant inflammatory cells involved in UC, regulatorycircuit.org was used. It is a comprehensive resource of 394 cell type- and tissue-specific gene regulatory networks, derived from 37 genome-wide mapping studies. These networks contain genome-wide connectivity among transcription factors, enhancers, promoters, and genes [222]. Using this resource, regulatory motif binding site information using general immune cells (high-level network '14_immune_organs.txt') was obtained.

3.4.2.3 OmniPath

To select TF-target genes OmniPath was used. OmniPath database contains a signalling network, enzyme-PTM connections, protein complexes, protein annotations, and intercellular communication functions [200].

3.4.3 Gene ontology (GO) terms

To extract sub-network from PPI connecting key TFs to surface receptors which are known UC risk genes and signal transduction genes, GO terms annotated for the cytokine-related signalling were used. For example, GO:0002768 (cell surface receptor signalling pathway) or GO:0019221 (cytokine-mediated signalling pathway) were used in paper-I which extract the subnetwork connecting key TF to only cytokine receptors present in the PPI network. The resulting network was used for all UC datasets used in training and model testing. Gene annotation was conducted using the Bioconductor 'org.Hs.eg.db' package version 3.12.0 [10.18129/B9.bioc.org.Hs.eg.db]. GO enrichment analysis was performed using cluster Profiler, an R package [223]. In paper II, genes specific to cell surface receptors signalling pathway were selected using the GO term (GO:0007166) which limits the network to known receptors.

3.5 Reference methods

In papers, I and II, nnet, a deep learning-based method [212] was used. The parameters of nnet were optimised over a grid of the number of hidden nodes (size) and regularization (decay) parameters. In nnet, 10-fold cross-validation was performed with 20 iterations using average accuracy to select the final model. This process was repeated to identify the optimum model structure for each dataset. Prediction results in the testing datasets were evaluated by area under the curve (AUC) analysis. The remaining parameters were kept at their default settings in the nnet. The available reference biomarker discovery method 'Linear Interpolation to Obtain Network Estimates for Single Samples (LIONESS)' was used to stratify treatment responder and non-responder UC patients. LIONESS is an algorithm for quantifying the network edge weights for individual samples in a population [196]. LIONESS approach was developed as a function within the PANDA (Passing Attributes between Networks for Data Assimilation) regulatory network reconstruction framework.

3.6 Data analytics

3.6.1 limma

The statistical significance testing was performed for differentially expressed genes using linear modelling (limma) [224]. limma analysis was performed to identify significantly changed edge weights between inflamed vs non-inflamed samples and anti-TNF responder vs non-responder samples obtained from LIONESS results.

3.6.2 PCA and PLS

Exploratory data analysis was performed using principal component analysis (PCA) and partial least square (PLS) regression [225]. PCA of diffusion score of all receptor-TF pairs and all samples in the training dataset was performed while PLS on the same dataset was performed to find better stratification between treatment response subgroups.

3.6.3 pandaR

pandaR (Passing Attributes between Networks for Data Assimilation) [226], was used to create a gene regulatory network (GRN) with weighted edges between transcription factors (TFs) and gene targets regulated by TFs. The regulatory network connections were calculated with weighted edges on the testing dataset. To evaluate which TFs significantly contributed to the variation in gene expression, a null distribution regulation network edge weight was computed by randomizing the TF-target gene connections in the input. Then, the value of null distribution was used to calculate an empirical p-value for each TF.

3.7 Statistical tests

Wilcoxon test, student T-test and empirical p-value testing were used for evaluating statistical significance. For correcting for multiple testing, the method of Benjamini and Hochberg was used [227]. Procrustes rotation was performed to compare new information obtained from the diffusion model to gene expression data. Procrustes rotation was also used to evaluate if the t_{50} data reflected the gene expression levels of just a few highly connected or 'hub' genes or global gene expression changes due to variations in proliferation rate or infiltration of immune cells.

3.8 Cell deconvolution method

To estimate the global change in gene expression in all samples by immune cell infiltration or proliferation rate, cell deconvolution tool [228] was used. This method gives an estimate of the diffusion score reflected in the gene expression levels of just a few highly connected or 'hub' genes or global gene expression changes due to variations in proliferation rate or infiltration of immune cells.

4 Summary of results

4.1 Paper-I

Identifying anti-TNF response biomarkers in ulcerative colitis using a diffusion-based signalling model

Amrinder Singh, Endre Anderssen, Christopher G Fenton, Ruth H Paulssen

Bioinformatics Advances, Volume 1, Issue 1, 2021, vbab017

In this study, we developed a model for quantitative network analysis of the connection between receptors and transcription factors (TFs). We used key TF identified using pandaR and known cytokine and chemokine receptors. Network connectivities between immune-specific receptor-TF pairs were computed using network diffusion in non-IBD controls and UC patients treated with anti-TNF drugs. The network connectivities between 83 receptors and 58 TFs were calculated using the gene expression of each patient to generate a signalling network. Using this signalling network, the diffusion model has identified connectivity between well-known pro-inflammatory receptors such as TNFRSF11B, OSMR, and CCR2 to inflammation-related TFs. The diffusion model generates a feature space that contains novel information as compared to the gene expression feature space alone.

We found three top-scoring receptor-TF pairs with AUC > 0.91 in 20 responders and 36 non-responder patients, improving on reference methods such as LIONESS and nnet for predicting anti-TNF response in UC. The receptor-TF pairs identified might be considered potential prognostic biomarkers of anti-TNF treatment responders. The model was further tested in rheumatoid arthritis (RA) where it successfully discriminated against responder and non-responder patients to tocilizumab treatment with 86 samples.

4.2 Paper-II

Identifying predictive signalling networks for Vedolizumab response in ulcerative colitis

Amrinder Singh, Christopher G Fenton, Endre Anderssen, Ruth H Paulssen

International Journal of Colorectal Disease (2022)

This study aims to identify signalling pathways that may predict the efficacy of anti-integrin therapy in UC patients. We used a network-based diffusion model to highlight genes and their interactions in signalling pathways which may be predictive in response to the anti-integrin targeted drugs. We used 12 non-IBD controls and 41 UC patients treated with Vedolizumab therapy. Initially, we tested 4 Vedolizumab (VDZ) specific genes, known for the VDZ drug inhibition mechanism using a linear model and found a significant p-value ~ 0.03 and AUC=0.68 for VCAM1 out of four integrin-specific genes.

To compare the diffusion model with the reference method; nnet, a deep learning approach was used. nnet uses gene expression data to stratify treatment responder and non-responder patients. However, nnet did not outperform the network-based diffusion model. Using the diffusion model, we found AUC ~ 0.78-0.80 for the top-scoring predictor receptor-TF pairs such as FFAR2-NRF1, FFAR2-RELB, FFAR2-EGR1, and FFAR2-NFKB1 in VDZ treatment subgroups before and after treatment. To estimate the predictive ability of the diffusion model, we compared the diffusion feature score with the gene expression alone. We found that the diffusion feature score has relatively better predictive power for the stratifying VDZ patients' subgroups than gene expression alone. Receptor-TF pairs with FFAR2 receptor demonstrate the best predictor ability with an AUC score (~ 0.81) in comparison to gene expression of FFAR2.

To characterise individualized pathways for treatment response, we identified a subnetwork that highlights the network interactions connecting receptors and transcription factors (TFs). Patients with VDZ non-response exhibit quicker signalling as compared to the patients with VDZ response and controls. We also tested the predictive power of the diffusion model on an alternate anti-integrin targeted drug (Etrolizumab) and found an AUC of 0.72. The obtained sub-networks feature genes involved in cytokine and fatty acid signalling. These findings may contribute to the development of a promising clinical decision-making tool for the stratification of UC patients.

4.3 Paper-III

Modelling individual variability in pattern recognition receptor pathway response in IBD

Endre Anderssen, Amrinder Singh, Christopher G Fenton, Ruth H Paulssen

This study focuses on quantifying individual responses to a stimulus that may contribute to the pathological immune response involved in UC. To model this, we have analysed publicly available gene expression data before and after stimulation with pattern recognition receptor (PRR) agonists. Two datasets were used: the first dataset (a) contains 26 donors with unstimulated *ex vivo* blood samples and the same samples were stimulated with ligands to PRRs such as TLR -2, -4, -7, -9, and NOD2. The donors are normal controls and systemic juvenile idiopathic arthritis (sJIA) patients. The second dataset (b) contains 34 donors that contain *ex vivo* samples of peripheral blood mononuclear cells (PBMCs) from IBD patients stimulated with the PRR ligands lipopolysaccharide (LPS) and muramyl dipeptide (MDP). Canonical pathway models and network diffusion were used to predict an individual's transcriptional response using the gene expression profile before stimulation.

To identify the individual variability in the response to the same PRR signal, we used the diffusion model. We adopt a network analysis method to evaluate network connectivity between PRRs and transcription factors (TFs). The method uses input gene expression data to personalise the signalling network connecting PRRs to downstream TFs. These individualized network connectivities are then compared to activation levels for known TF target genes following PRR stimulation. A statistically significant correlation was found between the estimated network connectivities and activation of TF target genes. The correlation for extracellular PRRs is higher, at a shorter time of stimulation and for exterior membrane PRRs.

This is a novel method to quantify individualized PRR response based on gene expression profiles and network models. Improved understanding of an individual's response to PRR ligands may have implications for treatment choices and may open treatment opportunities using PRR ligands as alternatives or adjuvants to current treatment options.

Whether you can observe a thing or not depends on the theory that you use.

Albert Einstein (1879–1955)

5 Discussion

5.1 Selection of patient datasets with treatment response

Why is a prediction of patient treatment response important in UC? Each patient has unique environmental, immunological, and genetic systems, which may contribute to their variable therapy responses. Evidence of induction trials with targeted therapies in UC suggests that there is a low remission rate (20-30%) in UC patients because of patient-to-patient variability [6,229–231]. A poor response to the therapy incurs an enormous financial burden on health care and insurance providers and adds more to patient suffering. Current therapeutic strategies in UC classify patients based on clinical symptoms and then select appropriate therapy. At present, there are various treatment strategies for managing UC based on the disease severity and clinical response to previous treatment (**Figure 6**). A moderate-to-severe patient with UC who fails to respond to these targeted therapies may require surgical interventions. To select a better therapy alternative, it is therefore important to predict and stratify non-responder from the responder patient subgroup at the earliest phase of disease development.

In this project, we initially searched for the ex-vivo stimulus-based experimental dataset on human samples. Unfortunately, due to the unavailability of a stimulus-based dataset, we selected the available IBD dataset on anti-TNFα therapy response to develop our diffusion model in paper I. When the dataset on stimulus-induced IBD response became publicly available in 2021, we used that dataset focusing on pattern recognition receptor (PRR) signalling to predict individualized variability in PRR pathway response in IBD. This dataset contains individual stimuli and stimuli-in-combinations to predict induction response. Using that dataset in the paper-III, we have studied how strongly transcription factors are activated by PRR in each patient. In papers I and II, we used patient datasets with targeted therapies such as Infliximab (anti-TNFα), and Vedolizumab (anti-integrin). The key consideration for selecting these datasets is that the drugs used in these datasets target a single molecule e.g., TNFα, integrin. Modelling the therapy efficacy may be arduous if the model involves a wide range of target molecules as compared to a single molecule. Additionally, both targeted therapies are often used for moderate-to-severe UC patients. Therefore, the prediction of the differential response following these targeted therapies is important.

5.2 Why study individual differences in UC using a network-based approach?

A key clinical goal of precision medicine is to implement effective individualized therapy. A complex and multifactorial disease such as UC [232]. Due to its complex aetiology, it is challenging to identify disease mechanisms that lead to UC pathogenesis. Inter-individual differences in how UC patients respond to a particular drug are dependent on their genetics, environmental exposure such as gut microbiota, and the proportion of cell types contributing to the disease pathogenesis [233,234]. Each individual has a difference in their regulatory wiring which influence the efficacy of the drug e.g., the drug target in individual A activates the key driver gene whereas the interaction between the drug target and key driver gene is absent in individual B. Therefore, there is a difference in drug efficacy in individual A vs individual B [235]. UC involves dysregulated immuno-inflammatory signalling pathways; it is, therefore, important to understand how perturbations in these pathways contribute to heterogeneous treatment outcomes. This could be because the signalling pathways involved in therapy response vary among patients despite the same diagnosis and induction therapy in UC [236] [**paper-II**]. Capturing individualized variability at the system level may enable the improvement of the identification of patient-specific pathomechanisms [237,238].

The disease aetiology of UC cannot merely be described by single-gene defects; instead, approaches that capture the combined effect of multiple genes at the system level are required [239]. Network-based approaches capture disease modules which contain genes interacting at the system level which otherwise is neglected in single-gene studies [163]. However, gaining novel insights into biological phenomenon through a network is not often trivial. Networks are perceived to be convoluted structures and quantifying the network connectivity using a biologically intuitive approach enables the extraction of subnetworks. These subnetworks are relatively easier to interpret as compared to the 'furball-like' structure of the original network. Many existing network-based methods generate network topologies that are difficult to interpret [240]. These methods also do not consider cell compartmentalization parameters as prior biological knowledge [241].

Using the t_{50} score obtained from the diffusion model, we estimated the network connectivities which helps to extract visually interpretable subnetworks. These subnetworks capture the most relevant interaction details and reduce the complexity of the convoluted network. These subnetworks characterise the individuals based on the network connectivities rather than merely generating a convoluted network visualisation. Our approach differs considerably from the previously published GRN-based approach in which sample-specific network connectivities were estimated [242]. This is because the simple methodology of our model that incorporates gene expression profiles with

networks does focus on signalling not on gene regulation. Our findings show that our model consistently outperformed existing GRN-based approaches to stratify the treatment response subpopulation of UC patients.

5.3 From network furball into patient information

The representation of a biological network is complex and densely connected as a furball e.g., PPI, which contains extensive information about interactions between proteins. However, it lacks context-specific information about in which cells, tissues, and specific mechanisms the interactions occur [240]. Most available databases contain context-free information which might lead to the false interpretation of biological results. On the other hand, existing biological networks are incomplete e.g., PPI data is yet 80 % complete despite the availability of high-throughput datasets [170]. Another example is signalling networks which have been compared to dumpling soup due to the lack of a formal schematic representation of multiple layers of complexity involved in signalling processes [240]. With around 700 publicly available pathway and molecular interaction repositories, it is still tedious to develop a context-specific patient-centric quantitative approach for clinical applications.

This abstruse structure of biological networks prevents interpretation of mechanistic interactions occurring in disease and normal state [240]. To extract patient information, patient-centric methods are required to link data to the context of interest such as patient disease, tissue, and cell type. GRN approaches are used to infer networks. However, the use of these networks beyond visualization is rarely demonstrated. Thus, GRN methods are of limited utility to model the therapy response of patients. We have tested the pandaR-LIONESS method which is a patient-centric method for generating sample-specific networks. However, it did not perform well to generate sample-specific networks to predict the therapy response of UC patients. Therefore, we utilised a diffusion model which uses the network-based approach and generates network connectivities to predict therapy response in UC patients.

The normal immune system functioning is determined by the interplay of pro-inflammatory and anti-inflammatory immunological activity, whilst excessive or inadequate activation of triggering factors is evident in many autoimmune diseases [243]. It is well studied that UC is induced by over-stimulation of pro-inflammatory signalling pathways and disruption of regulatory systems [103]. It is important to understand individual response and feedback variability that contribute to inflammation and the lack of resolution to inflammation over time. We tested the patient-wise variability (discussed in paper-III) using stimulation-induced PRR pathway response data. To date, there is no quantitative model

for quantifying the effect of stimuli in ex-vivo stimulation experiments to test inter-patient excitability using the PRR signalling pathway. We found that there is patient-specific variability in the PRR signalling pathway. This individualized variability was previously depicted in the study of anti-TNFα and anti-integrin therapy responses using a diffusion model to quantify patient-specific network connectivity [paper-I and II]. We hypothesised that there are alternate potential targets that might be associated with a potential compensatory inflammatory mechanism in a UC patient with non-response to anti-TNFα and anti-integrin therapy. As a result of other unidentified potential targets, there is non-response to targeted therapy such as anti-TNFα and anti-integrin agents.

5.4 Current network-based approaches and their drawbacks

Cellular processes are not merely governed by individual genes but rather by networks of interacting genes [163]. The network-based approaches have increasingly been used in the identification of disease genes and drug targets [244,245]. These approaches are gene-centric and generate a network-based inference based on regulatory relationships among genes using gene expression data. These biological networks can be used as an input to generate novel testable hypotheses in a patient-tailored manner.

Given numerous potential interactions comprising ~20,000 genes in humans, network-based approaches are perceived as an excellent tool to restrict merely context-relevant interactions. However, it is important to understand that network generated using these approaches is not the final result; these networks should also be utilised for modelling complex biological phenomena, catalysing actionable biological discoveries (biomarkers), and personalized disease interventions [170,246]. The assessment of inferred networks obtained from these network-based methods is important and challenging because the network is high-dimensional structured objects that can be used as input to model biological mechanisms. There are challenges while assessing the quality of inferred biological networks: i) identifying true interactions and ii) choice of statistical measures to quantitatively assess the quality of the network. Usually, published literature describing true interaction information can be utilised to identify true interactions [170]. However, there is a scarcity of approaches that focus on individualized network interactions.

Many network-based methods utilise a network to identify a cluster of genes present in a network. These methods are often used to generate metagene that is further used to relate to disease/treatment outcomes. The advantage of these methods is their ability to reduce random noise by averaging across multiple genes in a network which might allow a certain extent of model improvement by acting as a limited OR filter; however, network structures are no longer used after the initial clustering of genes

in a network. The network-based approaches e.g., Algorithm for the Reconstruction of Gene Regulatory Networks (ARACNE), Weighted Gene Correlation Network Analysis (WGCNA), C3Net, BC3Net, are based on gene correlation-based inference [247]. These tools use gene expression data to construct a network disregarding the gene function such as receptors, transcription factors or signal transducers. WGCNA describes transcriptional interaction with high confidence in the biological network using a weighted correlation pattern between gene expression profiles. For the construction of gene regulatory networks, these algorithms use indirect interactions inferred by co-expression methods to identify modules representing gene clusters (modules) of highly correlated genes. Several correlation-based network methods generate a network-based module containing a cluster of genes which no longer uses the network structure for the application e.g., biomarker discovery, and prediction of therapy response. These genes are then converted into scores or disease modules, which are linear transforms of the original gene expression data. Whereas methods like LPLS and diffusion model use the network and contribute to generating novel hypotheses which can be used to test individualized differences in disease subgroups. For instance, LPLS summarises the background information on regressor variables for improved predictive performance and the diffusion model uses a network which generates t_{50} score to estimate network connectivities in a patient-specific manner as compared to merely using original gene expression data (see paper II for more details).

The most sought application of network-based approaches for personalized medicine is to extract networks for individual samples in a patient population. The key characteristic of these methods is to model personalized networks for each sample in a dataset and capture heterogeneity in a population. LIONESS [242] is an example of this method which uses a network. LIONESS is used as a wrapper function on PANDA (Passing Attributes between Networks for Data Assimilation), regulatory network reconstruction framework [248]. However, the LIONESS-pandaR method could not predict biomarkers for therapy response in UC, which may be due to a large number of calculated network edge weights and consequential penalization for multiple testing. Thus, LIONESS-pandaR lacks the sensitivity to capture more subtle differences between subclasses of patient samples with therapy response as compared to large differences between normal and inflamed tissue. Whereas methods such as Partial Least Squares regression (LPLSR) [249] and diffusion-based signalling model [250] identify subtle differences between patient subgroups in disease populations. LPLSR algorithms function by the addition of an extra data matrix which summarises the background information on regressor variables for improved predictive performance. Whereas the diffusion model uses the network

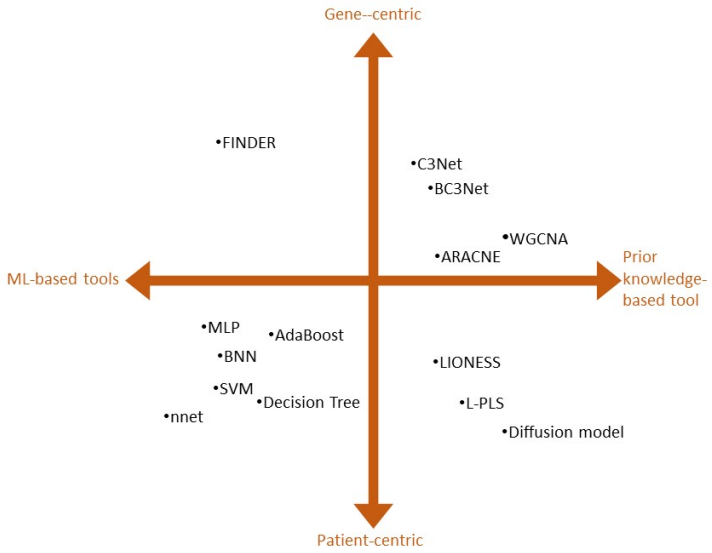


Figure 12. Illustration of various existing approaches/tools divided into four axes: Gene-centric to patient-centric and ML-based to prior knowledge-based approaches. FINDER [251]; C3Net [191]; BC3Net [252]; ARACNE [253]; WGCNA [189]; LIONESS [242]; L-PLS [249]; Diffusion Model [250], nnet [212]; SVM [254]; Decision Tree [255], BNN [256] ; AdaBoost [257]; MLP [258].

structure to define new variables for the prediction of patient subgroups. We believe that the potential use of these approaches will increase since the knowledge of genetic networks is rapidly expanding.

Machine learning (ML) approaches have emerged as an indispensable tool with rapidly increasing clinical applications [259–261] and demonstrated a better performance than that of other peer methods, particularly in the modelling of image-based data [212]. A study by Ghosh *et al.* has used seven deep learning models: Artificial neural network (ANN), convoluted neural network (CNN), gated recurrent unit (GRU), long short-term memory (LSTM), multi-layer perceptron (MLP) neural network, probabilistic neural network (PNN), and recurrent neural network (RNN) to study whole blood cell (WBC) dataset and found that the best performance among all DL models, achieving a performance accuracy over 99% for the early detection of Breast Cancer [262].

ML approaches are typically classified as prediction/modelling tools which are used for analysing the high-throughput heterogeneous data; predicting and identifying intricate patterns, and treatment outcomes [263,264]. However, ML-based approaches e.g., neural networks do not explain their prediction in a manner that humans can understand. This lack of explainability in clinical interpretation may lead to misleading findings. It is the widespread belief that there is a trade-off between accuracy and interpretability, however, it is not often the case that more complex models tend to have better accuracy and predictive performance [265]. Rather than developing a model which is interpretable, there has been an explosion of new 'explainable ML' where another model is developed to explain the first

'black box' model. In this study, we used nnet [212], a deep learning model that uses gene expression information alone as compared to diffusion model [250], a biologically interpretable model that utilises prior biological knowledge to predict treatment response in IBD.

5.5 Methodological consideration

Statistical evaluation of a new method to suggest complex biomarkers is crucial for the interpretation of the clinical data. The lack of statistical measures allows statistical bias, overfitting, and potential false positives to go undetected. In our study predicting the efficacy of the treatment outcome in UC, the method was evaluated to assess the predictive accuracy when stratifying the patient subgroups. Thus, statistical evaluation of the model can contribute to successful complex biomarker discovery. However, the procedure of clinical evaluation of complex biomarkers is arduous. Following are the important statistical considerations for clinical data analysis.

5.5.1 Feature Selection

While handling high-throughput datasets, it is important to reduce dimensionality to extract the relevant features of interest and remove irrelevant or redundant variables and identify confounding variables. There are several techniques for dimensionality reduction and some of them involve feature selection. The process of feature selection consists of the selection of the number of input variables that improve the model's performance and reduce the computational cost of predictive modelling.

In the era of big data, a myriad of data from high-throughput-omics approaches is increasingly being generated where the sample size is dwarfed by the number of available features. Clinical characteristics such as gender, ethnicity, age, and gene expression level are the features of the samples. A selection of these different clinical features defines a feature space. A predictive biomarker can be identified by training an algorithm to map out how these features can classify target groups. That trained algorithm can be further used to classify the new samples using the selected features [266].

A broad range of machine learning algorithms and statistical tests are available for feature selection. Examples include stepwise regression and support vector machines; decision tree-based models, e.g., classification, regression trees and random forest; and regularization-based models such as LASSO or ridge regression [177]. To date, no single method has been developed that has performed better than all other methods in all contexts. Therefore, applying combinations of methods may enhance the probability of obtaining improved model performance. An appropriate validation approach prevents

overfitting, and false assumptions about expected outcomes, thus increasing the generalizability of a model on future datasets. To prevent overfitting, regularization is a useful method that works on the principle of adding a penalty to the parameters of the model. This penalty is applied over a coefficient that is multiple products of the predictor variable. Lasso regularization enables the shrinking of the coefficients to zero and removes those reductant and irrelevant features from the model.

5.5.2 Overfitting

Various computational approaches are being developed that use multiple variables for predicting clinical parameters e.g., patient response to treatment [267,268]. However, these approaches are susceptible to overfitting. One of the major reasons for the occurrence of overfitting is the number of predictor variables is much larger than the number of patient samples. Overfitting leads to biassed estimates of classification accuracy (e.g., between patients responding or not responding to a given treatment). Overfitting can also be an issue of concern even for low-dimensional data, where the relationship between predictor variable and outcome is weak. Importantly, low sample numbers may also result in overfitting e.g., a model trained on a small dataset is more likely to generate patterns that do not exist in the new test dataset. To counter that issue, the classification accuracy of candidate predictor variables should be estimated using cross-validation, resampling or test sets for validation [269].

5.5.3 Multiple testing penalty

While conducting statistical validation in genomics or related fields, random events occur that cause variables to appear significant even if they have no real predictive ability. To avoid these issues, statistical confidence measures either analytical or empirical can be used to reduce the odds of being tricked by random variability. Confidence measures such as p-values, false discovery rates (FDR) or q-values are probability measures to reject null hypotheses in a given biological experiment. However, p-values can be problematic in a high-throughput experiment that generates many scores, making the p-value inappropriate to determine significance.

To prevent the chances of obtaining a false-positive variable (type I error/α) while performing multiple testing, Bonferroni correction (BC) is used. In BC, the p-value (P) for each test must be less than or greater than its alpha (probability of a Type I error) divided by the number of tests performed on the dependent variable ($P \leq \alpha / n$). It is the simplest and most conservative approach when multiple pairwise comparisons on a single set of data are involved simultaneously. In contrast to the Tukey

test (testing on a high number of means), BC has more power when a small number of comparisons are involved. Moreover, FDR estimates can be computed from p-values using Benjamini-Hochberg (BH) procedure. BH correction procedure relies on p-values to be uniformly distributed under the null hypothesis, implying the p-value of 5% of the sorted list should be ~ 0.05 to obtain an estimate of accurate FDRs [227]. Several methods have been developed for achieving improved and accurate FDR estimates for multiple testing correction [270]. Nevertheless, multiple testing corrections tend to introduce bogus significant results after performing thousands of tests [271]. Therefore, it is important to formulate experiment strategies accordingly to minimise the need for multiple tests required for a set of hypotheses [272]. Only a few biomarkers overcome statistical testing and clinical validation for successful translational use. The clinical therapeutic value of the biomarker depends on its predictive strength associated with disease characteristics.

5.5.4 Statistical power

The statistical power is the probability of correctly rejecting the null hypothesis. Power calculations help to detect true effect size by considering the number of samples and events in an experiment. Power estimation allows estimating a minimum number of samples required given desired significance level, effect size, and statistical power. The higher the statistical power in an experiment, the lower the probability of Type II error. Low statistical power may result in invalid conclusions about experiment results.

In all, using a panel of multiple candidate biomarkers can be useful to attain improved performance than a single biomarker. Multiple biomarkers which are in a non-dichotomized state provide optimal information for model development and clinical validation [273]. The choice of biomarker is challenging and depends on the sample size and clinical question. Involving multiple variable selection such as shrinkage reduces overfitting and increases the likelihood of validation.

5.6 Potential clinical implications

The clinically relevant question is how early “non-response” can be predicted with an acceptable level of reliability. Currently, it is feasible to conduct high-throughput sequencing within a week. How this would work in clinical practice is dependent on a variety of economical and organisational variables that vary between healthcare systems. However, rapidly developing technology and quicker turnaround times will certainly become achievable in the future. This is important to reduce the overall time of the experiment because therapy decisions need to be made within a few days.

Therefore, it is important to have a robust and reliable predictive tool that provides results in clinical routine.

We propose to reduce the risk of overfitting as compared to pure machine learning methods, using the diffusion model and identifying relevant genes by focusing on modelling gene combinations that connect extracellular signal receptors to transcription factors through PPI. The diffusion model can be used to generate testable hypotheses toward a better understanding of the treatment of the patients more effectively in UC and other autoimmune diseases such as RA, psoriasis, and asthma.

5.7 Future perspectives

We have used the UC patient datasets with frequently used targeted therapies such as anti-TNFα and anti-integrin treatment. We would like to implement our model to study therapy response outcomes on inhibitor-based therapies such as JAK-inhibitor targeted therapy. We assert that the predictive performance of our model might improve by incorporating the inducer, promotor, and inhibitor molecules in a newer version of the diffusion model.

We also would like to incorporate cell-and tissue context-specific data using single-cell transcriptomics and spatial transcriptomics to construct a context-specific network model. To obtain a broader picture of the pathogenesis of UC, omics analysis would be useful to estimate the network connectivity of patient's profiles with genomics, proteomics, and metabolism data. We are also interested in studying the influence of gut microbiota on heterogenous therapy response. Our findings have revealed the receptor-TF pairs and signalling pathways which can be validated as a prognostic biomarker to stratify subgroups of patients with UC. These future perspectives may open avenues to investigate therapy response outcomes in an unprecedented manner.

References

1. Ungaro, R., Mehandru, S., Allen, P. B., Peyrin-Biroulet, L. & Colombel, J. F. Ulcerative colitis. *Lancet* **389**, 1756–1770 (2017).

2. Griffel, L. H. & Das, K. M. Ulcerative colitis: pathogenesis, diagnosis, and current treatment. *J. Assoc. Acad. Minor. Physicians Off. Publ. Assoc. Acad. Minor. Physicians* **7**, 63–69 (1996).

3. Guindi, M. & Riddell, R. H. Indeterminate colitis. *J. Clin. Pathol.* **57**, 1233–1244 (2004).

4. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307–317 (2011).

5. Ordás, I., Eckmann, L., Talamini, M., Baumgart, D. C. & Sandborn, W. J. Ulcerative colitis. *Lancet* **380**, 1606–1619 (2012).

6. Jairath, V. *et al.* Placebo response and remission rates in randomised trials of induction and maintenance therapy for ulcerative colitis. *Cochrane database Syst. Rev.* **9**, CD011572 (2017).

7. Pineton de Chambrun, G. *et al.* The treatment of refractory ulcerative colitis. *Best Pract. Res. Clin. Gastroenterol.* **32**–**33**, 49–57 (2018).

8. Chen, A.-Y. & Oz, H. S. Rapid Induction and Maintenance of Remission in Refractory Ulcerative Colitis with Ustekinumab. *Diseases (Basel, Switzerland)* vol. 7 (2019).

9. Zhang, Y. Z. & Li, Y. Y. Inflammatory bowel disease: Pathogenesis. *World J. Gastroenterol.* **20**, 91–99 (2014).

10. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).

11. Porter, R. J., Kalla, R. & Ho, G.-T. Ulcerative colitis: Recent advances in the understanding of disease pathogenesis. *F1000Research* **9**, (2020).

12. Zammarchi, I. *et al.* Elderly-onset vs adult-onset ulcerative colitis: a different natural history? *BMC Gastroenterol.* **20**, 147 (2020).

13. Kobayashi, T. *et al.* Ulcerative colitis. *Nat. Rev. Dis. Prim.* **6**, 74 (2020).

14. Dignass, A. *et al.* Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 1: definitions and diagnosis. *J. Crohns. Colitis* **6**, 965–990 (2012).

15. Vavricka, S. R. *et al.* Frequency and risk factors for extraintestinal manifestations in the Swiss inflammatory bowel disease cohort. *Am. J. Gastroenterol.* **106**, 110–119 (2011).

16. Vavricka, S. R. *et al.* Chronological Order of Appearance of Extraintestinal Manifestations Relative to the Time of IBD Diagnosis in the Swiss Inflammatory Bowel Disease Cohort. *Inflamm. Bowel Dis.* **21**, 1794–1800 (2015).

17. Stolfi, C., Maresca, C., Monteleone, G. & Laudisi, F. Implication of Intestinal Barrier Dysfunction in Gut Dysbiosis and Diseases. *Biomedicines* **10**, (2022).

18. Monsén, U., Bernell, O., Johansson, C. & Hellers, G. Prevalence of inflammatory bowel disease

among relatives of patients with Crohn's disease. *Scand. J. Gastroenterol.* **26**, 302–306 (1991).

19. Childers, R. E. *et al.* Family history of inflammatory bowel disease among patients with ulcerative colitis: a systematic review and meta-analysis. *J. Crohns. Colitis* **8**, 1480–1497 (2014).

20. Appleyard, C. B., Hernández, G. & Rios-Bedoya, C. F. Basic epidemiology of inflammatory bowel disease in Puerto Rico. *Inflamm. Bowel Dis.* **10**, 106–111 (2004).

21. Lakatos, L. *et al.* Striking elevation in incidence and prevalence of inflammatory bowel disease in a province of western Hungary between 1977-2001. *World J. Gastroenterol.* **10**, 404–409 (2004).

22. Betteridge, J. D., Armbruster, S. P., Maydonovitch, C. & Veerappan, G. R. Inflammatory bowel disease prevalence by age, gender, race, and geographic location in the U.S. military health care population. *Inflamm. Bowel Dis.* **19**, 1421–1427 (2013).

23. Ng, S. C. *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet (London, England)* **390**, 2769–2778 (2017).

24. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *lancet. Gastroenterol. Hepatol.* **5**, 17–30 (2020).

25. Burisch, J., Jess, T., Martinato, M. & Lakatos, P. L. The burden of inflammatory bowel disease in Europe. *J. Crohn's Colitis* **7**, 322–337 (2013).

26. Walker, D. G. *et al.* Differences in inflammatory bowel disease phenotype between South Asians and Northern Europeans living in North West London, UK. *Am. J. Gastroenterol.* **106**, 1281–1289 (2011).

27. Beaugerie, L., Langholz, E., Nyboe-Andersen, N., Pigneur, B. & Sokol, H. Differences in epidemiological features between ulcerative colitis and Crohn's disease: The early life-programmed versus late dysbiosis hypothesis. *Med. Hypotheses* **115**, 19–21 (2018).

28. Loftus Jr, E. V *et al.* Ulcerative colitis in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival. *Gut* **46**, 336–343 (2000).

29. Loftus, E. V. J. *et al.* Crohn's disease in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival. *Gastroenterology* **114**, 1161–1168 (1998).

30. Farrokhyar, F., Swarbrick, E. T. & Irvine, E. J. A critical review of epidemiological studies in inflammatory bowel disease. *Scand. J. Gastroenterol.* **36**, 2–15 (2001).

31. Cosnes, J., Gower-Rousseau, C., Seksik, P. & Cortot, A. Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology* **140**, 1785–1794 (2011).

32. Souza, M. H. L. P. *et al.* [Trends in the occurrence (1980-1999) and clinical features of Crohn's disease and ulcerative colitis in a university hospital in southeastern Brazil]. *Arq. Gastroenterol.* **39**, 98–105 (2002).

33. Stowe, S. P. *et al.* An epidemiologic study of inflammatory bowel disease in Rochester, New York. Hospital incidence. *Gastroenterology* **98**, 104–110 (1990).

34. Kotze, P. G. *et al.* Review of the epidemiology and burden of ulcerative colitis in Latin America. *Therap. Adv. Gastroenterol.* **13**, 1–14 (2020).

35. Prideaux, L., Kamm, M. A., De Cruz, P. P., Chan, F. K. L. & Ng, S. C. Inflammatory bowel disease in Asia: a systematic review. *J. Gastroenterol. Hepatol.* **27**, 1266–1280 (2012).

36. Jess, T., Gamborg, M., Munkholm, P. & Sørensen, T. I. A. Overall and cause-specific mortality in ulcerative colitis: meta-analysis of population-based inception cohort studies. *Am. J. Gastroenterol.* **102**, 609–617 (2007).

37. Selinger, C. P. *et al.* Cause-specific mortality and 30-year relative survival of Crohn's disease and ulcerative colitis. *Inflamm. Bowel Dis.* **19**, 1880–1888 (2013).

38. Danese, S., Sans, M. & Fiocchi, C. Inflammatory bowel disease: the role of environmental factors. *Autoimmun. Rev.* **3**, 394–400 (2004).

39. Kim, D. H. & Cheon, J. H. Pathogenesis of Inflammatory Bowel Disease and Recent Advances in Biologic Therapies. *Immune Netw.* **17**, 25–40 (2017).

40. Gajendran, M. *et al.* A comprehensive review and update on ulcerative colitis(). *Dis. Mon.* **65**, 100851 (2019).

41. Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E. & Järnerot, G. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* **124**, 1767–1773 (2003).

42. Orholm, M., Binder, V., Sørensen, T. I., Rasmussen, L. P. & Kyvik, K. O. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand. J. Gastroenterol.* **35**, 1075–1081 (2000).

43. Loddo, I. & Romano, C. Inflammatory bowel disease: Genetics, epigenetics, and pathogenesis. *Front. Immunol.* **6**, 551 (2015).

44. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246–252 (2011).

45. Duerr, R. H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).

46. Ntunzwenimana, J. C. *et al.* Functional screen of inflammatory bowel disease genes reveals key epithelial functions. *Genome Med.* **13**, 181 (2021).

47. Darsigny, M. *et al.* Loss of hepatocyte-nuclear-factor-4alpha affects colonic ion transport and causes chronic inflammation resembling inflammatory bowel disease in mice. *PLoS One* **4**, e7609 (2009).

48. Silverberg, M. S. *et al.* Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat. Genet.* **41**, 216–220 (2009).

49. Bouma, G. *et al.* Genetic markers in clinically well defined patients with ulcerative colitis (UC). *Clin. Exp. Immunol.* **115**, 294–300 (1999).

50. De La Concha, E. G. *et al.* Positive and negative associations of distinct HLA-DR2 subtypes with ulcerative colitis (UC). *Clin. Exp. Immunol.* **108**, 392–395 (1997).

51. Barrett, J. C. *et al.* Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat. Genet.* **41**, 1330–1334 (2009).

52. Wheeler, J. M. *et al.* Hypermethylation of the promoter region of the E-cadherin gene (CDH1) in sporadic and ulcerative colitis associated colorectal cancer. *Gut* **48**, 367–371 (2001).

53. Zeng, Z., Mukherjee, A. & Zhang, H. From Genetics to Epigenetics, Roles of Epigenetics in Inflammatory Bowel Disease. *Front. Genet.* **10**, 1017 (2019).

54. Taman, H. *et al.* Transcriptomic landscape of treatment-naïve ulcerative colitis. *J. Crohn's Colitis* **12**, 327–336 (2018).

55. Kang, K. *et al.* A Genome-Wide Methylation Approach Identifies a New Hypermethylated Gene Panel in Ulcerative Colitis. *Int. J. Mol. Sci.* **17**, (2016).

56. Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504 (2009).

57. Celiberto, L. S. *et al.* Inflammatory bowel disease and immunonutrition: novel therapeutic approaches through modulation of diet and the gut microbiome. *Immunology* **155**, 36–52 (2018).

58. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).

59. Ortizo, R. *et al.* Exposure to oral contraceptives increases the risk for development of inflammatory bowel disease: a meta-analysis of case-controlled and cohort studies. *Eur. J. Gastroenterol. Hepatol.* **29**, 1064–1070 (2017).

60. Piovani, D. *et al.* Environmental Risk Factors for Inflammatory Bowel Diseases: An Umbrella Review of Meta-analyses. *Gastroenterology* **157**, 647-659.e4 (2019).

61. Vedamurthy, A. & Ananthakrishnan, A. N. Influence of Environmental Factors in the Development and Outcomes of Inflammatory Bowel Disease. *Gastroenterol. Hepatol. (N. Y).* **15**, 72–82 (2019).

62. Strachan, D. P. Hay fever, hygiene, and household size. *BMJ* **299**, 1259–1260 (1989).

63. Biedermann, L. *et al.* Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. *PLoS One* **8**, e59260 (2013).

64. Biedermann, L. *et al.* Smoking cessation alters intestinal microbiota: insights from quantitative investigations on human fecal samples using FISH. *Inflamm. Bowel Dis.* **20**, 1496–1501 (2014).

65. Allais, L. *et al.* Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. *Environ. Microbiol.* **18**, 1352–1363 (2016).

66. Lakatos, P. L., Szamosi, T. & Lakatos, L. Smoking in inflammatory bowel diseases: good, bad or ugly? *World journal of gastroenterology* vol. 13 6134–6139 (2007).

67. Cosnes, J. Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. *Best Pract. Res. Clin. Gastroenterol.* **18**, 481–496 (2004).

68. Kooij, I. A., Sahami, S., Meijer, S. L., Buskens, C. J. & Te Velde, A. A. The immunology of the vermiform appendix: a review of the literature. *Clin. Exp. Immunol.* **186**, 1–9 (2016).

69. Shaw, S. Y., Blanchard, J. F. & Bernstein, C. N. Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease. *Am. J. Gastroenterol.* **105**, 2687–2692 (2010).

70. Barclay, A. R. *et al.* Systematic review: the role of breastfeeding in the development of pediatric inflammatory bowel disease. *J. Pediatr.* **155**, 421–426 (2009).

71. Mawdsley, J. E. & Rampton, D. S. The role of psychological stress in inflammatory bowel disease. *Neuroimmunomodulation* **13**, 327–336 (2006).

72. Bonaz, B. L. & Bernstein, C. N. Brain-gut interactions in inflammatory bowel disease. *Gastroenterology* **144**, 36–49 (2013).

73. Jones, P. D. *et al.* Exercise decreases risk of future active disease in patients with inflammatory bowel disease in remission. *Inflamm. Bowel Dis.* **21**, 1063–1071 (2015).

74. Helander, H. F. & Fändriks, L. Surface area of the digestive tract - revisited. *Scand. J. Gastroenterol.* **49**, 681–689 (2014).

75. McGhee, J. R. & Fujihashi, K. Inside the mucosal immune system. *PLoS Biol.* **10**, e1001397 (2012).

76. Sartor, R. B. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **3**, 390–407 (2006).

77. Van Klinken, B. J., Van der Wal, J. W., Einerhand, A. W., Büller, H. A. & Dekker, J. Sulphation and secretion of the predominant secretory human colonic mucin MUC2 in ulcerative colitis. *Gut* **44**, 387–393 (1999).

78. Heller, F. *et al.* Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. *Gastroenterology* **129**, 550–564 (2005).

79. Kempski, J., Brockmann, L., Gagliani, N. & Huber, S. T(H)17 Cell and Epithelial Cell Crosstalk during Inflammatory Bowel Disease and Carcinogenesis. *Front. Immunol.* **8**, 1373 (2017).

80. Heller, F., Fromm, A., Gitter, A. H., Mankertz, J. & Schulzke, J.-D. Epithelial apoptosis is a prominent feature of the epithelial barrier disturbance in intestinal inflammation: effect of pro-inflammatory interleukin-13 on epithelial cell function. *Mucosal Immunol.* **1 Suppl 1**, S58-61 (2008).

81. Abraham, C. & Cho, J. H. Inflammatory bowel disease. *N. Engl. J. Med.* **361**, 2066–2078 (2009).

82. Niess, J. H. *et al.* CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science* **307**, 254–258 (2005).

83. Hart, A. L. *et al.* Characteristics of intestinal dendritic cells in inflammatory bowel diseases. *Gastroenterology* **129**, 50–65 (2005).

84. Cario, E. & Podolsky, D. K. Differential alteration in intestinal epithelial cell expression of toll-like receptor 3 (TLR3) and TLR4 in inflammatory bowel disease. *Infect. Immun.* **68**, 7010–7017 (2000).

85. Vamadevan, A. S. *et al.* Regulation of Toll-like receptor 4-associated MD-2 in intestinal epithelial cells: a comprehensive analysis. *Innate Immun.* **16**, 93–103 (2010).

86. Shen, X. *et al.* The Toll-like receptor 4 D299G and T399I polymorphisms are associated with Crohn's disease and ulcerative colitis: a meta-analysis. *Digestion* **81**, 69–77 (2010).

87. Zhang, F. X. *et al.* Bacterial lipopolysaccharide activates nuclear factor-kappaB through interleukin-1 signaling mediators in cultured human dermal endothelial cells and mononuclear phagocytes. *J. Biol. Chem.* **274**, 7611–7614 (1999).

88. Rogler, G. *et al.* Nuclear factor kappaB is activated in macrophages and epithelial cells of inflamed intestinal mucosa. *Gastroenterology* **115**, 357–369 (1998).

89. Spehlmann, M. E. & Eckmann, L. Nuclear factor-kappa B in intestinal protection and destruction. *Curr. Opin. Gastroenterol.* **25**, 92–99 (2009).

90. Eckmann, L. *et al.* Opposing functions of IKKbeta during acute and chronic intestinal inflammation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15058–15063 (2008).

91. Matsuda, R. *et al.* Quantitive cytokine mRNA expression profiles in the colonic mucosa of patients with steroid naïve ulcerative colitis during active and quiescent disease. *Inflamm. Bowel Dis.* **15**,

328–334 (2009).

92. Panés, J. & Granger, D. N. Leukocyte-endothelial cell interactions: molecular mechanisms and implications in gastrointestinal disease. *Gastroenterology* **114**, 1066–1090 (1998).

93. Hemmings, H. C. & Girault, J.-A. Chapter 3 - Cell signaling. in (eds. Hemmings, H. C. & Hopkins, P. M. B. T.-F. of A. (Second E.) 31–50 (Mosby, 2006). doi:https://doi.org/10.1016/B978-0-323-03707-5.50009-7.

94. Liu, T., Zhang, L., Joo, D. & Sun, S.-C. NF-κB signaling in inflammation. *Signal Transduct. Target. Ther.* **2**, 17023 (2017).

95. Maloy, K. J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474**, 298–306 (2011).

96. Nakase, H., Sato, N., Mizuno, N. & Ikawa, Y. The influence of cytokines on the complex pathology of ulcerative colitis. *Autoimmun. Rev.* **21**, 103017 (2022).

97. Osawa, Y. *et al.* TNF-alpha-induced sphingosine 1-phosphate inhibits apoptosis through a phosphatidylinositol 3-kinase/Akt pathway in human hepatocytes. *J. Immunol.* **167**, 173–180 (2001).

98. Digby-Bell, J. L., Atreya, R., Monteleone, G. & Powell, N. Interrogating host immunity to predict treatment response in inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 9–20 (2020).

99. Breese, E. J. *et al.* Tumor necrosis factor alpha-producing cells in the intestinal mucosa of children with inflammatory bowel disease. *Gastroenterology* **106**, 1455–1466 (1994).

100. Imam, T., Park, S., Kaplan, M. H. & Olson, M. R. Effector T Helper Cell Subsets in Inflammatory Bowel Diseases. *Front. Immunol.* **9**, 1212 (2018).

101. Danese, S. New therapies for inflammatory bowel disease: from the bench to the bedside. *Gut* **61**, 918–932 (2012).

102. Pedersen, J., Coskun, M., Soendergaard, C., Salem, M. & Nielsen, O. H. Inflammatory pathways of importance for management of inflammatory bowel disease. *World J. Gastroenterol.* **20**, 64–77 (2014).

103. Tatiya-Aphiradee, N., Chatuphonprasert, W. & Jarukamjorn, K. Immune response and inflammatory pathway of ulcerative colitis. *J. Basic Clin. Physiol. Pharmacol.* **30**, 1–10 (2019).

104. Wei, J. & Feng, J. Signaling Pathways Associated with Inflammatory Bowel Disease. *Recent Pat. Inflamm. Allergy Drug Discov.* **4**, 105–117 (2010).

105. Hazel, K. & O'Connor, A. Emerging treatments for inflammatory bowel disease. *Ther. Adv. Chronic Dis.* **11**, 2040622319899297 (2020).

106. Cario, E. & Podolsky, D. K. Toll-like receptor signaling and its relevance to intestinal inflammation. *Ann. N. Y. Acad. Sci.* **1072**, 332–338 (2006).

107. Kluwe, J., Mencin, A. & Schwabe, R. F. Toll-like receptors, wound healing, and carcinogenesis. *J. Mol. Med. (Berl).* **87**, 125–138 (2009).

108. Lu, Y., Li, X., Liu, S., Zhang, Y. & Zhang, D. Toll-like Receptors and Inflammatory Bowel Disease. *Front. Immunol.* **9**, 72 (2018).

109. Tatiya-Aphiradee, N., Chatuphonprasert, W. & Jarukamjorn, K. Immune response and inflammatory

pathway of ulcerative colitis. *J. Basic Clin. Physiol. Pharmacol.* **30**, 1–10 (2018).

110. Pizarro, T. T. & Cominelli, F. Cytokine therapy for Crohn's disease: advances in translational research. *Annu. Rev. Med.* **58**, 433–444 (2007).

111. Elson, C. O. *et al.* Monoclonal anti-interleukin 23 reverses active colitis in a T cell-mediated model in mice. *Gastroenterology* **132**, 2359–2370 (2007).

112. Maynard, C. L. & Weaver, C. T. Intestinal effector T cells in health and disease. *Immunity* **31**, 389–400 (2009).

113. Guijarro-Muñoz, I., Compte, M., Álvarez-Cienfuegos, A., Álvarez-Vallina, L. & Sanz, L. Lipopolysaccharide activates Toll-like receptor 4 (TLR4)-mediated NF-κB signaling pathway and proinflammatory response in human pericytes. *J. Biol. Chem.* **289**, 2457–2468 (2014).

114. Török, H.-P., Glas, J., Tonenchi, L., Mussack, T. & Folwaczny, C. Polymorphisms of the lipopolysaccharide-signaling complex in inflammatory bowel disease: association of a mutation in the Toll-like receptor 4 gene with ulcerative colitis. *Clin. Immunol.* **112**, 85–91 (2004).

115. De Jager, P. L. *et al.* The role of the Toll receptor pathway in susceptibility to inflammatory bowel diseases. *Genes Immun.* **8**, 387–397 (2007).

116. Söderholm, J. D. *et al.* Augmented increase in tight junction permeability by luminal stimuli in the non-inflamed ileum of Crohn's disease. *Gut* **50**, 307–313 (2002).

117. Curciarello, R., Canziani, K. E., Docena, G. H. & Muglia, C. I. Contribution of Non-immune Cells to Activation and Modulation of the Intestinal Inflammation. *Front. Immunol.* **10**, (2019).

118. Privitera, G. *et al.* Novel trends with biologics in inflammatory bowel disease: sequential and combined approaches. *Therap. Adv. Gastroenterol.* **14**, 17562848211006668 (2021).

119. Wei, J. & Feng, J. Signaling pathways associated with inflammatory bowel disease. *Recent Pat. Inflamm. Allergy Drug Discov.* **4**, 105–117 (2010).

120. Pedersen, J., Coskun, M., Soendergaard, C., Salem, M. & Nielsen, O. H. Inflammatory pathways of importance for management of inflammatory bowel disease. *World J. Gastroenterol.* **20**, 64–77 (2014).

121. BARON, J. H., CONNELL, A. M. & LENNARD-JONES, J. E. VARIATION BETWEEN OBSERVERS IN DESCRIBING MUCOSAL APPEARANCES IN PROCTOCOLITIS. *Br. Med. J.* **1**, 89–92 (1964).

122. Schroeder, K. W., Tremaine, W. J. & Ilstrup, D. M. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N. Engl. J. Med.* **317**, 1625–1629 (1987).

123. Geboes, K. *et al.* A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut* **47**, 404–409 (2000).

124. Satsangi, J., Silverberg, M. S., Vermeire, S. & Colombel, J.-F. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* **55**, 749–753 (2006).

125. Rogler, G., Vavricka, S., Schoepfer, A. & Lakatos, P. L. Mucosal healing and deep remission: what does it mean? *World J. Gastroenterol.* **19**, 7552–7560 (2013).

126. Boal Carvalho, P. & Cotter, J. Mucosal Healing in Ulcerative Colitis: A Comprehensive Review. *Drugs* **77**, 159–173 (2017).

127. Shah, S. C., Colombel, J.-F., Sands, B. E. & Narula, N. Mucosal Healing Is Associated With Improved Long-term Outcomes of Patients With Ulcerative Colitis: A Systematic Review and Meta-analysis. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **14**, 1245-1255.e8 (2016).

128. Jharap, B. *et al.* Randomised clinical study: discrepancies between patient-reported outcomes and endoscopic appearance in moderate to severe ulcerative colitis. *Aliment. Pharmacol. Ther.* **42**, 1082–1092 (2015).

129. Falvey, J. D. *et al.* Disease activity assessment in IBD: clinical indices and biomarkers fail to predict endoscopic remission. *Inflamm. Bowel Dis.* **21**, 824–831 (2015).

130. Regueiro, M. *et al.* Physician assessment of ulcerative colitis activity correlates poorly with endoscopic disease activity. *Inflamm. Bowel Dis.* **17**, 1008–1014 (2011).

131. Nakase, H. *et al.* Evidence-based clinical practice guidelines for inflammatory bowel disease 2020. *J. Gastroenterol.* **56**, 489–526 (2021).

132. Aloi, M., Nuti, F., Stronati, L. & Cucchiara, S. Advances in the medical management of paediatric IBD. *Nat. Rev. Gastroenterol. Hepatol.* **11**, 99–108 (2014).

133. Rawla, P., Sunkara, T. & Raj, J. P. Role of biologics and biosimilars in inflammatory bowel disease: current trends and future perspectives. *J. Inflamm. Res.* **11**, 215–226 (2018).

134. Billmeier, U., Dieterich, W., Neurath, M. F. & Atreya, R. Molecular mechanism of action of anti-tumor necrosis factor antibodies in inflammatory bowel diseases. *World J. Gastroenterol.* **22**, 9300–9313 (2016).

135. Reinisch, W. *et al.* Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. *Gut* **60**, 780–787 (2011).

136. Rutgeerts, P. *et al.* Infliximab for induction and maintenance therapy for ulcerative colitis. *N. Engl. J. Med.* **353**, 2462–2476 (2005).

137. Levin, A. D., Wildenberg, M. E. & van den Brink, G. R. Mechanism of Action of Anti-TNF Therapy in Inflammatory Bowel Disease. *J. Crohns. Colitis* **10**, 989–997 (2016).

138. Pallai, A. *et al.* Transmembrane TNF-α Reverse Signaling Inhibits Lipopolysaccharide-Induced Proinflammatory Cytokine Formation in Macrophages by Inducing TGF-β: Therapeutic Implications. *J. Immunol.* **196**, 1146–1157 (2016).

139. Danese, S. *et al.* Etrolizumab versus infliximab for the treatment of moderately to severely active ulcerative colitis (GARDENIA): a randomised, double-blind, double-dummy, phase 3 study. *lancet. Gastroenterol. Hepatol.* **7**, 118–127 (2022).

140. Rutgeerts, P. J. *et al.* A randomised phase I study of etrolizumab (rhuMAb β7) in moderate to severe ulcerative colitis. *Gut* **62**, 1122–1130 (2013).

141. Scaldaferri, F. & Pecere, S. Emerging Mechanisms of Action and Loss of Response to Infliximab in Ibd: A Broader Picture. *Biochem. Pharmacol. Open Access* **05**, (2015).

142. Gubatan, J. *et al.* Anti-Integrins for the Treatment of Inflammatory Bowel Disease: Current Evidence and Perspectives. *Clin. Exp. Gastroenterol.* **14**, 333–342 (2021).

143. Feagan, B. G. *et al.* Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N. Engl. J. Med.* **369**, 699–710 (2013).

144. Melde, M. *et al.* α4β7 integrin-dependent adhesion of T cells to MAdCAM-1 is blocked by

vedolizumab in patients with chronic refractory pouchitis. *Therap. Adv. Gastroenterol.* **14**, 17562848211054708 (2021).

145. Dai, B. *et al.* Dual targeting of lymphocyte homing and retention through α4β7 and αEβ7 inhibition in inflammatory bowel disease. *Cell Reports Med.* **2**, 100381 (2021).

146. Feagan, B. G. *et al.* Efficacy of Vedolizumab Induction and Maintenance Therapy in Patients With Ulcerative Colitis, Regardless of Prior Exposure to Tumor Necrosis Factor Antagonists. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **15**, 229-239.e5 (2017).

147. Loftus, E. V. J. *et al.* Long-term Efficacy of Vedolizumab for Ulcerative Colitis. *J. Crohns. Colitis* **11**, 400–411 (2017).

148. Danese, S. *et al.* Biological agents for moderately to severely active ulcerative colitis: a systematic review and network meta-analysis. *Ann. Intern. Med.* **160**, 704–711 (2014).

149. Pan, W.-J. *et al.* Clinical pharmacology of AMG 181, a gut-specific human anti-α4β7 monoclonal antibody, for treating inflammatory bowel diseases. *Br. J. Clin. Pharmacol.* **78**, 1315–1333 (2014).

150. Sandborn, W. J. *et al.* Efficacy and Safety of Abrilumab in a Randomized, Placebo-Controlled Trial for Moderate-to-Severe Ulcerative Colitis. *Gastroenterology* **156**, 946-957.e18 (2019).

151. Suzuki, Y. *et al.* Tofacitinib as Induction and Maintenance Therapy in Japanese Patients with Active Ulcerative Colitis. *Inflamm. Intest. Dis.* **4**, 131–143 (2019).

152. Sandborn, W. J. *et al.* Tofacitinib, an oral Janus kinase inhibitor, in active ulcerative colitis. *N. Engl. J. Med.* **367**, 616–624 (2012).

153. Yu, H. *et al.* Market share and costs of biologic therapies for inflammatory bowel disease in the USA. *Aliment. Pharmacol. Ther.* **47**, 364–370 (2018).

154. Okamoto, R. & Watanabe, M. Investigating cell therapy for inflammatory bowel disease. *Expert Opin. Biol. Ther.* **16**, 1015–1023 (2016).

155. Mayne, C. G. & Williams, C. B. Induced and natural regulatory T cells in the development of inflammatory bowel disease. *Inflamm. Bowel Dis.* **19**, 1772–1788 (2013).

156. Colman, R. J. & Rubin, D. T. Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. *J. Crohns. Colitis* **8**, 1569–1581 (2014).

157. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).

158. Barabási, A.-L. Network medicine--from obesity to the 'diseasome'. *The New England journal of medicine* vol. 357 404–407 (2007).

159. Pawson, T. & Linding, R. Network medicine. *FEBS Lett.* **582**, 1266–1270 (2008).

160. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623–627 (2000).

161. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006).

162. Sonawane, A. R., Weiss, S. T., Glass, K. & Sharma, A. Network Medicine in the Age of Biomedical Big Data . *Frontiers in Genetics* vol. 10 (2019).

163. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).

164. Charitou, T., Bryan, K. & Lynn, D. J. Using biological networks to integrate, visualize and analyze genomics data. *Genet. Sel. Evol.* **48**, 27 (2016).

165. Vousden, K. H. & Prives, C. Blinded by the Light: The Growing Complexity of p53. *Cell* **137**, 413–431 (2009).

166. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333–408 (2013).

167. Yang, Z., Algesheimer, R. & Tessone, C. J. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Sci. Rep.* **6**, 30750 (2016).

168. Subelj, L. & Bajec, M. Unfolding communities in large complex networks: combining defensive and offensive label propagation for core extraction. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **83**, 36103 (2011).

169. Shi, H. *et al.* Network Approaches for Dissecting the Immune System. *iScience* **23**, 101354 (2020).

170. Guo, M. G., Sosa, D. N. & Altman, R. B. Challenges and opportunities in network-based solutions for biological questions. *Brief. Bioinform.* **23**, bbab437 (2022).

171. Vandereyken, K., Van Leene, J., De Coninck, B. & Cammue, B. P. A. Hub Protein Controversy: Taking a Closer Look at Plant Stress Response Hubs. *Front. Plant Sci.* **9**, (2018).

172. Schadt, E. E. & Björkegren, J. L. M. NEW: network-enabled wisdom in biology, medicine, and health care. *Sci. Transl. Med.* **4**, 115rv1 (2012).

173. Kim, Y., Han, S., Choi, S. & Hwang, D. Inference of dynamic networks using time-course data. *Brief. Bioinform.* **15**, 212–228 (2013).

174. Carter, H., Hofree, M. & Ideker, T. Genotype to phenotype via network analysis. *Curr. Opin. Genet. Dev.* **23**, 611–621 (2013).

175. Kirouac, D. C. *et al.* Computational modeling of ERBB2-amplified breast cancer identifies combined ErbB2/3 blockade as superior to the combination of MEK and AKT inhibitors. *Sci. Signal.* **6**, ra68 (2013).

176. Yan, W., Xue, W., Chen, J. & Hu, G. Biological networks for cancer candidate biomarkers discovery. *Cancer Inform.* **15**, 1–7 (2016).

177. Han, L. *et al.* A probabilistic pathway score (PROPS) for classification with applications to inflammatory bowel disease. *Bioinformatics* **34**, 985–993 (2018).

178. Korcsmaros, T., Schneider, M. V. & Superti-Furga, G. Next generation of network medicine: interdisciplinary signaling approaches. *Integr. Biol. (Camb).* **9**, 97–108 (2017).

179. Whitcomb, D. C. Primer on Precision Medicine for Complex Chronic Disorders. *Clin. Transl. Gastroenterol.* **10**, e00067 (2019).

180. Denson, L. A. *et al.* Challenges in IBD Research: Precision Medicine. *Inflamm. Bowel Dis.* **25**, S31–S39 (2019).

181. Verstockt, B. *et al.* Results of the Seventh Scientific Workshop of ECCO: Precision Medicine in IBD-

Disease Outcome and Response to Therapy. *J. Crohns. Colitis* **15**, 1431–1442 (2021).

182. Vidal, M., Cusick, M. E. & Barabási, A.-L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).

183. Snider, J. *et al.* Fundamentals of protein interaction network mapping. *Mol. Syst. Biol.* **11**, 848 (2015).

184. Thomas, J. P., Modos, D., Korcsmaros, T. & Brooks-Warburton, J. Network Biology Approaches to Achieve Precision Medicine in Inflammatory Bowel Disease. *Front. Genet.* **12**, 760501 (2021).

185. Eguchi, R. *et al.* An integrative network-based approach to identify novel disease genes and pathways: A case study in the context of inflammatory bowel disease. *BMC Bioinformatics* **19**, 264 (2018).

186. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).

187. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770–780 (2008).

188. Tieri, P. & Castiglione, F. Modeling Macrophage Differentiation and Cellular Dynamics. in (ed. Wolkenhauer, O. B. T.-S. M.) 511–520 (Academic Press, 2021). doi:https://doi.org/10.1016/B978-0-12-801238-3.11644-7.

189. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

190. Margolin, A. A. *et al.* ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, 1–15 (2006).

191. Altay, G. & Emmert-Streib, F. Structural influence of gene networks on their inference: analysis of C3NET. *Biol. Direct* **6**, 31 (2011).

192. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, (2010).

193. Glass, K., Huttenhower, C., Quackenbush, J. & Yuan, G. C. Passing Messages between Biological Networks to Refine Predicted Interactions. *PLoS One* **8**, e64832 (2013).

194. Celiku, O. *et al.* Computational analysis of the mesenchymal signature landscape in gliomas. *BMC Med. Genomics* **10**, 13 (2017).

195. Min, L. *et al.* Gene regulatory pattern analysis reveals essential role of core transcriptional factors' activation in triple-negative breast cancer. *Oncotarget* **8**, 21938–21953 (2017).

196. Kuijjer, M. L., Hsieh, P. H., Quackenbush, J. & Glass, K. LionessR: Single sample network inference in R. *BMC Cancer* **19**, 4–9 (2019).

197. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

198. Verstockt, S. *et al.* Gene and Mirna Regulatory Networks During Different Stages of Crohn's Disease. *J. Crohns. Colitis* **13**, 916–930 (2019).

199. Aschenbrenner, D. *et al.* Deconvolution of monocyte responses in inflammatory bowel disease reveals an IL-1 cytokine network that regulates IL-23 in genetic and acquired IL-10 resistance. *Gut* 1–

14 (2020) doi:10.1136/gutjnl-2020-321731.

200. Türei, D. *et al.* Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **17**, e9923 (2021).

201. Csabai, L. *et al.* SignaLink3: a multi-layered resource to uncover tissue-specific signaling networks. *Nucleic Acids Res.* **50**, D701–D709 (2022).

202. Kamburov, A. *et al.* ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* **39**, D712-7 (2011).

203. Santra, T., Kolch, W. & Kholodenko, B. N. Navigating the multilayered organization of eukaryotic signaling: a new trend in data integration. *PLoS Comput. Biol.* **10**, e1003385 (2014).

204. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e20 (2019).

205. Brooks, J. *et al.* A systems genomics approach to uncover patient-specific pathogenic pathways and proteins in a complex disease. *bioRxiv* 692269 (2019).

206. Mignone, P., Pio, G., D'Elia, D. & Ceci, M. Exploiting transfer learning for the reconstruction of the human gene regulatory network. *Bioinformatics* **36**, 1553–1561 (2020).

207. Chen, K.-H., Wang, T.-F. & Hu, Y.-J. Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. *BMC Bioinformatics* **20**, 308 (2019).

208. Ching, T., Zhu, X. & Garmire, L. X. Inside the mucosal immune system. *PLOS Comput. Biol.* **14**, e1006076 (2018).

209. Wang, B., Zhu, J., Pierson, E., Ramazzotti, D. & Batzoglou, S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat. Methods* **14**, 414–416 (2017).

210. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

211. Venables WN, R. B. *Modern Applied Statistics with S*. (Springer, 2002). doi:2002.

212. Mallik, S., Seth, S., Bhadra, T. & Zhao, Z. A linear regression and deep learning approach for detecting reliable genetic alterations in cancer using dna methylation and gene expression data. *Genes (Basel).* **11**, 1–15 (2020).

213. Fiocchi, C. & Iliopoulos, D. IBD Systems Biology Is Here to Stay. *Inflamm. Bowel Dis.* **27**, 760–770 (2021).

214. Jean Philibert. One and a Half Century of Diffusion: Fick, Einstein, before and beyond. *diffusiion-fundamentals* (2005).

215. Patkar, S., Magen, A., Sharan, R. & Hannenhalli, S. A network diffusion approach to inferring sample-specific function reveals functional changes associated with breast cancer. *PLOS Comput. Biol.* **13**, e1005793 (2017).

216. Lee, M. S. & Kim, Y.-J. Signaling pathways downstream of pattern-recognition receptors and their cross talk. *Annu. Rev. Biochem.* **76**, 447–480 (2007).

217. Jeong, D. *et al.* Construction of Condition-Specific Gene Regulatory Network Using Kernel Canonical Correlation Analysis. *Front. Genet.* **12**, (2021).

218. Arijs, I. *et al.* Effect of vedolizumab (anti-α4β7-integrin) therapy on histological healing and mucosal gene expression in patients with UC. *Gut* **67**, 43–52 (2018).

219. Cepika, A.-M. *et al.* A multidimensional blood stimulation assay reveals immune alterations underlying systemic juvenile idiopathic arthritis. *J. Exp. Med.* **214**, 3449–3466 (2017).

220. Tew, G. W. *et al.* Association between Response to Etrolizumab and Expression of Integrin αe and Granzyme A in Colon Biopsies of Patients with Ulcerative Colitis. *Gastroenterology* **150**, 477-487.e9 (2016).

221. Veres, D. V. *et al.* ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* **43**, D485–D493 (2015).

222. Marbach, D. *et al.* Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **13**, 366–370 (2016).

223. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

224. Kuijjer, M. L. *et al.* limma powers differential expression analyses for {RNA-sequencing} and microarray studies. *Nucleic Acids Res.* **14**, 477--87.e9 (2015).

225. Gidskehaug, L., Anderssen, E., Flatberg, A. & Alsberg, B. K. A framework for significance analysis of gene expression data using dimension reduction methods. *BMC Bioinformatics* **8**, 346 (2007).

226. Schlauch, D., Paulson, J. N., Young, A., Glass, K. & Quackenbush, J. Estimating gene regulatory networks with pandaR. *Bioinformatics* **33**, 2232–2234 (2017).

227. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

228. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 1–20 (2016).

229. Hindryckx, P. *et al.* Clinical trials in ulcerative colitis: A Historical Perspective. *J. Crohn's Colitis* **9**, 580–588 (2015).

230. Alsoud, D., Verstockt, B., Fiocchi, C. & Vermeire, S. Breaking the therapeutic ceiling in drug development in ulcerative colitis. *The Lancet Gastroenterology and Hepatology* vol. 6 (2021).

231. Selin, K. A., Hedin, C. R. H. & Villablanca, E. J. Immunological Networks Defining the Heterogeneity of Inflammatory Bowel Diseases. *J. Crohn's Colitis* **15**, 1959–1973 (2021).

232. Zhang, Y.-Z. & Li, Y.-Y. Inflammatory bowel disease: pathogenesis. *World J. Gastroenterol.* **20**, 91 (2014).

233. de Souza, H. S. P., Fiocchi, C. & Iliopoulos, D. The IBD interactome: an integrated view of aetiology, pathogenesis and therapy. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 739–749 (2017).

234. Chang, J. T. Pathophysiology of Inflammatory Bowel Diseases. *N. Engl. J. Med.* **383**, 2652–2664 (2020).

235. van der Wijst, M. G. P., de Vries, D. H., Brugge, H., Westra, H.-J. & Franke, L. An integrative approach for building personalized gene regulatory networks for precision medicine. *Genome Med.* **10**, 96 (2018).

236. Singh, A., Fenton, C. G., Anderssen, E. & Paulssen, R. H. Identifying predictive signalling networks for

Vedolizumab response in ulcerative colitis. *Int. J. Colorectal Dis.* (2022) doi:10.1007/s00384-022-04176-w.

237. Seyed Tabib, N. S. *et al.* Big data in IBD: big progress for clinical practice. *Gut* **69**, 1520–1532 (2020).

238. Berger, S. I. & Iyengar, R. Network analyses in systems pharmacology. *Bioinformatics* **25**, 2466–2472 (2009).

239. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

240. Lewitzky, M., Simister, P. C. & Feller, S. M. Beyond 'furballs' and 'dumpling soups' - Towards a molecular architecture of signaling complexes and networks. *FEBS Lett.* **586**, 2740–2750 (2012).

241. Zaffaroni, G., Okawa, S., Morales-Ruiz, M. & del Sol, A. An integrative method to predict signalling perturbations for cellular transitions. *Nucleic Acids Res.* **47**, e72–e72 (2019).

242. Kuijjer, M. L., Hsieh, P.-H., Quackenbush, J. & Glass, K. lionessR: single sample network inference in R. *BMC Cancer* **19**, 1003 (2019).

243. Yan, J., Smyth, M. J. & Teng, M. W. L. Interleukin (IL)-12 and IL-23 and Their Conflicting Roles in Cancer. *Cold Spring Harb. Perspect. Biol.* **10**, (2018).

244. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

245. Fernald, G. H., Capriotti, E., Daneshjou, R., Karczewski, K. J. & Altman, R. B. Bioinformatics challenges for personalized medicine. *Bioinformatics* **27**, 1741–1748 (2011).

246. Emmert-Streib, F., Dehmer, M. & Haibe-Kains, B. Gene regulatory networks and their applications: Understanding biological and medical problems in terms of networks. *Front. Cell Dev. Biol.* **2**, 1–7 (2014).

247. Simoes, R. de M., Dehmer, M. & Emmert-Streib, F. B-cell lymphoma gene regulatory networks: Biological consistency among inference methods. *Front. Genet.* **4**, 281 (2013).

248. Glass, K., Huttenhower, C., Quackenbush, J. & Yuan, G. C. Passing Messages between Biological Networks to Refine Predicted Interactions. *PLoS One* **8**, e64832 (2013).

249. Sæbø, S., Almøy, T., Flatberg, A., Aastveit, A. H. & Martens, H. LPLS-regression: a method for prediction and classification under the influence of background information on predictor variables. *Chemom. Intell. Lab. Syst.* **91**, 121–132 (2008).

250. Singh, A., Anderssen, E., Fenton, C. G. & Paulssen, R. H. Identifying anti-TNF response biomarkers in ulcerative colitis using a diffusion-based signalling model. *Bioinforma. Adv.* **1**, 1–8 (2021).

251. Banerjee, S. *et al.* FINDER: an automated software package to annotate eukaryotic genes from RNA-Seq data and associated protein sequences. *BMC Bioinformatics* **22**, 205 (2021).

252. de Matos Simoes, R. & Emmert-Streib, F. Bagging statistical network inference from large-scale gene expression data. *PLoS One* **7**, (2012).

253. Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7 (2006).

254. Steinwart, I. & Christmann, A. Support Vector Machines. in *Information science and statistics* (2008).

255. James, G., Witten, D., Hastie, T. & Tibshirani, R. *Springer Texts in Statistics An Introduction to Statistical Learning*.

256. Bandopadhyay, S. Orebody modeling with : a Bayesian Neural Network Approach Orebody modeling with : a Bayesian Neural Network Approach. (2015).

257. Dangeti, P. *Statistics for Machine Learning*.

258. Crc, H. & Marsland, S. *Machine Learning: An Algorithmic Perspective*.

259. Lee, S. I. *et al.* A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat. Commun.* **9**, (2018).

260. Kraszewski, S., Szczurek, W., Szymczak, J., Reguła, M. & Neubauer, K. Machine Learning Prediction Model for Inflammatory Bowel Disease Based on Laboratory Markers. Working Model in a Discovery Cohort Study. *J. Clin. Med.* **10**, (2021).

261. Kumar, P. *et al.* Active and machine learning-based approaches to rapidly enhance microbial chemical production. *Metab. Eng.* **67**, 216–226 (2021).

262. Ghosh, P. *et al.* A Performance Based Study on Deep Learning Algorithms in the Effective Prediction of Breast Cancer. in *2021 International Joint Conference on Neural Networks (IJCNN)* 1–8 (2021). doi:10.1109/IJCNN52387.2021.9534293.

263. Weissler, E. H. *et al.* The role of machine learning in clinical research: transforming the future of evidence generation. *Trials* **22**, 537 (2021).

264. Muzio, G., O'Bray, L. & Borgwardt, K. Biological network analysis with deep learning. *Brief. Bioinform.* **22**, 1515–1530 (2021).

265. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, (2019).

266. Torres, R. & Judson-Torres, R. L. Research Techniques Made Simple: Feature Selection for Biomarker Discovery. *J. Invest. Dermatol.* **139**, 2068-2074.e1 (2019).

267. Hawtin, R. E. & Cesano, A. Immune monitoring technology primer: Single Cell Network Profiling (SCNP). *J. Immunother. cancer* **3**, 34 (2015).

268. Talhouk, A. *et al.* Single-Patient Molecular Testing with NanoString nCounter Data Using a Reference-Based Strategy for Batch Effect Correction. *PLoS One* **11**, e0153844 (2016).

269. Subramanian, J. & Simon, R. Overfitting in prediction models – Is it a problem only in high dimensions? *Contemp. Clin. Trials* **36**, 636–641 (2013).

270. Kerr, K. F. Comments on the analysis of unbalanced microarray data. *Bioinformatics* **25**, 2035–2041 (2009).

271. Groenwold, R. H. H., Goeman, J. J., Cessie, S. Le & Dekkers, O. M. Multiple testing: when is many too much? *European journal of endocrinology* vol. 184 E11–E14 (2021).

272. Sperrin, M. *Multiple Testing Procedures with Applications to Genomics*. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* vol. 173 (2010).

273. Ou, F.-S., Michiels, S., Shyr, Y., Adjei, A. A. & Oberg, A. L. Biomarker Discovery and Validation: Statistical Considerations. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **16**, 537–545 (2021).

Paper-I

Bioinformatics Advances, 2021, 1–8
doi: 10.1093/bioadv/vbab017
Advance Access Publication Date: 18 August 2021
Original paper

OXFORD

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

Systems biology

Identifying anti-TNF response biomarkers in ulcerative colitis using a diffusion-based signalling model

Amrinder Singh[1], Endre Anderssen [2], Christopher G. Fenton[2] and Ruth H. Paulssen[1,2,*]

[1]Clinical Bioinformatics Research Group, Department of Clinical Medicine, UiT-The Arctic University of Norway, Tromsø N-9037, Norway and [2]Genomics Support Centre Tromsø (GSCT), Department of Clinical Medicine, UiT-The Arctic University of Norway, Tromsø N-9037, Norway

*To whom correspondence should be addressed.

Associate Editor: Sofia Forslund
Received on May 21, 2021; editorial decision on July 21, 2021; accepted on July 21, 2021

Abstract

Motivation: Resistance to anti-TNF therapy in subgroups of ulcerative colitis (UC) patients is a major challenge and incurs significant treatment costs. Identification of patients at risk of nonresponse to anti-TNF is of major clinical importance. To date, no quantitative computational framework exists to develop a complex biomarker for the prognosis of UC treatment. Modelling patient-wise receptor to transcription factor (TF) network connectivity may enable personalized treatment.

Results: We present an approach for quantitative diffusion analysis between receptors and TFs using gene expression data. Key TFs were identified using pandaR. Network connectivities between immune-specific receptor-TF pairs were quantified using network diffusion in UC patients and controls. The patient-specific network could be considered a complex biomarker that separates anti-TNF treatment-resistant and responder patients both in the gene expression dataset used for model development and separate independent test datasets. The model was further validated in rheumatoid arthritis where it successfully discriminated resistant and responder patients to tocilizumab treatment. Our model may contribute to prognostic biomarkers that may identify treatment-resistant and responder subpopulations of UC patients.

Availability and implementation: Software is available at https://github.com/Amy3100/receptor2tfDiffusion.

Contact: ruth.h.paulssen@uit.no

Supplementary information: Supplementary data are available at *Bioinformatics Advances* online.

1 Introduction

1.1 About disease

This paper focuses on ulcerative colitis (UC) a subtype of inflammatory bowel disease (IBD) along with Crohn's disease. UC is a complex chronic inflammatory disease with dysregulation of the immune responses in the colonic mucosa. The disease features chronic acute relapsing disease activity, with intervals of remission (Khor *et al.*, 2011). Emerging evidence implicates immunological, microbial, environmental and genetic factors in the disease pathogenesis (Zhang and Li, 2014). Analysis of UC risk genes from genome-wide association studies (GWAS) implicates processes such as cell–cell communication, response to cytokine stimulus, and cell surface receptor intracellular signalling (Jostins *et al.*, 2012). Targeted treatments that induce remission in subpopulations of UC patients act by inhibiting signalling pathways between extracellular signalling molecules such as cytokines, and key transcriptional regulators of inflammatory processes (Schwartz *et al.*, 2017). However, there is significant patient-to-patient variability in treatment response, as shown by the low response rates in clinical trials (Hindryckx *et al.*, 2015; Jairath *et al.*, 2015). Therefore, we seek a method of quantifying patient-specific differences through receptor to transcription factor (TF) signalling.

1.2 Disease biomarkers

Successful personalized medicine for UC requires accurate biomarkers that can identify resistant and responders, but no individual molecular biomarker is currently recommended for clinical use to predict the treatment effects in UC (Kim *et al.*, 2017). Patient-specific biomarker discovery methods are prone to overfitting, resulting in the identification of clinically unreliable biomarkers (Hernández *et al.*, 2014). Embedding biological information from networks in the biomarker discovery process may reduce the risk of overfitting (Guo and Wan, 2014).

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

1.3 Proposed method

In this study, we propose quantifying patient-specific network connectivities between pairs of genes as complex biomarkers. However, with over 20 000 genes in the human genome, the number of potentials pairwise connections approaches 200 million. Therefore, it is necessary to identify a limited number of biologically relevant connections that explains a plausible biological mechanism central to UC aetiopathogenesis. This prevents overfitting caused by the large number of potential connections. Hence, we focus on the network connectivities between disease-relevant receptors and TFs that regulate the expression of genes involved in the inflammatory process (Fig. 1A and B). This connectivity can be quantified by network diffusion. Network diffusion describes the gradual spread of an abstract signal throughout a network. Diffusion is a global network process that considers all available paths, not just direct links or the shortest paths (Di Nanni *et al.*, 2020). Thus, the diffusion time represents the overall network connectivity between two genes, e.g. from a receptor to a TF (Fig. 1A).

2 Methods

The methods are briefly described (see the Supplementary Method for details). Statistical analysis and processing of the data were performed using R version 3.6.3 (www.r-project.org). To identify relevant TFs, the Bioconductor R package, pandaR (10.18129/B9.bioc.pandaR; Schlauch *et al.*, 2017) was used. IBD-relevant cytokines were selected from the list of GWAS risk genes for IBD (Supplementary Table S1). The comPPI database (Veres *et al.*, 2015) was used to create a signalling network connecting receptors to TFs. Network diffusion was performed on this network to estimate network connectivity between each receptor-TF pair. The differential connectivity between sample groups was tested using linear modelling (Ritchie *et al.*, 2015).

2.1 Initial data mining

The Gene expression Omnibus (GEO) was searched for datasets containing gene expression data from the colon biopsies obtained before treatment with anti-TNF and with treatment response data available. The detailed search protocol is available in the Supplementary Methods. Gene expression data for mucosal gene expression in IBD before and after treatment with anti-TNF (infliximab) were downloaded from the GEO (Supplementary Table S2). GSE16879 was used as a training dataset for model development (Arijs *et al.*, 2009a). The remaining datasets were used for testing.

Regulatory motif binding information was obtained from the regulatory circuits database (Marbach *et al.*, 2016), which contains available TF binding sites in several tissues and cell types. The binding motif-set representing general immune cells (high-level network '14_immune_organs.txt') was chosen from a regulatory circuits database as a relevant representation of the inflammatory cells involved in UC. Protein–protein interaction (PPI) data were obtained from the ComPPI database (Veres *et al.*, 2015). This is a cellular compartment-specific database of proteins and their interactions (http://ComPPI.LinkGroup.hu). Only interactions with a confidence score >0.6 were used in the network construction.

To identify key TFs, pandaR (Passing Attributes between Networks for Data Assimilation), was applied to the training Gene expression dataset (Schlauch *et al.*, 2017). pandaR creates a gene regulatory network (GRN) with weighted edges between TFs and gene targets regulated by these TFs. To evaluate which TFs significantly contributed to the variation in gene expression, a null distribution regulation network edge weight was computed by randomizing the TF gene target information. Then, the resulting null distribution was used to calculate an empirical *P*-value for each TF.

A sub-network was extracted from the PPI connecting key TFs to cytokine receptors. Genes annotated with the transducer Gene Ontology (GO) terms: GO:0002768 (cell surface receptor signalling pathway) or GO:0019221 (cytokine-mediated signalling pathway) were included in the intermediate network between the TFs and

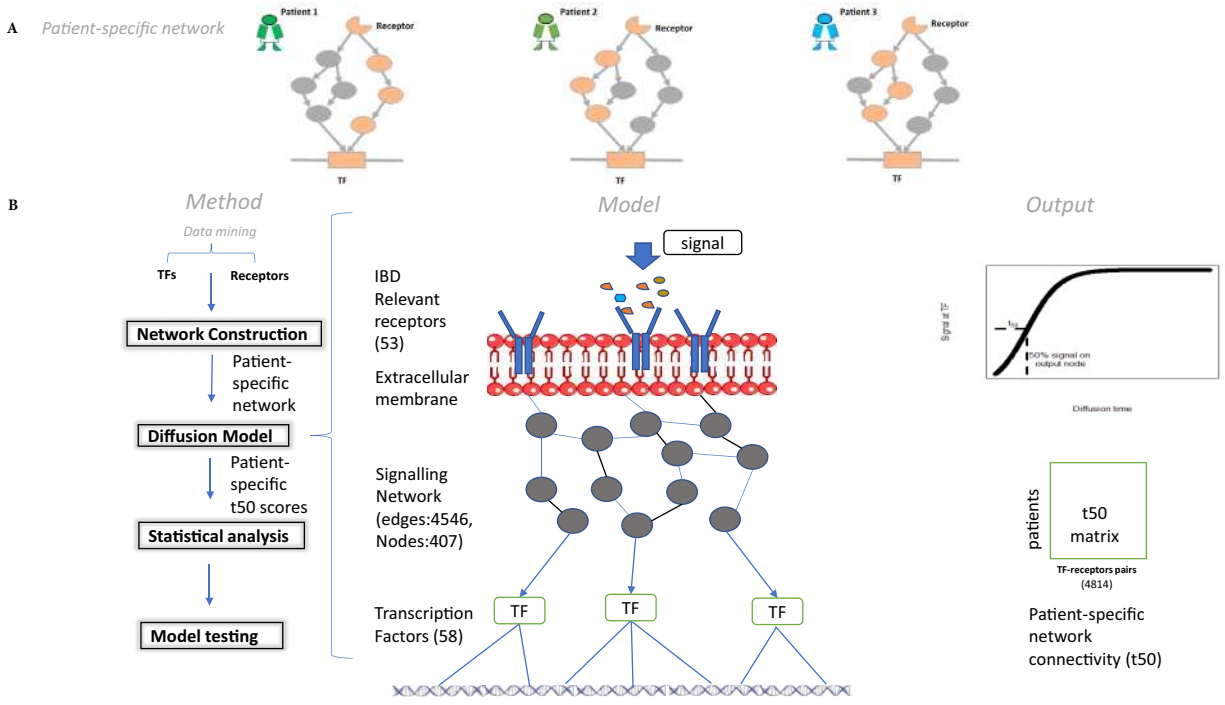


Fig. 1. Outline of diffusion model. (**A**) A schematic figure illustrating how the same biological pathway associated with a specific function may be perturbated leading to a different route of signal transduction from receptor to TF in different patients. This model can be adapted to data to build a patient-specific model (alternatively, the model can be completely generated from the data). The model can then be used to generate predictions of therapies for the patient. (**B**) Concept of diffusion model based on calculating patient-specific network edge weights in a network connecting cytokine receptors to TFs through a protein-PPI network. In the diffusion process, receptors receive a signal, and it diffuses through the network to the TFs. The model output is signal received by the TFs over time which is simplified as time to reach 50% of maximum signal at the TF (t_{50}). The output t_{50} data matrix contains t_{50} values for each receptors-TF pair per patient. This matrix can then be used for further statistical analysis or machine learning

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

receptors. In the resulting signalling network, the nodes represent the genes coding for the interacting proteins, and the edges represent physical interactions that may pass a biological signal. ComPPI was used to obtain PPIs. The network includes interactions involving the selected TFs, surface receptors which are known UC risk genes and signal transduction genes, such as kinases, that may contribute to passing information between the receptors and TFs. The same signalling network was used for all the UC datasets.

2.2 Reference methods for biomarker discovery

To separate treatment-resistant and responder patients, we initially tried the biomarker discovery tool LIONESS which quantifies patient-specific GRNs (Kuijjer *et al.*, 2019) and 'nnet' a deep learning-based method (Venables, 2002). The neural network parameters were optimized over a grid of the number of hidden nodes (size) and regularization (decay) parameters. We used 'nnet' with 10-fold cross-validation repeated 20 times using average accuracy to select the final model. This process was repeated across a grid of number of hidden nodes and the regularization parameter to identify the optimum model structure for each dataset. Prediction results in the testing datasets were evaluated by area under the receiver operating curve (AUC). In the algorithm, other parameters were kept at their default settings.

2.3 The diffusion model

We chose to model the results of the biochemical events that occur during signal transduction using a network connecting cell surface receptors to TFs in the nucleus. The model is adapted from Fick's law of chemical diffusion to a network structure. See e.g. Philibert (2005) for a review. Consider a patient-specific signalling network with nodes representing proteins e.g. cytokines, receptors and kinases create a signal transduction cascade. If a signal S, analogous to a concentration of a chemical in Fick's law, is placed on a node i, the signal flux F along a network edge connecting node i to node j at a time t is given by:

$$F(t)_{i\to j} = (S_i - S_j) \times E_i \times E_j.$$

Where the edge connectivity weight, analogous to the diffusion constant in Fick's law, is calculated using the patient's normalized gene expression values, E, of the genes coding for the proteins i and j. The signal present at each protein node i connected to J other protein nodes $j \in 1..J$ is then updated at time $t + 1$ using the sum of all fluxes:

$$S(t+1)_i = S(t)_i + \sum_{j=1}^{J} F(t)_{i\to j}.$$

The computation is initialized by setting all signal levels to zero and then placing one unit of signal on a starting receptor protein. The signal propagates through interconnected proteins throughout the network. To quantify the connectivity, we take the number of time steps to reach 50% of the maximum signal at the TF of interest (t_{50}). This methodology was implemented in R (4.1.0). Simulations were run for 2000 timesteps for all samples in each dataset, generating a new data matrix of t_{50} data with rows for each sample and a column for each receptor-TF pair. To evaluate if the obtained matrix of diffusion data contains new information or is merely a linear combination of the original gene expression data, the t_{50}–feature space was compared to the original gene expression matrix using Procrustes rotation (Peres-Neto and Jackson, 2001). This method was also used to test if the t_{50} data reflected the gene expression levels of just a few highly connected or 'hub' genes or global gene expression changes due to variations in proliferation rate or infiltration of immune cells. Cell deconvolution was used to estimate the infiltration of different immune cells in all samples (Becht *et al.*, 2016) and a gene expression signature (Sotiriou and Pusztai, 2009) was used to estimate the proliferation rate.

2.4 Statistical analysis

Significance testing for differentially expressed genes, regulatory network connectivities and diffusion (t_{50}) on the training dataset was performed using limma (Ritchie *et al.*, 2015).

Patients were grouped as normal controls, i.e. non-UC diagnosis. Responders, which attained a complete mucosal healing with a decrease of the Mayo endoscopic subscore and histological score to 0 or 1. Patients that did not attain the mentioned level of response were placed in the resistant group despite some of them showing endoscopic or histologic improvements (Arijs *et al.*, 2009b).

Three comparison tests were made:

1. Inflamed versus non-inflamed: To identify pathways that may be involved in active inflammation, we compared samples from patients with active endoscopic inflammation to non-UC controls and UC patients that had responded to treatment. The inflamed group comprised all patient samples taken from an active site of inflammation before treatment or from a treatment-resistant patient after treatment. The non-inflamed comprise normal control samples ($N = 6$) and responders after treatment ($N = 8$).
2. Resistant versus responder: To look for a biomarker of drug response, we compared samples from resistant and responder patients obtained before treatment.
3. Male versus female. As a negative control of samples concordant for inflammation, we compared samples obtained from males and female patients before treatment.

Correcting for multiple testing was done with the method of Benjamini and Hochberg (1995). Exploratory data visualization was done using principal component analysis (PCA) and Partial least squares (PLS) regression (Gidskehaug *et al.*, 2007). Gene annotation was performed using the Bioconductor org. Hs.eg.db package version 3.12.0 [10.18129/B9.bioc.org.Hs.eg.db]. GO enrichment analysis was performed using the clusterProfiler, Bioconductor package (Yu *et al.*, 2012).

3 Results

3.1 Data mining and network definition

Fifty-three IBD-relevant cytokines were selected from the 1067 identified GWAS risk genes for IBD (Supplementary Table S3). Key 58 TFs were identified using the sum of their regulatory network connections from pandaR (empirical P-value < 0.05). A list of the selected TFs and receptors is available in the supplements (Supplementary Table S3), and a full list of all TFs considered with their annotations and relevant target genes (Supplementary Tables S4 and S5). The comPPI database (Veres *et al.*, 2015) was used to create a signalling network connecting the cytokines to the TFs through 83 receptors and 266 intracellular signal transduction proteins generating a signalling network with 407 nodes and 4546 edges (Fig. 1B).

3.2 Diffusion model creates a feature space that contains novel information compared to gene expression

The diffusion model describes network connectivity from receptors to key TFs, using the time it takes a signal to diffuse from the receptor to the TF (Supplementary Fig. S1 for an example), generating a new feature space of 4814 receptor-TF pairs. Procrustes rotation was used to compare this feature space to the original gene expression space and estimate the fraction of t_{50}variability that is linearly dependent on gene expression (Table 1). Overall, in the four datasets examined, between 70% and 80% of the t_{50} information is directly linearly dependent on the gene expression data. To investigate if the t_{50} data were primarily driven by highly connected 'hub' genes in the signalling network, we extracted a subset of gene expression

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

Table 1. Dataset comparisons using Procrustes rotation

Dataset	GEO Acc#	Expression (%)[a]	Hub genes (%)[b]	Global (%)[c]	Expression and hub genes (%)[d]
Training	GSE16879	79	75	64	87
Test	GSE12251	79	75	64	87
Test	GSE23597	71	66	47	71
Test	GSE73661	69	60	48	77

Notes: Percentage of t_{50} dataset variability linearly explainable by (a) gene expression of all genes, (b) expression of genes that are highly connected in the signalling network (hub genes with more than 50 edges), (c) global expression changes due to changes in proliferation rate or immune cell infiltration. For comparison, (d) Percentage of gene expression data not explainable by the highly connected hub genes.

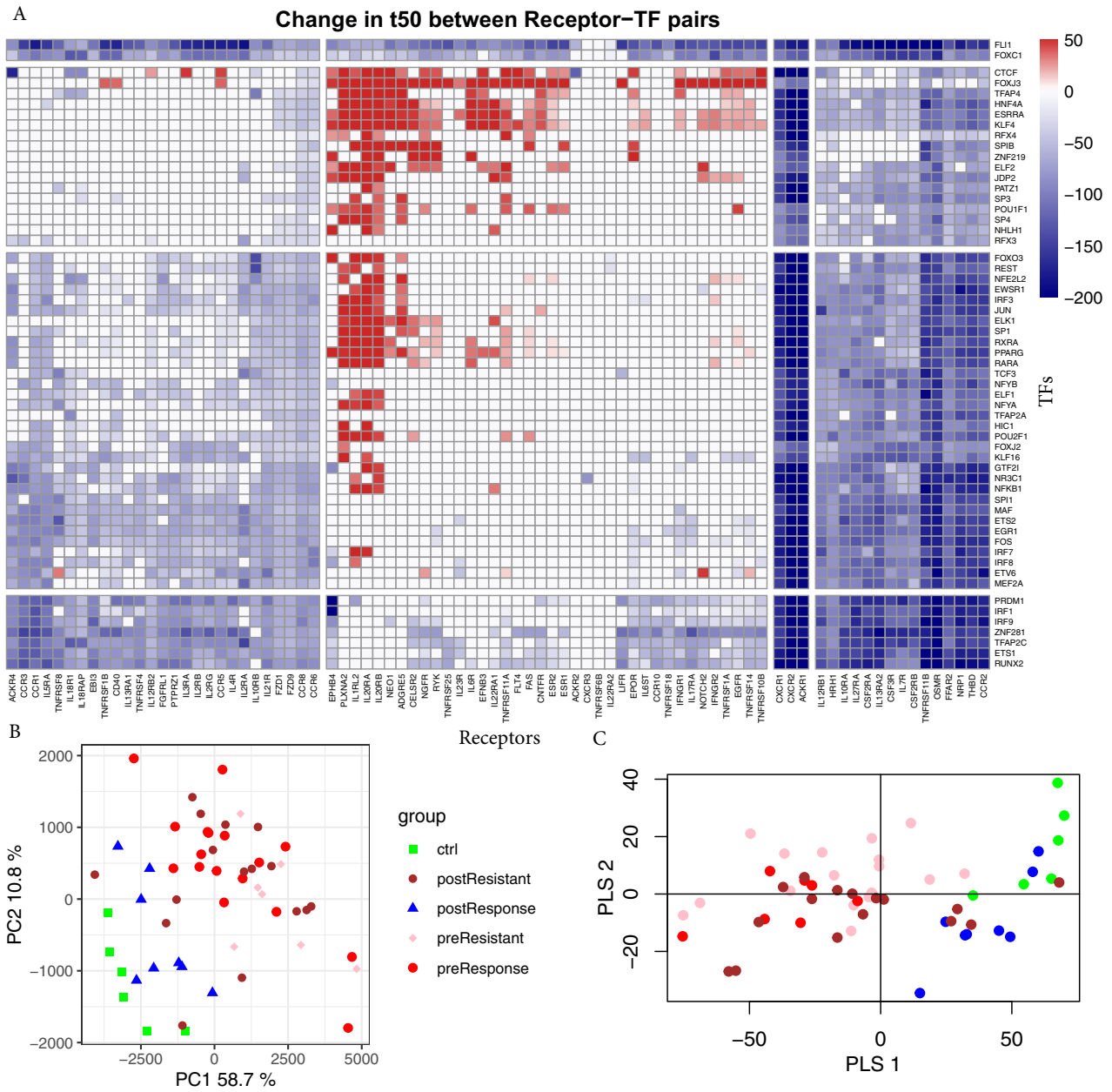


Fig. 2. Diffusion time (t_{50}) change between receptor-TF pairs in active UC and normal or responders after treatment. (**A**) Heatmaps shows the change in diffusion time in each pair of receptors (*x*-axis) and TFs (*y*-axis) between UC and normal or responders after treatment. Blue shows faster diffusion i.e. better signalling and red shows increased diffusion time i.e. a weakening strength of the connection between the receptor and the TF. (**B**) PCA of diffusion time (t_{50}) of all receptor–TF pairs and all samples in the training dataset. Normal controls and postresponse samples cluster together in contrast to the active UC samples. (**C**) PLS of the same dataset shows improved separation between preresponse and postresistant groups

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

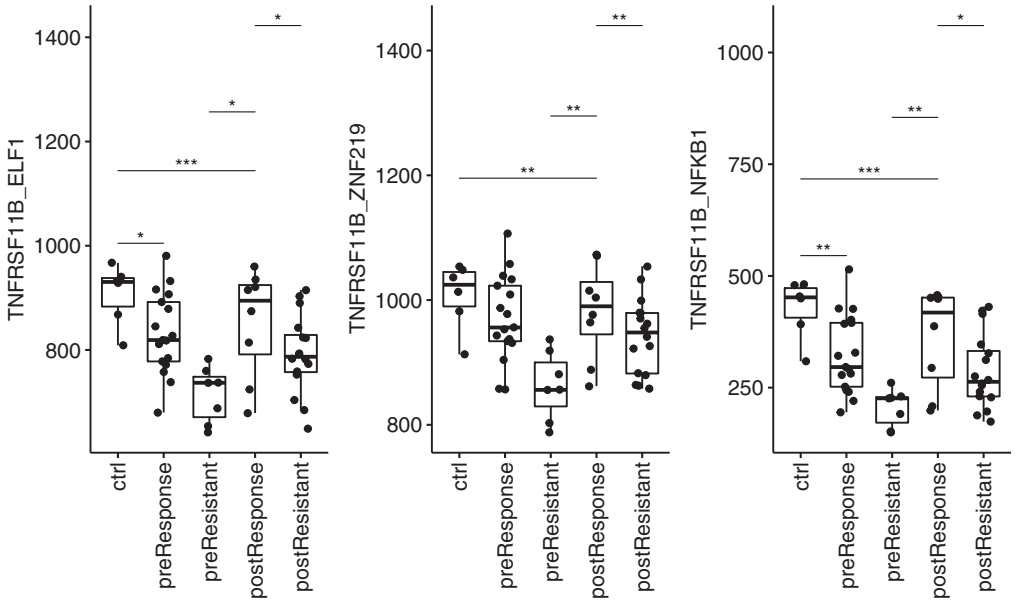


Fig. 3. Box plots of t_{50} data for the different patient groups. This figure indicates receptor–TF pairs with top AUC score in UC training dataset

Table 2. Testing of predictive ability of diffusion model compared to neural network modelling in Training and Test datasets

Dataset	GEO Acc#	nnet-AUC	TNFRSF11B-ELF1	TNFRSF11B-NFKB1
Training	GSE16879	0.80	0.91	0.91
Test	GSE12251	0.77	0.88	0.78
Test	GSE23597	0.72	0.66	0.59
Test	GSE73661	0.50	0.65	0.68

Notes: nnet-AUC shows AUC scores calculated by 'nnet'. Columns represent receptor-TF pairs TNFRSF11B-ELF, TNFRSF11B-NFKB1, TNFRSF11B-ZNF219 having the best AUC scores using diffusion model on the training dataset.

data for 46 genes with more than 50 edges. These hub genes could explain between 60% and 75% the variability in the t_{50} data. In comparison, the hub genes could explain between 71% and 87% of the total gene expression data variability.

Global changes in gene expression can be caused by large-scale tissue changes such as immune cell infiltration or changes in the proliferation rate. To investigate if this controlled the t_{50} data, we estimated the proliferation rate using a proliferation gene expression signature (Hamed *et al.*, 2015) and immune cell infiltration using a cell deconvolution tool developed for tissues analysed using Affymetrix data (Becht *et al.*, 2016). These data were compared to the t_{50} data in the same manner as the gene expression data, but it could explain only from 57% to 64% of the diffusion data (Table 1).

3.3 Diffusion model outperforms LIONESS and 'nnet' for predicting anti-TNF response in UC

We used a linear model to relate t_{50} to inflammation status and drug response in the training dataset. This enabled us to identify the receptor-TF pairs significantly related to active UC. We obtained 2362 receptor-TF pairs with adjusted *P*-values less than 0.01 (Fig. 2A).

PCA of network connectivities (t_{50}) shows postresponders clustered with the control group as expected but shows no clear separation between resistant and responders before treatment (Fig. 2B). Using PLS of t_{50}, we obtained some separation of treatment-resistant and responder patients (Fig. 2C). We used a linear model to relate the t_{50} data to the anti-TNF treatment-resistant and responder patient groups and identified 114 receptor-TF pairs with significant differences in network connectivity (adj. *P*.val < 0.05; Supplementary Table S6). Using the receiver operating characteristic curve (ROC), we evaluated the individual receptor-TF pairs for their ability to discriminate anti-TNF resistant from responders. We found 35 receptor-TF pairs with AUC higher than 0.84 in the training dataset (Supplementary Table S7). The top-scoring discriminators in the training UC dataset were the receptor-TF pairs TNFRSF11B-ELF1, TNFRSF11B-ZNF219 and TNFRSF11B-NFKB1, each with an AUC of $= 0.91$ (Supplementary Fig. S2). These pairs show distinct differences between treatment resistant and responders and between resistant and controls (Fig. 3, Supplementary Fig. S2 and Table 2). As a negative example, we compared male ($n = 14$) and female ($n = 10$) patient samples before treatment and found no significant differences in network connectivity (adj. $P < 0.05$). We then tested the ability of the top three receptor-TF pairs to predict anti-TNF response in the test datasets (Supplementary Table S2). The predictive ability of these receptor-TF pairs was compared with a deep learning method, 'nnet', a feed-forward neural network algorithm trained on the same dataset. We also tested LIONESS, a method for computing sample-specific GRNs. Surprisingly, the diffusion model outperformed the neural network, giving higher AUC scores (Table 2) in the UC training dataset and the majority of the independent test datasets. We compared inflamed versus noninflamed, i.e. normal controls and responders after treatment versus before treatment and resistant after treatment, in addition to the treatment resistant versus responders. LIONESS estimates a total of 2 678 095 regulatory edge weights per sample. Significantly changed edge weights were then identified using linear modelling (limma). Between the inflamed and noninflamed samples in the training set, 161 052 statistically significant edge weights (adj. *P*-value < 0.01) were found. However, no significant results were obtained (adj. *P*-value ~ 0.99) for the more important comparison of anti-TNF resistant versus responder comparison (Supplementary Table S8). LIONESS was therefore not applied to the test datasets.

3.4 Validation with rheumatoid arthritis

To assess our model's generalizability for other autoimmune diseases, we applied the diffusion model to the rheumatoid arthritis

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

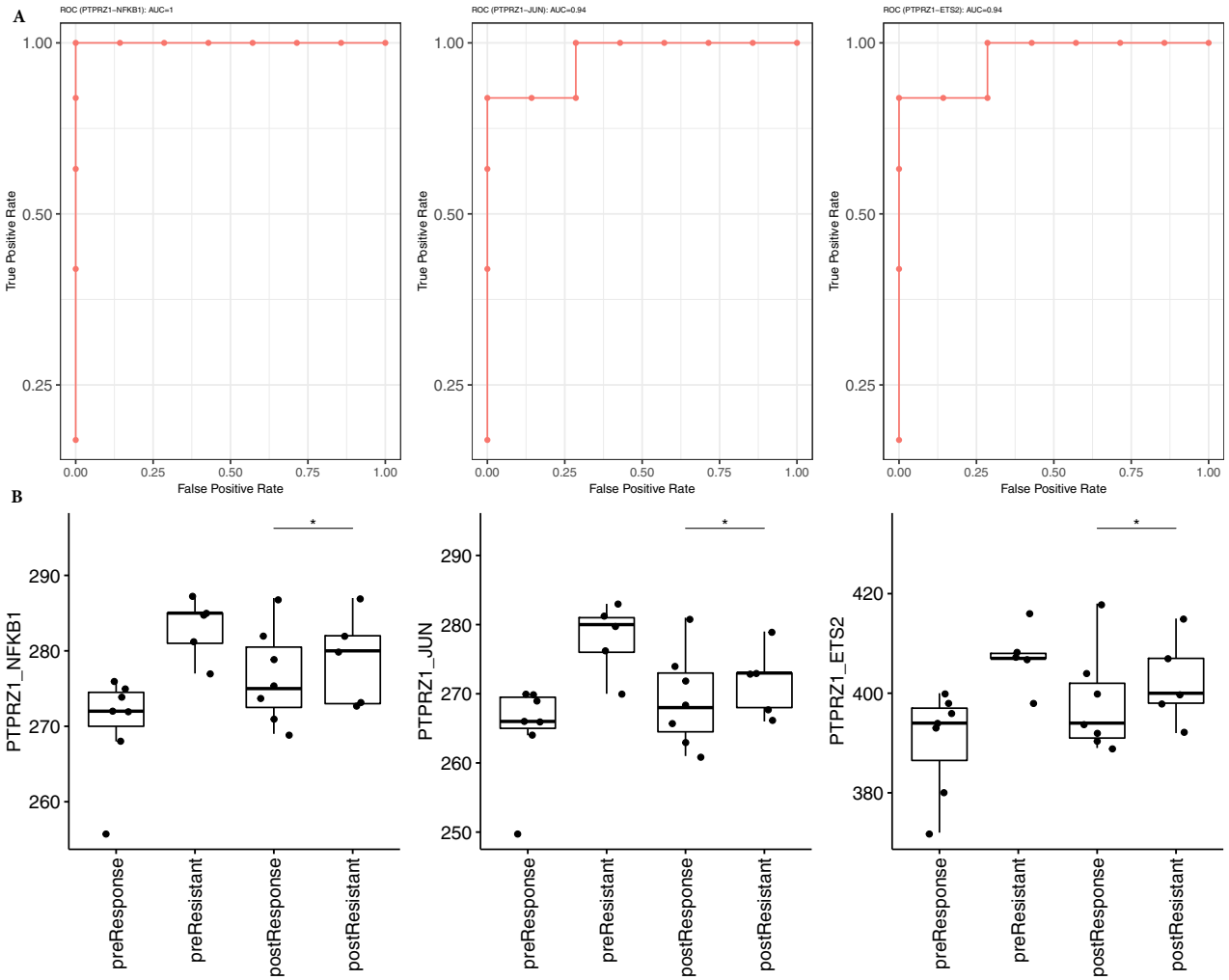


Fig. 4. Validation with RA dataset. ROC analysis shows the separation of treatment responder and resistant sample groups. It is calculated based on the t_{50} score of receptor–TF pairs (**A**) PTPRZ1-NFKB1 (**B**) PTPRZ1-JUN (**C**) PTPRZ1-ETS2 with the top AUC scores of 1, 0.94 and 0.94, respectively. Box plot illustrates receptor–TF pairs with top AUC score (**D**) PTPRZ1-NFKB1 (**E**) PTPRZ1-JUN (**F**) PTPRZ1-ETS2 which show distinction between treatment response groups based on t_{50} score calculated by diffusion model

(RA) dataset. UC and RA share many inheritable risk loci and have many overlapping pathogenic pathways (Bae *et al.*, 2017; Halling *et al.*, 2017; Hemminki *et al.*, 2009). Therefore, we selected and merged two RA gene expression datasets: GSE24742 (Ducreux *et al.*, 2014) and GSE45867 (Gutierrez-Roelens *et al.*, 2011) to validate our modelling method. The chosen dataset is a microarray gene expression study of paired synovial biopsy samples collected before therapy (T0) and after therapy (T12) from the affected knee of RA patients treated with tocilizumab (TCZ), methotrexate (MTX) or rituximab (RTX). The experiment design of the validation dataset was similar to the UC dataset in terms of before treatment biopsy, underlying disease mechanism (inflammation), and a sufficient number of samples used in the study. We used a dataset with 86 RA samples to validate the model for testing the performance of the model.

For validation, we used the same pipeline developed with the same score thresholds. We created a literature curated list of RA-relevant receptors (McInnes *et al.*, 2016; Mockridge *et al.*, 2017; Supplementary Table S3) which was subsequently integrated with expression data to create an RA-relevant diffusion model. Our ROC results found a remarkable AUC = 1 for receptor-TF pair PTPRZ1-NFKB1 (Fig. 4A), AUC = 0.94 each for PTPRZ1-JUN (Fig. 4B) and PTPRZ1-ETS2 (Fig. 4C) which accurately separates the TCZ treatment-resistant and responder patients (Fig. 4D–F). High AUC receptor-TF pairs are listed in Supplementary Data (Supplementary Table S5). GO analysis of the TFs identified in RA highlighted processes such as response to oxidative stress, cellular response to peptide, negative regulation of protein phosphorylation, etc. GO analysis highlights key immune processes associated with RA pathophysiology regulated by key TFs such as SPI1, RARA, PPARG, NFKB1, ETS1 and MAF (Supplementary Fig. S3; Giaginis *et al.*, 2009; Ikuta *et al.*, 2012; Kang *et al.*, 2017; Manuel Sánchez-Maldonado *et al.*, 2020; Zisakis et al., 2007).

4 Discussion

We have developed a diffusion model; a molecular pathway inspired method to model patient-specific treatment response. It creates a new feature space by using key TFs, receptors, biological prior information in the form of a PPI and gene expression data. This new feature space is a nonlinear transformation of the original gene

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

expression variables, designed with the goal of being more relevant for describing cytokine signalling in UC. We have compared the ability of the new features to predict anti-TNF response in UC to two other methods. We used 'nnet', a machine learning method that has recently been used in biomarker discovery (Mallik *et al.*, 2020) and LIONESS, a regulatory network reconstruction method that estimates patient-specific regulatory connections. These methods represent two extremes in the analysis of biological data. The 'nnet' is a general machine learning method that controls overfitting by regularization model parameters. LIONESS–pandaR, however, uses a large amount of biological background information about TF-targets and PPI to estimate patient-specific GRNs. Although both the 'nnet' and the t_{50} features may serve as useful biomarkers of drug response in independent datasets, the LIONESS-pandaR method fails to identify any biomarkers for drug response. This may be due to the large number of calculated network connection weights and the consequential penalty for multiple testing. Therefore, pandaR-LIONESS may lack the sensitivity to pick up the more subtle differences between subclasses of the patient samples, compared to the much larger difference between normal and inflamed tissue.

The diffusion-based features outperform the 'nnet' both in fitting to the training data and two out of three test datasets. This may be an indication that the combinations and transforms of the gene expression data derived from the signalling network topology might have more biological relevance than features obtained by a pure fit to the gene expression, despite the regularization penalties in 'nnet'. Our method identified well-known pro-inflammatory receptors such as TNFRSF11B, OSMR, NRP1 and CCR2 which exhibited stronger connectivity (low t_{50}) to most inflammation-related TFs in UC patients with active inflammation than in non-UC controls and responders after treatment (Fig. 2A and Supplementary Fig. S4). These results must be interpreted with a caution as the responder samples may still contain residual microscopic inflammation and have lasting changes to their epithelial cells (Fenton *et al.*, 2021; Planell *et al.*, 2013). However, the goal of this analysis is to identify the receptor-TF pairs involved in active inflammation that requires treatment. Notably, our model also identified TFs ESRRA and HNF4A, which play an important role in the regulation of intestinal homeostasis. ESRRA is a regulator of intestinal homeostasis (Kim *et al.*, 2020), and HNF4A modulates inflammation in UC and maintains epithelial barrier integrity in the normal intestine (Ahn *et al.*, 2008; Barrett *et al.*, 2009).

Despite the good predictive performance, it is also noteworthy that both 'nnet' and the diffusion model performed worse on test dataset 3 (GSE73661). A dataset analysed with a different array design than the training dataset. This highlights the importance of robust and repeatable measuring processes for the practical use of complex gene expression-based biomarkers. Unfortunately, no large-scale modern RNA-seq datasets are currently available to test for predicting anti-TNF response in UC.

The diffusion model may also be susceptible to predictive errors because of the assumptions made in the initial data mining. We have chosen to focus on cytokines as the source of the inflammatory signal (Chen and Sundrud, 2016), but inflammatory diseases may also involve other signalling systems such as pattern recognition receptors and metabolic factors. The method is also highly simplified, ignoring molecular functions such as activation, repression and feedback loops, which are not considered explicitly. In addition, biological molecules of unknown function that may influence true network connectivity are ignored. Moreover, epigenetic factors have a crucial role in determining the transcriptional activation of genes targeted by a specific TF (Gibney and Nolan, 2010). However, obtaining epigenetic signatures for every individual patient is currently cost-prohibitive. Additionally, the evaluation does not take into account changes in the patient's gene expression as the disease progresses. They may therefore be expected to give a more reliable prediction of short-term effects than in long-term remission. In conclusion, we assert that our diffusion model can be used to generate testable hypotheses applicable to UC and other autoimmune diseases such as RA, psoriasis and asthma. This framework outlines the receptor-TF-specific network connectivity which varies with the gene expression of each individual patient. Estimating the receptor-TF network connectivity associated with varied drug responses in disease subpopulations may yield valuable insights into a patient's treatment outcome.

Funding

This work has been supported by the Northern Norway Regional Health Authority [Strategisk-HN-10-16].

Conflict of Interest none declared.

References

Ahn,S.H. *et al.* (2008) Hepatocyte nuclear factor 4α in the intestinal epithelial cells protects against inflammatory bowel disease. *Inflamm. Bowel Dis.*, **14**, 908–920.

Arijs,I. *et al.* (2009a) Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PLoS One*, **4**, e7984.

Arijs,I. *et al.* (2009b) Mucosal gene signatures to predict response to infliximab in patients with ulcerative colitis. *Gut*, **58**, 1612–1619.

Bae,J.M. *et al.* (2017) Association of inflammatory bowel disease with ankylosing spondylitis and rheumatoid arthritis: a nationwide population-based study. *Mod. Rheumatol.*, **27**, 435–440.

Barrett,J.C. *et al.* (2009) Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat. Genet.*, **41**, 1330–1334.

Becht,E. *et al.* (2016) Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.*, **17**, 218–220.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B (Methodol.)*, **57**, 289–300.

Chen,M.L. and Sundrud,M.S. (2016) Cytokine networks and T-cell subsets in inflammatory bowel diseases. *Inflamm. Bowel Dis.*, **22**, 1157–1167.

Di Nanni,N. *et al.* (2020) Network diffusion promotes the integrative analysis of multiple omics. *Front. Genet.*, **11**, 106.

Ducreux,J. *et al.* (2014) Global molecular effects of tocilizumab therapy in rheumatoid arthritis synovium. *Arthritis Rheumatol.*, **66**, 15–23.

Fenton,C.G. *et al.* (2021) Transcriptional signatures that define ulcerative colitis in remission. *Inflamm. Bowel Dis.*, **27**, 94–105.

Giaginis,C. *et al.* (2009) Peroxisome proliferator-activated receptor-γ (PPAR-γ) ligands as potential therapeutic agents to treat arthritis. *Pharmacol. Res.*, **60**, 160–169.

Gibney,E.R. and Nolan,C.M. (2010) Epigenetics and gene expression. *Heredity*, **105**, 4–13.

Gidskehaug,L. *et al.* (2007) A framework for significance analysis of gene expression data using dimension reduction methods. *BMC Bioinformatics*, **8**, 346.

Guo,N.L. and Wan,Y.W. (2014) Network-based identification of biomarkers coexpressed with multiple pathways. *Cancer Inform.*, **13**, 37–47.

Gutierrez-Roelens,I. *et al.* (2011) Rituximab treatment induces the expression of genes involved in healing processes in the rheumatoid arthritis synovium. *Arthritis Rheum.*, **63**, 1246–1254.

Halling,M.L. *et al.* (2017) Patients with inflammatory bowel disease have increased risk of autoimmune and inflammatory diseases. *World J. Gastroenterol.*, **23**, 6137–6146.

Hamed,M. *et al.* (2015) Integrative network-based approach identifies key genetic elements in breast invasive carcinoma. *BMC Genomics*, **16** (**Suppl. 5**), S2.

Hemminki,K. *et al.* (2009) Familial associations of rheumatoid arthritis with autoimmune diseases and related conditions. *Arthritis Rheum.*, **60**, 661–668.

Hernández,B. *et al.* (2014) Why have so few proteomic biomarkers "survived" validation? (Sample size and independent validation considerations). *Proteomics*, **14**, 1587–1592.

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

Hindryckx,P. *et al.*; Clinical Trial Committee Clincom of the European Crohn's and Colitis Organisation (ECCO). (2015) Clinical trials in ulcerative colitis: a historical perspective. *J. Crohns Colitis*, **9**, 580–588.
Ikuta,K. *et al.* (2012) The Sp1 transcription factor is essential for the expression of gliostatin/thymidine phosphorylase in rheumatoid fibroblast-like synoviocytes. *Arthritis Res. Ther.*, **14**, R87.
Jairath,V. *et al.* (2015) Placebo response and remission rates in randomized trials of induction and maintenance therapy for ulcerative colitis. *Cochrane Database Syst. Rev.*, **2015**. doi:10.1002/14651858.CD011572.
Jostins,L. *et al.*; International IBD Genetics Consortium (IIBDGC). (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**, 119–124. [23128233]
Kang,K. *et al.* (2017) Interferon-γ represses M2 gene expression in human macrophages by disassembling enhancers bound by the transcription factor MAF. *Immunity*, **47**, 235–250.e4.
Khor,B. *et al.* (2011) Genetics and pathogenesis of inflammatory bowel disease. *Nature*, **474**, 307–317.
Kim,S. *et al.* (2020) ESRRA (estrogen related receptor alpha) is a critical regulator of intestinal homeostasis through activation of autophagic flux via gut microbiota. *Autophagy*, 1–20. doi:10.1080/15548627.2020.1847460.
Kim,W.M. *et al.* (2017) A role for oncostatin M in inflammatory bowel disease. *Nat. Med.*, **23**, 535–536.
Kuijjer,M.L. *et al.* (2019) LionessR: single sample network inference in R. *BMC Cancer*, **19**, 1003–1009.
Mallik,S. *et al.* (2020) A linear regression and deep learning approach for detecting reliable genetic alterations in cancer using DNA methylation and gene expression data. *Genes*, **11**, 931.
Manuel Sánchez-Maldonado,J. *et al.* (2020) NFKB2 polymorphisms associate with the risk of developing rheumatoid arthritis and response to TNF inhibitors: results from the REPAIR consortium. *Sci. Rep.*, **10**, 4316.
Marbach,D. *et al.* (2016) Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods*, **13**, 366–370.
McInnes,I.B. *et al.* (2016) Cytokines in rheumatoid arthritis-shaping the immunological landscape. *Nat. Rev. Rheumatol.*, **12**, 63–68.
Mockridge,J. *et al.* (2017) BMC rheumatology: a home for all rheumatology research within the BMC series. *BMC Rheumatol.*, **1**, 1–7.
Peres-Neto,P.R. and Jackson,D.A. (2001) How well do multivariate data sets match? The advantages of a procrustean superimposition approach over the Mantel test. *Oecologia*, **129**, 169–178.
Philibert,J. (2005) One and a half century of diffusion: Fick, Einstein, before and beyond. *Diffus. Fundam.*, **2**, 1.1–1.101.
Planell,N. *et al.* (2013) Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. *Gut*, **62**, 967–976.
RitchieM. *et al.* (2015) limma powers differential expression analyses for {RNA-sequencing} and microarray studies. *Nucleic Acids Res.*, **43**, e47.
Schlauch,D. *et al.* (2017) Estimating gene regulatory networks with pandaR. *Bioinformatics*, **33**, 2232–2234.
Schwartz,D.M. *et al.* (2017) Erratum: JAK inhibition as a therapeutic strategy for immune and inflammatory diseases (Nature reviews. Drug discovery (2017) 16 12 (843-862)). *Nat. Rev. Drug Discovery*, **17**, 78.
Sotiriou,C. and Pusztai,L. (2009) Gene-expression signatures in breast cancer. *N. Engl. J. Med.*, **360**, 790–800.
Venables,W.N. (2002) *Modern Applied Statistics with S*. 4th edn. Springer, New York, NY.
Veres,D.V. *et al.* (2015) ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.*, **43**, D485–D493.
Yu,G. *et al.* (2012) ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.
Zhang,Y.Z. and Li,Y.Y. (2014) Inflammatory bowel disease: pathogenesis. *World J. Gastroenterol.*, **20**, 91–99.
Zisakis,A. *et al.* (2007) Expression of retinoic acid receptor (RAR) α protein in the synovial membrane from patients with osteoarthritis and rheumatoid arthritis. *Int. J. Biomed. Sci.*, **3**, 46–49.

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/1/1/vbab017/6354348 by UiT The Arctic University of Norway user on 19 May 2022

Paper-II

International Journal of Colorectal Disease
https://doi.org/10.1007/s00384-022-04176-w

ORIGINAL ARTICLE

Identifying predictive signalling networks for Vedolizumab response in ulcerative colitis

Amrinder Singh[1] · Christopher G. Fenton[2] · Endre Anderssen[2] · Ruth H. Paulssen[1,2]

Accepted: 1 May 2022
© The Author(s) 2022

Abstract

Background In ulcerative colitis (UC), the molecular mechanisms that drive disease development and patient response to therapy are not well understood. A significant proportion of patients with UC fail to respond adequately to biologic therapy. Therefore, there is an unmet need for biomarkers that can predict patients' responsiveness to the available UC therapies as well as ascertain the most effective individualised therapy. Our study focused on identifying predictive signalling pathways that predict anti-integrin therapy response in patients with UC.

Methods We retrieved and pre-processed two publicly accessible gene expression datasets (GSE73661 and GSE72819) of UC patients treated with anti-integrin therapies: (1) 12 non-IBD controls and 41 UC patients treated with Vedolizumab therapy, and (2) 70 samples with 58 non-responder and 12 responder UC patient samples treated with Etrolizumab therapy without non-IBD controls. We used a diffusion-based signalling model which is mainly focused on the T-cell receptor signalling network. The diffusion model uses network connectivity between receptors and transcription factors.

Results The network diffusion scores were able to separate VDZ responder and non-responder patients before treatment better than the original gene expression. On both anti-integrin treatment datasets, the diffusion model demonstrated high predictive performance for discriminating responders from non-responders in comparison with 'nnet'. We have found 48 receptor-TF pairs identified as the best predictors for VDZ therapy response with AUC $\geq$ 0.76. Among these receptor-TF predictors pairs, FFAR2-NRF1, FFAR2-RELB, FFAR2-EGR1, and FFAR2-NFKB1 are the top best predictors. For Etrolizumab, we have identified 40 best receptor-TF pairs and CD40-NFKB2 as the best predictor receptor-TF pair (AUC = 0.72). We also identified subnetworks that highlight the network interactions, connecting receptors and transcription factors involved in cytokine and fatty acid signalling. The findings suggest that anti-integrin therapy responses in cytokine and fatty acid signalling can stratify UC patient subgroups.

Conclusions We identified signalling pathways that may predict the efficacy of anti-integrin therapy in UC patients and personalised therapy alternatives. Our results may lead to the advancement of a promising clinical decision-making tool for the stratification of UC patients.

Keywords T-cell receptor signalling · Ulcerative colitis · Immune regulation · Network connectivity · Therapy response · Vedolizumab · Infliximab · Signalling pathway · Personalised therapy

✉ Ruth H. Paulssen
ruth.h.paulssen@uit.no

Amrinder Singh
amrinder.singh@uit.no

Christopher G. Fenton
christopher.fenton@uit.no

Endre Anderssen
endre.anderssen@uit.no

[1] Clinical Bioinformatics Research Group, Department of Clinical Medicine, Faculty of Health Sciences, UiT- The Arctic University of Norway, Tromsø, Norway

[2] Genomics Support Centre Tromso, UiT- The Arctic University of Norway, Department of Clinical Medicine, Faculty of Health Sciences, UiT- The Arctic University of Norway, Tromso, Norway

Published online: 11 May 2022

Background

Ulcerative colitis (UC) is a multifaceted, chronic, immune-mediated inflammatory disorder. UC exhibits inflammation in the mucosa and submucosa, ranging from the rectum, and can spread to proximal segments of the colon [1–4].The patients may undergo periods of remission and relapses [5]. The immunopathogenesis of UC features exaggerated immune response inducing epithelial damage, microbial dysbiosis, abnormal activation of lymphocytes, and infiltration of innate immune cells [6]. The aetiology of UC is multifactorial and potentially caused by genetic, immunological, microbial, and environmental factors [3, 6]. Given the nature of the disease aetiology, there is no single effective therapy for all UC patients. Thus, the use of ineffective UC therapies for moderate-severe cases constitutes a significant burden on the healthcare system [7, 8].

Standard conventional therapies for UC are sulfasalazine, mesalazine (5-ASA), and corticosteroids for the mild-to-moderate disease activity. Some UC patients are unresponsive or intolerant to the standard therapies [4, 9–11] which have prompted the development of new drugs that target tumour necrosis factor (TNF), leukocyte adhesion, JAK-STAT pathway, IL-12 and IL-23, T-helper cell (Th)-1 polarisation, or T-cell activation [12, 13]. UC immunopathogenesis involves the altered immunoregulatory activity by crosstalk between T cell subsets that modulate inflammation [6, 14]. An example of T-cell directed therapy for IBD is a gut-selective anti-α4/β7 integrin heterodimers monoclonal antibody, Vedolizumab (VDZ). Integrin α4/β7 is expressed on immune cells such as T-cell(s). VDZ selectively inhibits the adhesion of integrin α4/β7 to the mucosal vascular address in cell adhesion molecule 1 (MAdCAM-1) which is expressed in the lamina propria [15–18]. Targeting integrin α4β7 prevents the influx of T-cells to the lamina propria, thereby suppressing the gut inflammation [19–21].

VDZ can be used as primary biologic therapy after failure of the standard therapy and also as a secondary therapy for UC patients showing primary non-response, loss of response, or intolerance to anti-TNFα (Infliximab) therapy. It can also be used for the maintenance of clinical remission and is considered a safer yet less efficacious alternative to infliximab [22]. VDZ reduces inflammation in the gut tissue as the gut expresses vascular cell adhesion molecule 1 (MADCAM1) and vascular cell adhesion molecule 1 (VCAM1) molecules [23, 24] In contrast anti-TNFα is associated with systemic immunosuppression [25]. About 30% of UC patients fail to respond to VDZ and suffer tissue damage, and leukocyte-driven inflammatory activity which is associated with TNF-dependent pathways [15]. Other targeted treatment alternatives are Ustekinumab and Tofacitinib. Ustekinumab is a monoclonal antibody biologic targeting both IL-12 and IL-23 to reduce chronic inflammation [26] while Tofacitinib works as an inhibitor which targets the JAK-STAT pathway by inhibiting phosphorylation and activation of JAKs to decrease the inflammatory response [27]. Ultimately, non-responders may require surgical interventions [28]. Therefore, it is important to identify non-responding patients as early as possible during disease development to provide better therapy alternatives.

Our major objective is to seek patient-specific networks that separate VDZ treatment responders and non-responders. In our recent work, we could successfully stratify infliximab responder vs non-responder patients using cytokine signalling network diffusion rates [29]. In this study, we implemented a diffusion model to discriminate UC patients (VDZ responders vs non-responders) to construct patient-specific subnetworks.

Methods

Data source

To identify VDZ related studies, a public dataset search on Gene Expression Omnibus (GEO) was performed using the keywords 'Vedolizumab' and 'Ulcerative colitis'. The only hit found was GSE73661 containing Affymetrix GeneChip Human Gene 1.0 ST arrays of UC patients before and after treatment with VDZ or IFX, and non-IBD controls for 178 samples [17]. This dataset contains patients who were recruited from two phase 3 VDZ trials (GEMINI & LTS) (Table 1). Biopsies were taken at week (W) 0 before, and W6, W12, and W52 after VDZ treatment giving a total of 124 colonic mucosal biopsies. The sampling location includes UC left-sided colitis/pancolitis biopsies collected at the edge of ulcers (if present) or at the most inflamed colon segment (absence of ulcers). The colonic biopsies for histological healing assessment were scored using the Geboes index [30, 31]. For endoscopic healing assessment, Mayo subscore was used [32] before and after treatment. VDZ-treated UC patients were classified as responders (n = 14) and non-responders (n = 27) (Table 2) based on the colonic healing sub-score of Geboes index for histological assessment.In this study, patients treated with VDZ all had a previous history of treatment with IFX. We used 112 VDZ-treated patient samples and 12 non-IBD control samples in the analysis. In this dataset, 27 patients did not respond to therapy (non-responder) while 14 had responded (responder) to VDZ therapy (Table 2). We also used a publicly available RNAseq dataset (GSE72819) of another anti-integrin

Table 1 Baseline characteristics of the patient treated with vedolizumab *

Baseline Characterstics	UC_VDZ (n=41)	non-IBD_controls (n=12)
Age (years) - Median (IQR)	40.5 (32-49.4)	68.2 (59–72.7)
Male/female (%)	21/20 (51.2/48.8)	6/6 (50/50)
Duration of disease (years) - Median (IQR)	10.2 (4.4–14.6)	NA
Extent of disease		
UC left-sided colitis/pancolitis (%)	18/23 (43.9/56.1)	NA
Histology (Geboes score) (2-5(%))	41 (100)	NA
Mayo Score (2-3 (%))	41 (100)	NA
Median (IQR) - total Mayo score (2-3(%))	10 (8–11)	NA
Concomitant medication -no. (%)		
5-Aminosalicylates	29 (70.7)	NA
Corticosteroids	17 (41.5)	NA
Methotrexate	1 (2)	NA
Azathioprine/6-mercaptopurine	7 (17.1)	NA
Prior anti-TNF treatment -no. (%)	0 (0)	NA
Active smoking (%)	5 (12.2)	0 (0)

*Data adapted from Arijs et al. 2018 [17]

biologic (Etrolizumab) containing UC patients with baseline biopsies (Table 4). This dataset contains a total of 70 samples with 58 non-responder and 12 responder UC patient samples but no non-IBD controls.

Treatment groups classification

Samples were classified into pre-resistant, post-resistant, pre-responder, and post-responder groups (Table 2). Samples collected at week (W) 0 from patients who did not respond to VDZ at W6, W12, and W52 were labelled as VDZ pre-resistant. Samples collected at W6, W12, and W52 from patients who did not respond to VDZ were labelled as VDZ post-resistant. Samples collected at week 0 from patients who did respond to VDZ at W6, W12, and W52 were labelled as VDZ pre-responders. Samples from patients who did respond to VDZ at W6, W12, and W52 were labelled as VDZ post-responders. Response in this context refers to endoscopic healing at any of the after-treatment time points (W6, W12, and W52) [17].

Table 2 VDZ response dataset including the number of controls, responder, and non-responder patients. The patient samples were classified into pre-resistant, post-resistant, pre-responder, and post-responder groups at the given time of biopsy of W0 and W (6–52), respectively

Patient classification	Number of patients (n = 41)
Non-responder patients	**27**
• Pre-resistant (patient biopsies) at W0	32
• Post-resistant (patient biopsies) at W (6–52)	49
Responder patients	**14**
• Pre-response (patient biopsies) at W0	09
• Post-response (patient biopsies) at W (6–52)	22
Controls	**12**

Gene expression data pre-processing

Data pre-processing was done using quantile normalisation. We annotated the gene expression matrix with matching gene symbols available in the metadata of the Affymetrix platform (https://doi.org/10.18129/B9.bioc.hgu133plus2.db). Non-coding genes, e.g. microRNAs, pseudogenes, and lncRNAs, were filtered from gene expression data. The final expression matrix contains row values with gene symbols and columns with patients' sample IDs.

Transcription factor (TF) identification

We used pandaR [33] to identify important TFs that potentially regulate the gene expression in the UC. The gene IQR (geneIQR > 0.30) was calculated on the gene expression to remove genes with low variance across the samples. To identify the TFs using pandaR, we input protein–protein interaction information using comPPI [34] and regulatory circuits [35] for the regulatory motif binding information in several tissues and cell types. The motif binding set was retrieved from regulatorycircuits.org representing CD4 and CD8 immune cells as relevant T-cell specific immune responses. We selected CD4 and CD8 regulatory circuit because it contains cell types which are surface markers for

T-cell. We used the threshold for pandaR result with a score cutoff > 0.01. The result is obtained from pandaR containing TF-gene target edge scores which define confidence for each TF regulating the corresponding target gene. To test which TFs significantly regulate gene expression, a null distribution of TF edge weight was calculated by randomising the TF-gene target 512 times and TFs were selected with an empirical p-value [36] ($p < 0.05$).

Generating signalling network

We used a previously constructed diffusion model that estimates network connectivity between receptors and TFs through a signalling network. For creating a signalling network, we used gene expression data, TFs, and receptors. A list of 80 receptors was used which includes IBD GWAS risk genes, cytokines-, chemokines-, pattern recognition-receptors as well as adhesion molecules (Table S1 and Figure S1). To select genes specific to cell surface receptors signalling pathway, we used GO term (GO:0007166) which limits the network to known receptors. Thus, we obtained 24 IBD-relevant cell-adhesion receptors and 37 cytokines receptors with a total of 61 extracellular signalling molecules (Table S1 and Figure S1). By applying significance testing with empirical p-value < 0.05, 34 key TFs were obtained using the sum of their regulatory network edge weights of a total of 643 TFs from pandaR [33]. A complete list of TFs selected for further analysis is provided with their annotations and target genes (Table S2). Genes such as TP53, HSP's, UBC which have a high number of protein–protein interactions were removed to reduce the complexity of the global network. To further restrict the network, we used GO terms which are associated with cell surface receptor signalling pathway, cytokine-mediated signalling pathway, and integrin α4β7 complex pathway (Table S3). Thus, the final signalling network contains only 309 nodes and 2645 edges for T-cell specific receptors, signal transducer proteins, and TFs.

Diffusion model

We used our previously published diffusion model [29] which uses gene expression data to generate edge-weighted signalling network graphs for quantifying connectivity from receptors to TFs for each patient. The edge weight is calculated by the product of the gene expression levels of the two genes connected by an edge. The signalling network was used to generate t_{50} variable with 2074 receptor-TF pairs for all samples. t_{50} is defined as the number of time steps needed to reach 50% of the maximum signal received by the TFs over time. The obtained t_{50} data matrix contains scores for each receptors-TF pair per sample representing sample-specific network connectivity. For each gene, a faster t_{50} (low t50 score) means that more signal is being transduced to the connected gene by connected genes.

Subnetwork identification

The subnetwork is a simplification of the entire global network. The subnetwork is created by shortest paths (using t_{50} values of a gene) between the receptor and the ten top TFs. Note that for each gene, a lower t_{50} indicates a greater accumulation of diffused signal. The top ten receptor-TF pairs (Table S4) all have an AUC > 0.78 (VDZ response vs non-response). The igraph (https://igraph.org/) package was used to plot the subnetworks. For each subnetwork, the sum of the gene t_{50} values for each branch (receptor to TF) for each sample was calculated. The group branch sums were used to compare the groups.

Statistical analysis

Statistical significance for differentially expressed genes was performed using linear modelling. Multiple testing correction was done with the method of Benjamini and Hochberg [37]. Exploratory data visualisation was done using principal component analysis (PCA). 'nnet' a deep learning-based method was used [38] with tenfold cross-validation repeated 20 times using average accuracy to select the final model. Prediction results were evaluated by area under the receiver operating curve (AUC). GO enrichment analysis was performed using the clusterProfiler, Bioconductor package [39]. For comparing the responder and non-responder sample groups, we used Wilcoxon test which is then evaluated using p-value measure.

Results

Testing VDZ-specific gene expression

We first assessed four VDZ-specific genes such as MADCAM1, VCAM1, Integrin Subunit Beta 4 (ITGB4), and Integrin Subunit Beta 7 (ITGB7) which are pivotal players in the VDZ drug inhibition of the interaction between α4β7 integrin on T cells with MAdCAM1. To test if these genes can predict VDZ response, we used a linear model on the four integrin-specific genes. Only VCAM1 obtained a significant *p*-value (0.003, AUC = 0.68) and no significance was found in the other three genes (Table 3).

Comparison with a reference method

We used a deep learning method, 'nnet', a feed-forward neural network algorithm to test the predictive ability to separate

Table 3 Adjusted p-value of VDZ-specific genes calculated using linear modelling

Genes	Adj. p-value
ITGB4	0.072
ITGB7	0.94
MADCAM1	0.49
VCAM1	0.003

treatment responder and non-responder patients. Using ROC (receiver operating characteristic curve), we obtained an AUC of 0.76 for the gene expression (Fig. 1). Similarly, we did a ROC analysis for the diffusion model to compare the predictive ability. We found 39 receptor-TF pairs with AUC > 0.78 (Table S4). The top-scoring discriminators were receptor-TF pairs free fatty acid receptor 2 (FFAR2) nuclear respiratory factor 1 (NRF1), colony-stimulating factor 3 receptor (CSF3R)-RELB, and integrin subunit beta 4 (ITGB4)-ETS proto-oncogene 1, transcription factor (ETS1) with AUC ~ 0.80. These pairs show distinct differences between VDZ responder and VDZ non-responder patients before and after treatment (Fig. 1A and C).

Comparing gene expression vs. diffusion model

To compare the predictive ability of the diffusion model against the gene expression data, we used ROC analysis to select the best predictor receptor-TF pairs. Then, we applied PCA on gene expression vs. diffusion model feature score (t_{50}) for all VDZ responders, non-responders, and controls samples (Fig. 2A and B). The diffusion model demonstrates an improved predictive power for the separation of patients with VDZ therapy response (see PC1, Fig. 2A) as compared to gene expression. In addition, we applied PCA on receptor FFAR2 which is the best predictor for a subnetwork using the diffusion model result with top receptor-TF AUC score (~ 0.81) and compared it with FFAR2 gene expression (Fig. 3A and B). We found that the PCA of the diffusion model using t_{50} branch sums of the shortest paths could separate VDZ responder and the non-responder group as compared to the gene expression.

Characterising individualised pathways for treatment response patient groups

Each UC patient exhibits heterogeneity in their network connectivity. Differences in patient-specific networks related to immunological pathways may cause UC pathogenesis. Anti-integrin therapies perturb immunological and inflammatory pathways besides cell trafficking interference [40]. For generating individualised subnetworks, we selected the top 10 receptor-TF pairs (AUC > 0.79), which show the best discriminatory ability (Figure S2). To discriminate VDZ treated responders and non-responders with pre-and post-treatment status, we used t_{50} scores of the diffusion model. We found that diffusion results with the FFAR2 receptor gene to TFs such as NRF1, ETS Like-1 (ELK1), RELB, ETS Like-1 (RFX3) and transcription factor AP-2 alpha (TFAP2A) demonstrate the best discriminatory ability for separating therapy responder and non-responder patients (Fig. 4B and Table S4). To test individual-level differences in the signalling pathway for the patient groups, we generated a simplified version of the overall signalling pathway into the subnetwork describing the diffusion of signal from receptor FFAR2 that passes transducers, to downstream TFs (Fig. 4A). For FFAR2 subnetwork, we selected the top 10 TFs selected with receptor-TF pairs (AUC > 0.77) that separate VDZ responders from non-responder UC patient groups. We found that patient with VDZ non-response exhibits quicker signalling as compared to the patients with VDZ response and controls.

Identification of patient-specific signalling pathways

To identify patient-specific signalling pathways, we use the shortest paths in the subnetwork of identified top receptor-TF pairs (Table S4). The top pairs were selected by ROC analysis (Figure S3). Using selected genes in the branch length of the shortest paths, we found the distinct separation of VDZ responders and controls from the non-responders by diffusion model in contrast with gene expression (Fig. 5A and B).

Testing on the alternate anti-integrin biologic drug (Etrolizumab)

For testing the diffusion model on an alternate anti-integrin biologic, we retrieved the publicly available published RNA seq data (GSE72819) with baseline biopsies from Etrolizumab-treated UC patients (Table 4). This dataset contains 70 samples with 58 non-responder and 12 responder UC patient samples with non-IBD controls. We used the same network generated in VDZ training dataset. We found an acceptable AUC of 0.72 for Etrolizumab dataset (Table S5). Next, we compared the obtained diffusion result with 'nnet' prediction, and we found an AUC of 0.69. To check the consistency of the predictive ability of the model on two separate datasets, we applied Pearson's product-moment correlation test [41] on two sets of

Table 4 Baseline characteristics of the patient treated with etrolizumab*

Patient baseline characteristics	Values - mean (SD) or number (%)
Age, years	40·3 (13·4)
Male	24 (60%)
White	38 (95%)
Bodyweight, kg	74·8 (17·1)
Duration of UC, years	8·0 (7·1)
Concomitant medication use	
Corticosteroids	18 (45%)
Dose, mg/day	14·5 (5·7)
Immunosuppressants	14 (35%)
Mesalazine	25 (63%)
Previous anti-TNF therapy	28 (70%)
No response to previous anti-TNF therapy	26 (65%)
Unacceptable adverse event	2 (5%)
Disease extent	
Rectosigmoid	8 (20%)
Left sided	14 (35%)
Pancolitis or extensive	18 (45%)
Non-specified	0
Mayo Clinic Score	9·2 (1·6)
CRP, mg/dL	1·8 (2·6)
Faecal calprotectin, µg/g	1301·3 (1482·6)

* Data adapted from reference Tew et al. [66]

AUCs obtained from Vedolizumab and Etrolizumab datasets. We found a significant correlation of 0.68 (95% Confidence interval) with p-value < 0.01.

Discussion

In this study, we focused on the signalling pathway signatures that stratify VDZ responders from non-responders. UC patients with non-response to VDZ and other biologic therapies have differences in their immune and inflammatory pathways [42]. Using a diffusion model, we generated subnetworks that represent the signalling signatures to discriminate VDZ treatment responders from non-responders. These subnetworks highlight the underlying UC-associated immuno-inflammatory pathways. The diffusion model uses the original gene expression data as well as prior biological knowledge for generating edge-weighted signalling network graphs to delineate the T-cell receptor signalling pathway.

After applying a linear model for gene expression of four pivotal VDZ-specific genes, only VCAM1 was found significantly different between the responder and non-responder patient groups (Table 3). In a previously published study on dextran sodium sulphate (DSS)-induced colitis, Soriano, A. et al. demonstrated the functional role of VCAM1 as a mediator of leukocyte adhesion in colitis and a potent therapeutic effect on immunoneutralisation as compared to MAdCAM-1 and intercellular adhesion molecule 1 (ICAM-1) [43]. Increased expression of VCAM1 in colonic biopsies from patients with IBD is associated with flare-ups leading to disease onset [44]. Gene expression variability is much higher in VCAM1 as compared to MADCAM1. VCAM1 has previously been shown to provide a reliable measure of predicting anti-TNFα therapy response [44].

Exploratory analysis using the diffusion model provides a better stratification of treatment response groups as compared to gene expression alone (Fig. 2). For testing alternate prediction tools for treatment response stratification, we used 'nnet' a deep learning algorithm that has recently been utilised for biomarker discovery [38]. We have found that 'nnet' and diffusion model features facilitate separation of the VDZ treatment response groups, however, 'nnet' fails to outperform the diffusion model (Fig. 1A and B). Here, we can argue that the higher predictive ability of the diffusion model is because of the non-linear transformation of gene expression which is derived from the nature of the signalling network. Another argument could be that the diffusion model enables the inclusion of relevant prior knowledge

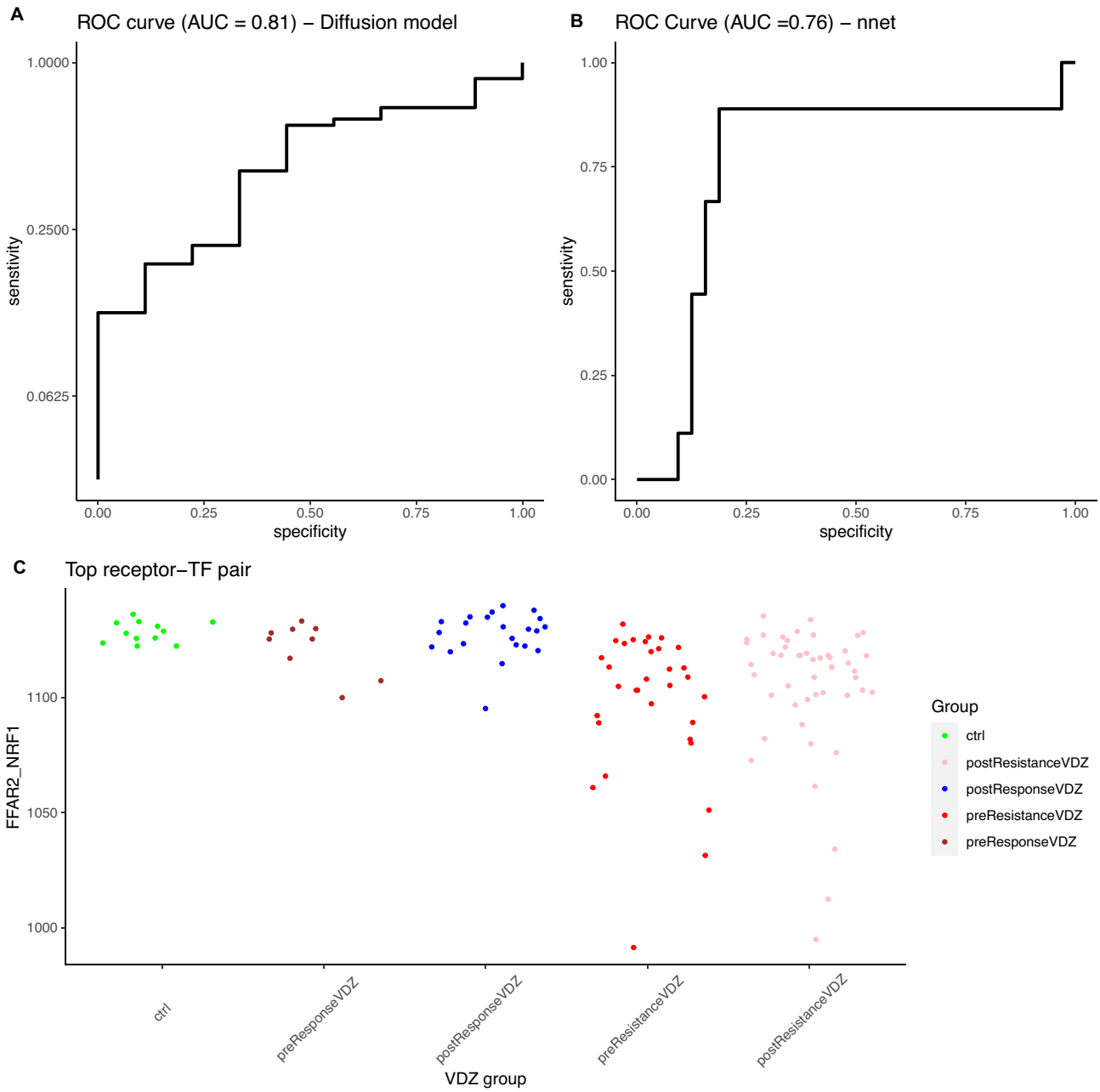


Fig. 1 Predictive analysis **(A)** ROC analysis using diffusion model represents AUC = 0.81 for receptor-TF pair (FFAR2-NRF1), which separates treatment responder and non-responder sample groups. **(B)** ROC analysis of 'nnet' method represents AUC = 0.76 with a complete gene expression matrix. **(C)** Dot plot represents the differences of t_{50} in the treatment response groups for best discriminant receptor-TF pair FFAR2-NRF1, separating responder pre-and post-treatment, non-IBD controls from the non-responder group. preResponseVDZ, postResponseVDZ represents biopsies obtained at W0 or W (6–52), respectively, from patients that respond to treatment. PreResistantVDZ, postResistantVDZ represents biopsies obtained at W0 and W (6–52), respectively, from patients with non-response

which explains the underlying biological determinants of VDZ therapeutic response as compared to 'nnet' which considers only gene expression.

A previous study has shown that reduction in faecal calprotectin after induction with anti-TNFα treatment, correlates with endoscopic remission. However, the calprotectin level at W0 is a poor predictor of therapy response [45]. Whereas diffusion model could predict the therapy response at W0 using gene expression data, therefore contributing a prognostic value at an early stage of UC.

A recently published study has found distinct signature genes with mucosal gene expression at baseline for VDZ treated UC patients [40]. While comparing these identified genes with the independent VDZ cohort, only about

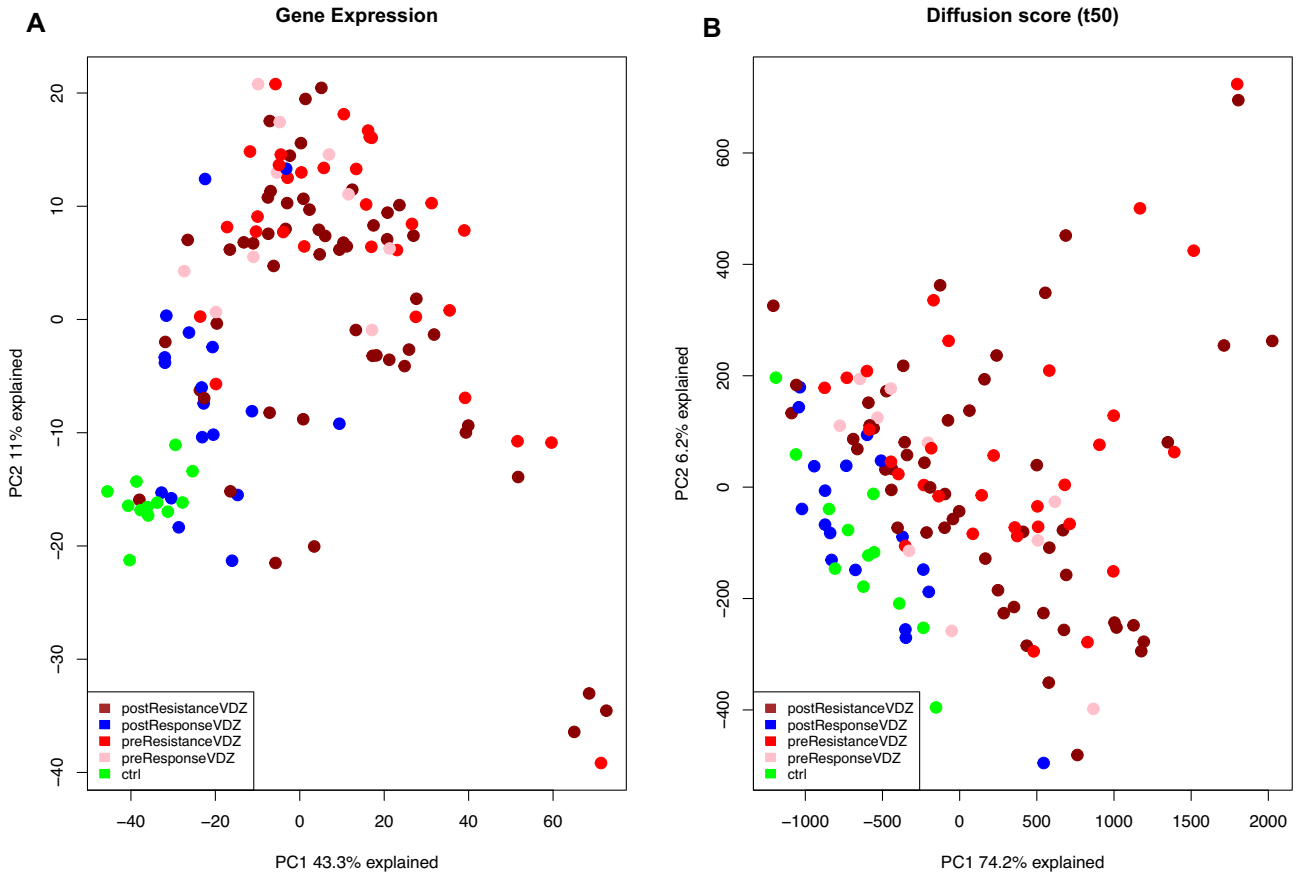


Fig. 2 PCA plot showing the difference in UC VDZ treatment response groups. **(A)** Gene expression of non-responder VDZ pre-treatment, Non-responder VDZ post-treatment, responder VDZ pre-treatment, non-responder VDZ post-treatment, and controls. **(B)** Diffusion score (t_{50}) of non-responder VDZ pre-treatment, non-responder VDZ post-treatment, responder VDZ pre-treatment, non-responder VDZ post-treatment, and controls

a quarter of the significantly differentially regulated genes were reproducible in the independent cohort. For assessing the model's reproducibility and generalisability, we used the alternate anti-integrin drug (Etrolizumab) comprising UC patients with treatment responders and non-responder without non-IBD controls. We found 40 best receptor-TF pairs with an AUC > 0.68 with a best receptor-TF pair CD40-NFKB2 (AUC = 0.72) (Table S4). Notably, on both anti-integrin treatment datasets, the diffusion model demonstrated high predictive performance for stratifying responders from non-responders in comparison to 'nnet'.

Our analysis revealed 48 receptors-TF pairs with AUC > 0.76 that separate the VDZ non-the responders from the responders' group before and after treatment (Table S4). As expected, the top identified receptor-TF pairs include genes that have a role in regulating intestinal inflammation and involvement in UC pathogenesis. In Fig. 4B, FFAR2 receptor gene to TFs such as NRF1, RELB, early growth response 1 (EGR1), and nuclear factor-kappa B subunit 1 (NFKB1) separates the pre-treatment non-responder vs. responders using branch sum (t_{50}) of the shortest path. FFAR2 is a G protein-coupled receptor (GPCR) reported to be a critical precursor of signalling molecules involved in regulating whole-body energy homeostasis, inflammatory and immune responses in the intestine [46, 47]. Non-responder patients have more signals from connected genes that result in quicker signals as compared to the controls and patients in the responder group. The subnetwork obtained using the best receptor-TF pair FFAR2-NRF1 shows the signal from receptor FFAR2 through transducer route 1 (TNF-RACK1-TRAF2) and transducer route 2 (TNF-CAV1-TRAF2) to TF NRF1 (Table S6).TRAF2 is a member of the TNF-receptor-associated factor (TRAF) protein family which directly associates with TNF as a major signal transducer for TNFα-mediated activation of JNK and NFκB [48, 49]. Through NFκB activation, TRAF2 regulates anti-apoptotic signalling by interacting with apoptosis inhibitors [50]. Adjacent to TRAF2 in the signalling network (Fig. 4), RACK1, which is an adaptor molecule that binds to the key signalling molecules

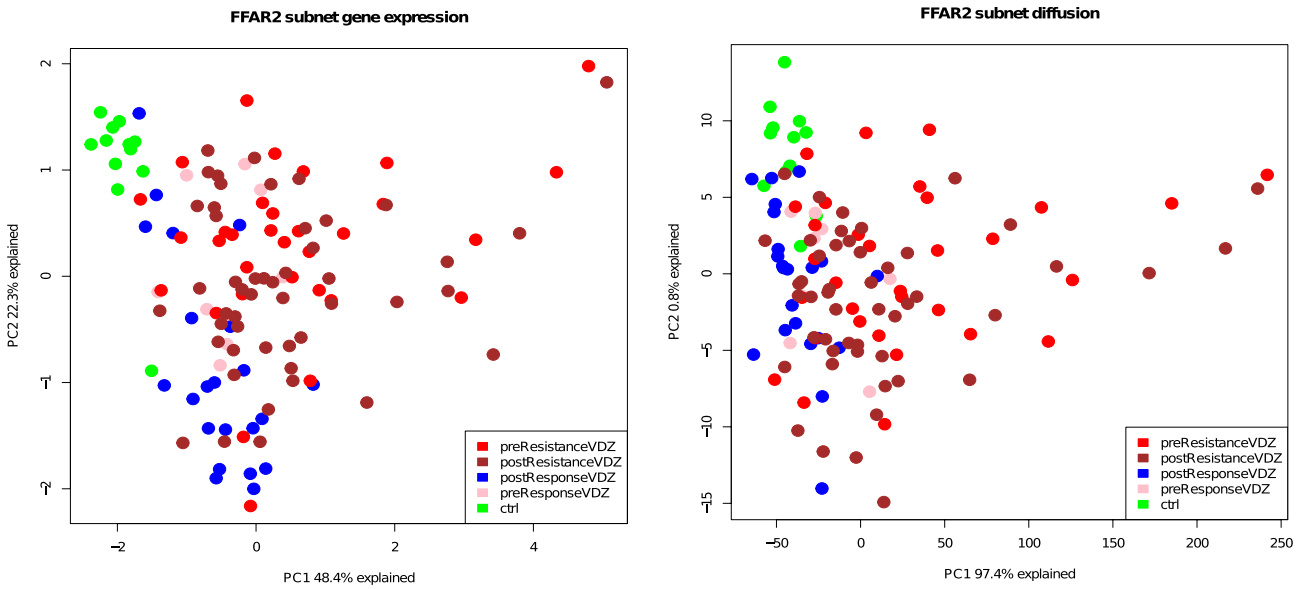


Fig. 3 PCA plot on FFAR2 receptor gene on the shortest path between receptor and TFs showing the difference in UC VDZ treatment response groups. **(A)** FFAR2 receptor gene subnet using gene expression of non-responder VDZ pre-treatment, non-responder VDZ post-treatment, responder VDZ pre-treatment, non-responder VDZ post-treatment, and non-IBD controls. **(B)** FFAR2 t_{50} sample branch sums of the shortest paths to top 10 AUC TFs from receptor FFAR2 of non-responder VDZ pre-treatment, non-responder VDZ post-treatment, responder VDZ pre-treatment, non-responder VDZ post-treatment, and non-IBD controls

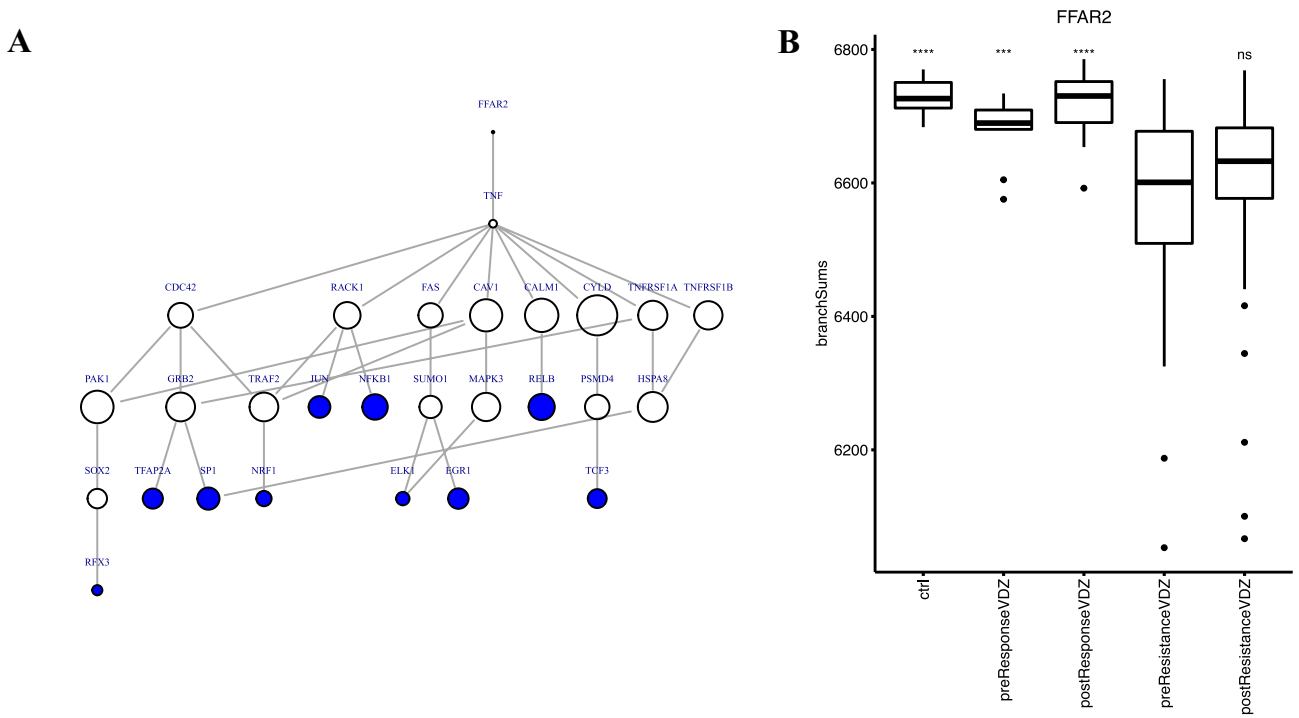


Fig. 4 Explanatory analysis **(A)** A simplified subnetwork generated by the shortest paths between FFAR2 to the top 10 AUC TFs. The shortest paths were defined by network diffusion values (t_{50}). The size of the node represents the differences between t_{50} between the responder and non-responder groups. A solid black line from FFAR2 to NRF1 represents the shortest path for the top receptor-TF pair FFAR2-NRF1. Red colour indicates receptors, white indicates transducers, and blue colour indicates TFs. **(B)** Box plot shows a separation of treatment response groups using the mean distance of branches in the sub-network. Higher branch length represents slower diffusion of signal in the responder and the non-IBD controls as compared to the non-responder group

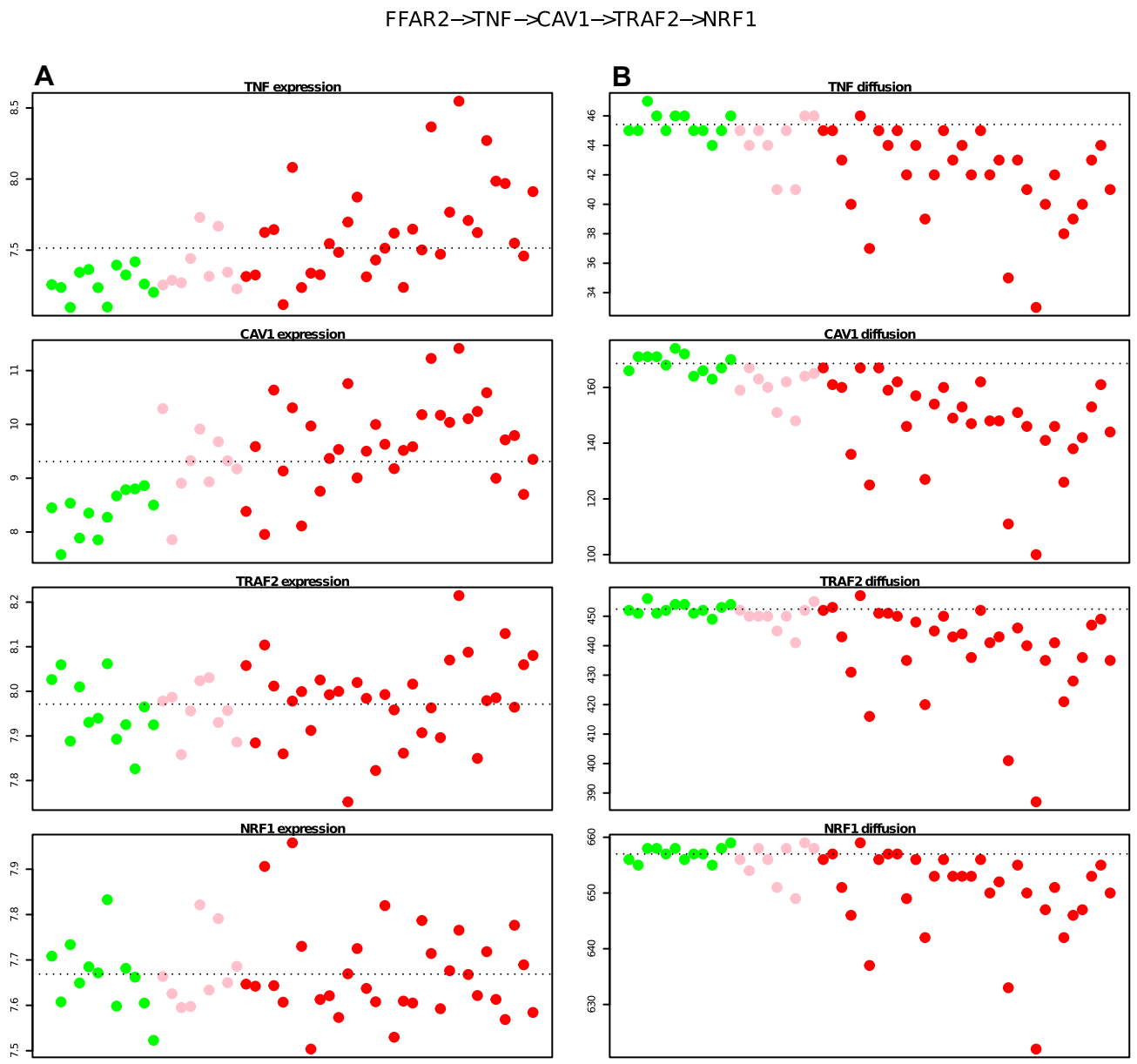


Fig. 5 Comparison of gene expression vs. t_{50} diffusion values in path 'FFAR2-TNF-CAV1-TRAF2-NRF1' **(A)** Gene expression of shortest path genes. **(B)** t_{50} diffusion values of path genes. The horizontal dotted line represents the mean values in controls. Pre-VDZ responders (pink), pre-VDZ non-responders (red), and controls (green) are indicated

involved in the cell migration, integrin adhesion and activity, and T-cell apoptosis [51–54]. A previous study demonstrates the role of RACK1 as a negative regulator of NF-κB signalling, NF-κB-mediated cytokine induction and inflammatory reactions [55]. Next, CAV1 is a gene involved in diverse signalling pathways and plays an essential role in cell proliferation, apoptosis, lipid migration, and exhibits a protective role in intestinal inflammation for IBD [56–58]. We hypothesise that TNF, RACK1, CAV1, and TRAF2 are part of the protein complex in which TNF and TRAF2 are connected to the TNFRSF1A. This might modulate downstream signalling to transcription factor NRF1 (Fig. 4) which is involved in maintaining organ integrity by regulating cytoprotective defences through cellular redox homeostasis [59, 60] by preventing cells against proteasome inhibition through regulation of proteasome gene expression. With a higher accumulation of proteasome inhibitors, NRF1 loses its potency to initiate transcription [61]. Some studies showed a reduction of intestinal inflammation by targeting immunoproteasome that attenuates proinflammatory signalling in DSS-induced colitis study on mice and IBD patients [62–64]. Targeting

NRF1-mediated endoplasmic reticulum-associated degradation (ERAD) pathway could increase the therapeutic efficacy of proteasome inhibitor drugs for providing readily actionable targets [65].

Conclusions

We used a network-based diffusion model to highlight genes and their interactions in signalling pathways which may be predictive in response to the anti-integrin treatment. In our case, the diffusion model outperformed a deep learning method (nnet) and can give comparable prognostic ability at initial diagnosis to longer-term monitoring of calprotectin. The obtained subnetworks feature genes involved in cytokine and fatty acid signalling. The results suggest that anti-integrin drug responses in cytokine and fatty acid signalling pathways can discriminate UC patient populations. As the availability of high throughput RNA sequencing in the clinic increases, these findings may offer useful insights into the development of clinical decision-making to aid in selecting UC treatment strategies.

Abbreviations UC: Ulcerative colitis; IBD: Inflammatory bowel disease; TNF: Tumour necrosis factor; IFX: Infliximab; VDZ: Vedolizumab; ETR: Etrolizumab; 5-ASA: 5-Aminosalicylic acid; GEO: Gene Expression Omnibus; IQR: Interquartile range; ROC: Receiver operating characteristic; AUC: Area under the ROC curve; GO: Gene ontology; TF: Transcription factor; MAdCAM-1: Mucosal vascular address in cell adhesion molecule 1; VCAM-1: Vascular cell adhesion molecule 1; ICAM-1: Intercellular adhesion molecule 1; ITGB4: Integrin subunit beta 4; ITGB7: Integrin subunit beta 7; lncRNAs: Long non-coding ribonucleic acid; PPI: Protein–protein interaction; FFAR2: Free fatty acid receptor 2; DSS: Dextran sodium sulphate

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00384-022-04176-w.

Acknowledgements Not applicable.

Authors' contributions AS, EA, CF, and RHP contributed to conception and design of the article. AS wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding Open Access funding provided by UiT The Arctic University of Norway. This work has been supported by the Northern Norway Regional Health Authority [Strategisk-HN-10–16].

Availability of data and material Public datasets used for analysis in this study are available in the Gene Expression Omnibus (GSE73661, GSE72819) repository.

Declarations

Ethics approval Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Ordás I, Eckmann L, Talamini M et al (2012) Ulcerative colitis. Lancet 380:1606–1619. https://doi.org/10.1016/S0140-6736(12)60150-0
2. Chen ML, Sundrud MS (2016) Cytokine networks and T-cell subsets in inflammatory bowel diseases. Inflamm Bowel Dis 22:1157–1167. https://doi.org/10.1097/MIB.0000000000000714
3. Zhang Y-Z, Li Y-Y (2014) Inflammatory bowel disease: pathogenesis. World J Gastroenterol 20:91
4. Ungaro R, Mehandru S, Allen PB et al (2017) Ulcerative colitis. Lancet 389:1756–1770. https://doi.org/10.1016/S0140-6736(16)32126-2
5. Liverani E, Scaioli E, John Digby R et al (2016) How to predict clinical relapse in inflammatory bowel disease patients. World J Gastroenterol 22:1017–1033. https://doi.org/10.3748/wjg.v22.i3.1017
6. De Souza HSP, Fiocchi C (2016) Immunopathogenesis of IBD: Current state of the art. Nat Rev Gastroenterol Hepatol 13:13–27. https://doi.org/10.1038/nrgastro.2015.186
7. Stawowczyk E, Kawalec P (2018) A Systematic Review of the Cost-Effectiveness of Biologics for Ulcerative Colitis. Pharmacoeconomics 36:419–434
8. Riaz N, Wolden SL, Gelblum DY, Eric J (2016) Market Share and Costs of Biologic Therapies for Inflammatory Bowel Disease in the United States 118:6072–6078. https://doi.org/10.1111/apt.14430.Market
9. Neurath MF (2017) Current and emerging therapeutic targets for IBD. Nat Rev Gastroenterol Hepatol 14:269–278. https://doi.org/10.1038/nrgastro.2016.208
10. Torres J, Danese S, Colombel JF (2013) New therapeutic avenues in ulcerative colitis: Thinking out of the box. Gut 62:1642–1652. https://doi.org/10.1136/gutjnl-2012-303959
11. Garud S, Peppercorn MA (2009) Ulcerative colitis: Current treatment strategies and future prospects. Therap Adv Gastroenterol 2:99–108. https://doi.org/10.1177/1756283X09102329
12. Coskun M, Steenholdt C, de Boer NK, Nielsen OH (2016) Pharmacology and Optimization of Thiopurines and Methotrexate in Inflammatory Bowel Disease. Clin Pharmacokinet 55:257–274. https://doi.org/10.1007/s40262-015-0316-9
13. Ardizzone Sandro and Porro Gabriele Bianchi (2005) Biologic therapy for Inflammatory Bowel Disease. Drugs 65:2253–2286. https://doi.org/10.1007/978-3-319-60276-9_11
14. Smids C, Horjus Talabur Horje CS, Drylewicz J et al (2017) Intestinal T Cell Profiling in Inflammatory Bowel Disease: Linking T Cell Subsets to Disease Activity and Disease Course. J Crohn's Colitis 12:465–475. https://doi.org/10.1093/ecco-jcc/jjx160

15. Rath T, Billmeier U, Ferrazzi F et al (2018) Effects of Anti-Integrin Treatment With Vedolizumab on Immune Pathways and Cytokines in Inflammatory Bowel Diseases. Front Immunol 9:1700. https://doi.org/10.3389/fimmu.2018.01700
16. Feagan BG, Rutgeerts P, Sands BE et al (2013) Vedolizumab as Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med 369:699–710. https://doi.org/10.1056/nejmoa1215734
17. Arijs I, De Hertogh G, Lemmens B et al (2018) Effect of vedolizumab (anti-α4β7-integrin) therapy on histological healing and mucosal gene expression in patients with UC. Gut 67:43–52. https://doi.org/10.1136/gutjnl-2016-312293
18. Dai B, Hackney JA, Ichikawa R et al (2021) Dual targeting of lymphocyte homing and retention through α4β7 and αEβ7 inhibition in inflammatory bowel disease. Cell Reports Med 2:100381. https://doi.org/10.1016/j.xcrm.2021.100381
19. Wyant T, Fedyk E, Abhyankar B (2016) An overview of the mechanism of action of the monoclonal antibody vedolizumab. J Crohn's Colitis 10:1437–1444. https://doi.org/10.1093/ecco-jcc/jjw092
20. Rosario M, Dirks NL, Milch C et al (2017) A Review of the Clinical Pharmacokinetics, Pharmacodynamics, and Immunogenicity of Vedolizumab. Clin Pharmacokinet 56:1287–1301. https://doi.org/10.1007/s40262-017-0546-0
21. Nourshargh S, Hordijk PL, Sixt M (2010) Breaching multiple barriers: Leukocyte motility through venular walls and the interstitium. Nat Rev Mol Cell Biol 11:366–378. https://doi.org/10.1038/nrm2889
22. Vermeire S, Loftus EV, Colombel JF et al (2017) Long-term efficacy of vedolizumab for Crohn's disease. J Crohn's Colitis 11:412–424. https://doi.org/10.1093/ecco-jcc/jjw176
23. Keir ME, Fuh F, Ichikawa R et al (2021) Regulation and Role of αE Integrin and Gut Homing Integrins in Migration and Retention of Intestinal Lymphocytes during Inflammatory Bowel Disease. J Immunol 207:2245–2254. https://doi.org/10.4049/jimmunol.2100220
24. de Krijger M, Wildenberg ME, Mookhoek A et al (2021) Expression of MAdCAM-1 and Gut-homing T Cells in Inflamed Pouch Mucosa. J Crohn's Colitis 15:1491–1499. https://doi.org/10.1093/ecco-jcc/jjab041
25. Bryant RV, Sandborn WJ, Travis SPL (2015) Introducing vedolizumab to clinical practice: Who, when, and how? J Crohn's Colitis 9:356–366. https://doi.org/10.1093/ecco-jcc/jjv033
26. Khanna R, Afif W (2021) Ustekinumab for Ulcerative Colitis. Gastroenterology 160:2184–2186. https://doi.org/10.1053/j.gastro.2021.02.002
27. D'Amico F, Parigi TL, Fiorino G et al (2019) Tofacitinib in the treatment of ulcerative colitis: efficacy and safety from clinical trials to real-world experience. Therap Adv Gastroenterol 12:1756284819848631–1756284819848631. https://doi.org/10.1177/1756284819848631
28. Witte J, Shivananda S, Lennard-Jones JE et al (2000) Disease outcome in inflammatory bowel disease: Mortality, morbidity and therapeutic management of a 796-person inception cohort in the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). Scand J Gastroenterol 35:1272–1277. https://doi.org/10.1080/003655200453610
29. Singh A, Anderssen E, Fenton CG, Paulssen RH (2021) Identifying anti-TNF response biomarkers in ulcerative colitis using a diffusion-based signalling model. Bioinforma Adv 1:1–8. https://doi.org/10.1093/bioadv/vbab017
30. Geboes K, Riddell R, Öst A et al (2000) A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. Gut 47:404–409. https://doi.org/10.1136/gut.47.3.404
31. Jauregui-Amezaga A, Geerits A, Das Y et al (2017) A Simplified Geboes Score for Ulcerative Colitis. J Crohns Colitis 11:305–313. https://doi.org/10.1093/ecco-jcc/jjw154
32. Lewis JD, Chuai S, Nessel L et al (2008) Use of the noninvasive components of the Mayo score to assess clinical response in ulcerative colitis. Inflamm Bowel Dis 14:1660–1666. https://doi.org/10.1002/ibd.20520
33. Schlauch D, Paulson JN, Young A et al (2017) Estimating gene regulatory networks with pandaR. Bioinformatics 33:2232–2234. https://doi.org/10.1093/bioinformatics/btx139
34. Veres DV, Gyurkó DM, Thaler B et al (2015) ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. Nucleic Acids Res 43:D485–D493. https://doi.org/10.1093/nar/gku1007
35. Marbach D, Lamparter D, Quon G et al (2016) Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods 13:366–370. https://doi.org/10.1038/nmeth.3799
36. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci 100:9440–9445. https://doi.org/10.1073/pnas.1530509100
37. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
38. Mallik S, Seth S, Bhadra T, Zhao Z (2020) A linear regression and deep learning approach for detecting reliable genetic alterations in cancer using dna methylation and gene expression data. Genes (Basel) 11:1–15. https://doi.org/10.3390/genes11080931
39. Yu G, Wang LG, Han Y, He QY (2012) ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi A J Integr Biol 16:284–287. https://doi.org/10.1089/omi.2011.0118
40. Gazouli M, Dovrolis N, Bourdakou MM et al (2021) Response to Anti-α4β7 Blockade in Patients With Ulcerative Colitis Is Associated With Distinct Mucosal Gene Expression Profiles at Baseline. Inflamm Bowel Dis 1–9. https://doi.org/10.1093/ibd/izab117
41. Rodgers JL, Nicewander WA (1988) Thirteen Ways to Look at the Correlation Coefficient. Am Stat 42:59–66. https://doi.org/10.1080/00031305.1988.10475524
42. Soendergaard C, Seidelin JB, Steenholdt C, Nielsen OH (2018) Putative biomarkers of vedolizumab resistance and underlying inflammatory pathways involved in IBD. BMJ Open Gastroenterol5. https://doi.org/10.1136/bmjgast-2018-000208
43. Soriano A, Salas A, Salas A et al (2000) VCAM-1, but Not ICAM-1 or MAdCAM-1, Immunoblockade Ameliorates DSS-Induced Colitis in Mice. Lab Investig 80:1541–1551. https://doi.org/10.1038/labinvest.3780164
44. Gu P, Theiss A, Han J, Feagins LA (2017) Increased Cell Adhesion Molecules, PECAM-1, ICAM-3, or VCAM-1, Predict Increased Risk for Flare in Patients With Quiescent Inflammatory Bowel Disease. J Clin Gastroenterol 51:522–527. https://doi.org/10.1097/MCG.0000000000000618
45. De Vos M, Dewit O, D'Haens G et al (2012) Fast and sharp decrease in calprotectin predicts remission by infliximab in anti-TNF naïve patients with ulcerative colitis. J Crohn's Colitis 6:557–562. https://doi.org/10.1016/j.crohns.2011.11.002
46. Hudson BD, Due-Hansen ME, Christiansen E et al (2013) Defining the molecular basis for the first potent and selective orthosteric agonists of the FFA2 free fatty acid receptor. J Biol Chem 288:17296–17312. https://doi.org/10.1074/jbc.M113.455337
47. Ichimura A, Hasegawa S, Kasubuchi M, Kimura I (2014) Free fatty acid receptors as therapeutic targets for the treatment of diabetes. Front Pharmacol 5:236. https://doi.org/10.3389/fphar.2014.00236
48. Au P-YB, Yeh W-C (2007) Physiological roles and mechanisms of signaling by TRAF2 and TRAF5. Adv Exp Med Biol 597:32–47. https://doi.org/10.1007/978-0-387-70630-6_3

49. Chung JY, Park YC, Hong Ye HW (2002) TRAFs: transducers for TNF and IL-1/Toll-like receptor superfamilies (p. 679). J Cell Sci 115:e402–e402
50. Vince JE, Pantaki D, Feltham R et al (2009) TRAF2 must bind to cellular inhibitors of apoptosis for tumor necrosis factor (tnf) to efficiently activate nf-{kappa}b and to prevent tnf-induced apoptosis. J Biol Chem 284:35906–35915. https://doi.org/10.1074/jbc.M109.072256
51. Hermanto U, Zong CS, Li W, Wang L-H (2002) RACK1, an insulin-like growth factor I (IGF-I) receptor-interacting protein, modulates IGF-I-dependent integrin signaling and promotes cell spreading and contact with extracellular matrix. Mol Cell Biol 22:2345–2365. https://doi.org/10.1128/MCB.22.7.2345-2365.2002
52. Mourtada-Maarabouni M, Kirkham L, Farzaneh F, Williams GT (2005) Functional expression cloning reveals a central role for the receptor for activated protein kinase C 1 (RACK1) in T cell apoptosis. J Leukoc Biol 78:503–514. https://doi.org/10.1189/jlb.0205070
53. Adams DR, Ron D, Kiely PA (2011) RACK1, A multifaceted scaffolding protein: Structure and function. Cell Commun Signal 9:22. https://doi.org/10.1186/1478-811X-9-22
54. Liliental J, Chang DD (1998) Rack1, a receptor for activated protein kinase C, interacts with integrin beta subunit. J Biol Chem 273:2379–2383. https://doi.org/10.1074/jbc.273.4.2379
55. Yao F, Long L-Y, Deng Y-Z et al (2014) RACK1 modulates NF-κB activation by interfering with the interaction between TRAF2 and the IKK complex. Cell Res 24:359–371. https://doi.org/10.1038/cr.2013.162
56. Liu P, Rudick M, Anderson RGW (2002) Multiple Functions of Caveolin-1*. J Biol Chem 277:41295–41298. https://doi.org/10.1074/jbc.R200020200
57. Weiss CR, Guan Q, Ma Y et al (2015) The potential protective role of caveolin-1 in intestinal inflammation in TNBS-induced murine colitis. PLoS ONE 10:e0119004–e0119004. https://doi.org/10.1371/journal.pone.0119004
58. Chidlow JH, Greer JJM, Anthoni C et al (2009) Endothelial Caveolin-1 Regulates Pathologic Angiogenesis in a Mouse Model of Colitis. Gastroenterology 136:575-584.e2. https://doi.org/10.1053/j.gastro.2008.10.085
59. Chang W-T, Huang A-M (2004) Alpha-Pal/NRF-1 regulates the promoter of the human integrin-associated protein/CD47 gene. J Biol Chem 279:14542–14550. https://doi.org/10.1074/jbc.M309825200
60. Zhang Y, Xiang Y (2016) Molecular and cellular basis for the unique functioning of Nrf1, an indispensable transcription factor for maintaining cell homoeostasis and organ integrity. Biochem J 473:961–1000. https://doi.org/10.1042/BJ20151182
61. Sha Z, Goldberg AL (2014) Proteasome-Mediated Processing of Nrf1 Is Essential for Coordinate Induction of All Proteasome Subunits and p97. Curr Biol 24:1573–1583. https://doi.org/10.1016/j.cub.2014.06.004
62. Schmidt N, Gonzalez E, Visekruna A et al (2010) Targeting the proteasome: partial inhibition of the proteasome by bortezomib or deletion of the immunosubunit LMP7 attenuates experimental colitis. Gut 59:896–906. https://doi.org/10.1136/gut.2009.203554
63. Visekruna A, Joeris T, Seidel D et al (2006) Proteasome-mediated degradation of IκBα and processing of p105 in Crohn disease and ulcerative colitis. J Clin Invest 116:3195–3203. https://doi.org/10.1172/JCI28804
64. Cleynen I, Vazeille E, Artieda M et al (2014) Genetic and microbial factors modulating the ubiquitin proteasome system in inflammatory bowel disease. Gut 63:1265–1274. https://doi.org/10.1136/gutjnl-2012-303205
65. Northrop A, Byers HA, Radhakrishnan SK (2020) Regulation of NRF1, a master transcription factor of proteasome genes: implications for cancer and neurodegeneration. Mol Biol Cell 31:2158–2163. https://doi.org/10.1091/mbc.E20-04-0238
66. Tew GW, Hackney JA, Gibbons D et al (2016) Association Between Response to Etrolizumab and Expression of Integrin alphaE and Granzyme A in Colon Biopsies of Patients With Ulcerative Colitis. Gastroenterology 150(477–87):e9. https://doi.org/10.1053/j.gastro.2015.10.041

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Paper-III

Modelling individual variability of pattern recognition receptor pathway response in IBD

Endre Anderssen [2], Amrinder Singh [1], Christopher G. Fenton [2], Ruth H. Paulssen [1,2]

[1]Clinical Bioinformatics Research Group, Department of Clinical Medicine, Faculty of Health Sciences, UiT- The Arctic University of Norway, Tromsø, Norway.

[2]Genomics Support Centre Tromsø, UiT- The Arctic University of Norway, Department of Clinical Medicine, Faculty of Health Sciences, UiT- The Arctic University of Norway, Tromsø, Norway.

Abstract

Background

Autoinflammatory diseases like inflammatory bowel diseases (IBD) have complex multifactorial etiologies that often include the interaction between the immune system and the microbiome. The immune system uses pattern recognition receptors (PRRs) to detect molecules associated with pathogens and damaged cells. These receptors are connected to transcriptional regulators of the immune system including Nuclear Factor κB (NF-κB), Activator protein 1 (AP1), and interferon-regulatory factor (IRF) protein family members. How strongly these transcriptional programs are activated by PRR can vary greatly from person to person.In this study, we aim to use canonical pathways to predict the magnitude activation of NFKB-, AP1- and IRF- target genes after stimulation with PRR ligands.

Results

A network analysis method to evaluate network connectivity between PRRs and transcription factors has been adapted. The method uses gene expression data to personalize signalling networks that connect PRRs to downstream transcription factors (TFs). The personalized network connectivities are compared to activation levels for known TF target genes after PRR stimulation. In two independent datasets, statistically significant correlations are found between the estimated network connectivities and activation of TF target genes. The correlations are higher for extracellular PRRs and at shorter times of stimulation.

Conclusion

To our knowledge, this is the first method to quantify individualized PRR responses based on gene expression data and network models. Improved understanding of an individual's response to PRR ligands may have implications for treatment choices and may open treatment opportunities using PRR ligands as alternatives or adjuvants to current treatment options.

Introduction

Ulcerative colitis (UC) is a chronic, inflammatory disease and the exact aetiology of UC is still elusive. The major factors inciting the pathogenesis of UC are genetic, immunological, environmental, and microbial [1]. Among these factors, cellular receptors with innate immune response functions recognise the foreign molecules which trigger the immune responses. Toll-like receptors (TLRs) are among the pattern recognition receptors with a role in innate as well as adaptive immunity (T cell activation). TLRs sense and recognize pathogen-associated molecular patterns (PAMPs) [2,3]. PAMPs are highly conserved microbial structures e.g., unmethylated double-stranded DNA, single-stranded RNA, lipoproteins, lipopolysaccharide (LPS), and flagellin. TLR 1, 2, 3, 4, 5, and 9 identified variants of TLRs expressed in the small and large intestines of humans, whereas TLR 6, 7, and 8 are expressed only in the colon [4]. TLR 1, 2, 4, 5, 6 are transmembrane innate receptors and deliver signalling through IRAK-4 while TLR 3, 4, 7, 8, 9 are expressed in intracellular endosomes which form homodimers after interacting with their ligands via IRAK-4 and IKK [5].

The activation of TLR signalling initiates a cascade of downstream events that may have a role in promoting UC. A study by Sánchez-Muñoz *et al.* have observed differential expression of TLRs in colonic mucosa from UC patients (both quiescent and active) as compared to healthy individuals ($p < 0.04$). This study showed that the expression of TLR 2, 4, 8, and 9 are upregulated in patients with active UC, whereas TLR5 is often overexpressed in patients with active UC as compared to quiescent UC disease, and downregulated in quiescent UC compared to controls with healthy colonic mucosa [6]. TLR4 is the first known important TLR that recognizes LPS in gram-negative bacteria and mediates inflammatory response against invading bacteria and promoting mucosal integrity in UC [7]. TLR4 is induced when derivates from the cell wall of gram-negative bacteria, like lipopolysaccharides, bind to TLR4 on the surface of mononuclear cells. Activation of TLR4 signalling induces activation of nuclear factor-κB (NF-κB) and mitogen-activated protein kinases (MAPK), causing increased cell proliferation and differentiation of macrophages as well as inducing expression of pro-inflammatory cytokines, e.g., TNF-α, interleukin (IL)-6 and IL-12. Increased TNF-α expression might decrease the mucosal barrier function in patients with IBD exacerbating the inflammation [8].

TLRs have multiple roles in UC aetiology such as immune responses, genetics as well as microbiota [2,5,6,9,10]. Besides the role of TLRs in UC pathogenies, there has been growing

attention to UC treatment strategies that highlight immune dysfunction pathways especially toll-like receptors (TLRs)-mediated innate immune dysfunction.

Stimulation of TLRs triggers the activation of signalling cascades, leading to the induction of immune and pro-inflammatory genes. After ligand binding, TLRs dimerize and undergo conformational changes. This is followed by recruitment to the receptor of TIR-domain-containing adaptors including myeloid differentiation primary response protein 88 (MyD88) and TIR-domain-containing adaptor protein-inducing IFN-β (TRIF), which are responsible for the activation of distinct signalling pathways. For appropriate homeostasis, the balance is maintained by multiple negative regulators. The expression of most negative regulators can be induced by the activation of TLRs which uses a negative feedback mechanism to terminate TLRs activation. The regulation of TLR signalling pathways constitutes a complex network. Although it has been studied for many years, additional new components and regulatory aspects of known components continue to be revealed. We believe that additions to the body of information concerning this regulation may provide added possibilities for therapeutic manipulation of these pathways and thus for more effective treatment of human diseases involving TLRs [11].

Another important nucleotide-binding oligomerization domain (NOD)-like receptor] (NOD-2) receptor has a vital role in recognizing pathogen and damage-associated molecular patterns (PAMPs and DAMPs, respectively) in the cytoplasm and eliciting innate immune responses. The expression of NOD2 is induced by bacterial components (e.g., LPS, muramyl dipeptide (MDP)), and thereby largely dependent on the presence of gut microbiota. Mutation of NOD-2 causes dysbiosis between ileal microbiota and mucosal immunity [12]. NOD2 and TLR work in a synergy that affects the polarization of T cell adaptive immunity[13,14].

This study aims at quantifying individual responses to signals that may contribute to the pathological immune response. To model this, we have analysed publicly available gene expression data where gene expression is available before and after stimulation with PRR agonists. Canonical pathway models and network diffusion was used to predict the activation of an individual's transcriptional response using the gene expression profile before stimulation.

Materials and Methods

Data sets

The Gene Expression Omnibus (GEO) was searched for gene expression data from human samples stimulated with pattern recognition ligands (Toll receptors and NOD2) obtained from a large number of different individuals. The stimuli include LPS, MDP R837 CpG-C, LTA, and peptidoglycan, datasets were considered candidates if they contained gene expression data from before and after stimulation of at least 20 subjects.

Signalling network definition

Network interactions were initially taken from a recent review of PRR signalling pathways with a focus on negative regulation [15]. The interaction was validated in the STRING database [16]. Gene symbols were updated using GeneCards [17]. Networks were created individually for TLR2, TLR4, TLR7, TLR9 and NOD2. Networks were stored as R data frames and visualised using the iGraph package in R [igraph.org].

Personalized network connectivity estimation

Network connectivity between PRRs and TFs was computed adopting a recently developed network diffusion algorithm [18]. In brief, the method works as follows: A network model connecting a receptor to one or more TFs is defined as a set of vertexes and edges where the vertexes represent genes/proteins, and the edges represent physical interactions that pass or inhibit signals. The network edges are then initialized with a connectivity value passed on the product of the expression level of the two interacting genes. This in effect personalizes the network as the network layout is shared between samples but the connection weights will depend on the gene expression levels of the individual sample.

At time zero, a unit of the signal is placed on the receptor node. Then at each time interval fluxes along vertexes are calculated. These fluxes are then used to move signal along edges between vertexes and the amount of signal at each vertex is updated. This process is repeated in several time steps. The network connectivity is then evaluated by identifying the time point of maximum variance at the transcription factor and using the accumulated signal from the receptor as a measure of connectivity.

The original implementation of the diffusion method for estimating network connectivity did not account for potential negative feedback mechanisms. All genes accumulated and passed on

signal to neighbours with less signal according to the diffusion law [19]. Negative feedback genes are included as signal sinks that contribute to fluxes but cannot accumulate signals. Consequently, the signal reaching negative regulator genes disappears, and will not be propagated further throughout the network.

Procrustes analysis of gene expression and signal diffusion

To assess if the network diffusion process creates new information or just acts as a linear transform of the original gene expression data, the TLR4 network was run to maximum overall variance and the signal accumulated on all nodes was compared to the original gene expression data on the nodes using Procrustes rotation [20].

Statistical methods

The linear models in the R stats package were used to create trend lines for all samples or just healthy controls. P values were computed to test the hypothesis that the slope of the fitted line is different from zero. Gene expression changes were determined using limma [21]. Genes with an FDR [22] adjusted p-value <0.05 were taken as significantly changed. TF target genes were obtained using OmniPath [23] with the encode-proximal dataset. TF activation was estimated as the average fold change of differentially expressed target genes.

Results

Data sets

The GEO was searched for datasets containing transcriptional data (microarrays or sequencing-based) both before and after stimulation with PRR ligands for a large number of human subjects. Two datasets were identified as suitable for the analysis. Dataset A (GSE103500) [24] contains microarray data for transcriptional response in pairs of unstimulated *ex vivo* blood samples and samples from the same donors stimulated with the PRR ligands for TLR 2,4,7,9. The donors are normal controls and systemic juvenile idiopathic arthritis (sJIA) patients that are either untreated or undergoing treatment with anakinra, an IL-1R antagonist.

Dataset B (GSE137680) [25] similarly contains *ex vivo* samples of peripheral blood mononuclear cells (PBMCs) from IBD patients stimulated with the PRR ligands LPS and MDP. Both data sets passed quality controlled by PCA density plots and expression of the gender-specific protein XIST was consistent with the listed gender of the participant (**Supplementary S1**). The datasets are summarized in **Table 1**.

Table 1 Overview of PRR stimulation experiments

Dataset	Stimuli	PRR	Stimulation time [H]	Number of donors	Diagnosis
A	LPS	TLR4	6	25	Ctrl:13/sJIA:12
A	LTA	TLR2	6	26	Ctrl:14/sJIA:12
A	R837	TLR7	6	24	Ctrl:14/sJIA:10
A	CpG-C	TLR9	6	26	Ctrl:14/sJIA:12
B	LPS	TLR4	16	34	IBD:34
B	MDP	NOD2	16	33	IBD:34

Initial TLR4 model

A signalling network model was created based on a review of Toll-like receptor signalling [15]. For LPS signalling through TLR4, MYD dependent and an MYD independent pathway was defined with connections to the transcription factors NFKB1, FOS, and IRF3. The signalling network is illustrated in **Figure 1**A.

The signal diffusion from TLR4 to the TFs was calculated for the control samples matching LPS stimulated individuals. One unit of signal on the TLR4 receptor and running the diffusion algorithm for 2000-time steps. To obtain a consistent measure of network connectivity from the time vs. signal curves, the time of maximum variability is identified and the accumulated signal at the TF for a given receptor at this time was used as a measure of the receptor to TF connectivity (**Figure 2 A-C**).

Activation of the downstream transcription factors after LPS stimulation was then estimated by first identifying differentially expressed genes (see Materials and methods). Then, the average log2 fold change of the differentially expressed genes that are target genes of NFKB1, FOS, and IRF3 was calculated to estimate the level of activation of each of these transcription factors. A linear model was then fitted relating These activation levels were then correlated to the network connectivity between TLR4 and each TF**.**

The estimated connectivity on the control samples was then related to the activation of NFKB1, FOS, and IRF3 target genes using a linear model. In all cases, the fitted line has a positive slope, and the correlation coefficients are highly significant (**Figure 2 D-F**).

Comparing network results to gene expression data

Procrustes analysis of accumulated signal on all pathway members at time step 1000. One matrix of gene expression data for the pathway members was compared to a matrix of an accumulated signal using Procrustes rotation. Only **0.39%** if the variability of the diffusion result is a linear fit from the gene expression.

Remaining receptors in dataset A

Network diffusion and TF activity estimation were then carried out in the same manner as for TLR4 for the networks connecting TLR2, TLR7 and TLR9 to the TF NFKB1 and FOS. In all cases, the slope of the linear fit between the network connectivity and the corresponding TF activation are all positive. However, a trend was identifiable among the normal controls only as a large number of the sJIA patients show little or no activation upon stimulation with R837 and CpG-C. Correlations and P-values for these Toll receptor agonists are therefore calculated for using the normal control samples only (**Figure 3**).

PBMCs from IBD patients

Finally, the method was applied to samples from Dataset B where the connectivity between TLR4 and NOD2 and activation of NFKB1, FOS, and IRF3 target genes were examined. Significant correlations were identified for LPS confirming the findings in Dataset A and a marginal significant finding for MDM via the NOD2 pathway (**Figure 4**). The correlations and significance values are all listed in **Table 2**.

Table 2 Summary of correlations p values and experimental conditions

Dataset	Stimuli	Receptor	TF	Correlation	P-value	Extra/Intra -cellular
A	LPS	TLR4	NFKB1	0.71	0.00007	Extra
A	LPS	TLR4	FOS	0.6	0.002	Extra
A	LPS	TLR4	IRF3	0.33	0.1	Extra
A	LTA	TLR2	NFKB1	0.71	0.00006	Extra
A	LTA	TLR2	FOS	0.58	0.002	Extra
A	R837	TLR7	NFKB1	0.51*	0.06	Intra
A	R837	TLR7	FOS	0.38*	0.18	Intra
A	CpG-C	TLR9	NFKB1	0.28*	0.34	Intra

A	CpG-C	TLR9	FOS	0.16*	0.59	Intra
B	LPS	TLR4	NFKB1	0.35	0.04	Extra
B	LPS	TLR4	FOS	0.36	0.03	Extra
B	LPS	TLR4	IRF3	0.10	0.58	Extra
B	MDP	NOD2	NFKB1	0.34	0.055	Intra

* sIJA patients excluded from statistical analysis

Discussion

There is a highly significant correlation between the network connectivities and TF activation for TLR4 and TLR2 but only marginally significant for TLR7 and NOD2 and quite weak for TLR9.

It is noteworthy that the diffusion model works quite well for the two examples of extracellular receptors TLR4 and TLR2. While showing the worst performance with receptors canonically located to the membrane of liposomes. The current formulation of the method ignores the transport of the ligand into a liposome. Individual differences in drug transport into the cell, liposome formation and any mechanisms preventing the ligand from getting to the receptor will therefore be ignored.

Dataset A overall gives the best correlations. This could be due to the difference in stimulation time between Dataset A (6H) and Dataset B (16H). The longer period allows more time for feedback mechanisms and secondary effects to act, thereby reducing the predictive ability of the network connectivity at the time of stimulation. Indeed, macrophages stimulated with LPS for 4,8,12, and 24H showed that the biggest increase in TNF-α level occurred between 4 and 8 hours, and the TNF-α level remained largely constant thereafter [26].

It is also apparent in Dataset A, that the sIJA patients show less transcriptional activation than predicted. This can be most strongly observed in the case of TLR7, but the effect is also apparent for TLR2-FOS, TLR9-NFKB1, TLR4-FOS, and TLR4 IRF3. In the case of TLR7 stimulation, the degree of TF activation is essentially zero for these patients. This may be due to saturation effects, where the sIJA patients are already responding to the PRR stimuli to a greater or lesser extent. Further stimulation, therefore, gives limited additional effect, despite the genes of the signalling pathways being expressed. It has been suggested that single-strand RNA from the synovial fluid of RA patients can function as a TLR7 endogenous agonist [27].

The worst model performance is seen in the case of TLR9. This receptor is canonically expressed in dendritic cells [28], a much less in more abundant cell types in the blood. Indeed, the gene expression levels of TLR9 in Dataset A are close to the microarray background level (**Supplementary S2**). TLR9 is also unique among PRRs as it requires extensive post-translational modification before it can function as a receptor. It interacts with numerous proteins that mediate trafficking to the endosomes, whereupon cleavage of the protein is necessary before it can interact with CpG DNA [29]. An expanded signalling network model including the proteins responsible for TLR9 trafficking and processing may therefore be necessary to obtain good correlations. This may be of interest because it has been suggested that TLR9 agonists may be able to induce wound healing in UC [30].

Pattern recognition receptors are the eyes and ears of the innate immune system. They are central to the interaction between the microbiota and the host and play an important role in autoimmune and autoinflammatory diseases. To our knowledge, this is the first attempt at modelling individual patients' response to PRR ligands based on their gene expression data. Analysing the gene expression data using signalling networks and diffusion generates new information that is not merely a linear transform of the original gene expression data. Currently, there are considerable individual differences in response to therapy with pro-biotics and faecal microbiota transplantation [31]. Therefore, an improved understanding of an individual's response to PRR ligands may have implications for treatment choices.

References

1. Ungaro, R., Mehandru, S., Allen, P. B., Peyrin-Biroulet, L. & Colombel, J. F. Ulcerative colitis. *Lancet* **389**, 1756–1770 (2017).

2. Lim, K.-H. & Staudt, L. M. Toll-like receptor signaling. *Cold Spring Harb. Perspect. Biol.* **5**, a011247 (2013).

3. Kawasaki, T. & Kawai, T. Toll-Like Receptor Signaling Pathways. *Front. Immunol.* **5**, (2014).

4. Moossavi, S. & Rezaei, N. Toll-like receptor signalling and their therapeutic targeting in colorectal cancer. *Int. Immunopharmacol.* **16**, 199–209 (2013).

5. Lu, Y., Li, X., Liu, S., Zhang, Y. & Zhang, D. Toll-like Receptors and Inflammatory Bowel Disease. *Front. Immunol.* **9**, 72 (2018).

6. Sánchez-Muñoz, F. *et al.* Transcript levels of Toll-Like Receptors 5, 8 and 9 correlate with inflammatory activity in Ulcerative Colitis. *BMC Gastroenterol.* **11**, 138 (2011).

7. Toiyama, Y. *et al.* The expression patterns of Toll-like receptors in the ileal pouch mucosa of postoperative ulcerative colitis patients. *Surg. Today* **36**, 287–290 (2006).

8. Pedersen, J., Coskun, M., Soendergaard, C., Salem, M. & Nielsen, O. H. Inflammatory pathways of importance for management of inflammatory bowel disease. *World J. Gastroenterol.* **20**, 64–77 (2014).

9. Takeda, K. & Akira, S. Toll-like receptors. *Curr. Protoc. Immunol.* **109**, 14.12.1-14.12.10 (2015).

10. De Jager, P. L. *et al.* The role of the Toll receptor pathway in susceptibility to inflammatory bowel diseases. *Genes Immun.* **8**, 387–397 (2007).

11. Wang, J., Hu, Y., Deng, W. W. & Sun, B. Negative regulation of Toll-like receptor signaling pathway. *Microbes Infect.* **11**, 321–327 (2009).

12. Sidiq, T., Yoshihama, S., Downs, I. & Kobayashi, K. S. Nod2: A Critical Regulator of Ileal Microbiota and Crohn's Disease. *Front. Immunol.* **7**, 367 (2016).

13. van Heel, D. A. *et al.* Synergy between TLR9 and NOD2 innate immune responses is lost in genetic Crohn's disease. *Gut* **54**, 1553–1557 (2005).

14. Maisonneuve, C., Bertholet, S., Philpott, D. J. & De Gregorio, E. Unleashing the potential of NOD- and Toll-like agonists as vaccine adjuvants. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12294–12299 (2014).

15. Lee, M. S. & Kim, Y.-J. Signaling pathways downstream of pattern-recognition receptors and their cross talk. *Annu. Rev. Biochem.* **76**, 447–480 (2007).

16. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

17. Rebhan, M., Chalifa-Caspi, V., Prilusky, J. & Lancet, D. GeneCards: integrating information about genes, proteins and diseases. *Trends Genet.* **13**, 163 (1997).

18. Singh, A., Anderssen, E., Fenton, C. G. & Paulssen, R. H. Identifying anti-TNF response biomarkers in ulcerative colitis using a diffusion-based signalling model.

Bioinforma. Adv. **1**, 1–8 (2021).

19. Jean Philibert. One and a Half Century of Diffusion: Fick, Einstein, before and beyond. *diffusiion-fundamentals* (2005).

20. Peres-Neto, P. R. & Jackson, D. A. How well do multivariate data sets match? The advantages of a procrustean superimposition approach over the Mantel test. *Oecologia* **129**, 169–178 (2001).

21. Ritchie, M. E. *et al.* limma powers differential expression analyses for {RNA-sequencing} and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

22. Tsai, C.-A., Hsueh, H. & Chen, J. J. Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* **59**, 1071–1081 (2003).

23. Türei, D. *et al.* Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **17**, e9923 (2021).

24. Cepika, A.-M. *et al.* A multidimensional blood stimulation assay reveals immune alterations underlying systemic juvenile idiopathic arthritis. *J. Exp. Med.* **214**, 3449–3466 (2017).

25. Aschenbrenner, D. *et al.* Deconvolution of monocyte responses in inflammatory bowel disease reveals an IL-1 cytokine network that regulates IL-23 in genetic and acquired IL-10 resistance. *Gut* 1–14 (2020) doi:10.1136/gutjnl-2020-321731.

26. Reis, J. *et al.* LPS-induced formation of immunoproteasomes: TNF-α and nitric oxide production are regulated by altered composition of proteasome-active sites. *Cell Biochem. Biophys.* **60**, 77–88 (2011).

27. Chamberlain, N. D. *et al.* Ligation of TLR7 by rheumatoid arthritis synovial fluid single strand RNA induces transcription of TNFα in monocytes. *Ann. Rheum. Dis.* **72**, 418–426 (2013).

28. Krieg, A. M. Toll-like receptor 9 (TLR9) agonists in the treatment of cancer. *Oncogene* **27**, 161–167 (2008).

29. Fukui, R. *et al.* Cleavage of Toll-Like Receptor 9 Ectodomain Is Required for In Vivo Responses to Single Strand DNA. *Front. Immunol.* **9**, 1491 (2018).

30. Schmitt, H. *et al.* The TLR9 Agonist Cobitolimod Induces IL10-Producing Wound Healing Macrophages and Regulatory T Cells in Ulcerative Colitis. *J. Crohns. Colitis* **14**, 508–524 (2020).

31. Shen, Z.-H. *et al.* Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation. *World J. Gastroenterol.* **24**, 5–14 (2018).

Figure Descriptions

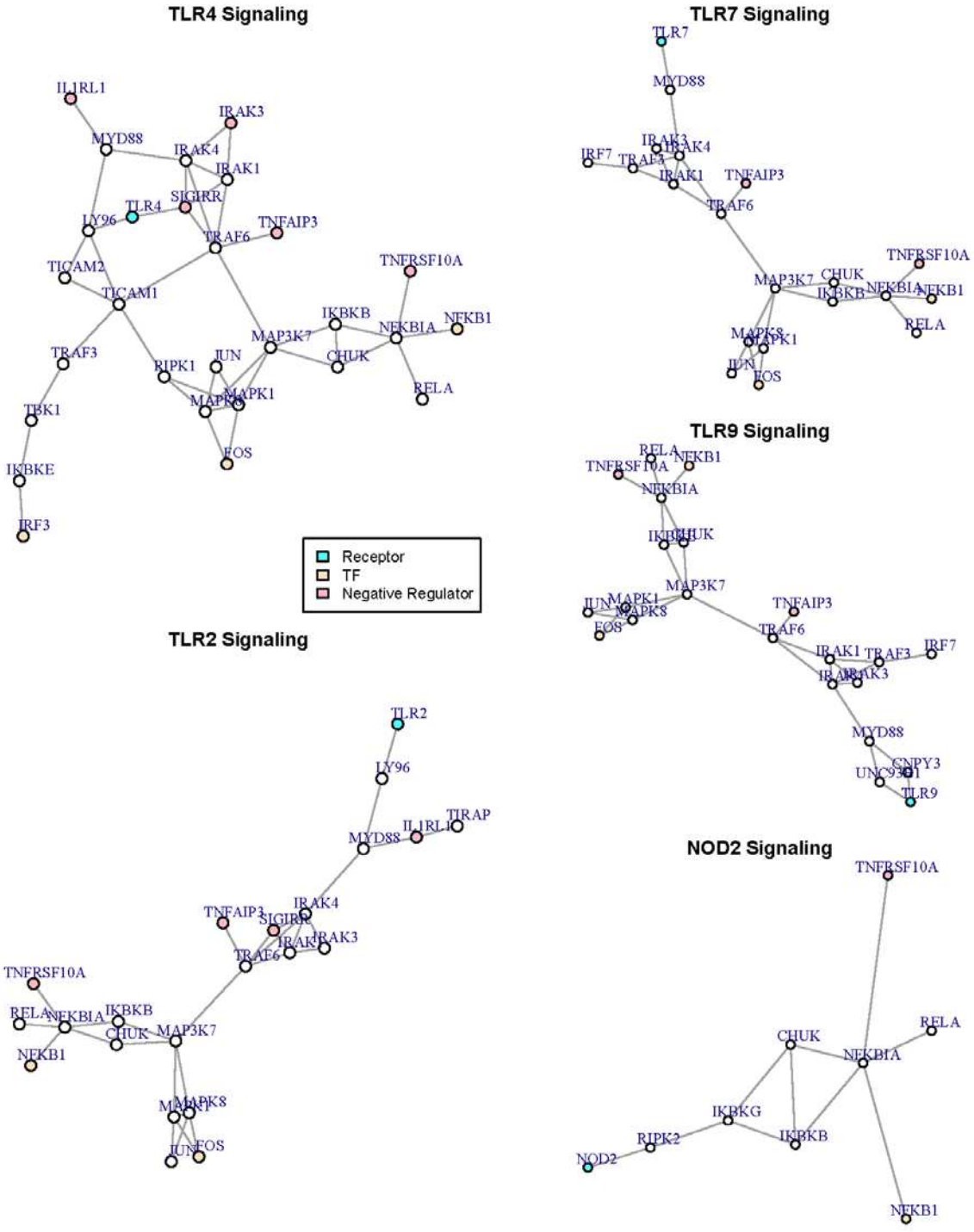


Figure 1 (A-B) Network models for PRR stimulation. Signalling network model connecting TLR-2, -4,-7,-9, and NOD2 to the transcription factors NFKB1, FOS, and IRF3. Nodes or vertexes represent proteins/genes and edges represent interaction that signal can travel along. Receptor nodes are given in cyan, signal transduction molecules in white and transcription factors in beige and negative regulators in pink.

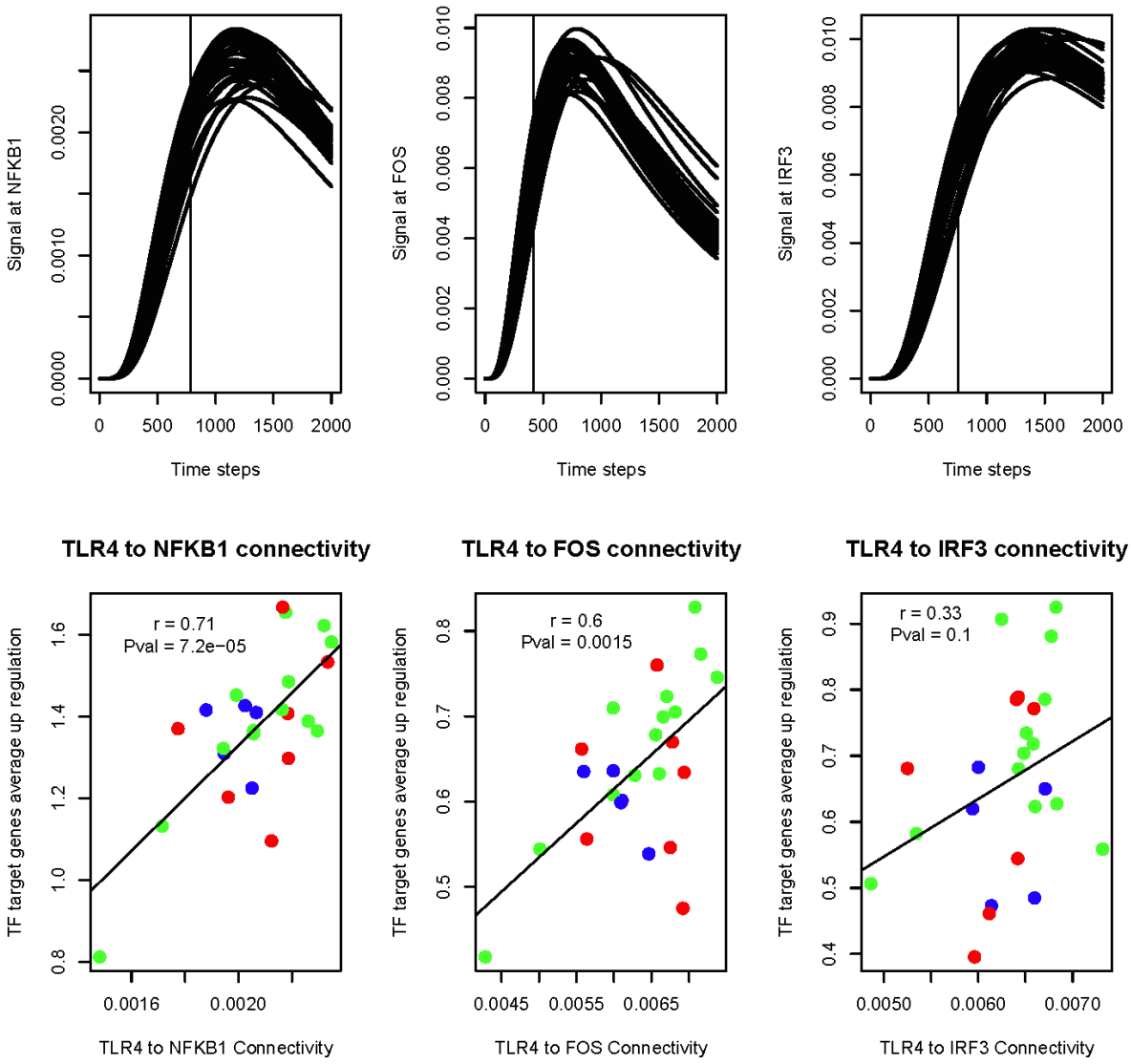


Figure 2 (A-C) Starting a simulation with one unit of signal on TLR4. The graph shows the signal accumulated on the transcription factors NFKB1, FOS, and IRF3 as a function of time. The time point of maximum between sample variance is indicated with a solid vertical line, and the signal intensity of this TF at this time is used as a measure of overall connectivity. (D-G) Scatter plot showing estimated network connectivity between TLR4 estimated as signal accumulated at the TF at the time of maximum variability vs. activation of TF targets after stimulation with LPS for 6 hours.

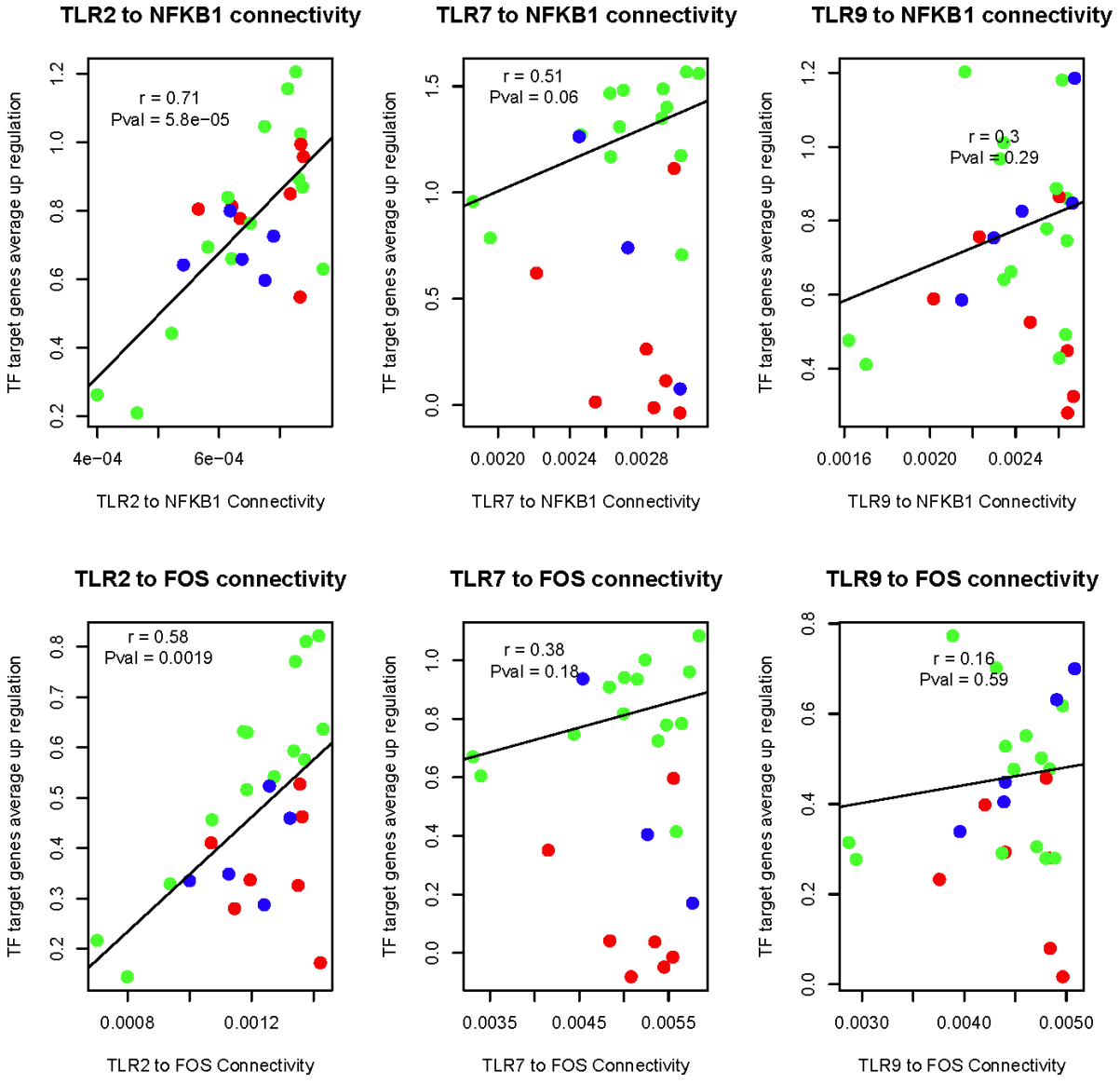


Figure 3 Network connectivity between receptors TLR -2, -7 and -9 and transcription factors NFKB1 and FOS vs. activation of the target genes of NFKB1 and FOS after stimulation with LTA, R837, and CpG-C respectively for six hours.

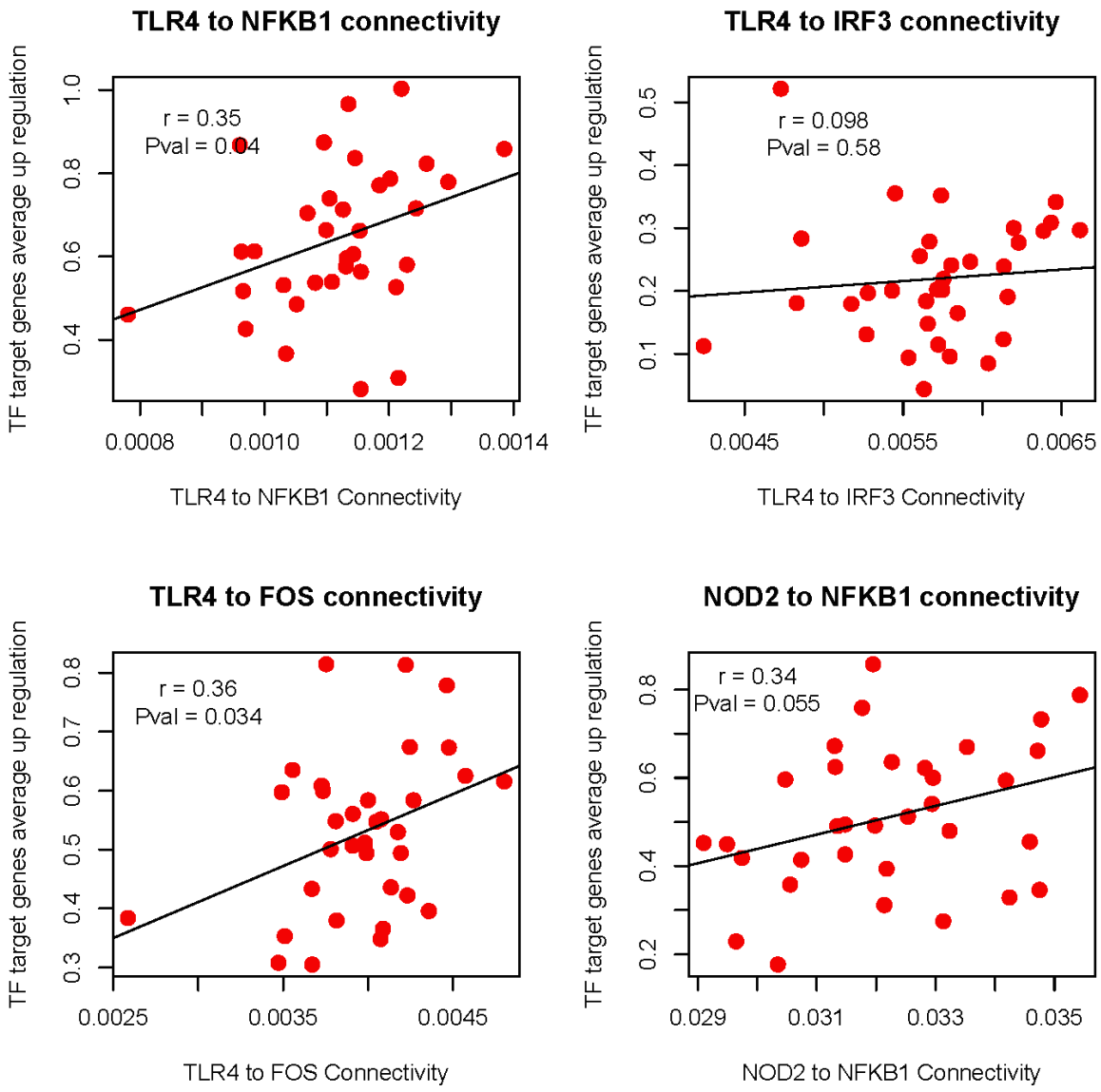


Figure 4. Network connectivity between receptors TLR4 and NOD2 and transcription factors NFKB1, FOS, and IRF3 vs transcription factor target gene activation after stimulation with LPS and MDP for sixteen hours.