



HAL
open science

Caractérisation multi-omique des maladies inflammatoires de l'intestin

Kevin Perez

► **To cite this version:**

Kevin Perez. Caractérisation multi-omique des maladies inflammatoires de l'intestin. Médecine humaine et pathologie. Université Paris Cité, 2021. Français. NNT : 2021UNIP7339 . tel-04114667

HAL Id: tel-04114667

<https://theses.hal.science/tel-04114667>

Submitted on 2 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Paris

Ecole doctorale HOB 561

*Laboratoire U1160 "Ecotaxie, Microenvironnement et
développement Lymphocytaire" - EMiLy*

Multi-omics characterization of Inflammatory Bowel Diseases

Par Kevin Perez

Thèse de doctorat de Biothérapies

Dirigée par Matthieu Allez

Présentée et soutenue publiquement le 26 mars 2021

Devant un jury composé de :

Edouard Louis (Rapporteur, Université de Liège, Belgium), Philippe Seksik (Examineur, Hopital Saint-Antoine), Marion Espelli (Examineur, IRSL Saint-Louis), Encarnita Mariotti (Rapporteur, Sorbonne Université), Lionel le Bourhis (Membre invité, IUH Saint-Louis)



Except where otherwise noted, this work is licensed under
<https://creativecommons.org/licenses/by-nd/3.0/fr/>

Titre : Characterization multi-omique des maladies inflammatoires de l'intestin.

Resumé :

Les maladies inflammatoires de l'intestin (MICI) restent un problème majeur pour la prise en charge cliniques et peut parfois altérer considérablement la vie de jeunes patients à travers des symptômes et complications potentielles. Bien que de nouveaux traitements tels que les anti TNF ont révolutionné la prise en charge des MICI, prédire le cours de la maladie et la réponse aux traitements restent un enjeu majeur de la pratique médicale. Les études 'omiques', tels que le transcriptome, le séquençage du microbiome ou la génétique ont changé notre manière de faire de la science, et révèlent de nouvelles caractéristiques des maladies étudiées par le passé. Ici nous souhaitons utiliser nos cohortes internes et externes de patients atteints de MICI, à travers le séquençage multi-omique pour répondre à ces enjeux, identifier des nouvelles cibles thérapeutiques, améliorer notre compréhension de la maladie and mieux prédire son cours. Dans la première partie, nous effectuons une méta-analyse de l'expression génique de la muqueuse des patients atteints de MICI, accessible publiquement, totalisant plus de 1000 échantillons, pour montrer qu'il existe des marqueurs spécifiques des maladies iléales et coliques. Ces marqueurs pourraient déboucher sur l'identification de nouvelles cibles thérapeutiques. Dans la deuxième partie, nous utilisons la puissance de l'analyse multi-omique sur notre cohorte de patients chirurgicaux (REMIND), pour prédire la récurrence postopératoire après résection iléo-caecale dans la maladie de Crohn (MC). Ces résultats pourront être associés à des données cliniques pour améliorer la prédiction du cours de la maladie et guider les décisions cliniques de manière impartiale. En utilisant la génétique, le transcriptomique et le microbiome, nous démontrons la puissance de l'analyse multi-omique pour atteindre une médecine de précision dans les MICI.

Mots clés : MICI ; bioinformatique ; transcriptome ; génétique ; RNAseq ; Crohn ; RCH

Title : Multi-omics characterization of Inflammatory Bowel Diseases

Abstract :

Inflammatory bowel diseases (IBD) remain a constant challenge in clinical practice and sometimes alter young patients' lives through symptoms and potential complications. Although new treatments like anti-TNF therapies have revolutionized the management of IBD, predicting the course of disease and its response to treatment remains an unmet need in medical practice. 'OMICS' studies, such as transcriptomic and next-generation sequencing of microbial or host genetic information, revolutionize how we do science and shed new light on diseases that have been studied extensively in the past. We aim to use internal and external cohorts of IBD patients with associated 'omics' to answer unmet needs; identifying new drug targets, improving our understanding of the disease and predicting its varying course. In the first section, we use a meta-analysis of publicly available IBD mucosa gene expression totaling more than 1000 samples to show that there are specific markers of IBD in the ileum and in the colon. These markers harness the potential to yield new drug targets. In the second section, we show the power of multi-omics on our cohort of patients undergoing surgery (REMIND) to predict postoperative recurrence after ileocecal resection in Crohn's disease (CD). These findings can be coupled with clinical data to improve the prediction of outcomes and guide clinical decisions in an unbiased fashion. Overall using genetics, transcriptome, and microbiome, we demonstrate the power of Omics to approach our goal of precision medicine in IBD.

Keywords : IBD ; bioinformatics ; transcriptomics ; genetics ; RNAseq ; Crohn ; UC

Acknowledgments

This work was made possible thanks to the help of several people to whom I would like to express my gratitude. First, I would like to thank Pr. Matthieu Allez, head of the gastroenterology unit at Hopital Saint-Louis and head of the lab that hosted my PhD. He offered me the opportunity to work in his team, and showed interest in my unusual profile from the beginning. A special thank, to Lionel Le Bourhis, senior researcher in the lab, for guiding me through my PhD work, for his insights, and the many discussions we had on how to analyze data in a meaningful way.

Thank you to Marjolaine Ngollo and Nassim Hammoudi, whom I have had the pleasure to collaborate during this PhD. Marjolaine had a wet lab background when I joined the lab but quickly learned bioinformatics techniques alongside me. Nassim brought his clinical background, and our discussions have always been very productive. Thank you to Pr. Mark Daly, Pr. Ramnik Xavier who supervised me during my 1 year stay at the Broad Institute in Boston, MA. I learned many things at the Broad notably how to handle and organize large scale datasets. I'm lucky to have been able to go to the Broad, one of the best places in the world for bioinformatics.

Thank you to the Helmsley Charitable Trust for funding this research, and my PhD salary. Helmsley meetings in Paris, Barcelona or Munich have always been a unique experience throughout my PhD years. Thank you to Philippe Seksik, who has supported all these years in my conversion from a software engineer to a bioinformatics researcher. He introduced me to microbiome research and supported me all these years. He's now a member of my jury. Thank you to the other members of my PhD committee: Marion Espeli, Hugues Aschard for continuously supporting me through this work. Hugues provided expertise in statistical genetics and Marion helped with her background in immunology.

Abstract

Inflammatory bowel diseases (IBD) remain a constant challenge in clinical practice and sometimes alter young patients' lives through symptoms and potential complications. Although new treatments like anti-TNF therapies have revolutionized the management of IBD, predicting the course of disease and its response to treatment remains an unmet need in medical practice. 'OMICS' studies, such as transcriptomic and next-generation sequencing of microbial or host genetic information, revolutionize how we do science and shed new light on diseases that have been studied extensively in the past. We aim to use internal and external cohorts of IBD patients with associated 'omics' to answer unmet needs; identifying new drug targets, improving our understanding of the disease and predicting its varying course. In the first section, we use a meta-analysis of publicly available IBD mucosa gene expression totaling more than 1000 samples to show that there are specific markers of IBD in the ileum and in the colon. These markers harness the potential to yield new drug targets. In the second section, we show the power of multi-omics on our cohort of patients undergoing surgery (REMIND) to predict postoperative recurrence after ileocecal resection in Crohn's disease (CD). These findings can be coupled with clinical data to improve the prediction of outcomes and guide clinical decisions in an unbiased fashion. Overall using genetics, transcriptome, and microbiome, we demonstrate the power of Omics to approach our goal of precision medicine in IBD.

Table of Contents

Acknowledgments	1
Abstract	2
Table of Contents	3
Chapter 1.	6
1. INTRODUCTION	6
1.1 Inflammatory Bowel Diseases	6
1.1.1 Ulcerative colitis and Crohn's disease	6
1.1.2 Disease localization: ileal, colonic, ileo-colonic	7
1.1.3 Disease complications and post-operative recurrence	8
1.2 Epidemiology of Inflammatory Bowel Disease	9
1.3 Genetics in IBD	9
1.3.1 Familial aggregation	9
1.3.2 Disease aggregation in twins	10
1.3.3 Susceptibility Genes	10
1.4 Environmental Factors	10
1.4.1 Geographic and temporal variability	11
1.4.2 Breast feeding and nutrition	12
1.4.3 Sanitation, hygiene and antibiotics	12
1.4.4 Smoking	13
1. 5. Intestinal Biology and Immunotherapies	13
1.6 Multi-omics Methods in Human Diseases	16
1.6.1 Why multi-omics?	17
1.6.2. Genome-wide association studies	17
1.6.3. Transcriptomic studies	18
1.6.4. Epigenome-wide association studies	19
1.6.5. Proteomic studies of human biology and diseases	20
1.6.6 Metabolome-wide association studies	21
1.6.7 Microbiome studies	21
1.6.8 Single-cell studies	22
1.6.9 Issues and limitations of omics studies	23
1.7 Multi-omics in IBD	24

1.7.1	Goal of multi-omics in IBD	24
1.6.2	IBD genetics	25
1.7.3	Transcriptomics in IBD	26
1.8.4	Proteomics in IBD	26
1.8.5	Metabolomics in IBD	27
1.6.5	Epigenetics in IBD	27
1.8.6	Microbiome in IBD	28
1.6.7	Single-cell in IBD	28
1.8	Multi-omics in IBD	29
1.8.1	Multi-omics studies in IBD	29
	Outlined Methods	32
	Chapter 2.	36
	2.1 Meta-analysis of IBD gut samples identifies specific markers of disease in Ileal and Colonic Diseases	37
	2.1. Other approaches and extension to Pouchitis	51
	2.2.2. Molecular classification	51
	2.2.3. Extension to pouchitis	53
	Chapter 3.	56
	Genetic, multi-omics analysis of a surgical IBD cohort	56
	3.1. The REMIND cohort	56
	3.2. Transcriptome analysis of the REMIND cohort	57
	3.3 Genetics analysis of the REMIND cohort	59
	3.3.1. Ancestries	60
	3.3.2 Distribution of IBD susceptibility SNPs in the REMIND cohort.	61
	3.3.3 GWAS for Post-Operative Recurrence	63
	3.3.4 IBD susceptibility SNPs, smoking and post-operative recurrence	65
	Chapter 4.	72
	Conclusion and Discussion	72
	4.1 Summary of my work	72
	4.2 Future Perspectives	73
	4.2.1 Follow-up on the meta-analysis of IBD samples	64
	4.2.2 Follow-up on the REMIND study	64
	4.2.3 Multi-omics analysis of the REMIND study	73
	4.2.4 Perspectives	77
	BIBLIOGRAPHY	78

List of publications	91
Conferences attended	91
Appendix	92
A. 1. The ELYP study	92

Chapter 1.

1. INTRODUCTION

1.1 Inflammatory Bowel Diseases

1.1.1 Ulcerative colitis and Crohn's disease

Inflammatory Bowel Diseases (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), are characterized by chronic inflammatory lesions of segments within the lining of the digestive tract. These lesions may occur in the small intestine, colon and/or anus during CD, but are restricted to the rectum and colon in UC. IBDs often progress with periods of relapse and remission. During IBD flare-ups, it is not uncommon for patients to present common digestive disorder symptoms such as diarrhea or abdominal pain. Other generalized symptoms may include fever, fatigue, and/or weight loss. In some cases, extra-digestive manifestations arthritis and/or erythema may also be exhibited. These inflammatory lesions, often ulcerated, persist if the disease is not controlled and can be responsible of complications, such as abscesses, fistulas or stenosis that may require surgical intervention. In addition, chronic inflammation increases the risk of cancer in affected segments (Lennard-Jones, 1989, Jones et al., 1973). These gastroenterology diseases affect an estimated 1 million individuals in the US and 2.5 million in Europe (Giardiello et al., 1989, Okayasu et al., 1990).

IBD is a multifactorial disease whose etiology is not yet clearly established but has a clear genetic component. More than 200 susceptibility loci have been identified such as mutations in the genes coding for NOD2 / CARD15 (detector of intracellular bacteria), ATG16L1 (linked to the degradation of these bacteria) or IL23R (implicated in the polarization of the effector T lymphocytes) (Podosky, 1991, Harig et al., 1989). Environmental factors are known to impact disease onset and progression, such as smoking behavior and/or air pollution. Abnormalities in the

microbiota (dysbiosis) are also observed in patients with IBD (Madara et al., 1985). Thus, IBD is a multifactorial disease which aspects we will detail in the following sections.

1.1.2 Disease localization: ileal, colonic, ileo-colonic

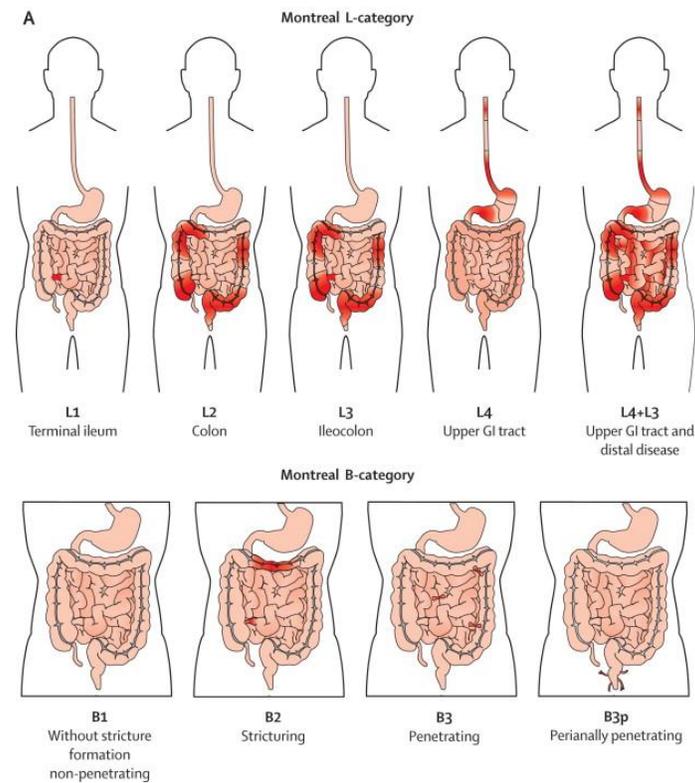


Figure 1. Montreal classification. (top) location of the affected segments (bottom) behavior of the disease. From Lancet, Crohn’s disease, 2012

CD is typically classified using the Montreal (or Vienna) classification [Figure 1]. The location can be terminal ileum (L1), colon (L2), ileocolonic (L3), and upper GI location (L4). That classification also includes a behavior component: B1 non-stricturing, non-penetrating, B2 stricturing, B3 penetrating. Usually, the localization is stable over time. The initial classification and localization can influence greatly the risk of complications over time, and disease management. UC is classified as follows: E1 Ulcerative proctitis (Involvement limited to the

rectum), E2 Left-sided UC (Involvement limited to a proportion of the colorectum distal to the splenic flexure), E3 Extensive UC (pancolitis), Involvement extends proximal to the splenic flexure. However, several gastroenterology specialists have recently started to question these classifications and claim that we should perhaps establish a more general distinction between ileal and colonic IBD (Dulai et al., 2019). This is an idea that we will develop in this thesis, based on findings at the molecular levels in different localizations.

1.1.3 Disease complications and post-operative recurrence

CD can progress to strictures, abscesses and fistula. When this occurs, patients often require surgical intervention to remove the diseased segment. New therapies like anti-TNF have delayed the need for surgery but this risk is still important. Surgery is required in more than 50% of CD patients with an involvement of the ileum [Figure 2]. Unfortunately, the surgery is not curative and the disease can come back, usually in the neo-terminal ileum. Risk factors for post-operative recurrence in CD patients include being of male gender, penetrating diseases, smoking and prior surgeries. Preventive anti-TNF treatment after surgery have been shown to be efficient on disease recurrence (Auzolle et al, 2017). Later on, we will explore the molecular mechanism at play during this post-operative recurrence, and try to build a predictive model for post-operative recurrence.

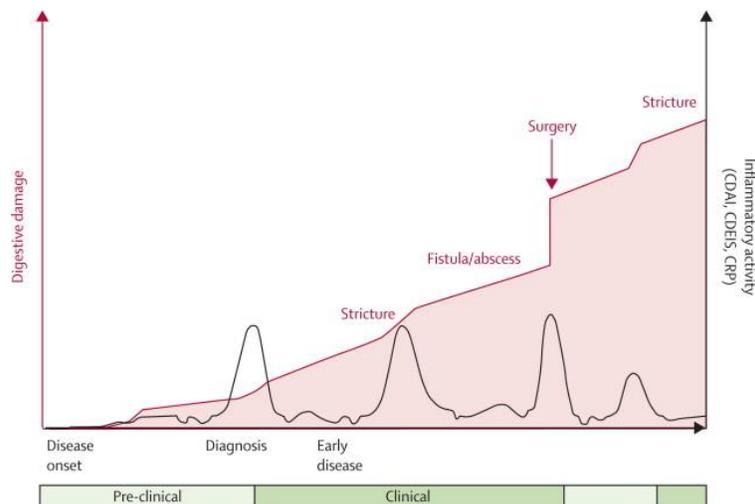


Figure 2. ‘Lemann score’: evolution of the digestive damage over time. From Lancet, Crohn’s disease, 2012

1.2 Epidemiology of Inflammatory Bowel Disease

Internationally, in 2017, there were 6.8 million (95% UI 6.4–7.3) cases of IBD globally. The age-standardized prevalence rate increased from 79.5 (75.9–83.5) per 100 000 population in 1990 to 84.3 (79.2–89.9) per 100 000 population in 2017 (Collaborators, 2020). The incidence rate of inflammatory bowel diseases is higher in Northern Europe, the United Kingdom, and North America. These rates are seemingly stabilizing (Baumgart and Carding, 2007). However, the countries with low incidence rates of inflammatory bowel diseases are showing a rise in these rates (Ng et al., 2017); these include southern Europe, Asia, north Africa and other developing countries (Loftus Jr, 2004). Crohn's disease is more prevalent in Hispanic and Asian population than white and African-American people in North America (Kurata et al., 1992). Another study supports the significance of racial and ethnic characteristics in the onset of IBD as well as associated complications (Nguyen et al., 2006).

Regarding ethnic groups, Jewish populations are more prone to developing Crohn's disease (Heresbach et al., 1998, Hampe et al., 2000). Jewish people also show three to four-time more incidence rates of ulcerative colitis (Niv et al., 2000). The living location of the Jewish people also shows marked differences between Jews living in Israel and living anywhere else. As with other migrant populations, their risk usually equals the prevalence rate of the location they are settling in (Grossman et al., 1989). As we will see, in addition to geographical influences, the effects of environment and lifestyles also play a role on the epidemiology, prevalence and incidence rate of IBD (Zheng et al., 2005, Leong et al., 2004), as does genetics.

1.3 Genetics in IBD

1.3.1 Familial aggregation

The familial relationship of inflammatory bowel disease incidence was first observed in 1930 (Russell and Satsangi, 2004). One of the major risk factors of these diseases is having a family member who was diagnosed with IBD. If a person has an inflammatory bowel disease, then there is a major risk to any family member of acquiring this disorder. Many people with Crohn's disease have some first relatives already suffering or having a history of CD. Moreover, the risk of the disease is imposed at a greater level to the sibling than any other first relative of the patient.

Regarding ethnic groups, it is found or observed that the risk of developing Crohn's disorder for first degree relatives throughout life is more in Jews than in other people belonging to other ethnic groups. Similar findings are observed for ulcerative colitis (Russell and Satsangi, 2004), and the occurrence of extraintestinal complications are 70% more reported in Jews than non-Jews (Satsangi et al., 1996).

1.3.2 Disease aggregation in twins

The genetic component of IBD has been shown by studying risk in monozygotic and dizygotic twins (Tysk et al., 1988, Thompson et al., 1996, Orholm et al., 2000). 37.3% concordance is shown in monozygotic twins for Crohn's disease, 10% for ulcerative colitis; while dizygotic twins show concordance of about 7% for CD and 3% in UC. These results imply that genetic factors and various genetic products contribute more to the onset of Crohn's disease than in the onset of ulcerative colitis (Gaya et al., 2006). It also shows that heritability is rather low compared to other inflammatory diseases, clearly demonstrating the importance of the environment for IBD onset.

1.4 Environmental Factors

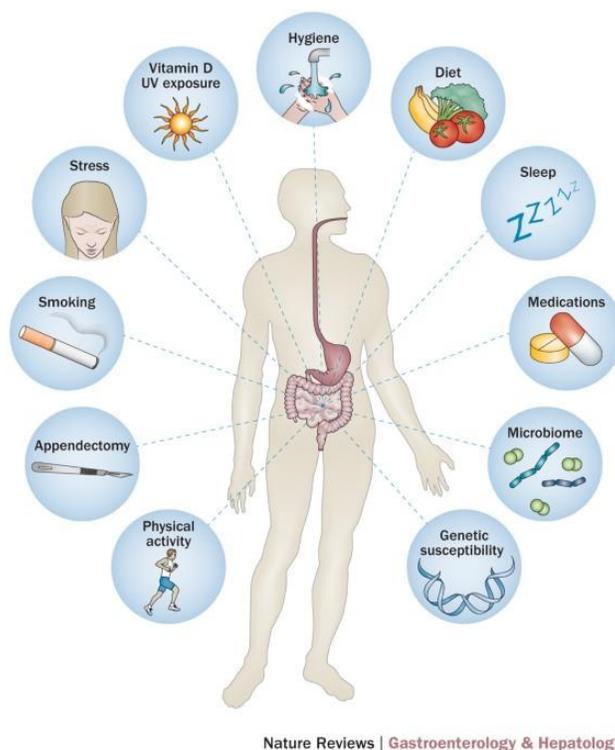


Figure 3. Contribution of multiple environmental factors to IBD (Ananthakrishnan et al., 2015)

1.4.1 Geographic and temporal variability

As mentioned in the epidemiology of inflammatory bowel diseases, North America and North Europe show increase incidence as well as prevalence of the diseases. Likewise, less incidence and prevalence are reported from areas like South America, South Asia, Australia, and Africa (Loftus Jr, 2004). Such observations suggest that prevalence and incidence rate increase moving from south to north. These observations are attributed to many variables such as the availability of health care, sanitation, industrialization, and similar other aspects. Indeed, there is a lower incidence in developing countries. The association of IBD with environmental factors is demonstrated by the increase incidence of immigrants from low incidence countries after one or two generations after their arrival in high incidence regions (Zheng et al., 2005). Other studies conducted in North America and Europe have also shown that IBD are more common in urban areas than rural areas of these geographical regions, but this can also be attributed to other environmental factors and health care access, rather than the geographical attributes (Moum et al., 1996, Sonnenberg et al., 1994). These latter results are not supported by further studies conducted in other regions (Soncini et al., 2006, Lewis et al., 2004).

1.4.2 Breast feeding and nutrition

While the immunity of a child is being developed, immunity is provided by breastfeeding (Schack-Nielsen and Michaelsen, 2006). A number of studies have been conducted so far regarding the correlative risk factors between IBD and breastfeeding, but the results have been heterogeneous. The severity of IBD is not directly associated with breastfeeding, but the discontinuation of any prescription drug treatments while breastfeeding is associated (Kane and Lemieux, 2005). Many observational studies have shown that a high carbohydrate intake imposes a greater risk to the onset of IBD, especially Crohn's disease (Riordan et al., 1998). A few studies say that intake of polyunsaturated fatty acids is directly proportional to the risk of IBD (Geerling et al., 2000). Arachidonic acid derivatives were also thought to be pro-inflammatory in nature, but further research contradicted these results. No specific food product has been identified to date as directly responsible for IBD, but clinicians would generally recommend a healthy diet rich in fruits and vegetables outside of flare periods.

1.4.3 Sanitation, hygiene and antibiotics

As with other auto-immune diseases, hygiene and sanitation are thought to be major factors with respect to the risk of IBD. As it is seen that in developing countries, the incidence rate was low, which is now increasing can be attributed to socioeconomic values that also affect hygiene and sanitation (Desai and Gupte, 2005). Conditions like consumption of contaminated food, unavailability of certain things like hot water, consumption of contaminated water, low-income families living in poor conditions like congested space with a greater number of family members, help in maturing the intestinal immunity of the person. However, when mucosal immunity is not developed in childhood, later exposure to antigens can lead to IBDs, particularly Crohn's disease (Hampe et al., 2003, Kinra et al., 2006). Similarly, childhood antibiotic exposures are associated with an increased risk of developing IBD (Kronman et al., 2012). In addition, psychological stress can also contribute to developing IBDs, as there is a close relationship between the nervous and immune systems (Mawdsley and Rampton, 2005).

1.4.4 Smoking

Smoking cigarettes has a variable effect on Crohn's disease and ulcerative colitis [Figure 4]. Smoking aggravates Crohn's disease but protects from ulcerative colitis. For CD patients, smoking cessation is highly recommended. In CD, smokers generally have a more severe disease, including strictures and fistula formation, the need for using corticosteroids, and the need for surgical intervention to handle the disorder, as well as increased risk for post-operative recurrence (Cosnes, 2004). The reason for this dual effect is unclear and is a highly active area of investigation.

1.5. Intestinal Biology and Immunotherapies

The intestinal epithelium is a single layer of cells containing crypts and villi. It renews itself rapidly thanks to intestinal epithelial stem cells. The cells that line the intestinal lumen perform the primary functions of digestion, water and nutrient absorption, and form a barrier against luminal pathogens. Different intestinal cell types include enterocytes (the majority), Paneth cells, goblet cells, and neuroendocrine cells [Figure 5]. Briefly, Paneth cells secrete antimicrobial peptides, and goblet cells secrete mucus. Trillions of bacteria inhabit the gut. While some are pathogenic, others are commensals that are usually harmless and can also provide important functions for the organism like regulating immunity. The intestinal lamina propria is also full of immune cells including macrophages, dendritic cells, B and T cells, which can be recruited to the site of inflammation, notably in the context of IBD (Mowat and Agace, 2014). The microbiota also plays an important role in the training of the immune system during development and infancy. What results is an equilibrium between host and microbiome in a tolerogenic state, by protecting against pathogens but allowing the presence of commensal bacteria. The use of antibiotics and changes in diet have been proposed as culprits for the breaking of this equilibrium thus allowing the development of IBD (Belkaid and Hand, 2014).

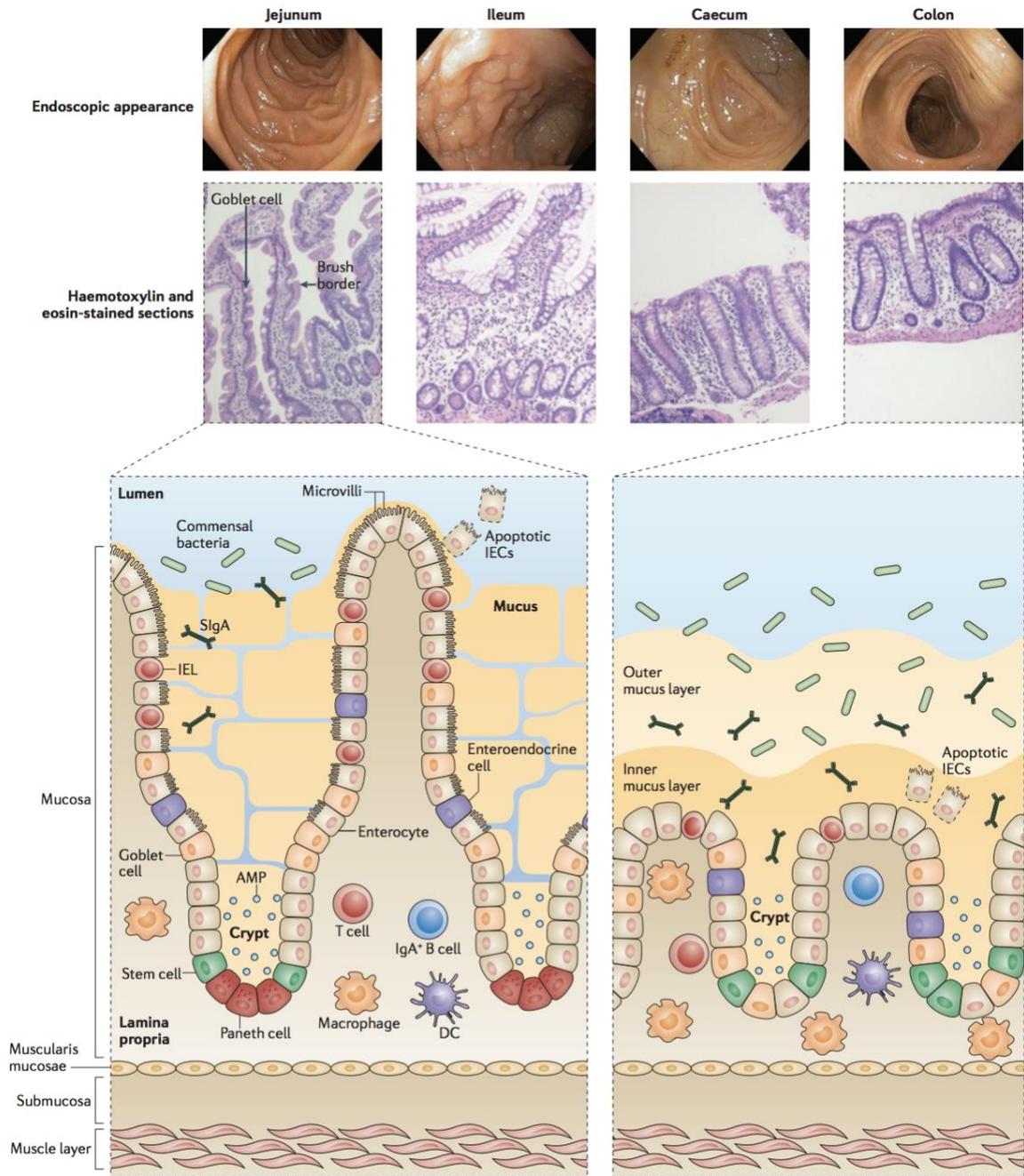


Figure 5. (top) Endoscopic pictures and immunostaining of different sections of the gut. (bottom) Intestinal biology in the ileum (left) and the colon (right), including immune cells, epithelial cells, mucus, bacteria, and secreted factors (Mowat et al., 2014).

This understanding of intestinal biology has led, in part, to the discovery of new therapeutics for IBD. This includes drugs that might target pathways that we see dysregulated in IBD [Figure 6]. Biologic monoclonal antibody therapy, especially Anti-TNF, has been a major improvement for the treatment of IBD. It has been shown to decrease disease-related morbidity substantially (Su et al., 2002). Trials of therapeutic strategies have shown that early initiation of such therapy is associated with greater response. Consequently, treatment paradigms have evolved from a step-up strategy to favor up-front biologic therapy to prevent damage. However, about 40% of patients do not respond to treatment, which unnecessarily puts them at risk of infection, carries a high cost, and gives them false hopes. It has brought forward the importance of personalized treatments, which requires a priori prediction of the response to each mechanism of action (Adegbola et al., 2018). Several current treatment strategies target pathways involved in the T cell response: thiopurines, inducing T cell apoptosis, anti- $\alpha4\beta7$ targeting the traffic of gut homing T cells, and anti-IL12/23 p40, interfering with the Th1 and Th17 differentiation. Hence, the characterization of the T cell compartment and its evolution during treatment in IBD patients could provide predictive markers of efficacy and/or new therapeutic strategies (Podosky, 1991, Madara et al., 1985, Okayasu et al., 1990, Hammer et al., 1990).

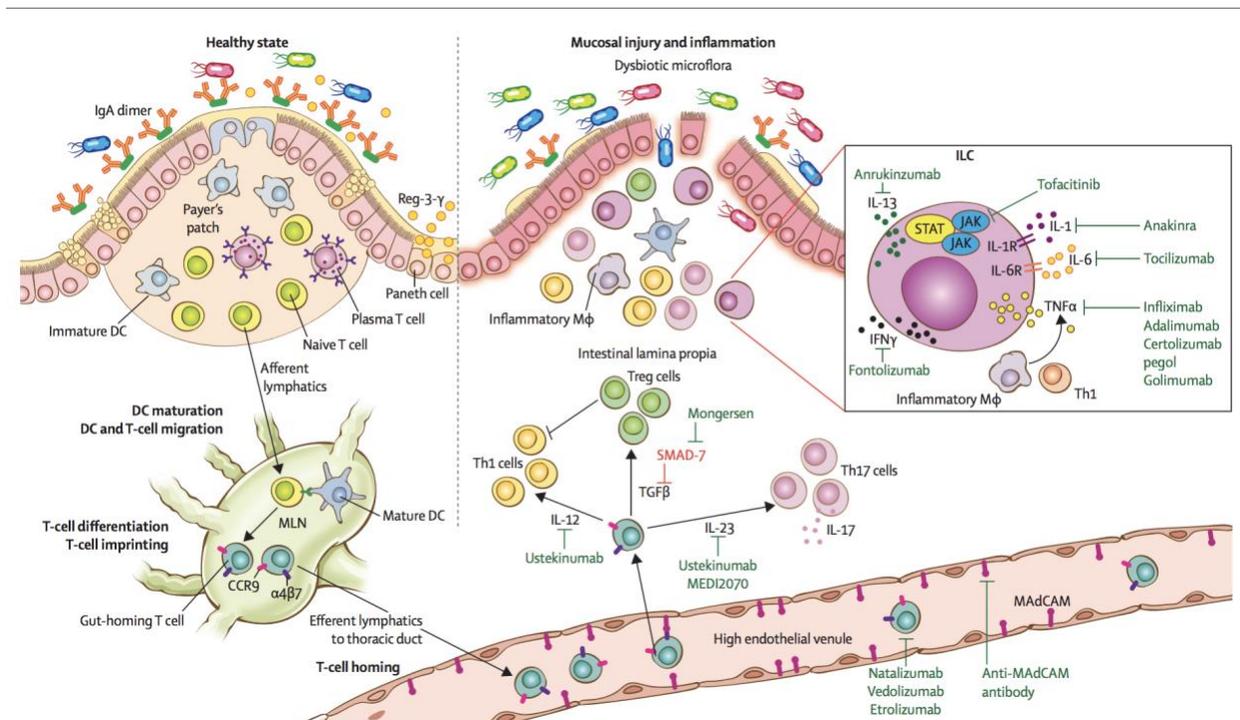


Figure 6. Intestinal immunity in health and inflammation, including mucosa, lymph nodes and blood vessels. Treatments blocking the different inflammatory molecules are shown by arrows. (from Lancet, Crohn’s disease, 2016)

1.6 Multi-omics Methods in Human Diseases

Human disease development is a highly dynamic and interactive biological process involve genes, epigenetic modifications, transcripts, proteins, and metabolites in the human body (Consortium, 2015, Kim et al., 2014). A biological process can be better understood by completely analyzing the information in these different molecular levels (Civelek and Lusis, 2014). A multi-omics data approach for studying a disease can be proved to be helpful in a thorough understanding of it as well as early detection, prevention, and proper treatment (Consortium, 2015, Kim et al., 2014). A biological process can be better understood by completely analyzing the information in these molecular levels (Civelek and Lusis, 2014) [Figure 7].

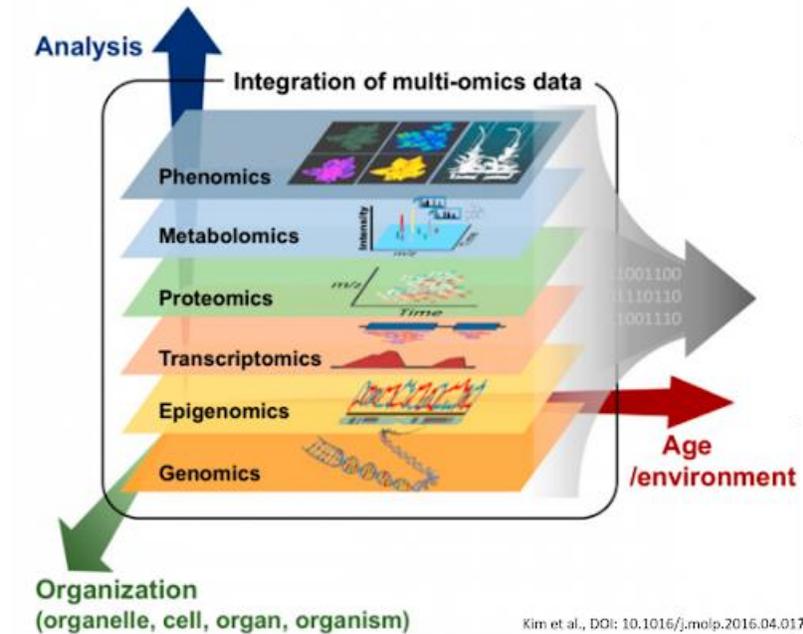


Figure 7. Layers of the multi-omics ‘model’. Each omics can be analyzed at the cell, organ or organism level. Omics can also be studied longitudinally with age, or modified by environmental factors (Kim et al., 2016).

1.6.1 Why multi-omics?

Biomedical research has been successful in identifying genes and loci on the chromosomes responsible for the development of some human diseases through genome studies. Through these approaches, the high complexity of biological systems got revealed for advancements in disease diagnosis and the handling of diseases. However, it quickly became clear that genetic studies alone were not sufficient to understand the global complexity of IBD. Driven by the progress in technologies, scientists discovered other molecular systems in disease development, including epigenome, proteome, and metabolome. These omics studies analyze a particular molecule, such as DNA, RNA, proteins, or metabolites and generate multi-level large complex datasets. Furthermore, analyzing how all these systems interact together is an approach now commonly known as ‘systems biology’. DNA transcribes into RNA which translates into proteins. Hence, missing any molecular step in a system can lead to a gap in our understanding of the biological process. After identifying a GWAS hit, we may need to know its impact on the transcriptome or the proteome in the tissue, and then perhaps start to attribute a causal role to this genetic variation (Sun and Hu, 2016). This is where multi-omics studies come into place to push our understanding of complex diseases.

1.6.2. Genome-wide association studies

For decades, profiling of genetic variants has been used in human disease studies (Klein et al., 2005). The goal of genome-wide association studies (GWAS) is to associate genetic variations called Single Nucleotide Polymorphisms (SNPs) (Nielsen et al., 2011) to a particular phenotype (height, weight, BMI) or disease (IBD, Type 1 diabetes, breast cancer). By now, GWAS has identified a huge number of genetic variants that are associated with specific phenotypes in human diseases. As of today millions of people have had their genome sequenced either through research studies or personal genomics companies (Welter et al., 2014). One question today is what more information can we gain by increasing more and more the sample size, when we already have robust studies with hundreds of thousands of people. Another issue with these studies is that the SNPs often fall into non-coding regions of the genome, making it harder to identify the gene involved. Also, SNPs tend to be associated (linkage disequilibrium), making it harder to identify

causal genes for one disease. Once a gene has been identified it can be confirmed using functional validation of its effects in cell lines, organoids or mice models, and then potentially be the target for drug compounds. Current areas of interest in the genetic field include increasing the diversity of the populations sampled, currently most studies include Caucasian populations. Another approach is whole exome sequencing (WES) and whole genome sequencing (WGS), which instead of using SNP arrays, sequence the whole genome or the whole exome to identify genetic mutations. These might solve the problem of missing heritability in GWAS compared to twin studies. New efforts like the UK Biobank are making available the genome of millions of people together with carefully annotated clinical records, biological information (Metzker, 2010).

1.6.3. Transcriptomic studies

Transcriptomic studies have become the go-to technique to look at gene expression in tissue and cell culture. Initially microarray chips were used to detect the expression of thousands of genes using fluorescent probes. These have been largely replaced by RNA sequencing techniques, which uses next generation sequencing techniques to quantify the number of reads corresponding to each gene in the genome (Ozsolak and Milos, 2011). RNA-seq is more quantitative and precise than microarray. RNA-seq can also be used to look at gene isoforms and alternative splicing. The issue with transcriptomic studies of tissues is that the signal comes from all different cell contained within the tissue biopsy or blood sample, it can therefore be noisy and hard to interpret. Single-cell studies are trying to tackle this problem. Another shortcoming of transcriptomic studies is how to make sense of hundreds of differentially expressed genes, and pathway analyses methods have been developed to condense the information (Trapnell et al., 2012). The Cancer Genome Atlas (TCGA) which looks at the expression of different organ tumors, and the Genome Tissue Expression (GTEx) project are great examples of insightful gene expression studies (Marx, 2015).

onset, its susceptibility, and development along with providing insights about how to handle diseases (Relton and Davey Smith, 2012).

Epigenome-wide association studies are the analysis of epigenetic markers related to a specific phenotype by examining the epigenomic markers in a population (Sun, 2014). EWAS technique has been recently affordable and available to scientists for research, and a lot of studies have found DNA methylation markers that are leukocyte-based and are associated with the phenotypic characteristics of diseases. The technique has facilitated the ability for scientists to identify many of the molecules that mediate the genetic and environmental factors responsible for disease onset. EWAS can include ATAC-seq, Hi-C and Chip-seq studies, all targeting different epigenome modifications (chromatin accessibility, histone markers, methyl groups) (Demerath et al., 2015, Dick et al., 2014, Liang et al., 2015). Along with providing amazing results, numerous flaws in the technique have also been found, including issues of sample size, specimen types from humans, imperfections in epigenome editing techniques, specificity for cell type, and framework for analyzing data. Further improvements in these areas are required (Heijmans and Mill, 2012).

1.6.5. Proteomic studies of human biology and diseases

Proteomic studies include the comprehensive analysis of all the protein products of a single genome (Wilkins et al., 1996). Proteome knowledge leads to an understanding of normal protein or amino acid structure and function that can help access abnormal behaviors. Proteomic studies tell about the isomorphous forms of peptides, modifications that take place after the translational process, and any mutations that occur in amino acids. This knowledge base contributes to a greater understanding of certain diseases and their specific mechanism of action (Nilsson et al., 2010). A unique character of proteomic studies is that it can also point out the localization position and timing of a protein, highlighting the disease process that cannot be told by any other genomic study.

This approach of studying a disease is very valuable and is being practiced in studying diseases like cancer (Tsai et al., 2015), multiple sclerosis (Farias et al., 2014), and schizophrenia (Al Awam et al., 2015). A study done by Petricoin et al. was successful in distinguishing the ovarian cancer

cases from the normal ones using just the proteomic technique (Petricoin III et al., 2002). Many biomarkers have been identified so far for various types of cancers. Overall, these proteomic studies enhance the knowledge about the disease markers providing an integrated picture of pathogenesis, which can help develop more effective treatment strategies.

1.6.6 Metabolome-wide association studies

The term metabolome refers to the group of all metabolites of low molecular weight that are produced by the cells. Their metabolism can depict the complete physiological status and activity of the cell. Metabolome depicts all the endogenous and exogenous characters and factors of a cell, including genetics, developmental factors, pathological aspects, lifestyle effects as well as the influence of the environment on biological processes. Metabolomics is a narrow field of study within the broad context of biological studies that involves profiling of complete metabolome of a cell or a particular biological sample. It focuses on the processes at a biochemistry level. Unlike other omic studies, it relates more closely to environmental influences and the effects on cellular traits. Several databases have been made available to trace the effect on exposome to metabolites (ex. Exposome Explorer). Contrary to RNA and proteins, metabolites are not a direct product of DNA, as described by the concept of the 'Central Dogma' (Sun and Hu, 2016).

This approach has yielded interesting findings in human studies. Altmaier et al. researched metabolome and disease, their targeted approach found out the association of antihypertensives drugs and lipid-lowering drugs with effects on serum (Altmaier et al., 2014). Untargeted mass-spectrometry based techniques, and MWAS has now become an effective approach to understanding any biological questions (Jones et al., 2012). Infectious diseases also get covered by metabolomic analysis, such as a study done by Weiner et al. found an association of eight metabolites with tuberculosis disease (Staff, 2016).

1.6.7 Microbiome studies

High throughput sequencing of the microbiome allows identifying quantitatively and qualitatively the different microbial species present in a sample. Two main techniques exist: 16S ribosomal RNA (rRNA) gene amplicons and shotgun metagenomics. These enable us to assess within-sample

diversity (alpha-diversity), between-sample diversity (beta-diversity) and more generally the species differentially abundant in health and disease (Org et al., 2015). The ratio of Firmicutes to Bacteroidetes (F/B ratio) can also be determined. Perturbation of this ratio has been associated with diseases including obesity, and differences in composition have been identified notably in diabetes. With the meta-genome and even now the meta-transcriptome, we can assess the function of the different microbial species present in a sample. The Human Microbiome Project (Huttenhower et al., 2012) is a great example of integrated microbiome analysis in humans.

1.6.8 Single-cell studies

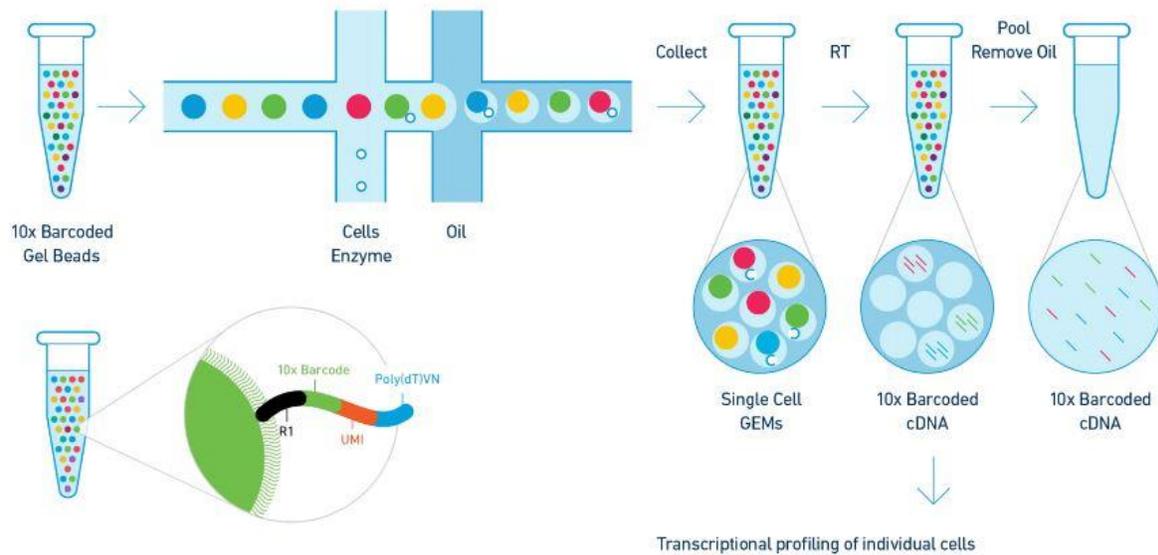


Figure 9. Overview of single cell sequencing through the 10x Chromium system protocol. Each cell is barcoded and encapsulated through a fluidics system to a specific GEM that gives it a unique molecular identifier (UMI). Then all reads are pooled together to infer the transcriptional profile of individual cells.

A major issue in bulk tissue RNA-seq studies is that it doesn't account for the heterogeneity in the cell types contained in the biopsies. Although some deconvolution techniques have been proposed, it is mostly impossible to know which cell types contributed to the signal that is being captured. For this reason, and others, single cell techniques have been developed where transcriptome for each cell can be profiled [Figure 9]. This allows us to look at the changes in cell types proportions

between health and disease, and the changes in gene expression from each specific cell type. Single-cell is currently enabling the confirmation and discovery of hypotheses regarding the contribution of different cell types to pathologies. Single-cell technologies are being developed to go further than transcriptomic: i.e. epigenetic (ATAC-seq), proteomics, spatial transcriptomics, immune repertoire.

1.6.9 Issues and limitations of omics studies

There are several limitations in using omics studies in human diseases. Remarkable differences in omics data can be the consequence of the 'batch effect.' Indeed, there are documented differences in data generated by various labs, technicians, or techniques. But when collaborative research is being done, and the subjective population is large, then these kinds of differences are hard to rectify. Other than this, the differences also occur because of various platforms available for a single omics analysis, like many manufacturers are working for transcriptomic and epigenomic analysis. They can have various genomic sites and features to work on, like sequencing techniques. Methods have been developed to resolve these batch effect, but these remain imperfect and the risk of losing biological information remains high, especially if the variable of interest is confounded by the batch, or if the different groups are not well represented in each cohort (Sun and Hu, 2016).

Another major problem is the concordance between omics results. The correlation between RNA-seq results, and proteomics results for example is generally poor; the reasons why are still not completely understood. Similarly, the methods to integrate different omics results are still debated. One can pool them into a model or process them separately prior to integration steps. Another major issue in the omics analysis is biological variation. Most commonly, blood samples are used for collecting omics data. Blood is typically used as a surrogate endpoint for looking at diseases. However, it could prove more informative to look at tissue samples. Yet, these also have several issues including heterogeneity of cell types known as 'sampling bias', and availability of fresh samples. Statistical methods have been developed to rectify the issues related to the heterogeneity of cell numbers and proportions (Abbas et al., 2009), notably the CIBERSORT algorithm (Newman et al., 2019), but with limited success.

One of the main challenges of multi-omics analysis is also to distinguish causation from correlation. We can identify correlations between markers and disease at multiple omics layer but they may only be the consequence of the disease, not the cause. It is possible that only longitudinal omics studies where each omics is assessed at several timepoints, and we can identify which changes precedes the occurrence of a disease will answer that question (Hasin et al., 2017).

Over the years many multi-omics methods will be developed for disease subtyping and classification, biomarker discovery for diagnostics and disease status. These will include Bayesian models (iCluster), multivariate models (Mixomics), correlation networks. It will become more and more difficult for non-experts to choose the right models to perform their analysis (Subramanian et al., 2020).

1.7 Multi-omics in IBD

Inflammatory Bowel Diseases (IBDs) are still very complex to understand in terms of the causes, its different forms, difficulty in the possible diagnosis, proper treatment, need for surgical interventions, risk of cancer development, and other risk factors associated with IBD outcomes. Ulcerative colitis and Crohn's disease both are in the category of complex diseases that are caused by many factors, including environment, genetics, immune responses, and microbiota. These factors and their interactions define the onset of diseases and possible outcomes (Fiocchi, 2014) (Xavier and Podolsky, 2007). If we consider the environment, then there are innumerable environmental agents that we encounter in our daily life. Genetics consists of numerous genes and their interactions. The microbiome consists of trillions of microbial species that take part in disease development. There are an unlimited number of complex biomolecules functioning and interacting in the human body (Fiocchi, 2014).

1.7.1 Goal of multi-omics in IBD

Various studies conducted on inflammatory bowel diseases have revealed that not just a single biological compartment plays a role in developing IBD. Instead, all mentioned biological levels and systems come up to take part and interact to trigger the disease. Exposome, microbiome, genome, and immunome all play a role in IBD. Immunome and microbiome depend upon the

environment, and change in genome occurs in about 100 years. In one of the studies, numerous components affecting the onset of IBD were studied along with various interactions of those components and resulted in explaining the need for a more comprehensive approach to studying the disease with all its components (Morris et al., 2010). The significance of the multi-omics study approach for IBD is also justified by the fact that every individual is unique in behavior or response to a certain factor (Fiocchi, 2014).

Many emphases have been put on the interactions between omics, but how can these interactions be studied simultaneously is still in question. Considering IBD, all its components, i.e., genome, microbiome, immunome, and exposome, all are reciprocally interrelated. Every single interaction can have a positive or negative effect. According to a recent GWAS, there are about one hundred and sixty-three variants of both manifestations of IBD (Jostins et al., 2012), and more variants may later be discovered. Seemingly not one single omic will be able to capture the complexity of IBD, and multi-omics approaches should provide a more comprehensive picture. Still, there is a need for such approaches and techniques to take all omics under consideration (Fiocchi, 2014).

1.7.2 IBD genetics

The first IBD related gene identified in Crohn's disease was the NOD2 gene (Nucleotide-binding Oligomerization Domain-Containing protein 2), which has a major contribution to Crohn's disease risk (Hugot et al., 2001). This gene codes for a protein that acts as an intracellular receptor for bacterial products, mainly in monocytes, and transduces signals leading to NF- κ B activation (Loddo and Romano, 2015). To date, more than 200 susceptibility loci have been identified in the genome for IBD (Jostins et al., 2012, Liu et al., 2015, Huang et al., 2017). Some of these are for CD, and some are for UC. The coding variants occupy about 30% of loci (Knights et al., 2013). Functions of many of these loci are still unknown. Other notable genes of interest include ATG16L1, IRGM or IL23R. ATG16L1 and IRGM are responsible for a defect in autophagy that can lead to reduced pathogen clearance (Salem et al., 2015, Ajayi et al., 2019). IL23R is a receptor for IL23, a proinflammatory cytokine, highly expressed in the membrane of T cells notably (Naser et al., 2012). A more recent study added 25 risk loci, including 3 integrin genes (de Lange et al., 2017). Recent fine-mapping efforts have tried to pinpoint confident causal variants among these numerous SNPs (Huang et al., 2014). Causal variants may also be identified through expression

quantitative trait loci (eQTL) studies (Momozawa et al., 2018). Other ongoing studies include Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) of IBD patients and controls. Thanks to these new approaches, a new variant in the ADCY7 gene has been shown to double the risk of UC (Luo et al., 2017). In addition to disease susceptibility, other studies have looked at disease prognosis as an endpoint for GWAS. Although generally less powered, one study notably identified four SNPs associated with prognosis but not with disease susceptibility (Lee et al., 2017).

1.7.3 Transcriptomics in IBD

The next most widely used omics technique in IBD after genomics has been gene expression studies via microarray or RNA-seq (Wang et al., 2009). One of the first, and still one of the main transcriptomic studies in IBD comes from the RISK cohort. This is a pediatric cohort of new-onset ileal CD. Early transcriptomics studies in CD have identified markers of inflammation in IBD: upregulated genes including DUOX2, and down-regulated ones including APOA1 (Haberman et al., 2014). In a more recent analysis, it was shown the ileal signatures, specifically of extracellular matrix genes, can predict strictures in this same cohort (Kugathasan et al., 2017). A first attempt at a molecular classification of CD on a low number of samples hinted at the idea that there would be two subtypes of samples, one ‘colon-like’ and one ‘ileum-like’ (Weiser et al., 2018). Lately, in the PROTECT trial including pediatric UC patients, transcriptome profiles have highlighted mitochondrial defects in UC (Haberman et al., 2019). Limitations of transcriptomic studies in IBD are that it is generally not possible to know which cell types are responsible for the differences we see in bulk tissue. These could come from changes in diverse immune cells or epithelial cells. The problem can be partly solved by using a single-cell multi-omics technique (Huang et al., 2014).

1.7.4 Proteomics in IBD

Proteomics studies do not always corroborate findings from RNA-seq studies, notably because of post-translational modifications (Pandey and Mann, 2000, Mann and Jensen, 2003, Bennike et al., 2014). Mass Spectrometry is a widely used technique in proteomics (Ong and Mann, 2005). Several proteomic studies have been conducted in IBD using either blood or mucosal tissues (Jin et al., 2019, Schniers et al., 2019, Chan et al., 2016, Pierre et al., 2020b). Pathways analyses have

revealed alterations in immune and inflammatory responses, complement cascade, PPAR signaling in UC. In CD, increased T-helper cell differentiation and elevated toll-like receptor and JAK/STAT signaling were observed. Another goal of these studies was to identify additional biomarkers in IBD, in addition to the existing CRP, and fecal Calprotectin. In that sense these studies have been somewhat unsuccessful given that to date CRP and calprotectin are still the only two markers used in the clinic. But a more recent had identified protein biomarkers associated with a risk of relapse after infliximab removal in CD (Pierre et al., 2020a).

1.7.5 Metabolomics in IBD

In the metabolomics field, a number of small studies have been conducted on fecal samples as well as on serum samples, plasma samples (Williams et al., 2012), colon tissue biopsy samples (Bjerrum et al., 2010), and urinary samples (Bjerrum et al., 2010) in mouse models of colitis and human IBD. But the dysregulation of gut metabolites in IBD is not well understood (Storr et al., 2013). A recent study by Fransoza and al. aimed to fill this gap. Using untargeted metabolomic analysis on more than 200 samples of CD, UC and non-IBD controls, they found that sphingolipids and bile acids were increased in IBD, while triacylglycerols and tetrapyrroles were depleted. These profiles were generally correlated with the level of fecal calprotectin. One challenge is that the majority of these metabolites are still uncharacterized, but their function could be identified by ‘guilt-by-association’ methods. These metabolites could prove very informative to the pursuit of small molecule compounds reducing inflammation in IBD.

1.7.5 Epigenetics in IBD

As with other inflammatory diseases, the field of epigenetics in IBD is fairly new but holds a lot of promises. Histone modifications, methylation marks and chromatin rearrangements have been studied in IBD, for either case-control, disease activity or disease behavior studies. Yet, to date the consensus on epigenetic modifications in IBD is fairly limited. Furthermore, it is generally difficult to target epigenetic modifications, making it seem like an out-of-reach target for therapeutics (Zeng et al., 2019, Ray and Longworth, 2019, Mateos et al., 2019). The same can be said for

microRNA (miRNA), which also fall into the category of epigenetic modifications (Schaefer, 2016).

1.7.6 Microbiome in IBD

Early 16s microbial studies in IBD have identified dysbiosis, the imbalance of different species in the gut, and reduced alpha-diversity as a marker of IBD. The magnitude of this dysbiosis is debated, as well as its potential for tracking inflammation and predicting disease relapse. Several species have also been identified as beneficial (*Faecalibacterium prausnitzii*) or detrimental (*Ruminococcus gnavus*) (Khan et al., 2019, Sokol et al., 2008). More recently shotgun metagenomic sequencing and strain analysis has been conducted on IBD patients and might yield new insights into the disease. Some strains have been shown to interact with T cells in the gut and could be targeted. Fecal microbiota transplantation is another area of great interest. So far it has shown promises in *Clostridium difficile* infection (Liubakka and Vaughn, 2016), but results in IBD are conflicting (Sokol et al., 2020b). Poor engraftment of the donor microbiome could be hindering these approaches.

1.7.7 Single-cell in IBD

Two major studies (Smillie et al., 2019, Martin et al., 2019) have been conducted in IBD using single-cell techniques, one for Crohn's disease, one for Ulcerative Colitis. Both as part of the new Human Cell Atlas initiative. The UC study has revealed new subsets of immune and epithelial cell types and showed an association between the presence of inflammatory fibroblast and resistance to anti-TNF treatment. This study has also mapped GWAS susceptibility loci to identify in which cell type they were expressed. The CD study, utilizing both high-dimensional cytometry (Cytof) and scRNA-seq on ileal CD samples, has shown that the presence of IgG plasma cells, inflammatory mononuclear phagocytes, activated T cells, and stromal cells was associated with resistance to anti-TNF treatment. Yet, these studies are usually costly and laborious, restricting them for now to a low number of patients analyzed.

1.8 Multi-omics in IBD

The complex traits of IBD are attributed to an abnormality in immune responses, environmental effects, microbiome, diet, and drug histories as well as immune functions that can be impaired by genetic mutations leading to inflammatory bowel diseases (Tyler et al., 2013). Dysbiosis or imbalance in the gut microbiome can occur because of the bodies' immune response towards commensal bacterial species. Dysbiosis conditions can also change the immune response or microbiome structure as an increased number of opportunistic bacteria results in more pathogens invading the lining of the gut. Diet also plays a role in changing the gut environment as well as the microbiome (Kugathasan et al., 2017). Drugs that are antimicrobial and immunosuppressants in nature are also thought to bring changes in the microbiome and the immune responses. Response to some IBD drugs can also be predicted by the genetics of the host (Gevers et al., 2014). So, it is seen that there are wide associations of genetics, medication, environment, microbiome, and immune responses in IBD. All such aspects of omics can be studied via technologies of high throughput sequencing. Genomics, transcriptomics, and immunomics in general give an overview of the human immune response and genome of the host. Various gene sequencing technologies, metagenomics, meta-transcriptomics give an overview of the microbiota. Signaling between host and microbiome and simultaneous metabolism can be studied via metabolomics and proteomics (Huang et al., 2014). The dual RNA sequencing technique can be used to study the transcriptome of the host and microbiome at the same time (Wang et al., 2009).

1.8.1 Multi-omics studies in IBD

A cohort study was conducted by Kugathasan et al. (2017) focusing on stricturing and penetrating diseases, as they lead to the morbidity of Crohn's disease in adults and children and that there was less information available on the risks of these complications and the effects of treatment on the disease progression (Kugathasan et al., 2017). The study analyzed multi-omics ("genotypes, antimicrobial serologies, ileal gene expressions, the microbiota of ileal, rectal and fecal samples") of the patients and found that anti-TNF α therapy reduces the risk of penetration and diminishes the risk of stricturing. Moreover, it was found out that ileal genes that control the regulation of the

extracellular matrix are associated with stricturing complications (Kugathasan et al., 2017). Another study combined GWAS and eQTL to establish a transcriptional risk score predictive of complications in Crohn's disease (Marigorta et al., 2017).

Another cohort study was conducted by Gevers et al. (2014) about the treatment of IBD in the new onset of Crohn's disease in patients. The study found out through microbiome studies that the use of antibiotic drugs adds to the microbial dysbiosis in Crohn's disease. Also, early diagnosis of Crohn's disease can be made through analysis of the microbiota of rectal mucosal sample (Gevers et al., 2014). A study on mucosal transcriptomes in ulcerative colitis, done by Haberman et al. (2019) revealed that in active patients, increase in mitochondrial genes along with innate immune genes are related to the severity of the disease. A subset of these severity genes can help predict the remission of disorder following corticosteroids use. An association between this gene subset and anti-TNF α and anti- α 4 β 7 was also found. These changes are also associated with the gut microbiota composition (Haberman et al., 2019).

One of the best examples to date of a multi-omics study in IBD comes from the Human Microbiome Project 2 (HMP2) cohort (Lloyd-Price et al., 2019). This huge effort analyzed genome, transcriptome, methylome of the host in blood and biopsy, and meta-genome, meta-transcriptome, metabolome, proteome and virome of the fecal microbiota. This including CD, UC, and non-IBD samples from adult and pediatric subjects from 5 different cohorts over the US, in different locations of the gut, at different timepoints. One major concern for this study is analysis power. By multiplying so many conditions and combinations it is hard to see how we can ascertain the robustness of some of the results. Still the authors were able to identify major disruptions in the microbiome (clostridia) and the metabolome (bile acids), as well as temporals shifts with disease activity.

Last, an in-depth study conducted by Smillie et al. (2019) assessed the pathways and cell type specifications of risk alleles of ulcerative colitis, found by GWAS, using 51 subsets of cells from colon mucosa. It was found that in the disease course, expansion is seen in M-like cells, inflammatory monocytes and fibroblasts, as well as CD8 T cells expressing IL-17. Major findings also include that drug response in UC can be affected by the presence of pro-inflammatory

fibroblasts (Smillie et al., 2019). Each omics study in IBD provides insights about a single aspect of the disease that is much complex in nature. So, multi-omics studies can be key to develop novel and comprehensive insights on the disease in order to devise various treatment strategies. However, this is an uphill task to collect all omics data on a cohort of the same patients and analyze it simultaneously. Meticulous planning of such a study is often required (Huang et al., 2014).

Outlined Methods

The methods for each specific study are described in the results section separately. I give here, a brief overview of the methods used.

Cohort selection

For the meta-analysis study, we performed a search for IBD gene expression studies on GEO, in the literature and in our different centers. To maximize robustness, we selected 5 independent cohorts with the greatest sample size, including patients with Crohn's ileitis, Crohn's colitis, Ulcerative colitis, as well as ileum control, and colon control [Supplementary Figure 1, Table 1]. Cohorts with lack of basic clinical information on samples were excluded. The RISK cohort⁶, the REMIND study⁷, two cohorts from Leuven^{8,9} and the PROTECT trial¹⁰ were chosen. In each cohort, we selected only samples obtained from patients with active disease, and no prior to treatment. In each study, biopsies were collected from IBD patients and controls either during endoscopy or from a surgical specimen. RNA was extracted and analyzed with microarray chips or RNA sequencing.

Gene expression analyses

For public cohorts, RNA-seq count and microarray expression data were downloaded from the Gene expression omnibus (GEO) website. For the REMIND cohort microarray expression data were generated in-house. We collected patient characteristics for each study, including gender, location of biopsies, and disease status from GEO. We collected relevant clinical factors for the REMIND study in the Case Report Form (CRF). RNA-seq count matrix was processed and normalized using DESeq (Anders and Huber, 2010). Microarray data was processed and normalized using Robust Multichip Average (RMA) method. To calculate differentially expressed genes (DEG) we fitted a linear model on each gene using the log-transformed, library-size normalized, counts for RNA-seq, and the gene expression of microarray data. Then, for each comparison, FDR 5% significant genes were selected. Finally, the top 20 DEGs per comparison were selected to generate the heatmaps, ranked by absolute log Fold-Change (logFC).

GWAS study

We mapped our samples to the HapMap Project (International HapMap et al., 2010) to obtain ancestries for our cohort. The GWAS for post-operative recurrence was performed using standard methods. SNPs were imputed using Michigan Imputation Server (imputationserver.sph.umich.edu), using the 1000 genomes reference. Association testing between SNP and outcome was run using plink 1.9 (Purcell et al., 2007), with Minor Allele Frequency (MAF) < .05, Hardy-Weinberg equilibrium 10^{-6} . We adjusted for Sex, smoking, post-operative anti-TNF and PC1, PC2 as covariate.

Microbiome analyses

16s microbiome FASTQ were normalized and aligned using the Qiime (Bolyen et al., 2019) software. OUT assignment was done in the Qiime software. Count taxa OTU tables were then generated and log-transformed for downstream analyses.

Mutli-omics analyses

Correlations between each distinct omic (microbiome-transcriptome) were calculated using the Spearman coefficient. P-value and multi-correction testing (Bonferroni) were also performed to obtain significance.

Batch correction

To correct for batch effect we used Combat, a widely used batch correction algorithm, based on empirical Bayes regression (Johnson et al. 2007). After batch correction, with cohort as a batch and location (ileum, colon) as a covariate, we were able to obtain distinct clusters for the ileum samples and the colon samples.

eQTL, mQTL

For eQTL and mQTL studies a linear model was fitted to expression or microbiome data like so: 'Gene expression' ~ SNP or 'Microbiome abundance' ~ SNP. Results were corrected for multiple comparison using Bonferroni correction.

Pathway analyses

To derive the pathways containing differentially regulated genes, we performed a hyper-geometric test to assess over-representation. We queried Gene-Ontology (GO) (Mi et al., 2019), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), Reactome (Jassal et al., 2020). The top 100 upregulated and downregulated genes were selected from the bulk RNA-seq study to execute the test. Pathways with p-value <0.05 were reported.

Predictive modelling

For predictive modelling the cohort was split into training (70%) and testing set (30%). Then a regularized linear model was fit on the gene expression matrix with a 10-fold cross-validation in the package glmnet. The model accuracy was assessed on the testing set using Area-Under-the-Curve (AUC).

Genetic risk score

To characterize the individual genetic risk score for IBD, we calculated an IBD polygenic risk score (PRS) for each patient defined as follows:

$$IBD_risk_score = \sum beta_i * SNP_i$$

By capturing multiple SNPs in one metric, such score has been able to identify groups of individuals with substantially increased risks in multiple diseases (Crouch et al., 2020)

Differential expression

We performed a differential analysis in each cohort applying three different contrasts: Crohn's ileitis vs. ileum control, Crohn's colitis vs. colon control and Ulcerative colitis vs. colon control. Differential expression was obtained via a linear model on each gene using limma's lmFit, and empirical Bayes statistics (eBayes). We used a threshold value of FDR < 5%.

Meta-analysis

We performed a meta-analysis of each differentially expressed genes in the different studies. Log Fold-Change (LogFC) values were combined using invariance variance model in the R package metafor. Genes were included in the meta-analysis only if they were differentially expressed in the same direction in at least two cohorts.

Data availability & browser

The raw data from each study can be accessed via GEO (GSE117993, GSE75214, GSE16879, GSE109142). The curated differential expression tables resulting from the meta-analysis for each disease subtype are available as supplementary. We also provide an online browser where the results of the meta-analysis can be queried (https://ibdot.shinyapps.io/IBD_meta_analysis/). The tool can take a gene or a list of gene (Entrez gene symbol) and will display a heatmap of DEGs in the three conditions Crohn's ileitis vs. ileum control, Crohn's colitis vs. colon control, Ulcerative colitis vs. colon control.

Immuno-staining

Three micrometers thickness ileal and colonic resection margins; tissue sections fixed on polarized slides were first deparaffinized and rehydrated with successive baths of Xylenes isomeric mixtures and Ethanol at different concentrations. Antigen retrieval was performed using Citrate buffer (pH 6,0; Sigma Aldrich®) heated in a microwave. For immunohistochemistry, endogenous peroxidase activity was inhibited using hydrogen peroxide. After blocking unspecific bonds with blocking buffer (prepared with a 5% species's serum of the secondary antibody), staining with primary antibodies was performed. For CA2 immunohistochemistry, primary antibodies were detected with an iVIEW DAB Detection Kit (Ventana ®) using an automated system. CA2 staining was performed using primary nuclear antibody from Abcam (ab182611, Dilution 1/50, Rabbit polyclonal Antibody).

Scientific questions and thesis aim.

IBD are highly heterogenous diseases with a varying course, and challenging medical and surgical management. In this work we ask whether we can use multi-omics studies to improve our understanding of the pathophysiology of IBD, identify gene drug targets, and predict disease course. We will use publicly available IBD mucosal gene expression data, totaling more than 1000 samples, to ask whether ileal or colonic diseases share common or distinct gene signatures. These datasets have been generated using different platforms (microarray and RNAseq), and it remains unknown whether we can effectively combine both. We will also compare our results with another IBD manifestation called pouchitis. In the second section, we will ask whether multi-omics in a cohort of patients undergoing surgery (REMIND) can accurately predict postoperative recurrence after ileocecal resection in Crohn's disease (CD), and help us understand the physiopathology of recurrence. We will demonstrate that using both 'open-data' publicly available data and our own cohorts we can leverage multi-omics analysis, notably transcriptomics and genetics to improve our understanding of IBD.

Chapter 2.

2.1 Meta-analysis of IBD gut samples identifies specific markers of disease in Ileal and Colonic Diseases

Here we present a first example of how to leverage publicly available gene expression data in IBD to improve our understanding of the disease. Analyzing more than 1000 samples from 6 different studies including ileal CD, colonic CD, UC, control ileum and colon, we robustly identify specific markers of ileal and colonic diseases. These results suggest that, at the gene expression level, these diseases are better characterized by their localization (ileum vs. colon) than their classification. The markers identified could also lead to specific therapies for ileal CD, colonic CD and UC.

Meta-analysis of IBD gut samples gene expression identifies specific markers of ileal and colonic diseases

Kevin Perez¹, Marjolaine Ngollo¹, Keren Rabinowitz², Nassim Hamoudi¹, Philippe Seksik⁴,
Ramnik J. Xavier³, Mark J. Daly^{3,4}, Iris Dotan², Lionel Le Bourhis¹, Matthieu Allez¹

¹Hopital Saint-Louis APHP, Université Denis Diderot Paris 7, Inserm U1160, Paris, France

²Division of Gastroenterology, and the Felsenstein Medical Research Center, Rabin Medical Center, Petah Tikva, Israel, and the Sackler Faculty of Medicine, Tel Aviv University, Israel

³Broad Institute of MIT and Harvard. Massachusetts General Hospital. Harvard Medical School, Boston, United States

⁴Hopital Saint-Antoine APHP. Université Sorbonne. UPMC Paris 6, Paris, France

⁵Finnish Institute for Molecular Medicine, Helsinki, Finland

Corresponding author: Matthieu Allez (matthieu.allez@aphp.fr)

Conflict of interest:

The authors disclose no conflict of interest

Acknowledgments

We are grateful to the patients and their families who contributed to these studies. This work was supported by the Helmsley Charitable Trust.

Summary (40 words). We meta-analyzed more than 1000 samples from 5 different cohorts of mucosal gene expression from patients with IBD and controls, and identified specific markers of ileal and colonic diseases. We showed that colonic CD resembles UC, and is distinct from ileal CD, which is in turn similar to pouchitis.

Abstract (250 words).

Background

Inflammatory Bowel Diseases (IBD) are chronic inflammatory diseases that can affect any segment of gastrointestinal tract. Ulcerative colitis (UC) is limited to the colon and the rectum, while Crohn's disease (CD) may involve any part of the gastrointestinal tract and is most prevalent in the terminal ileum and the colon. The pathogenesis and disease course can vary greatly according to the location of the disease, asking the question whether ileal and colonic CD are indeed two separate entities.

Methods

We utilized meta-analysis to identify specific signatures of ileal and colonic diseases, and to compare gene expression profiles in CD ileitis, CD colitis, UC and pouchitis. We meta-analyzed 1047 samples from 5 different studies of mucosal gene expression of patients with IBD and healthy individuals, in ileal and colonic biopsies. In each cohort separately, we compared CD ileitis to normal ileum, and CD colitis and UC compared to normal colon, and meta-analyzed the results.

Results

We identified specific markers of ileal (FOLH1, CA2) and colonic (REG3A) inflammation, and showed that, with disease, some cells from the ileum start to express colonic markers. We confirmed by immune-histochemistry (IHC) that these markers were specifically present in ileal

or colonic diseases. We highlighted that, overall, colonic CD resembles UC, and is distinct from ileal CD, which is in turn similar to pouchitis.

Conclusions

We demonstrated that ileal and colonic diseases exhibit specific signatures, independent of their initial clinical classification. This supports molecular, rather than clinical disease stratification, and may be used to design drugs specifically targeting ileal or colonic diseases.

Keywords (3-5)

IBD, transcriptome, disease location, meta-analysis

Introduction

Inflammatory bowel diseases (IBD) are chronic inflammatory diseases that can affect any segment of gastrointestinal tract. Ulcerative colitis (UC) is limited to the colon and the rectum, while Crohn's disease (CD) may involve any part of the gastrointestinal tract and is most prevalent in the terminal ileum and the colon.¹ There is a high heterogeneity in the clinical presentation and course of these debilitating diseases, thus presenting a continuous challenge to medical care. In addition to the CD/UC distinction, CD is often categorized by the segment of the GI tract being affected: as upper gastrointestinal, ileal, ileo-colic, or colonic disease². The pathogenesis and disease course of these subtypes can vary drastically, pointing to the notion that perhaps CD should be divided into ileum-dominant and isolated colonic diseases³. Furthermore, genetics studies have shown that the genetic risk for IBD is not uniform, but rather a continuum of disorders, from UC to colonic CD and ileal CD⁴. Another type of inflammation, pouchitis, develop in the majority of patients with UC undergoing total proctocolectomy and ileal pouch-anal anastomosis restoration (pouch surgery). The ‘*de novo*’ inflammation occurring in the ileal pouch resembles that of ileitis in CD⁵.

Multiple studies of mucosal gene expression in IBD have now been published⁵⁻⁶. Meta-analysis is a powerful tool, notably in epidemiology and genetics, to combine results from different studies and highlight the most significant signals. Here we aimed to utilize meta-analysis to identify specific signatures of ileal and colonic diseases, and to assess whether colonic CD has

a similar profile to UC, and pouchitis-to ileal CD. We meta-analyzed 1047 samples from 5 different studies of mucosal gene expression of ileal and colonic biopsies of patients with IBD and healthy individuals. We analyzed differential gene expression in CD ileitis compared to normal ileum, and in CD colitis and UC compared to normal colon. In addition to assessing the reproducibility of results from studies utilizing different techniques (RNA-seq and microarray), we identified specific markers of ileal and colonic inflammation, and showed that the segment affected determines the expression profile rather than the type of disease. We further confirmed by immune-histochemistry (IHC) that these markers were specifically present in ileal or colonic diseases.

Methods

Cohort selection

We performed a search for IBD gene expression studies on GEO, in the literature and in our different centers. To maximize robustness, we selected 5 independent cohorts with the greatest sample size, including patients with Crohn's ileitis, Crohn's colitis, Ulcerative colitis, as well as ileum control, and colon control [Supplementary Figure 1, Table 1]. Cohorts with lack of basic clinical information on samples were excluded. The RISK cohort⁶, the REMIND study⁷, two cohorts from Leuven^{8,9} and the PROTECT trial¹⁰ were chosen. In each cohort, we selected only samples obtained from patients with active disease, and no prior to treatment. In each study, biopsies were collected from IBD patients and controls either during endoscopy or from a surgical specimen. RNA was extracted and analyzed with microarray chips or RNA sequencing.

Data gathering & QC

We used Gene Expression Omnibus (GEO) tools to gather gene expression matrices from each study. Microarray data was preprocessed, normalized and quality controlled using Robust Multi-array Average (RMA). RNA-seq data was aligned using the STAR pipeline. Each sample's annotation was standardized for disease type (CD, UC) and biopsy location (ileum, colon).

Differential expression

We performed a differential analysis in each cohort applying three different contrasts: Crohn's ileitis vs. ileum control, Crohn's colitis vs. colon control and Ulcerative colitis vs. colon control. Differential expression was obtained via a linear model on each gene using limma's lmFit, and empirical Bayes statistics (eBayes). We used a threshold value of $FDR < 5\%$.

Meta-analysis

We performed a meta-analysis of each differentially expressed genes in the different studies. Log Fold-Change (LogFC) values were combined using invariance variance model in the R package metafor. Genes were included in the meta-analysis only if they were differentially expressed in the same direction in at least two cohorts.

Data availability & browser

The raw data from each study can be accessed via GEO (GSE117993, GSE75214, GSE16879, GSE109142). The curated differential expression tables resulting from the meta-analysis for each disease subtype are available as supplementary [Suppl.]. We also provide an online browser where the results of the meta-analysis can be queried (https://ibdot.shinyapps.io/IBD_meta_analysis/).

Immuno-staining

Three micrometers thickness ileal and colonic resection margins; tissue sections fixed on polarized slides were first deparaffinized and rehydrated with successive baths of Xylenes isomeric mixtures and Ethanol at different concentrations. Antigen retrieval was performed using Citrate buffer (pH 6,0; Sigma Aldrich®) heated in a microwave. For immunohistochemistry, endogenous peroxidase activity was inhibited using hydrogen peroxide. After blocking unspecific bonds with blocking buffer (prepared with a 5% species's serum of the secondary antibody), staining with primary antibodies was performed. For CA2 immunohistochemistry, primary antibodies were detected with an iVIEW DAB Detection Kit (Ventana ®) using an automated system. CA2 staining was performed using primary nuclear antibody from Abcam (ab182611, Dilution 1/50, Rabbit polyclonal Antibody).

Results

Similar results observed in each cohort, according to disease type and location

To identify robust markers of IBD in each location, we selected four studies of mucosal gene expression from IBD patients and controls, this included samples from Crohn's ileitis, Crohn's colitis, Ulcerative colitis, Ileum controls and Colon controls. The RISK⁵ cohort comprised of 322 samples of RNA-seq from newly diagnosed pediatric Crohn's disease patients. The REMIND¹¹ cohort consisted of 415 microarray data from surgical patients undergoing ileo-caecal resection. Two cohorts from Leuven^{9,8} of microarray data studying mucosal gene expression in IBD, and response to Infliximab (IFX) treatment were included. The PROTECT¹² trial looked at pre-treatment rectal gene expression in 206 pediatric UC patients. Characteristics of the patients including age, gender and disease duration are detailed in Suppl Figure 1.

In each cohort, we selected samples from inflamed tissue (active disease) and non-IBD controls only. Since the ileum and the colon harbor different gene expression in normal conditions, we compared Crohn's ileitis to ileum controls, and Crohn's colitis and UC to colon controls. In each distinct cohort we observed similar differentially expressed genes, according to disease type and location [Fig. 1]. We then performed a meta-analysis of all studies to obtain a more definitive assessment of the differences in each disease location and subtype.

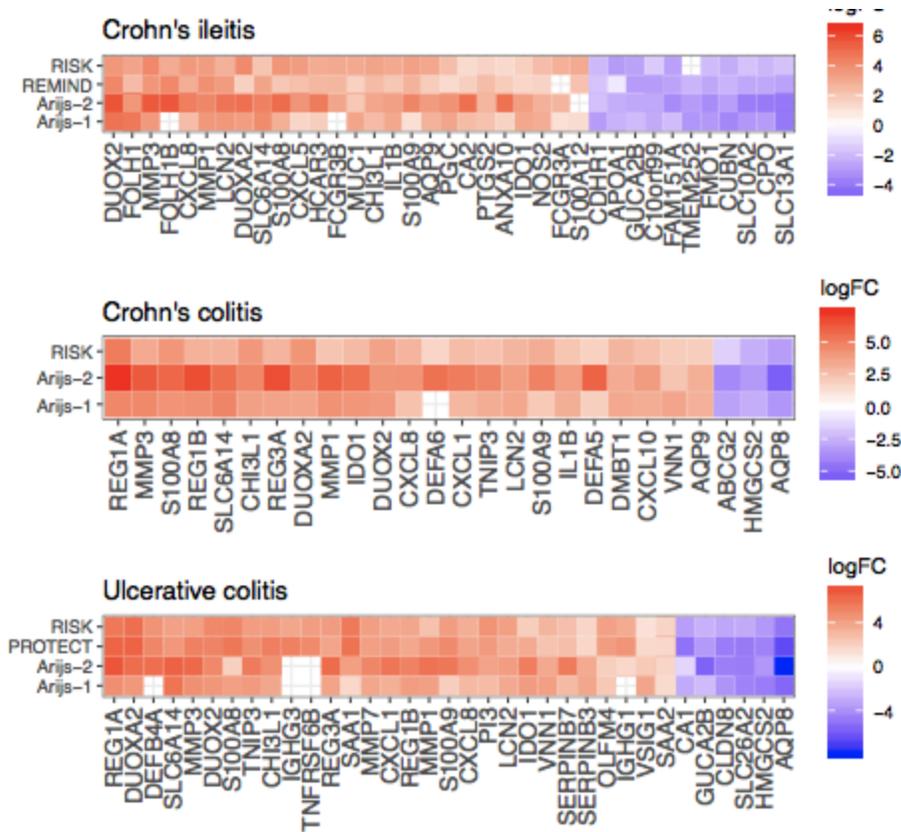


Figure. 1. Top differentially expressed genes in Crohn's ileitis vs. ileum control, Crohn's colitis vs. Colon control and Ulcerative colitis vs. colon control. Union of the top DEGs for each cohort, disease type and location, colored by Fold-Change. Non-significant comparisons are shown in grey tiles. FDR < 5%. All top DEGs are shown for each cohort.

Meta-analysis of IBD gut samples gene expression identifies location specific markers in ileal and colonic diseases

We combined the expression fold change from each comparison, using inverse variance method, for each contrast. [Fig. 2]. First, there was a common inflammatory gene signature to all three subtypes, including several markers of innate immunity DUOX2, DUOX2, MMP1, MMP3, S100A8, SLC6A14, CHI3L1, CXCL8, LCN2, and IL1B. Similarly, there were a number of genes that were downregulated in all three comparisons, including gut mucosa and tight junction genes GUCA2B, PRKG2, MT1M, CLDN8.

Overall, the pattern of gene expression in Crohn's colitis was similar to Ulcerative colitis, but partially different from Crohn's ileitis. This could indicate that the tissue type may be a determining factor of observed differences rather than the distinction between CD and UC. Among the differences between ileal and colonic diseases, we saw that genes that were primarily expressed in the ileum were downregulated in Crohn's ileitis, like CUBN, SLC13A1 and genes that were primarily expressed in the colon were downregulated with Crohn's colitis or UC, like AQP8 and HMGCS2. Conversely, a number of genes were upregulated in Crohn's ileitis only (FOLH1, CA2, ...); anti-microbial peptides REG genes (REG3A, REG1A), and defensins (DEF5A) were upregulated in colonic diseases (colonic CD and UC) only. [Fig. 3]

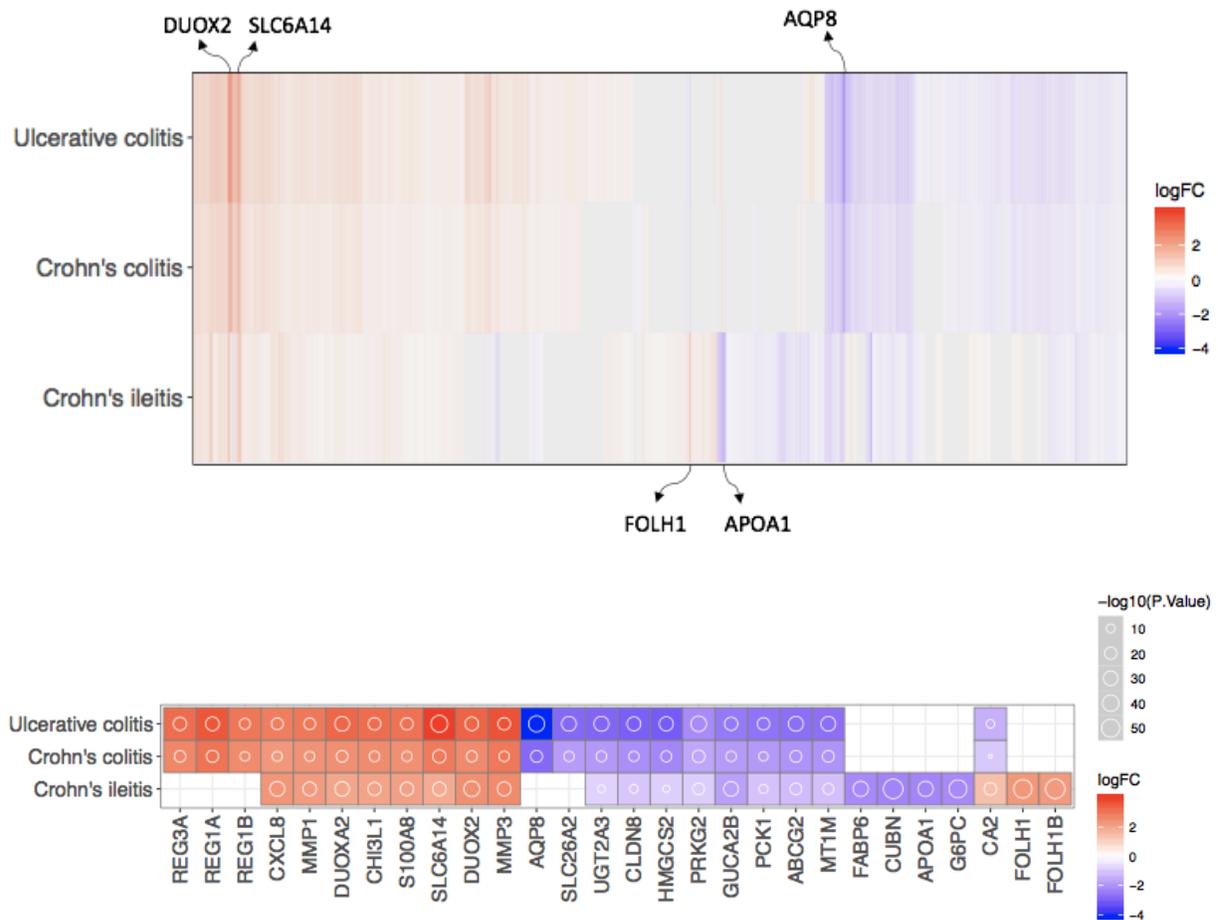


Figure. 2. Combined results of the meta-analysis in Crohn's ileitis vs. ileum control, Crohn's colitis vs. Colon control and Ulcerative colitis vs. colon control. All top DEGs are shown for each comparison, colored by Fold-Change. Non-significant comparisons are shown in grey tiles. logFC

were combined using an inverse variance method. Genes were kept only if they showed significant differential expression in at least two cohorts, with the same directionality. FDR < 5%.

FOLH1, REG3A and CA2 exhibit distinct patterns of expression in CD ileum and in CD and UC colon

Using gene expression studies, we were able to identify specific markers of ileal and colonic disease, particularly upregulation of FOLH1 and CA2 in ileal CD, and upregulation of REG3A in colonic diseases. We next looked at the actual level of expression of these markers in each cohort separately. We first see that our findings are replicated at the individual level, in each cohort separately. Interestingly, FOLH1 was upregulated in Crohn's ileitis only. REG3A was upregulated in colonic diseases, not expressed in the normal colon, while its level was high in both normal ileum and Crohn's ileitis without noticeable difference [Suppl Figure 2]. In contrast, CA2 was upregulated in Crohn's ileitis samples, with no expression in normal ileum, but highly expressed in normal colon.

To investigate this pattern of gene expression further, we stained CA2 in the normal colon, inflamed ileum in CD, non-inflamed ileum in CD from the REMIND cohort, and non-IBD controls. Consistent with our findings, CA2 was ubiquitously expressed in the colon crypts, not expressed in normal and non-inflamed ileum, and exhibited sporadic expression in some regions of the inflamed ileum in CD. [Fig. 4].

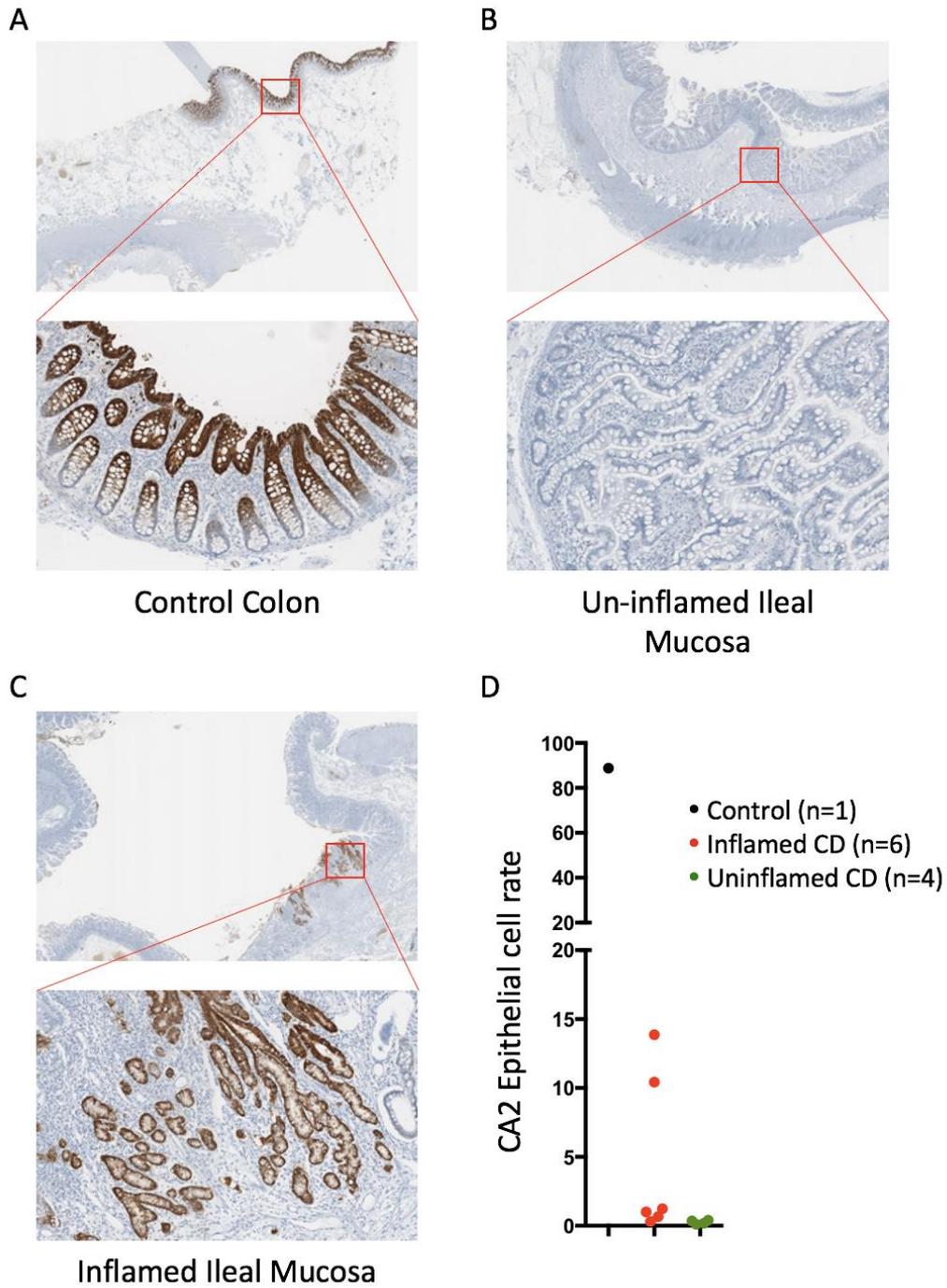


Figure. 4. A. B. C. Immuno-histochemistry (IHC) of CA2 in control colon, inflamed ileum in CD, non-inflamed ileum in CD and control ileum. D. Quantification of CA2 epithelial cell rate in Control, inflamed ileal mucosa and un-inflamed ileal mucosa

Discussion

We performed a meta-analysis of gene expression in IBD, which included more than 1000 samples from 6 different cohorts. While each study taken separately yielded a separate message, here we identified similar results across studies coming from different centers, and using various techniques. We showed that the clinical distinction between CD and UC is only partially reflected in the transcriptome. On the contrary, colonic CD resembles UC, whereas pouchitis resembles CD, suggesting that the segment affected is more important in determining the disease signature. We therefore established a distinct signature for ileal and colonic diseases. Taken together, our results highlighted specific patterns of disease in CD ileitis (FOLH1, CA2), CD colitis, and UC (REG genes). We also established an inflammatory gene signature (CXCL8, MMP3, DUOX2, S100A8) common to all IBD subtypes. Folate hydrolase FOLH1/PSMA/GCPII is upregulated in CD ileitis only. FOLH1 is responsible for cleaving polyglutamyl folate and enable the freeing and uptake of folate¹⁴. FOLH1 has been shown by immuno-staining to be expressed on the apical part of the villi in IBD¹⁵. Rais et al¹⁵ had previously reported that FOLH1 is elevated in patients with IBD and that its inhibition seems to improve inflammation in preclinical murine models. Here we show that this finding is specific to CD ileitis, and pouchitis. This not only supports the similarities between ileal inflammation regardless of phenotypic definitions, but may also be important to guide future clinical studies. Indeed, it is for example already known that anti-TNF therapies show better efficacy in colonic than ileal diseases in CD¹⁶.

Another interesting finding is that REG genes are upregulated in colonic IBD, and they are highly expressed in the normal or diseased ileum. REG3A had been shown by immuno-staining¹² to be expressed in both ileum control, and ileal CD, not in colon control, but expressed in the crypts of the inflamed colon. These antimicrobial peptides are secreted by Paneth cells in the ileal crypts, which may suggest that inflammation in the colon drives the differentiation of secretory cells¹⁷, normally absent from this segment of the GI tract. Interestingly, Paneth cell metaplasia has previously been described in UC patient's mucosa¹⁸. Lastly, Carbonic Anhydrase (CA2) was upregulated in colonic IBD. CA2 is expressed in the colon, the stomach, and the duodenum particularly, where it helps neutralize gastric acid. However, it is normally not expressed in the ileum¹⁹. The reason why these ileum-specific genes or colon-specific genes start to be expressed in the colon or the ileum with disease is unclear. We could hypothesize that either inflammation is a driver for the expression of these genes, or that perhaps that this is a response to aberrant

microbial populations in each specific segment (an ileal microbial population that colonizes the colon, or vice-versa).

Generally, we used meta-analysis to highlight importance of specific ileal or colonic signatures by combining several studies. Using similar techniques, Kaiko¹² et al. recently identified PAI-1 as a key factor in mucosal damage studying gene expression data of colon-only samples from several cohorts of patients with IBD. Some of these results (DUOX2, APOA1) had also already been described in mono center cohorts⁷. There are several limitations to this study. First, even in patients with active disease, mucosal inflammation may vary between samples, particularly in CD. Second, patients recruited in these studies are at different stages of their disease (early vs late, behavior: inflammatory vs penetrating and stricturing disease), and may be receiving different treatments. We believe that those issues were addressed by the large sample size, as well as confirmatory sub-analysis in each separate study. Also, clinical features and outcome were not available for all samples. Hence, it was not possible to associate the expression of certain markers with a better or worse disease course, or a specific clinical phenotype. Moreover, gene expression studies on whole tissue do not account for heterogeneity in the cellular composition of the biopsy samples. In that regard, single cell studies may contribute information, tracking changes at the cellular level.

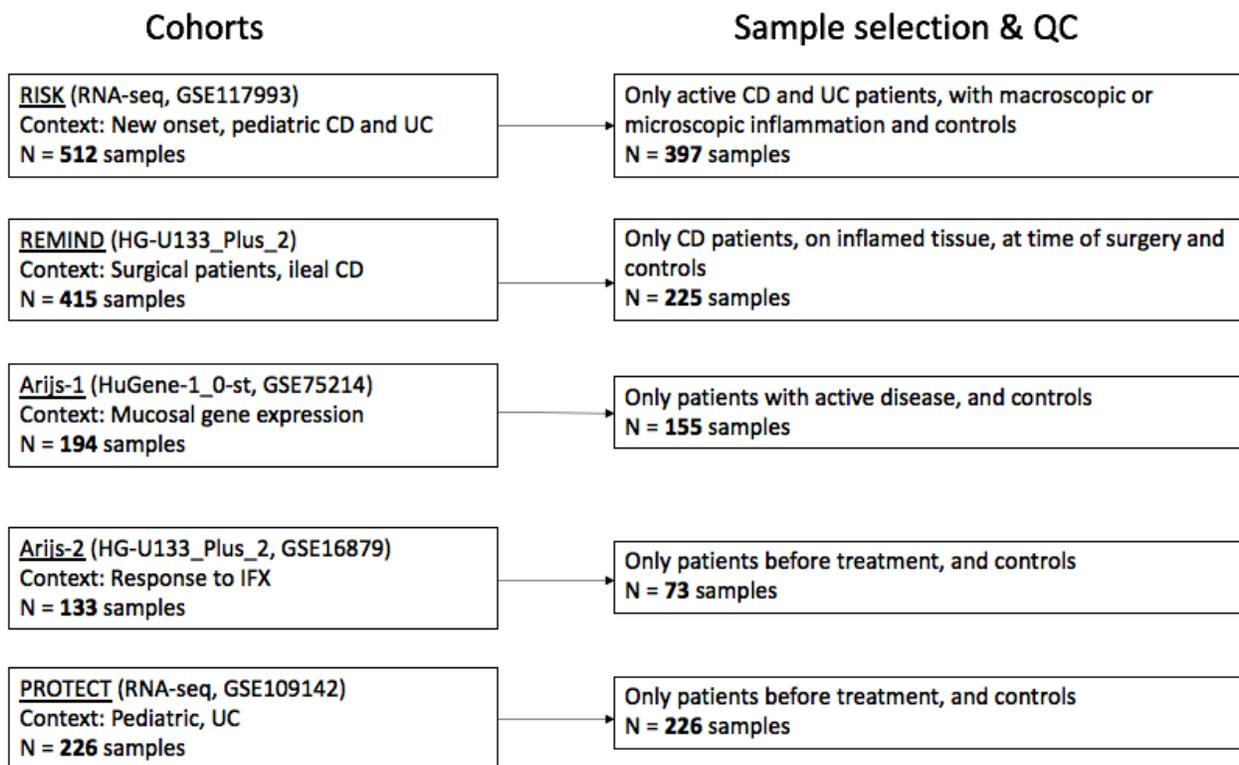
Conclusion

In summary, we hereby show that ileal and colonic IBD exhibit specific signatures, that may be more informative than their clinical classification. In addition to potentially highlighting different pathogenetic pathways, these findings may be relevant for drug design, targeting specifically either ileal or colonic IBD.

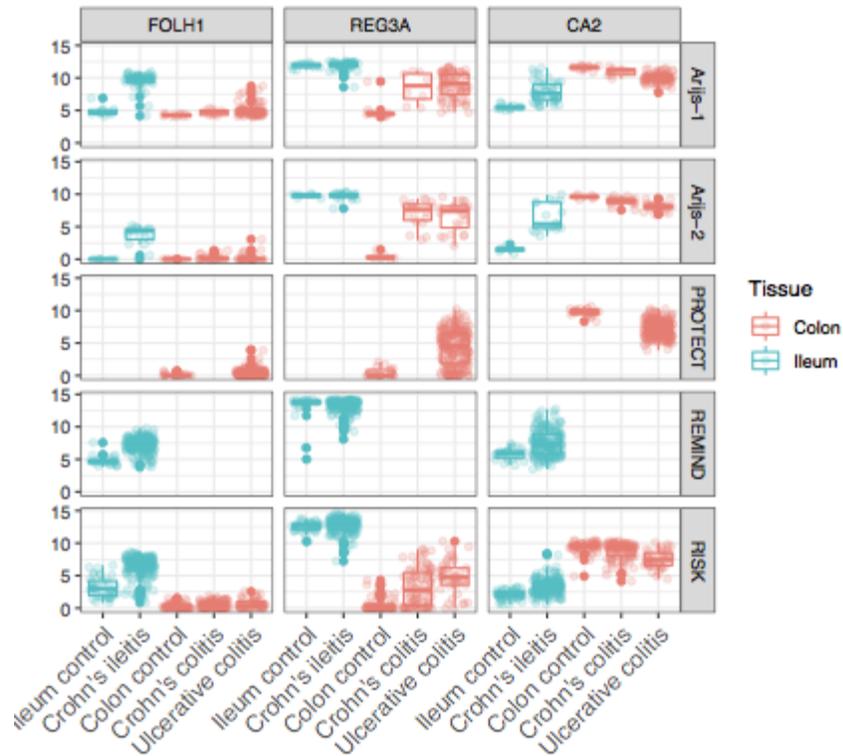
References

1. Abraham, C. & Cho, J. H. Inflammatory Bowel Disease. *N. Engl. J. Med.* **361**, 2066–2078 (2009).
2. Dulai, P. S. *et al.* Should We Divide Crohn's Disease Into Ileum-Dominant and Isolated Colonic Diseases? *Clin. Gastroenterol. Hepatol.* (2019). doi:10.1016/j.cgh.2019.04.040
3. Haberman, Y. *et al.* Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J. Clin. Invest.* **124**, 3617–3633 (2014).
4. Gaujoux, R. *et al.* Cell-centred meta-analysis reveals baseline predictors of anti-TNF α non-response in biopsy and blood of patients with IBD. *Gut* gutjnl-2017-315494 (2018). doi:10.1136/gutjnl-2017-315494
5. Kugathasan, S. *et al.* Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet* **389**, 1710–1718 (2017).
6. Haberman, Y. *et al.* Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J. Clin. Invest.* **124**, 3617–3633 (2014).
7. Auzolle, C. *et al.* Male gender, active smoking and previous intestinal resection are risk factors for post-operative endoscopic recurrence in Crohn's disease: results from a prospective cohort study. *Aliment. Pharmacol. Ther.* **48**, 924–932 (2018).
8. Vancamelbeke, M. *et al.* Genetic and Transcriptomic Bases of Intestinal Epithelial Barrier Dysfunction in Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **23**, 1718–1729 (2017).
9. Arijs, I. *et al.* Effect of vedolizumab (anti- α 4 β 7-integrin) therapy on histological healing and mucosal gene expression in patients with UC. *Gut* **67**, 43–52 (2018).
10. Haberman, Y. *et al.* Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nat. Commun.* **10**, 38 (2019).
11. Fumery, M. *et al.* Postoperative Complications after Ileocecal Resection in Crohn's Disease: A Prospective Study From the REMIND Group. *Am. J. Gastroenterol.* **112**, 337–345 (2017).
12. Kaiko, G. E. *et al.* PAI-1 augments mucosal damage in colitis. *Sci. Transl. Med.* **11**, (2019).
13. Rais, R. *et al.* FOLH1/GCPII is elevated in IBD patients, and its inhibition ameliorates murine IBD abnormalities. *JCI insight* **1**, (2016).

14. Sasaki, N. *et al.* Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5399-407 (2016).
15. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.)*. (2015). doi:10.1126/science.1262110
16. LÖNNERHOLM, G., KNUTSON, L., WISTRAND, P. J. & FLEMSTRÖM, G. Carbonic anhydrase in the normal rat stomach and duodenum and after treatment with omeprazole and ranitidine. *Acta Physiol. Scand.* **136**, 253–262 (1989).



Suppl. Figure. 1. Flowchart of the study. Description of selected cohorts, including technique, context and sample size. Criteria for sample selection, including disease activity and patient status. Quality control.



Suppl. Figure. 2. Expression of disease-specific genes CA2, FOLH1, REG3A, in each study in Crohn's ileitis, ileum control, Crohn's colitis, Colon control and Ulcerative colitis.

2.2. Other approaches and extension to Pouchitis

2.2.2. Pooled analysis of all samples from separate studies

As an alternative approach, we also tried to analyze all the samples from all cohorts together. This approach can be complex because of the heterogeneity in the cohorts (localization, disease), how the samples were processed, and the technique used to obtain gene expression (RNA-seq or microarray). Indeed, when we first tried to cluster all samples together, we saw each cohort was clustering separately indicating a clear batch effect [Figure 1]. To correct for batch effect we used Combat, a widely used batch correction algorithm, based on empirical Bayes regression (Johnson et al. 2007). After batch correction, with cohort as a batch and location (ileum, colon) as a covariate, we were able to obtain distinct clusters for the ileum samples and the colon samples. These clusters also recapitulated the disease subtypes we observed previously: ileum control, ileal

disease (Crohn's ileitis), colon control, colonic disease (Crohn's colitis, and Ulcerative colitis). Here again there was no marked difference between Crohn's colitis and UC [Figure 2]. We also saw that these clusters displayed a gradient of Calprotectin (S100A8/S100A9) gene expression, compatible with different levels of inflammation in the samples [Figure 3].

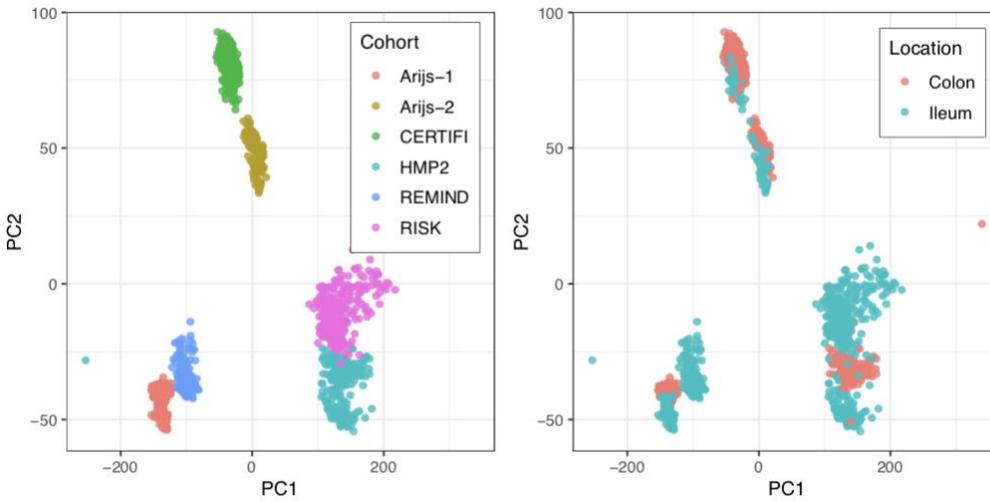


Figure 1. PCA of all samples **before** batch correction. Colored by cohort (left), and location (right).

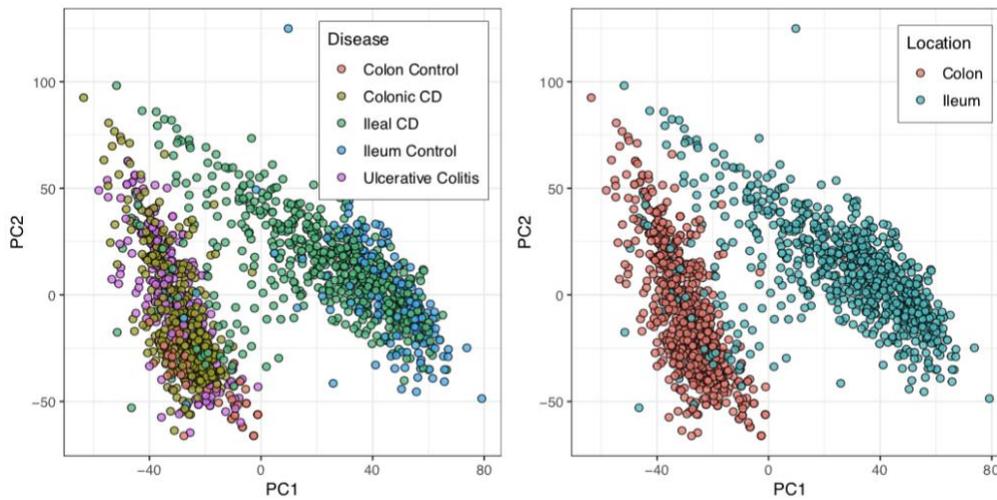


Figure 2. PCA of all samples **after** batch correction. Colored by disease subtype (left), and location (right).

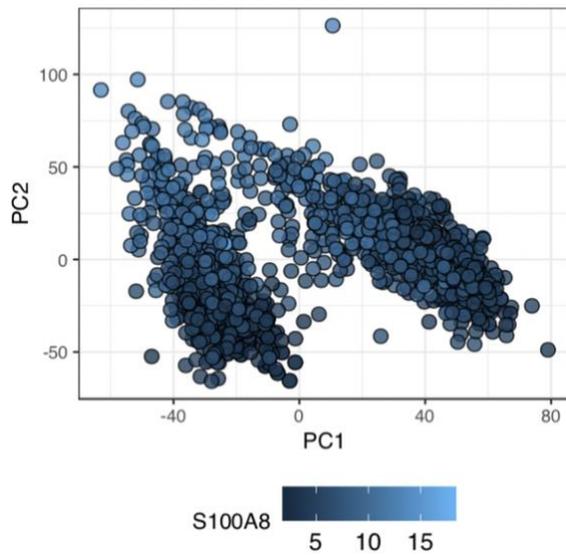
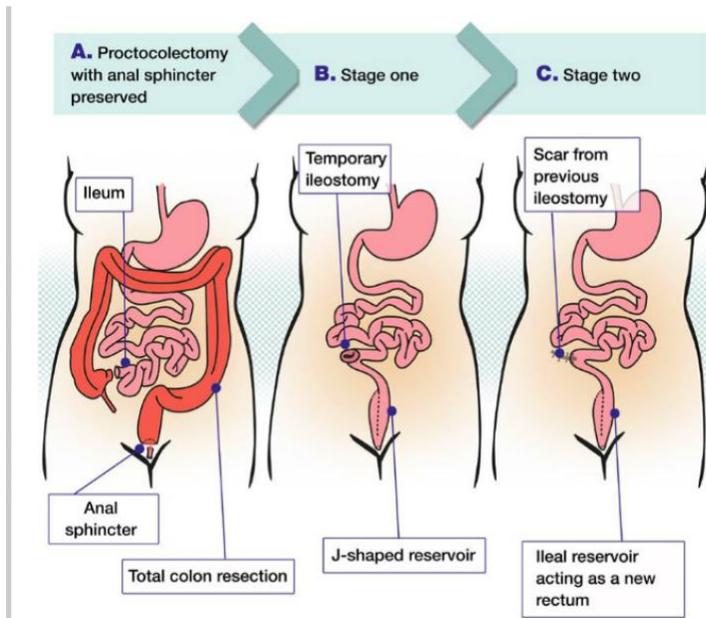


Figure 3. Gradient of Calprotectin (S100A8) gene expression in all samples.

2.2.3. Extension to pouchitis



Figure

Lastly, we sought to compare the gene expression changes in another instance of gut inflammation, Pouchitis. The Pouchitis is an inflammation that occurs in the lining of the ileal pouch created during surgical removal of the colon to treat ulcerative colitis (Mayo Clinic). To this end we

included a cohort of samples obtained patients after pouch surgery from the Rabin Medical Center, Israel (Dubinsky et al., 2020). The samples included both pouch inflammation as well as ileal control. By comparing the gene expression changes in Pouchitis and in all other IBD subtypes we saw that Pouchitis was clearly similar to Crohn's ileitis, and not Ulcerative Colitis or Crohn's colitis. Notably over-expression of FOLH1, CA2 and CEACAM7 were observed in Pouchitis, and REG1A/B was not upregulated, which is characteristic of an ileal disease. This finding adds to our postulate that the segment affected by the disease, and not the type of disease, will determine the gene expression changes with inflammation.

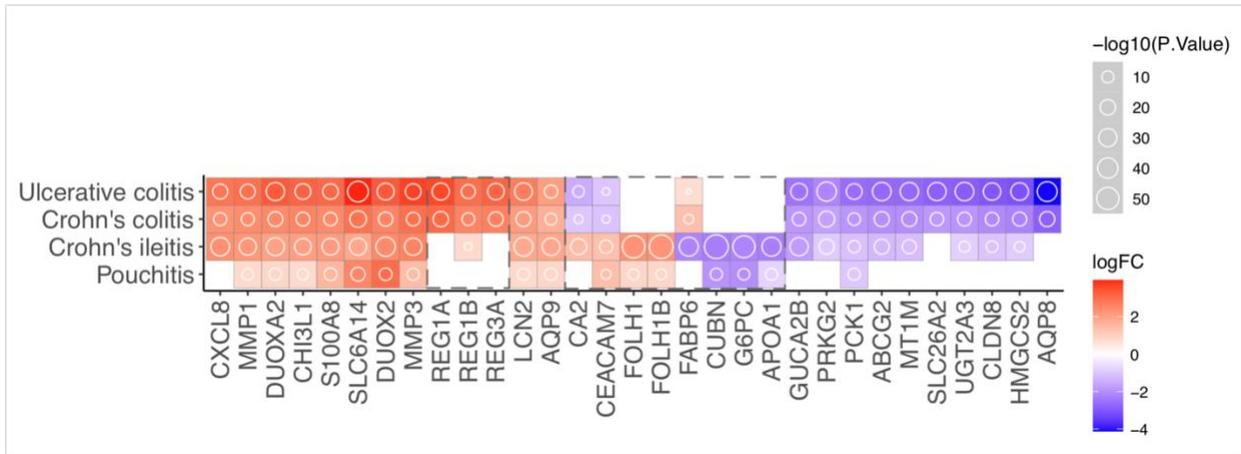


Figure 1. Gene expression changes in Pouchitis, Crohn's ileitis, Crohn's colitis and Ulcerative colitis. Last 3 come from the results of the meta-analysis, Pouchitis is an external cohort.

2.4 Conclusion.

We have used internal and external cohorts to identify specific markers of IBD in ileal (FOLH1, CA2) and colonic diseases (REG3A/B). These markers should be functionally validated to understand their role in disease, perhaps using organoid culture. We could see if the expression of these genes is triggered by inflammation on organoids. We could try to knock down or knock in (sh-RNA, AAV, or small molecules) some of the genes in mice models and see their impact on colitis. Some compounds are already available. Antimicrobial peptide treatments could also be used to boost the response. Anti-FOLH1 compounds are currently tested in pre-clinical trials (REF?). Seeing that some genes expressed in the ileum are expressed in the colon with inflammation and vice-versa, we could explore the interplay between the ileum and the colon. One ambitious idea would be to expose ileal organoids to microbiota isolated from the colon. We have

also identified new 'Paneth-like' secretory cells in the colon; perhaps we could try to isolate them and characterize their transcriptome, compared to normal ileal Paneth cells.

Chapter 3.

Genetic, multi-omics analysis of a surgical IBD cohort

3.1. The REMIND cohort

The REMIND cohort is a multi-omics cohort of surgical ileal CD patients. Biopsies were taken from surgical specimens after ileocecal resection. We performed gene expression, and microbiome analysis on the inflamed part (M0) of the mucosa and distant part (MI) at time of surgery. We also performed the same analysis on the biopsies gathered at follow-up endoscopy 6 months after surgery (M6). During this endoscopy about $\frac{2}{3}$ of the patients show signs of endoscopic recurrence. We also have clinical records for all of these patients. And genotyping is available for the majority of patients.

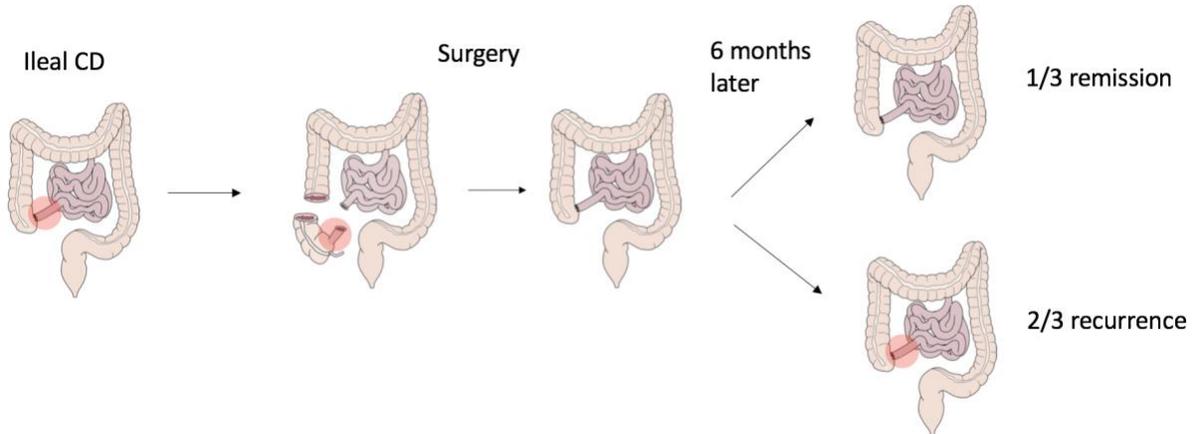


Figure 1. The REMIND cohort.

We first identified male gender as a risk factor for postoperative recurrence (POR), in addition to known risk factors: smoking, anti-TNF (protecting), previous surgery (Auzolle et al., 2018). We also showed via TCR sequencing that clonal expansions of T cells in CD are associated with smoking and POR (Allez et al., 2019). A microbiome study found bacteria in the ileal mucosa

associated with POR (Sokol et al., 2020a). Pathology studies identified microscopic lesions in the ileal margin predictive of POR (Hammoudi et al., 2020).

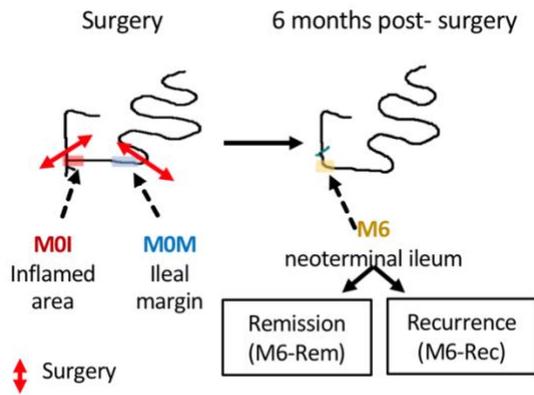


Figure 2. Sketch describing the different localizations (MOI, MOM, M6), and study design (Ngollo et al., 2020).

3.2. Transcriptome analysis of the REMIND cohort

In this section we analyzed the mucosal transcriptome of the REMIND cohort. Biopsies were available at time of surgery (inflamed part and margin) and 6 month later during follow-up endoscopy. We used bioinformatics tools to identify genes and pathways predictive of post-operative recurrence.

Crohn's disease mucosal transcriptomes highlight JAK/STAT pathway involvement in postoperative recurrence.

Marjolaine Ngollo, Kevin Perez, Nassim Hammoudi, Yuri Gorelik, Marc Delord, Claire Auzolle, Hugo Bottois, Dominique Cazals-Hatem, Madeleine Bezault, Stéphane Nancey, Benjamin Pariente, Xavier Treton, Mathurin Fumery, Antony Buisson, Nicolas Barnich, Philippe Seksik, REMIND study group investigators, Shai S. Shen-Orr, Lionel Le Bourhis, Matthieu Allez

Summary

Background: Crohn's disease displays heterogeneity in terms of disease location, progression and response to treatment. Most patients require an intestinal resection due to complicated disease, and a majority of them will experience disease recurrence. We aimed to characterize biological pathways driving postoperative recurrence and to identify predictors which could guide therapeutic management.

Methods: In this multi-center prospective study, we conducted transcriptome analysis on ileal mucosa of CD patients undergoing ileocolonic resection. Samples were collected from the most inflamed part of the ileum (n=200), from the ileal resection margin (n=149) and in the neo-terminal ileum 6 months after surgery (n=122). The primary endpoint was postoperative endoscopic recurrence at month 6. We applied a regression model to identify gene signatures predicting endoscopic recurrence.

Findings: Molecular patterns of early endoscopic recurrence significantly differed from those observed in chronic inflammation. Gene expression analysis of the inflamed area of the surgical specimens identified clusters of CD patients. However, pathway gene signatures at the ileal margin were superior to those of the inflamed area to predict postoperative outcome. Several pathways, including JAK/STAT signaling, cell adhesion molecules and extracellular matrix protein were associated with a higher risk of endoscopic recurrence. JAK/STAT pathway activation at the ileal margin predicted early postoperative endoscopic recurrence (AUC of 0.74) and was associated with a higher rate of clinical relapse in the long term. A significant increase of p-STAT3 levels at the ileal margin was associated with severe endoscopic recurrence (i3, i4) compared to no recurrence (i0) (p=0.02) and moderate recurrence (i1, i2) (p=0.04).

Interpretation: Gene expression analysis at the ileal margin of the surgical specimen may strongly predict postoperative outcome. The inflammatory processes associated with postoperative recurrence differ from those engaged in chronic inflammation. Activation of JAK/STAT signaling at the ileal margin of the surgical specimen is predictive of postoperative recurrence and should be specifically targeted.

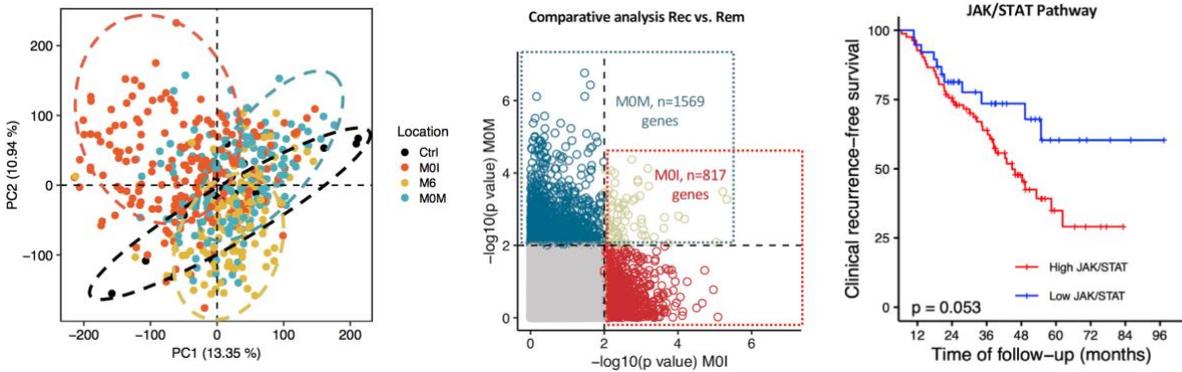


Figure 1: (left to right) A. PCA of all samples: Ctrl, M0, MI, M6. B. Genes predictive of Post-Operative Recurrence from biopsies at M0I and M0M. P-value is shown from a linear model on each gene. C. Kaplan-Meier curves of recurrence-free survival of JAK/STAT gene signature. Red line is the group with an increase activation of the pathway and blue line is the group with a low activation. The cut-off value between high and low group was computed by ROC AUC analysis.

3.3 Genetics analysis of the REMIND cohort

Here we analyze the genetics of the REMIND cohort and its relations to other available omics. 246 patients of the REMIND cohort were genotyped from peripheral blood, using the sequencing platform at the Broad Institute in Cambridge, MA for SNP array chip (GWAS). The baseline characteristics of the cohort, and clinical predictors of POR are listed below:

Table 1: Cohort characteristics, at baseline, and clinical predictors of Post-Operative Recurrence.

TABLE 1 Patient characteristics at time of surgery

	N = 225 (%)
Men	104 (47)
Mean age (y, SD)	34.6 (12)
Age at diagnosis	
≤16 y (A1)	26 (12)
17-40 y (A2)	176 (78)
>40 y (A3)	23 (10)
Mean disease duration (y, SD)	8.8 (9)
Mean time between resection and colonoscopy (mo, SD)	7.3 (1.9)
Active smoker at surgery	73 (32)
Smoking cessation	14 (19)
Previous intestinal resection	42 (19)
<10 y	14 (6)
Number of previous resection(s)	
0	183 (81)
1	29 (13)
2	9 (4)
3	4 (2)
Surgical indication	
Stricture complication	126 (56)
Penetrating complication	85 (38)
Failure of drug therapy	14 (6)
Disease phenotype (Montreal classification)	
Inflammatory phenotype (B1)	9 (4)
Stricture phenotype (B2)	117 (52)
Penetrating phenotype (B3)	99 (44)
Disease location (Montreal classification)	
Ileal (L1)	139 (62)
Ileocolonic (L3)	83 (37)
Anoperineal lesion	51 (23)
Extra-digestive symptoms	
Joint manifestations	41 (18)
Skin manifestations	12 (5)
Eye manifestations	4 (2)
Median Harvey-Bradshaw index (Q1-Q3)	5 (2-7)
Mean body mass index (kg/m ² , min-max)	21.8 (14-39)
Pre-operative enteral nutrition	37 (16)
Pre-operative parenteral nutrition	33 (15)
Previous exposure to anti-TNF therapy	142 (63)
Anti-TNF therapy <3 mo before surgery	112 (50)
Adalimumab	70 (62)
Infliximab	42 (38)
Previous exposure to thiopurines	163 (72)
Thiopurines <3 mo before surgery	71 (32)

TABLE 2 Bivariate analysis: predictors of early endoscopic recurrence

	No endoscopic recurrence (I0-I1), n = 117 (%)	Endoscopic recurrence (I2-I4), n = 107 (%)	P-value
Gender (male)	43 (37)	60 (56)	0.006
Mean age (y, SD)	33.8 (11.8)	35.6 (12.4)	0.30
BMI (kg/m ² , SD)	21.5 (4.8)	22.0 (3.5)	0.40
Mean time between diagnosis and surgery (y, SD)	8.2 (8.4)	8.5 (9.6)	0.70
Previous intestinal resection	15 (13)	27 (25)	0.03
Indication for surgery			
Stricture disease	63 (54)	63 (59)	0.30
Penetrating disease	49 (42)	36 (34)	
Other reasons	5 (5)	8 (7)	
Active smokers at surgery	26 (22)	46 (43)	0.001
Smoking cessation	4 (4)	10 (9)	0.80
Perianal disease	28 (24)	23 (22)	0.80
Penetrating phenotype	58 (50)	41 (38)	0.10
Anastomosis technique			
Side-to-side anastomosis	99 (85)	81 (76)	0.20
End-to-end anastomosis	16 (14)	24 (22)	
Side-to-end anastomosis	2 (1)	2 (2)	
Length of small bowel resection (cm)	22.6 (19.3)	23.5 (17.3)	0.70
Healthy resection margin (>5 cm)	64 (55)	48 (44)	0.10
Pre-operative antibiotics	41 (35)	35 (33)	0.80
Pre-operative steroids	39 (33)	38 (36)	0.80
Preoperative IS ^a	40 (34)	31 (29)	0.50
Pre-operative anti-TNF α ^a	59 (50)	53 (50)	1
Post-operative antibiotics	8 (7)	7 (6)	1
Post-operative IS	18 (15)	22 (21)	0.20
Post-operative anti-TNF α	40 (34)	26 (24)	0.09

We first sought to identify the ancestry population structure of our cohort. This information is important in the GWAS when adjusting for PCA components notably. We mapped our samples to the HapMap Project to obtain ancestries for our cohort.

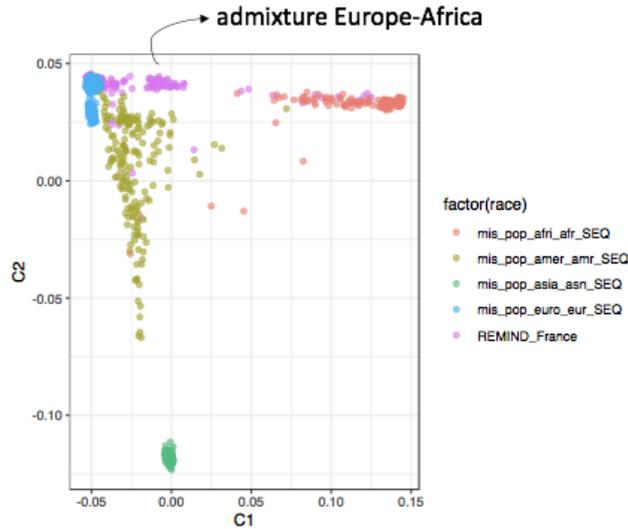


Figure 1. Population structure of the REMIND cohort. In red, yellow, blue and green, dots represent the various ethnic populations in the HapMap Project, projected on PC1 (C1), PC2 (C2). The REMIND cohort is shown in purple. The REMIND cohort lies in between European and African populations.

The Remind cohort was a mix of patients of European and African ancestry, with some admixed population between Europeans and Africans.

3.3.2 Distribution of IBD susceptibility SNPs in the REMIND cohort.

Here we focus only on IBD susceptibility SNPs, selected according to the IBD Genetics consortium data (Jostins et al., 2012; Liu et al. 2015(Huang et al., 2017)). We first looked at the general distribution of IBD susceptibility SNPs in the REMIND cohort, compared to the general population. We also distinguished between SNPs that have been associated with CD, UC or IBD.

Variant	Gene	IBD type	P-value (IBD GWAS)	Minor allele frequency (Gen. Pop.)	Minor allele frequency (REMIND)	Ratio (REMIND/Gen. Pop.)
rs5743293 (fs1007insC)	NOD2	CD	0	0.036	0.085	2.36
rs2066844 (R702W)	NOD2	CD	4.47E-140	0.063	0.127	2.02
rs2066845 (G908R)	NOD2	CD	2.34E-92	0.022	0.071	3.23
rs6752107 (T300A)	ATG16L1	CD	1.41E-73	0.453	0.446	0.98
rs6426833	RNF186	UC	1.51E-67	0.445	0.417	0.94
rs11236797	C11orf30	IBD	2.24E-50	0.467	0.46	0.99
rs6017342	HNF4A	UC	9.77E-48	0.456	0.429	0.94
rs3812565	CARD9	IBD	6.46E-46	0.402	0.375	0.93
rs3197999	MST1	IBD	1.38E-44	0.303	0.31	1.02
rs1887428	JAK2	IBD	2.57E-42	0.397	0.363	0.91
rs33993564	IRGM	IBD	8.32E-37	0.093	0.204	2.19
rs56167332	IL12B	IBD	3.72E-36	0.353	0.325	0.92
rs1801274	FCGR2A	IBD	9.77E-34	0.477	0.466	0.98
rs11614178	IFNG	IBD	2.24E-32	0.398	0.345	0.87
rs10775412	KSR1	CD	2.51E-25	0.316	0.222	0.70
rs61839660	IL2RA	CD	6.46E-24	0.094	0.125	1.33
rs28449958	IL27	IBD	7.59E-24	0.344	0.306	0.89
rs10748781	NKX2-3	IBD	6.46E-23	0.45	0.427	0.95
rs1260326	GCKR	CD	1.74E-21	0.41	0.442	1.08
rs395157	OSMR	IBD	2.22E-20	0.48	0.46	0.96
rs6561151	LACC1	CD	1.05E-19	0.239	0.292	1.22
rs2816958	NR5A2	IBD	1.14E-18	0.11	0.131	1.19
rs601338	FUT2	CD	1.78E-15	0.478	0.435	0.91
rs10220572	GPR65	IBD	2.14E-14	0.092	0.111	1.21

Table 1. Distribution of IBD susceptibility SNPs in the REMIND cohort.

Pos: SNP genomic coordinate (Chr:Bp), variant: SNP rsid, gene: implicated gene, trait: variant concerns IBD, UC or CD only, P: p-value of the association with IBD, UC or CD, MAF: minor allele frequency, Freq: frequency in the REMIND cohort, ratio: ratio between the REMIND cohort and the general population, log(ratio).

Overall, the REMIND cohort was enriched for NOD2 variants (fs1007insC, 2.36x, R702W, 2.02x, G908R, 3.23x), which is to be expected of ileal CD cohort. The other SNPs were mostly on par with the general population, except for an enrichment in IRGM (rs33993564, 2.19x). Also, the pairwise correlation between the main SNPs NOD2 and ATG16L1, was very low ($cor < 5\%$), they were mostly decorrelated. We also derived a combined NOD2 phenotype by summing all 3 NOD2 variants. In this combined NOD2 score, 59% of patients had 0 NOD2 risk allele, 25% of patients had 1 NOD2 risk allele, and 15.9% of patients had 2 NOD2 risk allele. This number is compatible with previous reports, indeed between 30% and 50% of CD patients in the Western hemisphere carry Nod2 mutations of at least one allele (Yamamoto et al., 2009).

3.3.3 GWAS for Post-Operative Recurrence

We next performed a GWAS for our primary endpoint, Post-Operative Recurrence, in the REMIND cohort. The GWAS for post-operative recurrence was performed using standard methods. SNPs were imputed using Michigan Imputation Server (imputationserver.sph.umich.edu), using the 1000 genomes reference. Association testing between SNP and outcome was run using plink 1.9, with Minor Allele Frequency (MAF) < .05, Hardy-Weinberg equilibrium 10^{-6} . We adjusted for Sex, smoking, post-operative anti-TNF and PC1, PC2 as covariate.

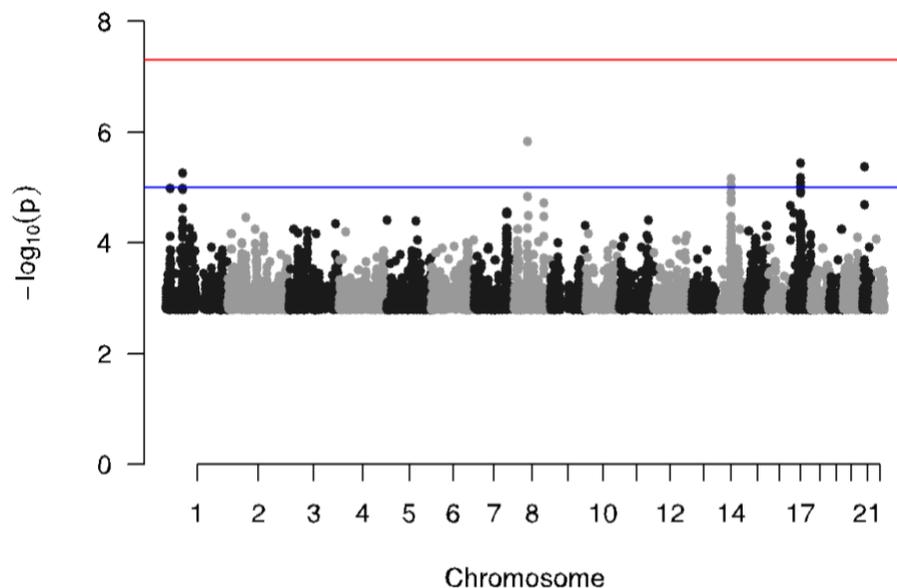


Figure 2. GWAS Post-Operative Recurrence

Adjusted for Sex, Smoking, Post-Op Anti-TNF, (+ PC1, PC2)

No SNP reached genome wide significance ($p < 5 \times 10^{-8}$). 10 loci reached the 10^{-5} p-value threshold, indicating a trend towards significance. Notably 5 SNPs on Chromosome 17 in the q12 subunit, in the vicinity of the genes MED1 and CDK12. Also 2 SNPs on Chromosome 14, close to the PRKCH gene.

Chromosome	Position	Allele	P-value
8	8:53073253	T	1.45E-06
17	17:39504593	C	3.67E-06
21	21:20536091	A	4.25E-06
1	1:67073314	C	5.53E-06
17	17:39432017	C	6.68E-06
14	14:61421132	G	6.93E-06
17	17:39450988	T	7.91E-06
14	14:61462565	C	8.83E-06
17	17:39419525	A	9.37E-06
17	17:39507970	G	9.71E-06

Table 1. SNPs associated with POR at p-value < 10⁻⁵.

MED1 is a mediator of transcription, not previously associated with IBD. CDK12 and PRKCH are two protein kinases involved in transcription and signaling, but neither of them have been previously associated with IBD.

As few studies have looked at the predictive value of genetics to post-operative recurrence and/or reoperation (Seiderer et al., 2006, Alvarez-Lobos et al., 2005) (Büning et al., 2004). Some found Nod2/Card15 loci to be associated with post-operative recurrence. It is likely that previous studies were less powered (lower sample size), or the statistical analysis was not done rigorously (selecting only SNPs of interest, not correcting for multiple comparisons, ...).

We had a relatively low sample size for a GWAS study. Generally, genome wide significant results can be obtained with a sample size of N > 1000 samples. With N = 300 we have a relatively low power to detect associations (http://csg.sph.umich.edu/abecasis/gas_power_calculator/). It will be interesting to run this GWAS again when we have reached double this number of samples with the next Remind cohort.

3.3.4 IBD susceptibility SNPs, smoking and post-operative recurrence

We next investigated the relationship between IBD susceptibility SNPs, smoking and the risk for post-operative recurrence. Smoking is a major environmental risk factor for both Crohn's disease and Post-Operative Recurrence. It is thus interesting to interrogate the relationships between this environmental factor and the genetic susceptibility to IBD. To characterize the individual genetic risk score for IBD, we calculated an IBD polygenic risk score (PRS) for each patient defined as follows:

$$IBD_risk_score = \sum beta_i * SNP_i$$

By capturing multiple SNPs in one metric, such score has been able to identify groups of individuals with substantially increased risks in multiple diseases (Crouch et al., 2020). In IBD, CD patients with higher risk scores demonstrated clinical characteristics typically associated with a more severe disease course including ileal location and earlier age at diagnosis (Chen et al., 2017).

There was no clear distinction between the risk score distribution in recurrence vs. non-recurrence, however a subset of patients seemed to have a higher risk score ($IBD_score > .75$) than the rest of the population. More males were in the high-risk group than females (70% vs. 42%, $p = 0.03$). All others characteristics were similarly distributed between high-risk and non-high-risk patients.

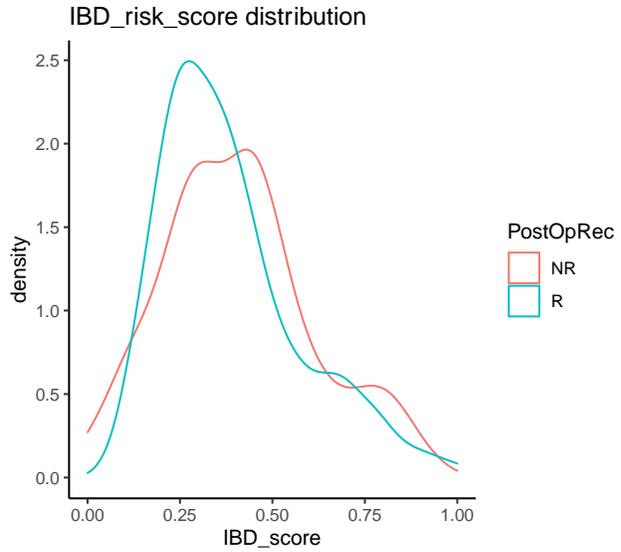


Figure 1. IBD risk score in recurrence and no-recurrence.

We looked at the distribution of the IBD risk score in both smokers and non-smokers, with or without post-operative recurrence.

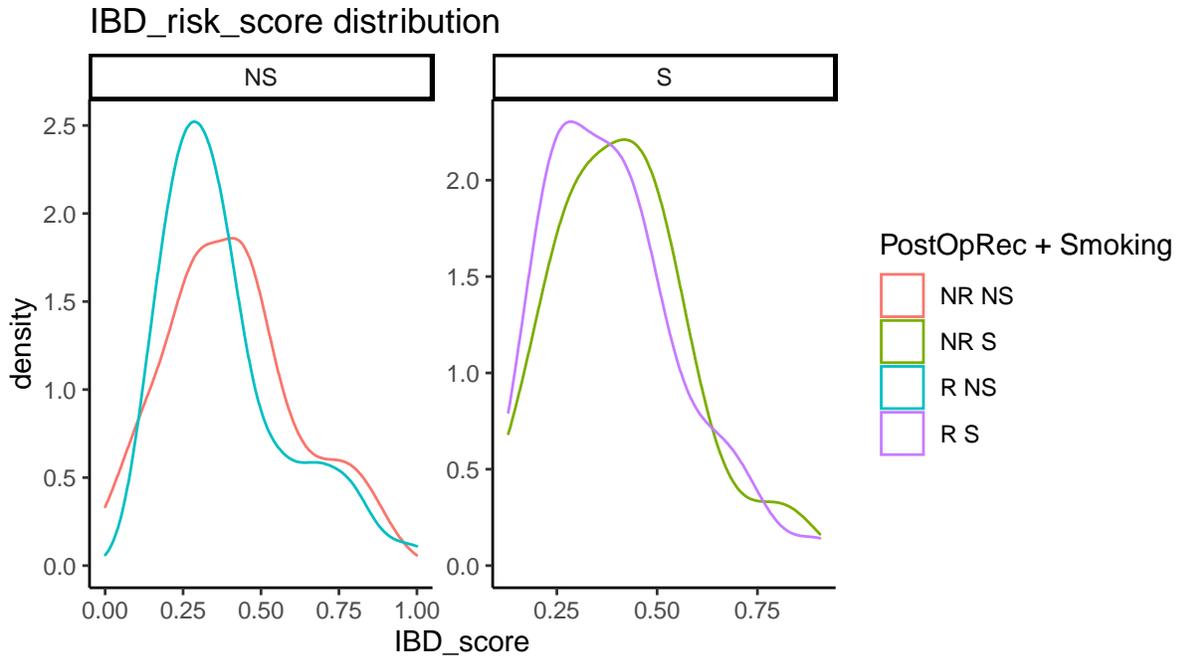


Figure 2. IBD risk score distribution. According to smoker status (smoker: S, non-smoker: NS) and post-operative recurrence (POR: R, no POR: NR). Probability density distribution of the IBD

risk score is shown in smokers, and non-smokers separately. The IBD risk has been calculated according to established formula listed above.

We then tested the association of individual IBD SNPs with both smoking behavior and post-operative recurrence. To maximize power, we selected SNPs that have both a strong association with IBD, and a frequency between 30% and 50% so that we can have equally sized positive and negative groups. We tested associations for SNPs in the genes ATG16L1, NOD2 (at least one allele), CARD9 and JAK2.

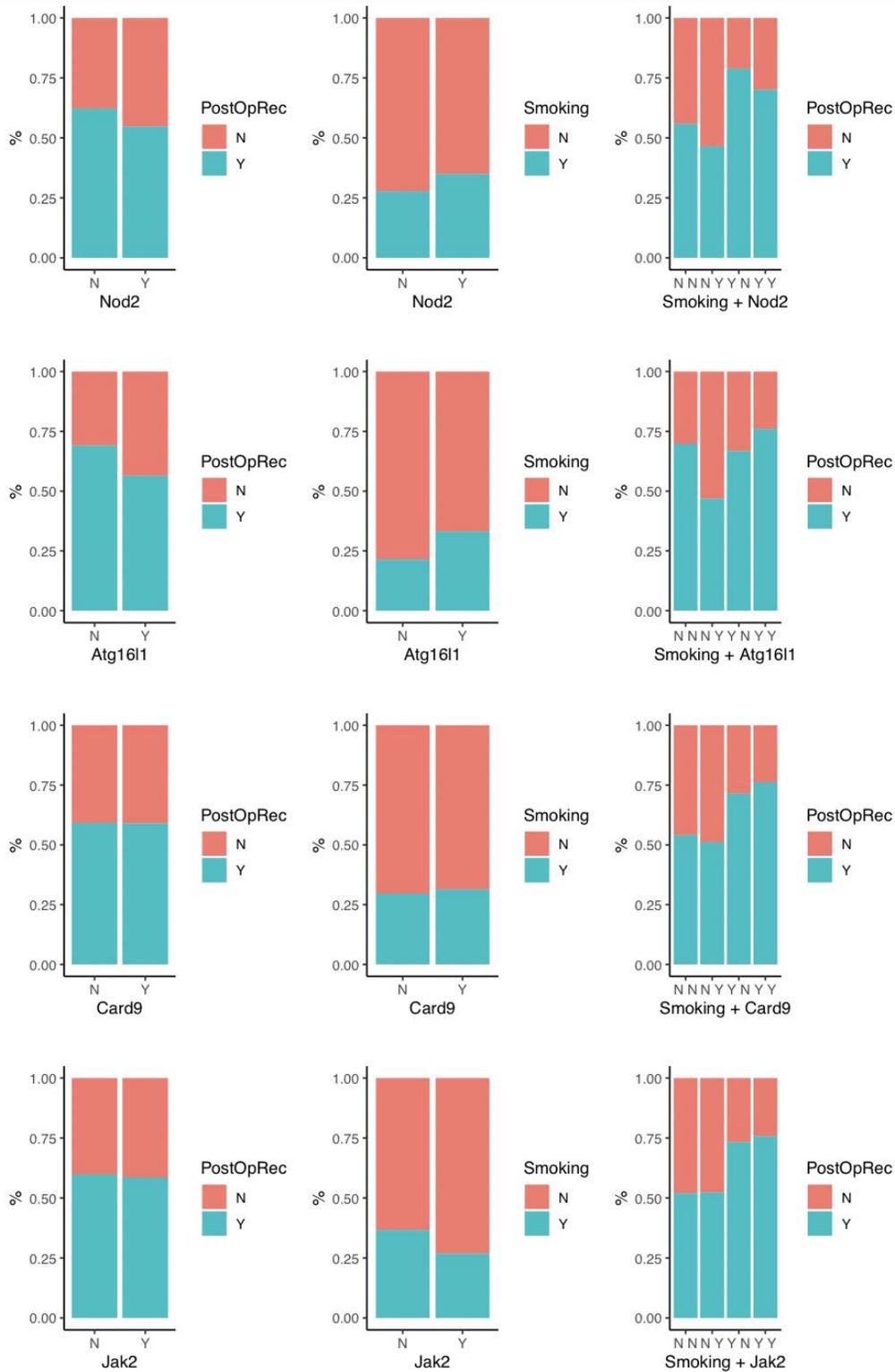


Figure 4. Association between smoking, POR and IBD SNPs

Association of smoking status, post-operative recurrence and IBD SNPs.

Out of the four genes we tested it seems that Atg16l1 was the one for which the interaction with smoking and post-operative recurrence was the strongest. Indeed, the effect of Atg16l1 on POR seemed to be stronger in non-smokers than smokers. Although the result was not significant.

Interestingly a study by (Liu and Stappenbeck, JCI 2018) found that “CD subjects who were of the ATG16L1T300A genotype and who were also smokers (T300A-smoking group) showed significantly shorter time to recurrence after surgery”. However, the outcome is not exactly similar since the authors looked at the time to recurrence instead of recurrence at 6 months. Overall, these results hint at possible interesting interactions between environmental factors, genetic factors and post-operative recurrence but need to be validated in an independent cohort.

3.3.5 Associations between IBD susceptibility SNPs and the transcriptome.

We next looked at the relationship between genetics and transcriptomics in the Remind cohorts. For each timepoint, inflamed ileum at time of surgery (M0), ileal margin at time of surgery (MI) and neo-terminal ileum (during follow-up endoscopy at 6 months), we computed the differentially expressed genes with 2 of the main IBD associated SNPs: ATG16L1_rs6752107, and NOD2_rs2066844.

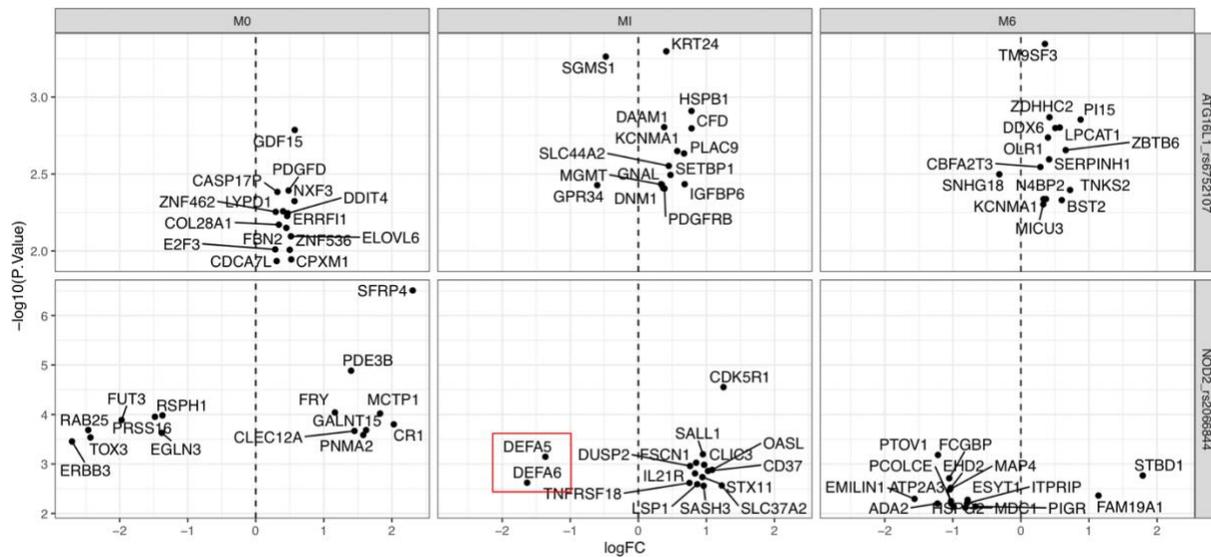


Figure 5. Association between genetics and transcriptomics

Differentially expressed genes (DEGs) with Atg16l1 or NOD2 genotype, at the three timepoints M0, MI, M6. Top 15 DEGs ranked by absolute logFC are shown and labeled. P-value < .01.

DEGs with SNPs were computed using a linear model on each gene. 1 or 2 risk alleles were treated as positive, 0 as negative. Bonferroni correction was applied for multiple comparison testing.

We saw that NOD2_rs2066844 allele was associated with a decrease in Paneth cell function, characterized by the 2 alpha-defensins genes DEFA5 and DEFA6, only for the ileal margin. Interestingly this association has already been reported: ‘NOD2 (CARD15) mutations in Crohn’s disease are associated with diminished mucosal α -defensin expression’(Wehkamp et al., 2004). Another study has found that Atg16l1 and smoking in combination alter the Paneth cell function in CD (Liu et al., 2018)

These results should be validated in a second cohort (Remind cohort 2). Or both cohorts should be pooled together to increase power.

3.4 Conclusion.

The work on post-operative recurrence will be validated in a second cohort (REMIND 2). Doubling the sample size will allow us to strengthen the power of our association with post-

operative recurrence if we chose to pool the samples or validate our findings in a validation cohort. Another option would be to identify an external validation cohort with a model of post-operative recurrence. A long-term goal for this study could be to build a point-of-care diagnostic tool to improve the prediction of risk for recurrence. Given the involvement of the JAK-STAT pathway in post-operative recurrence, we could also envision using anti-JAK drugs to prevent post-operative recurrence. Ustekinumab is currently in trial at our center for this indication (NCT?). Studying the transcriptome in response to biotherapy treatment after surgery, together with our longitudinal analysis of response to biotherapies in our ELYP cohort are promising areas of research.

Chapter 4.

Conclusion and Discussion

4.1 Summary of my work

In the first section of our work, we performed a meta-analysis of IBD affected gut samples and identified specific markers of disease in ileal and colonic diseases. This study analyzed gene expression in IBD using more than 1000 samples from 6 contrastive studies. By meta-analyzing each study's results, we highlighted specific gene expression patterns of Crohn's ileitis (FOLH1, CA2, etc.), Crohn's colitis, and Ulcerative colitis (REG gene). We showed that at the transcriptional level, signatures of diseases in the ileum and in the colon were distinct, regardless of disease classification (CD/UD). We also identified a common inflammatory gene signature (CXCL8, MMP3, DUOX2, S100A8) present in all three subtypes. We also highlighted some cross-talk between the colon and the ileum with disease. Some genes that were expressed in normal colon, but not normal ileum, started to be expressed in inflamed ileum. We lastly showed that this work can be extended to other inflammatory diseases such as Psoriasis.

The second part of our work was the multi-omics analysis of a surgical IBD cohort, the REMIND cohort. After studying the clinical factors associated with post-operative recurrence (POR), we showed that the engagement of the JAK-STAT pathway at the ileal margin was predictive of post-operative recurrence. At the genetic level, we saw trends associated with post-operative recurrence but these need to be confirmed with increased sample size and power. Looking specifically at the IBD susceptibility genes we saw that ATG16L1 was protective of POR, and had a link with smoking behavior. We also confirmed that NOD2 polymorphism was associated with a reduction of defensins. We also outlined some multi-omics comparisons between genetics, transcriptome and microbiome.

4.2 Future Perspectives

4.2.1 Multi-omics analysis of the REMIND study

In this work we presented genetics, transcriptomic and combined genetic and transcriptomic of the REMIND cohort. Given the ongoing interest in studies of the gut microbiome in IBD, and its relationship with the host transcriptome and immunome, it will be of primary interest to try to combine these analyses with 16s microbiome of the gut, as well as TCR sequencing studies previously performed in the REMIND cohort. Here we present preliminary results of our ongoing effort to perform multi-omics analysis of the REMIND cohort.

We first start by showing the completion of each omics in the cohort at each timepoint. We had almost a complete cohort for DNA genotyping. We also had a decent cohort completion in the other omics (transcriptome, microbiome) given that gathering all these samples in different omics with matching IDs requires a great organization.

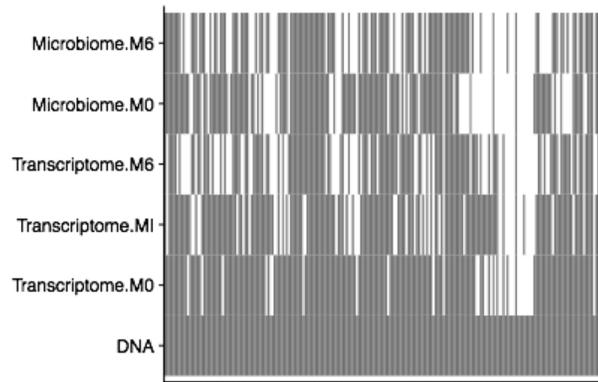


Figure 1. Cohort completion. All patients in X-axis. Grey bar if a sample is available for the specific omics for this patient.

Contribution of each omics to clinical characteristics

As a first step we tried to understand the contribution of each clinical factor to the variance of the different omics. This will allow us to see the impact of the clinical factors on the omics, and identify potential cofounders.

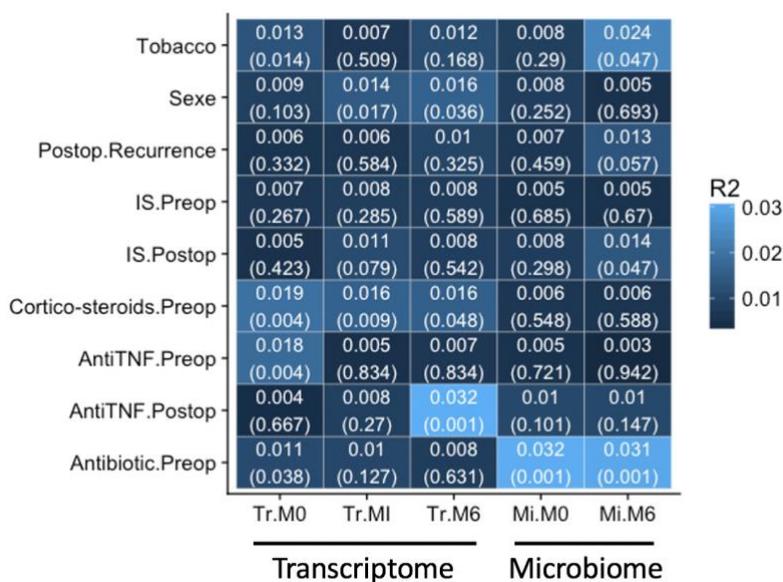


Figure 2. Contribution of clinical factor to omics. Analysis of variance (Permanova). Colored and labeled by R2, in parenthesis significance levels.

We conducted an analysis of variance using the Adonis R package, which implements a permutational multivariate analysis of variance (Permanova). We saw that antibiotics had the biggest impact on the microbiome at M0 and M6. The association at M0 would be expected given that antibiotics are known to alter the gut flora, the one 6 months later less so since one could expect a recovery of the microbiome after such period of time. Post-op Anti-TNF treatment had an impact on the transcriptome at M6, which can suggest an effect of the post-operative treatment on the transcriptome 6 months later, potentially bringing it down a less inflammatory state, and knowing that post-op anti-TNF also has protective effects for post-operative recurrence. No other

significant association was detected in our analysis, notably preoperative treatments (immune-suppressive, steroids, ...).

Correlations between genetics, transcriptome and microbiome

We next wanted to investigate the relationship between each omics. The transcriptome may have an impact the gut microbiome, and vice-versa. The genetic background can also alter both the transcriptome and the gut microbiome of the host. Here we show the correlation between transcriptome and microbiome, and genetics and transcriptome (eQTL) and microbiome (mQTL).

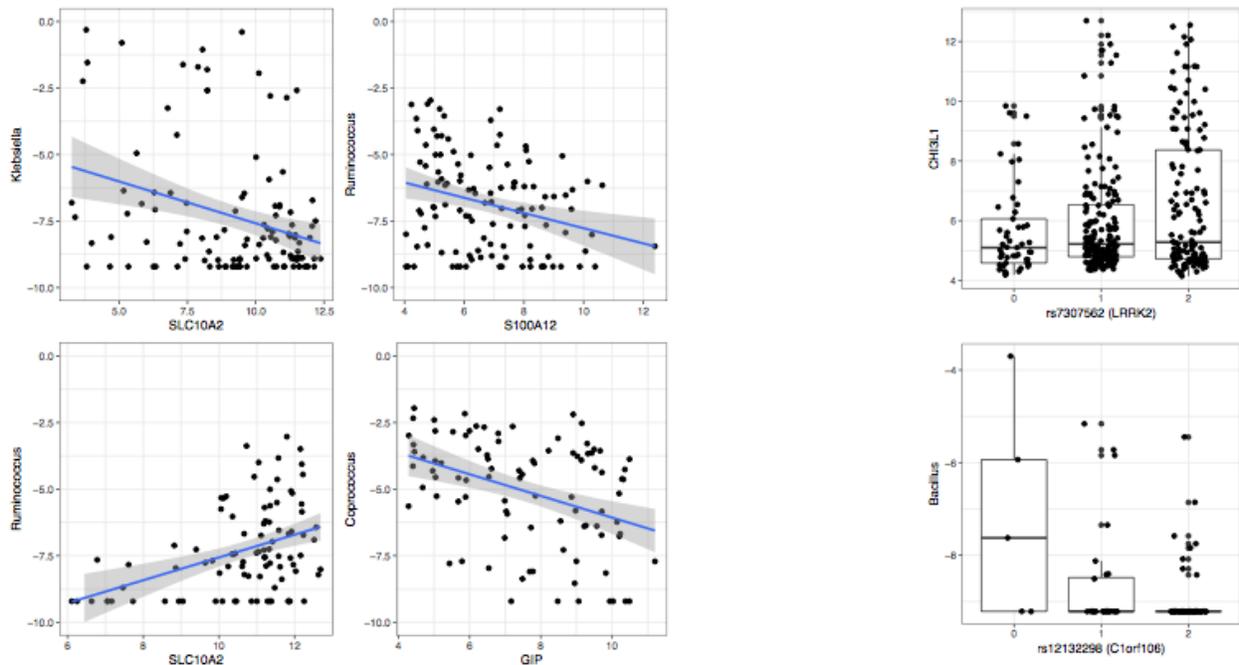


Figure 3. Correlation between genes and bacteria. Correlation between SNPs and genes/bacteria.

By performing pair-wise correlation using the Spearman coefficient, we identified an interesting correlation with a solute carrier SLC10A2, negatively correlated with a detrimental bacteria (Klebsiella), and positively correlated with a beneficial one (*Ruminococcus*). We also found a correlation between the IBD and Parkinson's susceptibility gene LRRK2 and the inflammatory marker CHI3L1. A SNP in C1orf106, a gene recently shown to regulate the stability of epithelial adherents junctions (Mohan et al., 2018), was negatively correlated with the presence of the

bacteria *Bacillus*. Lastly the inflammatory marker S100A12 was negatively correlated with the anti-inflammatory bacteria *Ruminococcus*. The gastric inhibitor GIP was inversely correlated with *Coprococcus*. These all show potential interplay between the genetic background, the host transcriptome and gut microbiome. A pro-inflammatory bacteria is shown to be inversely associated with an anti-inflammatory gene and vice-versa. The key question here would be to identify causality, which could only be answered by either functional experiment or longitudinal data where we could try to identify which change precedes the other (changes in microbiome precede changes in transcriptome or vice-versa).

Multi-omics patient subtypes

Lastly, we tried to identify clusters of patients using both transcriptome and microbiome at M0 and M6, and to relate them to clinical subtypes. Instead of identifying clusters in only the transcriptome or the microbiome we used a combination of both to derive subtypes of patients.

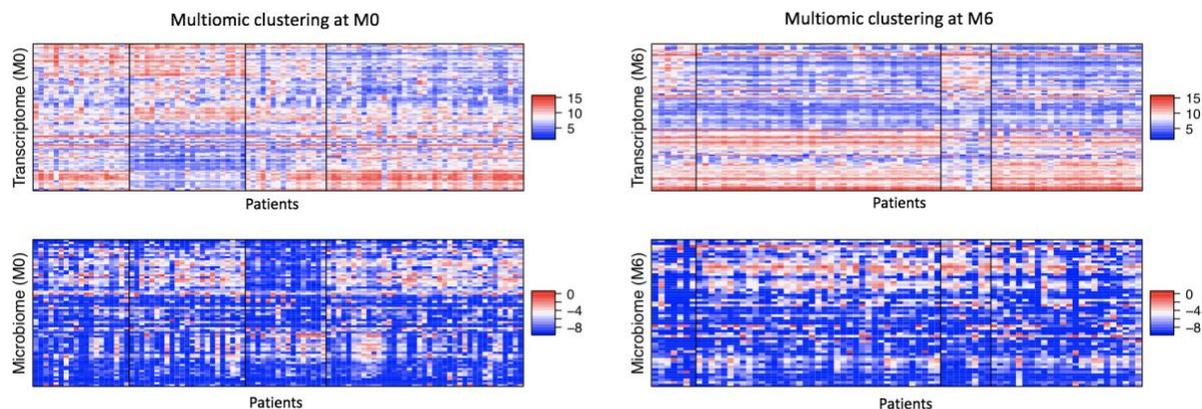


Figure 4. Multi-omics clustering at M0 and M6. 10% most variable genes/microbes are shown.

For that we used the R package iCluster, that fits a regularized latent variable model-based clustering that generates an integrated cluster assignment based on joint inference across data types. We identified 4 clusters at M0 and 4 clusters at M6. These clusters did not correlate with any clinical marker we had but one of them did correlate with the level of inflammation in the

samples, assessed by Calprotectin (S100A8). This result emphasizes that what we see at the transcriptomic level in the gut do not necessarily reflect clinical characterization, or that additional clinical information is required, something we also showed in the meta-analysis of IBD samples (chapter 2.). Nevertheless, with a larger cohort and more detailed phenotypes we cannot exclude that we would see clearer associations.

Overall, these preliminary results set the ground for a multi-omics framework of the REMIND cohort aimed at better understanding and predicting post-operative recurrence in CD. In the future we can also try new approaches for analyzing multi-omics data. For example, the MixOmics R package, which offers a wide range of multivariate models for multi-omic data (PCA, PLS-DA, CCA), as well as visualization tools. Another key question when doing predictive modellings with several omics, is whether to treat each omic matrix separately and then combine the results maybe with ensemble model, or combine the matrices column-wise, thus creating one single matrix later used for modelling.

4.2.4 Perspectives

This work highlights also the power of data-sharing in IBD. We could not have conducted our analyses without investigators agreeing to share their genomic data publicly. One major goal for multi-omics study is to establish a clear path to therapeutics. So far, most of the treatments on the market for IBD do not come from multi-omics studies in IBD. Most treatments come from general studies on inflammation, sometimes in other diseases: rheumatoid arthritis (anti-TNF, anti-JAK), psoriasis (Ustekinumab). If we summarize multi-omics findings in IBD: from genetics (NOD2, ATG16L1), to the transcriptome (DUOX2, APOA1) and microbiota (*F. prausnitzii*), so far none of them have yielded treatments. This can be for many reasons, one of them being the time it takes to develop drugs. Another one may be the reproducibility of the results; however, the majority have now been validated in large studies. Yet we might still see them come through in the near future. But we need to think about how to follow-up on these targets. This would probably require a partnership between academia and industry. Multi-omics methods will constantly improve in accuracy and resolution. Our ability to follow-up on these discoveries may not follow the same pace.

BIBLIOGRAPHY

- ABBAS, A. R., WOLSLEGEL, K., SESHASAYEE, D., MODRUSAN, Z. & CLARK, H. F. 2009. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PloS one*, 4, e6098.
- ADEGBOLA, S. O., SAHNAN, K., WARUSAVITARNE, J., HART, A. & TOZER, P. 2018. Anti-TNF Therapy in Crohn's Disease. *Int J Mol Sci*, 19.
- AJAYI, T. A., INNES, C. L., GRIMM, S. A., RAI, P., FINETHY, R., COERS, J., WANG, X., BELL, D. A., MCGRATH, J. A., SCHURMAN, S. H. & FESSLER, M. B. 2019. Crohn's disease IRGM risk alleles are associated with altered gene expression in human tissues. *Am. J. Physiol. Gastrointest. Liver Physiol.*, 316, G95-G105.
- AL AWAM, K., HAUSSLEITE, I. S., DUDLEY, E., DONEV, R., BRÜNE, M., JUCKEL, G. & THOME, J. 2015. Multiplatform metabolome and proteome profiling identifies serum metabolite and protein signatures as prospective biomarkers for schizophrenia. *Journal of neural transmission*, 122, 111-122.
- ALLEZ, M., AUZOLLE, C., NGOLLO, M., BOTTOIS, H., CHARDINY, V., CORRALIZA, A. M., SALAS, A., PEREZ, K., STEFANESCU, C., NANCEY, S., BUISSON, A., PARIENTE, B., FUMERY, M., SOKOL, H., TRETON, X., BARNICH, N., SEKSIK, P., LE BOURHIS, L. & GROUP, R. S. 2019. T cell clonal expansions in ileal Crohn's disease are associated with smoking behaviour and postoperative recurrence. *Gut*, 68, 1961-1970.
- ALTMAIER, E., FOBO, G., HEIER, M., THORAND, B., MEISINGER, C., RÖMISCH-MARGL, W., WALDENBERGER, M., GIEGER, C., ILLIG, T. & ADAMSKI, J. 2014. Metabolomics approach reveals effects of antihypertensives and lipid-lowering drugs on the human metabolism. *European journal of epidemiology*, 29, 325-336.
- ALVAREZ-LOBOS, M., AROSTEGUI, J. I., SANS, M., TASSIES, D., PLAZA, S., DELGADO, S., LACY, A. M., PIQUE, J. M., YAGÜE, J. & PANÉS, J. 2005. Crohn's disease patients carrying Nod2/CARD15 gene variants have an increased and early need for first surgery due to stricturing disease and higher rate of surgical recurrence. *Ann. Surg.*, 242, 693-700.
- ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biol*, 11, R106.
- AUZOLLE, C., NANCEY, S., TRAN-MINH, M. L., BUISSON, A., PARIENTE, B., STEFANESCU, C., FUMERY, M., MARTEAU, P., TRETON, X., HAMMOUDI, N., INVESTIGATORS, R. S. G., JOUVEN, X., SEKSIK, P. & ALLEZ, M. 2018. Male gender, active smoking and previous intestinal resection are risk factors for post-operative endoscopic recurrence in Crohn's disease: results from a prospective cohort study. *Aliment Pharmacol Ther*, 48, 924-932.
- BAUMGART, D. C. & CARDING, S. R. 2007. Inflammatory bowel disease: cause and immunobiology. *The Lancet*, 369, 1627-1640.
- BELKAID, Y. & HAND, T. W. 2014. Role of the microbiota in immunity and inflammation. *Cell*, 157, 121-41.

- BENNIKE, T., BIRKELUND, S., STENSALLE, A. & ANDERSEN, V. 2014. Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies. *World Journal of Gastroenterology: WJG*, 20, 3231.
- BIRD, A. 2007. Perceptions of epigenetics. *Nature*, 447, 396.
- BJERRUM, J. T., NIELSEN, O. H., HAO, F., TANG, H., NICHOLSON, J. K., WANG, Y. & OLSEN, J. 2010. Metabonomics in ulcerative colitis: diagnostics, biomarker identification, and insight into the pathophysiology. *Journal of proteome research*, 9, 954-962.
- BOCKLANDT, S., LIN, W., SEHL, M. E., SÁNCHEZ, F. J., SINSHEIMER, J. S., HORVATH, S. & VILAIN, E. 2011. Epigenetic predictor of age. *PLoS one*, 6, e14821.
- BOLYEN, E., RIDEOUT, J. R., DILLON, M. R., BOKULICH, N. A., ABNET, C. C., AL-GHALITH, G. A., ALEXANDER, H., ALM, E. J., ARUMUGAM, M., ASNICAR, F., BAI, Y., BISANZ, J. E., BITTINGER, K., BREJNROD, A., BRISLAWN, C. J., BROWN, C. T., CALLAHAN, B. J., CARABALLO-RODRIGUEZ, A. M., CHASE, J., COPE, E. K., DA SILVA, R., DIENER, C., DORRESTEIN, P. C., DOUGLAS, G. M., DURALL, D. M., DUVALLET, C., EDWARDSON, C. F., ERNST, M., ESTAKI, M., FOUQUIER, J., GAUGLITZ, J. M., GIBBONS, S. M., GIBSON, D. L., GONZALEZ, A., GORLICK, K., GUO, J., HILLMANN, B., HOLMES, S., HOLSTE, H., HUTTENHOWER, C., HUTTLEY, G. A., JANSSEN, S., JARMUSCH, A. K., JIANG, L., KAEHLER, B. D., KANG, K. B., KEEFE, C. R., KEIM, P., KELLEY, S. T., KNIGHTS, D., KOESTER, I., KOSCIOLEK, T., KREPS, J., LANGILLE, M. G. I., LEE, J., LEY, R., LIU, Y. X., LOFTFIELD, E., LOZUPONE, C., MAHER, M., MAROTZ, C., MARTIN, B. D., MCDONALD, D., MCIVER, L. J., MELNIK, A. V., METCALF, J. L., MORGAN, S. C., MORTON, J. T., NAIMEY, A. T., NAVAS-MOLINA, J. A., NOTHIAS, L. F., ORCHANIAN, S. B., PEARSON, T., PEOPLES, S. L., PETRAS, D., PREUSS, M. L., PRUESSE, E., RASMUSSEN, L. B., RIVERS, A., ROBESON, M. S., 2ND, ROSENTHAL, P., SEGATA, N., SHAFFER, M., SHIFFER, A., SINHA, R., SONG, S. J., SPEAR, J. R., SWAFFORD, A. D., THOMPSON, L. R., TORRES, P. J., TRINH, P., TRIPATHI, A., TURNBAUGH, P. J., UL-HASAN, S., VAN DER HOOFT, J. J. J., VARGAS, F., VAZQUEZ-BAEZA, Y., VOGTMANN, E., VON HIPPEL, M., WALTERS, W., et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*, 37, 852-857.
- BREITLING, L. P., YANG, R., KORN, B., BURWINKEL, B. & BRENNER, H. 2011. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *The American Journal of Human Genetics*, 88, 450-457.
- BÜNING, C., GENSCHER, J., BÜHNER, S., KRÜGER, S., KLING, K., DIGNASS, A., BAIER, P., BOCHOW, B., OCKENGA, J., SCHMIDT, H. H. J. & LOCHS, H. 2004. Mutations in the NOD2/CARD15 gene in Crohn's disease are associated with ileocecal resection and are a risk factor for reoperation. *Aliment. Pharmacol. Ther.*, 19, 1073-1078.
- CHAN, P. P., WASINGER, V. C. & LEONG, R. W. 2016. Current application of proteomics in biomarker discovery for inflammatory bowel disease. *World J. Gastrointest. Pathophysiol.*, 7, 27-37.
- CIVELEK, M. & LUSIS, A. J. 2014. Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*, 15, 34-48.
- COLLABORATORS, G. B. D. I. B. D. 2020. 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.
- CONSORTIUM, G. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348, 648-660.
- COSNES, J. 2004. Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. *Best practice & research Clinical gastroenterology*, 18, 481-496.
- DE LANGE, K. M., MOUTSIANAS, L., LEE, J. C., LAMB, C. A., LUO, Y., KENNEDY, N. A., JOSTINS, L., RICE, D. L., GUTIERREZ-ACHURY, J., JI, S. G., HEAP, G., NIMMO, E. R., EDWARDS, C., HENDERSON, P., MOWAT, C., SANDERSON, J., SATSANGI, J., SIMMONS, A., WILSON, D. C., TREMELLING, M., HART, A., MATHEW, C. G., NEWMAN, W. G., PARKES, M., LEES, C. W., UHLIG, H., HAWKEY, C., PRESCOTT, N. J., AHMAD, T., MANSFIELD, J. C., ANDERSON, C. A. & BARRETT, J. C. 2017. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*, 49, 256-261.
- DEMERATH, E. W., GUAN, W., GROVE, M. L., ASLIBEKYAN, S., MENDELSON, M., ZHOU, Y.-H., HEDMAN, Å. K., SANDLING, J. K., LI, L.-A. & IRVIN, M. R. 2015. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Human molecular genetics*, 24, 4464-4479.
- DESAI, H. & GUPTA, P. 2005. Increasing incidence of Crohn's disease in India: is it related to improved sanitation? *Indian journal of gastroenterology: official journal of the Indian Society of Gastroenterology*, 24, 23-24.
- DICK, K. J., NELSON, C. P., TSAPROUNI, L., SANDLING, J. K., AÏSSI, D., WAHL, S., MEDURI, E., MORANGE, P.-E., GAGNON, F. & GRALLERT, H. 2014. DNA methylation and body-mass index: a genome-wide analysis. *The Lancet*, 383, 1990-1998.
- DUBINSKY, V., RESHEF, L., BAR, N., KEIZER, D., GOLAN, N., RABINOWITZ, K., GODNY, L., YADGAR, K., ZONENSAIN, K., TULCHINSKY, H., GOPHNA, U. & DOTAN, I. 2020. Predominantly Antibiotic-resistant Intestinal Microbiome Persists in Patients With Pouchitis Who Respond to Antibiotic Therapy. *Gastroenterology*, 158, 610-624 e13.
- DULAI, P. S., SINGH, S., VANDE CASTEELE, N., BOLAND, B. S., RIVERA-NIEVES, J., ERNST, P. B., ECKMANN, L., BARRETT, K. E., CHANG, J. T. & SANDBORN, W. J. 2019. Should We Divide Crohn's Disease Into Ileum-Dominant and Isolated Colonic Diseases? *Clin Gastroenterol Hepatol*, 17, 2634-2643.
- FARIAS, A. S., PRADELLA, F., SCHMITT, A., SANTOS, L. M. & MARTINS-DE-SOUZA, D. 2014. Ten years of proteomics in multiple sclerosis. *Proteomics*, 14, 467-480.
- FIOCCHI, C. 2014. Integrating omics: the future of IBD? *Digestive Diseases*, 32, 96-102.
- GAYA, D. R., RUSSELL, R. K., NIMMO, E. R. & SATSANGI, J. 2006. New genes in inflammatory bowel disease: lessons for complex diseases? *The Lancet*, 367, 1271-1284.
- GEERLING, B., DAGNELIE, P., BADART-SMOOK, A., RUSSEL, M., STOCKBRÜGGER, R. & BRUMMER, R.-J. 2000. Diet as a risk factor for the development of ulcerative colitis. *The American journal of gastroenterology*, 95, 1008-1013.
- GEVERS, D., KUGATHASAN, S., DENSON, L. A., VÁZQUEZ-BAEZA, Y., VAN TREUREN, W., REN, B., SCHWAGER, E., KNIGHTS, D., SONG, S. J. & YASSOUR,

- M. 2014. The treatment-naive microbiome in new-onset Crohn's disease. *Cell host & microbe*, 15, 382-392.
- GIARDIELLO, F. M., LAZENBY, A. J., BAYLESS, T. M., LEVINE, E. J., BIAS, W. B., LADENSON, P. W., HUTCHEON, D. F., DEREVJANIK, N. L. & YARDLEY, J. H. 1989. Lymphocytic (microscopic) colitis. *Digestive diseases and sciences*, 34, 1730-1738.
- GROSSMAN, A., FIREMAN, Z., LILOS, P., NOVIS, B., ROZEN, P. & GILAT, T. 1989. Epidemiology of ulcerative colitis in the Jewish population of central Israel 1970-1980. *Hepato-gastroenterology*, 36, 193.
- HABERMAN, Y., KARNS, R., DEXHEIMER, P. J., SCHIRMER, M., SOMEKH, J., JURICKOVA, I., BRAUN, T., NOVAK, E., BAUMAN, L. & COLLINS, M. H. 2019. Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nature communications*, 10, 1-13.
- HABERMAN, Y., TICKLE, T. L., DEXHEIMER, P. J., KIM, M. O., TANG, D., KARNS, R., BALDASSANO, R. N., NOE, J. D., ROSH, J., MARKOWITZ, J., HEYMAN, M. B., GRIFFITHS, A. M., CRANDALL, W. V., MACK, D. R., BAKER, S. S., HUTTENHOWER, C., KELJO, D. J., HYAMS, J. S., KUGATHASAN, S., WALTERS, T. D., ARONOW, B., XAVIER, R. J., GEVERS, D. & DENSON, L. A. 2014. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest*, 124, 3617-33.
- HAMMOUDI, N., AUZOLLE, C., TRAN MINH, M. L., BOSCHETTI, G., BEZAULT, M., BUISSON, A., PARIENTE, B., TRETON, X., SEKSIK, P., FUMERY, M., LE BOURHIS, L., NANCEY, S. & ALLEZ, M. 2020. Postoperative Endoscopic Recurrence on the Neoterminal Ileum But Not on the Anastomosis Is Mainly Driving Long-Term Outcomes in Crohn's Disease. *Am J Gastroenterol*, 115, 1084-1093.
- HAMPE, J., HEYMANN, K., KRAWCZAK, M. & SCHREIBER, S. 2003. Association of inflammatory bowel disease with indicators for childhood antigen and infection exposure. *International journal of colorectal disease*, 18, 413-417.
- HAMPE, J., HEYMANN, K., KRUIS, W., RAEDLER, A., FÖLSCH, U. R. & SCHREIBER, S. 2000. Anticipation in inflammatory bowel disease: a phenomenon caused by an accumulation of confounders. *American journal of medical genetics*, 92, 178-183.
- HARIG, J. M., SOERGEL, K. H., KOMOROWSKI, R. A. & WOOD, C. M. 1989. Treatment of diversion colitis with short-chain-fatty acid irrigation. *New England Journal of Medicine*, 320, 23-28.
- HASIN, Y., SELDIN, M. & LUSIS, A. 2017. Multi-omics approaches to disease. *Genome Biol*, 18, 83.
- HEIJMANS, B. T. & MILL, J. 2012. Commentary: The seven plagues of epigenetic epidemiology. *International journal of epidemiology*, 41, 74-78.
- HERESBACH, D., GULWANI-AKOLKAR, B., LESSER, M., AKOLKAR, P. N., LIN, X.-Y., HERESBACH-LE BERRE, N., BRETAGNE, J. F., KATZ, S. & SILVER, J. 1998. Anticipation in Crohn's disease may be influenced by gender and ethnicity of the transmitting parent. *The American journal of gastroenterology*, 93, 2368-2372.
- HUANG, H., FANG, M., JOSTINS, L., UMICEVIC MIRKOV, M., BOUCHER, G., ANDERSON, C. A., ANDERSEN, V., CLEYNEN, I., CORTES, A., CRINS, F., D'AMATO, M., DEFFONTAINE, V., DMITRIEVA, J., DOCAMPO, E., ELANSARY,

- M., FARH, K. K., FRANKE, A., GORI, A. S., GOYETTE, P., HALFVARSON, J., HARITUNIANS, T., KNIGHT, J., LAWRENCE, I. C., LEES, C. W., LOUIS, E., MARIMAN, R., MEUWISSEN, T., MNI, M., MOMOZAWA, Y., PARKES, M., SPAIN, S. L., THEATRE, E., TRYNKA, G., SATSANGI, J., VAN SOMMEREN, S., VERMEIRE, S., XAVIER, R. J., INTERNATIONAL INFLAMMATORY BOWEL DISEASE GENETICS, C., WEERSMA, R. K., DUERR, R. H., MATHEW, C. G., RIOUX, J. D., MCGOVERN, D. P. B., CHO, J. H., GEORGES, M., DALY, M. J. & BARRETT, J. C. 2017. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*, 547, 173-178.
- HUANG, H., VANGAY, P., MCKINLAY, C. E. & KNIGHTS, D. 2014. Multi-omics analysis of inflammatory bowel disease. *Immunology letters*, 162, 62-68.
- HUGOT, J.-P., CHAMAILLARD, M., ZOUALI, H., LESAGE, S., CÉZARD, J.-P., BELAICHE, J., ALMER, S., TYSK, C., O'MORAIN, C. A. & GASSULL, M. 2001. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, 411, 599-603.
- HUTTENHOWER, C., GEVERS, D., KNIGHT, R., ABUBUCKER, S., BADGER, J. H., CHINWALLA, A. T., CREASY, H. H., EARL, A. M., FITZGERALD, M. G. & FULTON, R. S. 2012. Structure, function and diversity of the healthy human microbiome. *nature*, 486, 207.
- INTERNATIONAL HAPMAP, C., ALTSHULER, D. M., GIBBS, R. A., PELTONEN, L., ALTSHULER, D. M., GIBBS, R. A., PELTONEN, L., DERMITZAKIS, E., SCHAFFNER, S. F., YU, F., PELTONEN, L., DERMITZAKIS, E., BONNEN, P. E., ALTSHULER, D. M., GIBBS, R. A., DE BAKKER, P. I., DELOUKAS, P., GABRIEL, S. B., GWILLIAM, R., HUNT, S., INOUE, M., JIA, X., PALOTIE, A., PARKIN, M., WHITTAKER, P., YU, F., CHANG, K., HAWES, A., LEWIS, L. R., REN, Y., WHEELER, D., GIBBS, R. A., MUZNY, D. M., BARNES, C., DARVISHI, K., HURLES, M., KORN, J. M., KRISTIANSOON, K., LEE, C., MCCARROL, S. A., NEMESH, J., DERMITZAKIS, E., KEINAN, A., MONTGOMERY, S. B., POLLACK, S., PRICE, A. L., SORANZO, N., BONNEN, P. E., GIBBS, R. A., GONZAGA-JAUREGUI, C., KEINAN, A., PRICE, A. L., YU, F., ANTTILA, V., BRODEUR, W., DALY, M. J., LESLIE, S., MCVEAN, G., MOUTSIANAS, L., NGUYEN, H., SCHAFFNER, S. F., ZHANG, Q., GHORI, M. J., MCGINNIS, R., MCLAREN, W., POLLACK, S., PRICE, A. L., SCHAFFNER, S. F., TAKEUCHI, F., GROSSMAN, S. R., SHLYAKHTER, I., HOSTETTER, E. B., SABETI, P. C., ADEBAMOWO, C. A., FOSTER, M. W., GORDON, D. R., LICINIO, J., MANCA, M. C., MARSHALL, P. A., MATSUDA, I., NGARE, D., WANG, V. O., REDDY, D., ROTIMI, C. N., ROYAL, C. D., SHARP, R. R., ZENG, C., BROOKS, L. D. & MCEWEN, J. E. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467, 52-8.
- JASSAL, B., MATTHEWS, L., VITERI, G., GONG, C., LORENTE, P., FABREGAT, A., SIDIROPOULOS, K., COOK, J., GILLESPIE, M., HAW, R., LONEY, F., MAY, B., MILACIC, M., ROTHFELS, K., SEVILLA, C., SHAMOVSKY, V., SHORSER, S., VARUSAI, T., WEISER, J., WU, G., STEIN, L., HERMJAKOB, H. & D'EUSTACHIO, P. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res*, 48, D498-D503.
- JIN, L., LI, L., HU, C., PAEZ-CORTEZ, J., BI, Y., MACORITTO, M., CAO, S. & TIAN, Y. 2019. Integrative Analysis of Transcriptomic and Proteomic Profiling in Inflammatory Bowel Disease Colon Biopsies. *Inflamm. Bowel Dis.*, 25, 1906-1918.

- JONES, D. P., PARK, Y. & ZIEGLER, T. R. 2012. Nutritional metabolomics: progress in addressing complexity in diet and health. *Annual review of nutrition*, 32, 183-202.
- JONES, J. H., LENNARD-JONES, J., MORSON, B., CHAPMAN, M., SACKIN, M., SNEATH, P., SPICER, C. & CARD, W. 1973. Numerical taxonomy and discriminant analysis applied to non-specific colitis. *QJM: An International Journal of Medicine*, 42, 715-732.
- JOSTINS, L., RIPKE, S., WEERSMA, R. K., DUERR, R. H., MCGOVERN, D. P., HUI, K. Y., LEE, J. C., SCHUMM, L. P., SHARMA, Y. & ANDERSON, C. A. 2012. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, 491, 119-124.
- KANE, S. & LEMIEUX, N. 2005. The role of breastfeeding in postpartum disease activity in women with inflammatory bowel disease. *American Journal of Gastroenterology*, 100, 102-105.
- KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28, 27-30.
- KHAN, I., ULLAH, N., ZHA, L., BAI, Y., KHAN, A., ZHAO, T., CHE, T. & ZHANG, C. 2019. Alteration of Gut Microbiota in Inflammatory Bowel Disease (IBD): Cause or Consequence? IBD Treatment Targeting the Gut Microbiome. *Pathogens*, 8.
- KIM, M.-S., PINTO, S. M., GETNET, D., NIRUJOGI, R. S., MANDA, S. S., CHAERKADY, R., MADUGUNDU, A. K., KELKAR, D. S., ISSERLIN, R. & JAIN, S. 2014. A draft map of the human proteome. *Nature*, 509, 575-581.
- KINRA, S., SMITH, G. D., JEFFREYS, M., GUNNELL, D., GALOBARDES, B. & MCCARRON, P. 2006. Association between sibship size and allergic diseases in the Glasgow Alumni Study. *Thorax*, 61, 48-53.
- KLEIN, R. J., ZEISS, C., CHEW, E. Y., TSAI, J.-Y., SACKLER, R. S., HAYNES, C., HENNING, A. K., SANGIOVANNI, J. P., MANE, S. M. & MAYNE, S. T. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science*, 308, 385-389.
- KNIGHTS, D., LASSEN, K. G. & XAVIER, R. J. 2013. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. *Gut*, 62, 1505-1510.
- KRONMAN, M. P., ZAOUTIS, T. E., HAYNES, K., FENG, R. & COFFIN, S. E. 2012. Antibiotic exposure and IBD development among children: a population-based cohort study. *Pediatrics*, 130, e794-803.
- KUGATHASAN, S., DENSON, L. A., WALTERS, T. D., KIM, M.-O., MARIGORTA, U. M., SCHIRMER, M., MONDAL, K., LIU, C., GRIFFITHS, A. & NOE, J. D. 2017. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *The Lancet*, 389, 1710-1718.
- KURATA, J. H., KANTOR-FISH, S., FRANKL, H., GODBY, P. & VADHEIM, C. M. 1992. Crohn's disease among ethnic groups in a large health maintenance organization. *Gastroenterology*, 102, 1940-1948.
- LEE, J. C., BIASCI, D., ROBERTS, R., GEARRY, R. B., MANSFIELD, J. C., AHMAD, T., PRESCOTT, N. J., SATSANGI, J., WILSON, D. C., JOSTINS, L., ANDERSON, C. A., CONSORTIUM, U. I. G., TRATHERNE, J. A., LYONS, P. A., PARKES, M. & SMITH, K. G. 2017. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet*, 49, 262-268.

- LENNARD-JONES, J. 1989. Classification of inflammatory bowel disease. *Scandinavian Journal of Gastroenterology*, 24, 2-6.
- LEONG, R. W., LAU, J. Y. & SUNG, J. J. 2004. The epidemiology and phenotype of Crohn's disease in the Chinese population. *Inflammatory bowel diseases*, 10, 646-651.
- LEWIS, J. D., ABERRA, F. N., LICHTENSTEIN, G. R., BILKER, W. B., BRENSINGER, C. & STROM, B. L. 2004. Seasonal variation in flares of inflammatory bowel disease. *Gastroenterology*, 126, 665-673.
- LIANG, L., WILLIS-OWEN, S. A., LAPRISE, C., WONG, K. C., DAVIES, G. A., HUDSON, T. J., BINIA, A., HOPKIN, J. M., YANG, I. V. & GRUNDBERG, E. 2015. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature*, 520, 670-674.
- LIU, J. Z., VAN SOMMEREN, S., HUANG, H., NG, S. C., ALBERTS, R., TAKAHASHI, A., RIPKE, S., LEE, J. C., JOSTINS, L., SHAH, T., ABEDIAN, S., CHEON, J. H., CHO, J., DAYANI, N. E., FRANKE, L., FUYUNO, Y., HART, A., JUYAL, R. C., JUYAL, G., KIM, W. H., MORRIS, A. P., POUSTCHI, H., NEWMAN, W. G., MIDHA, V., ORCHARD, T. R., VAHEDI, H., SOOD, A., SUNG, J. Y., MALEKZADEH, R., WESTRA, H. J., YAMAZAKI, K., YANG, S. K., INTERNATIONAL MULTIPLE SCLEROSIS GENETICS, C., INTERNATIONAL, I. B. D. G. C., BARRETT, J. C., ALIZADEH, B. Z., PARKES, M., BK, T., DALY, M. J., KUBO, M., ANDERSON, C. A. & WEERSMA, R. K. 2015. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*, 47, 979-986.
- LIU, T.-C., KERN, J. T., VANDUSSEN, K. L., XIONG, S., KAIKO, G. E., WILEN, C. B., RAJALA, M. W., CARUSO, R., HOLTZMAN, M. J., GAO, F., MCGOVERN, D. P., NUNEZ, G., HEAD, R. D. & STAPPENBECK, T. S. 2018. Interaction between smoking and ATG16L1T300A triggers Paneth cell defects in Crohn's disease. *J. Clin. Invest.*, 128, 5110-5122.
- LIUBAKKA, A. & VAUGHN, B. P. 2016. Clostridium difficile Infection and Fecal Microbiota Transplant. *AACN Adv Crit Care*, 27, 324-337.
- LLOYD-PRICE, J., ARZE, C., ANANTHAKRISHNAN, A. N., SCHIRMER, M., AVILA-PACHECO, J., POON, T. W., ANDREWS, E., AJAMI, N. J., BONHAM, K. S., BRISLAWN, C. J., CASERO, D., COURTNEY, H., GONZALEZ, A., GRAEBER, T. G., HALL, A. B., LAKE, K., LANDERS, C. J., MALLICK, H., PLICHTA, D. R., PRASAD, M., RAHNAVARD, G., SAUK, J., SHUNGIN, D., VAZQUEZ-BAEZA, Y., WHITE, R. A., 3RD, INVESTIGATORS, I., BRAUN, J., DENSON, L. A., JANSSON, J. K., KNIGHT, R., KUGATHASAN, S., MCGOVERN, D. P. B., PETROSINO, J. F., STAPPENBECK, T. S., WINTER, H. S., CLISH, C. B., FRANZOSA, E. A., VLAMAKIS, H., XAVIER, R. J. & HUTTENHOWER, C. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, 569, 655-662.
- LODDO, I. & ROMANO, C. 2015. Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Front. Immunol.*, 6, 551.
- LOFTUS JR, E. V. 2004. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology*, 126, 1504-1517.
- LUO, Y., DE LANGE, K. M., JOSTINS, L., MOUTSIANAS, L., RANDALL, J., KENNEDY, N. A., LAMB, C. A., MCCARTHY, S., AHMAD, T., EDWARDS, C., SERRA, E. G., HART, A., HAWKEY, C., MANSFIELD, J. C., MOWAT, C., NEWMAN, W. G.,

- NICHOLS, S., POLLARD, M., SATSANGI, J., SIMMONS, A., TREMELLING, M., UHLIG, H., WILSON, D. C., LEE, J. C., PRESCOTT, N. J., LEES, C. W., MATHEW, C. G., PARKES, M., BARRETT, J. C. & ANDERSON, C. A. 2017. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat Genet*, 49, 186-192.
- MADARA, J. L., PODOLSKY, D. K., KING, N. W., SEHGAL, P. K., MOORE, R. & WINTER, H. S. 1985. Characterization of spontaneous colitis in cotton-top tamarins (*Saguinus oedipus*) and its response to sulfasalazine. *Gastroenterology*, 88, 13-19.
- MANN, M. & JENSEN, O. N. 2003. Proteomic analysis of post-translational modifications. *Nature biotechnology*, 21, 255-261.
- MARIGORTA, U. M., DENSON, L. A., HYAMS, J. S., MONDAL, K., PRINCE, J., WALTERS, T. D., GRIFFITHS, A., NOE, J. D., CRANDALL, W. V., ROSH, J. R., MACK, D. R., KELLERMAYER, R., HEYMAN, M. B., BAKER, S. S., STEPHENS, M. C., BALDASSANO, R. N., MARKOWITZ, J. F., KIM, M. O., DUBINSKY, M. C., CHO, J., ARONOW, B. J., KUGATHASAN, S. & GIBSON, G. 2017. Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet*, 49, 1517-1521.
- MARTIN, J. C., CHANG, C., BOSCHETTI, G., UNGARO, R., GIRI, M., GROUT, J. A., GETTLER, K., CHUANG, L. S., NAYAR, S., GREENSTEIN, A. J., DUBINSKY, M., WALKER, L., LEADER, A., FINE, J. S., WHITEHURST, C. E., MBOW, M. L., KUGATHASAN, S., DENSON, L. A., HYAMS, J. S., FRIEDMAN, J. R., DESAI, P. T., KO, H. M., LAFACE, I., AKTURK, G., SCHADT, E. E., SALMON, H., GNJATIC, S., RAHMAN, A. H., MERAD, M., CHO, J. H. & KENIGSBURG, E. 2019. 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.
- MARX, V. 2015. The DNA of a nation. Nature Publishing Group.
- MATEOS, B., PALANCA-BALLESTER, C., SAEZ-GONZALEZ, E., MORET, I., LOPEZ, A. & SANDOVAL, J. 2019. Epigenetics of Inflammatory Bowel Disease: Unraveling Pathogenic Events. 1.
- MAWDSLEY, J. & RAMPTON, D. 2005. Psychological stress in IBD: new insights into pathogenic and therapeutic implications. *Gut*, 54, 1481-1491.
- METZKER, M. L. 2010. Sequencing technologies—the next generation. *Nature reviews genetics*, 11, 31-46.
- MI, H., MURUGANUJAN, A., EBERT, D., HUANG, X. & THOMAS, P. D. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*, 47, D419-D426.
- MOHANAN, V., NAKATA, T., DESCH, A. N., LEVESQUE, C., BOROUGHS, A., GUZMAN, G., CAO, Z., CREASEY, E., YAO, J., BOUCHER, G., CHARRON, G., BHAN, A. K., SCHENONE, M., CARR, S. A., REINECKER, H. C., DALY, M. J., RIOUX, J. D., LASSEN, K. G. & XAVIER, R. J. 2018. C1orf106 is a colitis risk gene that regulates stability of epithelial adherens junctions. *Science*, 359, 1161-1166.
- MOMOZAWA, Y., DMITRIEVA, J., THEATRE, E., DEFFONTAINE, V., RAHMOUNI, S., CHARLOTEAUX, B., CRINS, F., DOCAMPO, E., ELANSARY, M., GORI, A. S., LECUT, C., MARIMAN, R., MNI, M., OURY, C., ALTUKHOV, I., ALEXEEV, D., AULCHENKO, Y., AMININEJAD, L., BOUMA, G., HOENTJEN, F., LOWENBERG, M., OLDENBURG, B., PIERIK, M. J., VANDER MEULEN-DE JONG, A. E.,

- JANNEKE VAN DER WOUDE, C., VISSCHEDIJK, M. C., INTERNATIONAL, I. B. D. G. C., LATHROP, M., HUGOT, J. P., WEERSMA, R. K., DE VOS, M., FRANCHIMONT, D., VERMEIRE, S., KUBO, M., LOUIS, E. & GEORGES, M. 2018. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat Commun*, 9, 2427.
- MORRIS, T., STABLES, M., COLVILLE-NASH, P., NEWSON, J., BELLINGAN, G., DE SOUZA, P. M. & GILROY, D. W. 2010. Dichotomy in duration and severity of acute inflammatory responses in humans arising from differentially expressed proresolution pathways. *Proceedings of the National Academy of Sciences*, 107, 8842-8847.
- MOUM, B., AADLAND, E., EKBOM, A. & VATN, M. 1996. Seasonal variations in the onset of ulcerative colitis. *Gut*, 38, 376-378.
- MOWAT, A. M. & AGACE, W. W. 2014. Regional specialization within the intestinal immune system. *Nat Rev Immunol*, 14, 667-85.
- NASER, S. A., ARCE, M., KHAJA, A., FERNANDEZ, M., NASER, N., ELWASILA, S. & THANIGACHALAM, S. 2012. Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World J. Gastroenterol.*, 18, 412-424.
- NEWMAN, A. M., STEEN, C. B., LIU, C. L., GENTLES, A. J., CHAUDHURI, A. A., SCHERER, F., KHODADOUST, M. S., ESFAHANI, M. S., LUCA, B. A., STEINER, D., DIEHN, M. & ALIZADEH, A. A. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*, 37, 773-782.
- NG, S. C., SHI, H. Y., HAMIDI, N., UNDERWOOD, F. E., TANG, W., BENCHIMOL, E. I., PANACCIONE, R., GHOSH, S., WU, J. C. Y., CHAN, F. K. L., SUNG, J. J. Y. & KAPLAN, G. G. 2017. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*, 390, 2769-2778.
- NGUYEN, G. C., TORRES, E. A., REGUEIRO, M., BROMFIELD, G., BITTON, A., STEMPAK, J., DASSOPOULOS, T., SCHUMM, P., GREGORY, F. J. & GRIFFITHS, A. M. 2006. Inflammatory bowel disease characteristics among African Americans, Hispanics, and non-Hispanic Whites: characterization of a large North American cohort. *American Journal of Gastroenterology*, 101, 1012-1023.
- NIELSEN, R., PAUL, J. S., ALBRECHTSEN, A. & SONG, Y. S. 2011. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12, 443-451.
- NILSSON, T., MANN, M., AEBERSOLD, R., YATES, J. R., BAIROCH, A. & BERGERON, J. J. 2010. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nature methods*, 7, 681-685.
- NIV, Y., ABUKSIS, G. & FRASER, G. M. 2000. Epidemiology of ulcerative colitis in Israel: a survey of Israeli kibbutz settlements. *The American journal of gastroenterology*, 95, 693-698.
- OKAYASU, I., HATAKEYAMA, S., YAMADA, M., OHKUSA, T., INAGAKI, Y. & NAKAYA, R. 1990. A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology*, 98, 694-702.
- ONG, S.-E. & MANN, M. 2005. Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology*, 1, 252-262.
- ORG, E., PARKS, B. W., JOO, J. W. J., EMERT, B., SCHWARTZMAN, W., KANG, E. Y., MEHRABIAN, M., PAN, C., KNIGHT, R. & GUNSALUS, R. 2015. Genetic and

- environmental control of host-gut microbiota interactions. *Genome research*, 25, 1558-1569.
- ORHOLM, M., BINDER, V., SØRENSEN, T., RASMUSSEN, L. & KYVIK, K. 2000. Concordance of inflammatory bowel disease among Danish twins: results of a nationwide study. *Scandinavian journal of gastroenterology*, 35, 1075-1081.
- OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics*, 12, 87-98.
- PANDEY, A. & MANN, M. 2000. Proteomics to study genes and genomes. *Nature*, 405, 837-846.
- PETRICCOIN III, E. F., ARDEKANI, A. M., HITT, B. A., LEVINE, P. J., FUSARO, V. A., STEINBERG, S. M., MILLS, G. B., SIMONE, C., FISHMAN, D. A. & KOHN, E. C. 2002. Use of proteomic patterns in serum to identify ovarian cancer. *The lancet*, 359, 572-577.
- PIERRE, N., BAIWIR, D., HUYNH-THU, V. A., MAZZUCHELLI, G., SMARGIASSO, N., DE PAUW, E., BOUHNİK, Y., LAHARIE, D., COLOMBEL, J. F., MEUWIS, M. A., LOUIS, E. & GETAID 2020a. Discovery of biomarker candidates associated with the risk of short-term and mid/long-term relapse after infliximab withdrawal in Crohn's patients: a proteomics-based study. *Gut*.
- PIERRE, N., SALEE, C., MASSOT, C., BLETARD, N., MAZZUCHELLI, G., SMARGIASSO, N., MORSA, D., BAIWIR, D., DE PAUW, E., REENAERS, C., VAN KEMSEKE, C., LOLY, J. P., DELVENNE, P., MEUWIS, M. A. & LOUIS, E. 2020b. Proteomics Highlights Common and Distinct Pathophysiological Processes Associated with Ileal and Colonic Ulcers in Crohn's Disease. *J Crohns Colitis*, 14, 205-215.
- PODOSKY, D. 1991. Inflammatory bowel disease (first of two parts). *N Engl J Med*, 325, 928-37.
- PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I., DALY, M. J. & SHAM, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81, 559-75.
- RAY, G. & LONGWORTH, M. S. 2019. Epigenetics, DNA Organization, and Inflammatory Bowel Disease. *Inflamm. Bowel Dis.*, 25, 235-247.
- RELTON, C. L. & DAVEY SMITH, G. 2012. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology*, 41, 161-176.
- RIORDAN, A., RUXTON, C. & HUNTER, J. 1998. A review of associations between Crohn's disease and consumption of sugars. *European journal of clinical nutrition*, 52, 229-238.
- RUSSELL, R. & SATSANGI, J. 2004. IBD: a family affair. *Best practice & research Clinical gastroenterology*, 18, 525-539.
- SALEM, M., AMMITZBOELL, M., NYS, K., SEIDELIN, J. B. & NIELSEN, O. H. 2015. ATG16L1: A multifunctional susceptibility factor in Crohn disease. *Autophagy*, 11, 585-594.
- SATSANGI, J., GROOTSCHOLTEN, C., HOLT, H. & JEWELL, D. 1996. Clinical patterns of familial inflammatory bowel disease. *Gut*, 38, 738-741.
- SCHACK-NIELSEN, L. & MICHAELSEN, K. F. 2006. Breast feeding and future health. *Current Opinion in Clinical Nutrition & Metabolic Care*, 9, 289-296.

- SCHAEFER, J. S. 2016. MicroRNAs: how many in inflammatory bowel disease? *Curr. Opin. Gastroenterol.*, 32, 258-266.
- SCHNIERS, A., GOLL, R., PASING, Y., SØRBYE, S. W., FLORHOLMEN, J. & HANSEN, T. 2019. Ulcerative colitis: functional analysis of the in-depth proteome. *Clin. Proteomics*, 16, 4.
- SEIDERER, J., SCHNITZLER, F., BRAND, S., STAUDINGER, T., PFENNIG, S., HERRMANN, K., HOFBAUER, K., DAMBACHER, J., TILLACK, C., SACKMANN, M., GÖKE, B., LOHSE, P. & OCHSENKÜHN, T. 2006. Homozygosity for the CARD15 frameshift mutation 1007fs is predictive of early onset of Crohn's disease with ileal stenosis, entero-ental fistulas, and frequent need for surgical intervention with high risk of re-stenosis. *Scand. J. Gastroenterol.*, 41, 1421-1432.
- SMILLIE, C. S., BITON, M., ORDOVAS-MONTANES, J., SULLIVAN, K. M., BURGIN, G., GRAHAM, D. B., HERBST, R. H., ROGEL, N., SLYPER, M. & WALDMAN, J. 2019. Intra-and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell*, 178, 714-730. e22.
- SOKOL, H., BROT, L., STEFANESCU, C., AUZOLLE, C., BARNICH, N., BUISSON, A., FUMERY, M., PARIENTE, B., LE BOURHIS, L., TRETON, X., NANCEY, S., ALLEZ, M., SEKSIK, P. & INVESTIGATORS, R. S. G. 2020a. Prominence of ileal mucosa-associated microbiota to predict postoperative endoscopic recurrence in Crohn's disease. *Gut*, 69, 462-472.
- SOKOL, H., LANDMAN, C., SEKSIK, P., BERARD, L., MONTIL, M., NION-LARMURIER, I., BOURRIER, A., LE GALL, G., LALANDE, V., DE ROUGEMONT, A., KIRCHGESNER, J., DAGUENEL, A., CACHANADO, M., ROUSSEAU, A., DROUET, E., ROSENZWAJG, M., HAGEGE, H., DRAY, X., KLATZMAN, D., MARTEAU, P., SAINT-ANTOINE, I. B. D. N., BEAUGERIE, L. & SIMON, T. 2020b. Fecal microbiota transplantation to maintain remission in Crohn's disease: a pilot randomized controlled study. *Microbiome*, 8, 12.
- SOKOL, H., PIGNEUR, B., WATTERLOT, L., LAKHDARI, O., BERMUDEZ-HUMARAN, L. G., GRATADOUX, J. J., BLUGEON, S., BRIDONNEAU, C., FURET, J. P., CORTHER, G., GRANGETTE, C., VASQUEZ, N., POCHART, P., TRUGNAN, G., THOMAS, G., BLOTTIERE, H. M., DORE, J., MARTEAU, P., SEKSIK, P. & LANGELLA, P. 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A*, 105, 16731-6.
- SONCINI, M., TRIOSSI, O., LEO, P., MAGNI, G., GIGLIO, L., MOSCA, P., BERTELÈ, A., POMPEO, F., PIETRINI, L. & MURATORI, R. 2006. Seasonal patterns of hospital treatment for inflammatory bowel disease in Italy. *Digestion*, 73, 1-8.
- SONNENBERG, A., JACOBSEN, S. J. & WASSERMAN, I. H. 1994. Periodicity of hospital admissions for inflammatory bowel disease. *American Journal of Gastroenterology*, 89.
- STAFF, P. O. 2016. Correction: Biomarkers of Inflammation, Immunosuppression and Stress Are Revealed by Metabolomic Profiling of Tuberculosis Patients. *PloS one*, 11, e0153050.
- STORR, M., VOGEL, H. J. & SCHICHO, R. 2013. Metabolomics: is it useful for IBD? *Current opinion in gastroenterology*, 29, 378.
- SU, C., SALZBERG, B. A., LEWIS, J. D., DEREN, J. J., KORNBLUTH, A., KATZKA, D. A., STEIN, R. B., ADLER, D. R. & LICHTENSTEIN, G. R. 2002. Efficacy of anti-tumor

- necrosis factor therapy in patients with ulcerative colitis. *Am J Gastroenterol*, 97, 2577-84.
- SUBRAMANIAN, I., VERMA, S., KUMAR, S., JERE, A. & ANAMIKA, K. 2020. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights*, 14, 1177932219899051.
- SUN, Y. V. 2014. The influences of genetic and environmental factors on methylome-wide association studies for human diseases. *Current genetic medicine reports*, 2, 261-270.
- SUN, Y. V. & HU, Y.-J. 2016. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Advances in genetics*. Elsevier.
- THOMPSON, N. P., DRISCOLL, R., POUNDER, R. E. & WAKEFIELD, A. J. 1996. Genetics versus environment in inflammatory bowel disease: results of a British twin study. *Bmj*, 312, 95-96.
- TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7, 562-578.
- TSAI, T. H., SONG, E., ZHU, R., DI POTO, C., WANG, M., LUO, Y., VARGHESE, R. S., TADESSE, M. G., ZIADA, D. H. & DESAI, C. S. 2015. LC-MS/MS-based serum proteomics for identification of candidate biomarkers for hepatocellular carcinoma. *Proteomics*, 15, 2369-2381.
- TYLER, A. D., MILGROM, R., STEMPAK, J. M., XU, W., BRUMELL, J. H., MUISE, A. M., SEHGAL, R., COHEN, Z., KOLTUN, W. & SHEN, B. 2013. The NOD2insC polymorphism is associated with worse outcome following ileal pouch-anal anastomosis for ulcerative colitis. *Gut*, 62, 1433-1439.
- TYSK, C., LINDBERG, E., JÄRNEROT, G. & FLODERUS-MYRHED, B. 1988. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut*, 29, 990.
- UDDIN, M., KOENEN, K., AIELLO, A., WILDMAN, D., DE LOS SANTOS, R. & GALEA, S. 2011. Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. *Psychological medicine*, 41, 997.
- WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10, 57-63.
- WATERLAND, R. A. & JIRTLE, R. L. 2003. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Molecular and cellular biology*, 23, 5293-5300.
- WEHKAMP, J., HARDER, J., WEICHENTHAL, M., SCHWAB, M., SCHÄFFELER, E., SCHLEE, M., HERRLINGER, K. R., STALLMACH, A., NOACK, F., FRITZ, P., SCHRÖDER, J. M., BEVINS, C. L., FELLERMANN, K. & STANGE, E. F. 2004. NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut*, 53, 1658-1664.
- WEISER, M., SIMON, J. M., KOCHAR, B., TOVAR, A., ISRAEL, J. W., ROBINSON, A., GIPSON, G. R., SCHANER, M. S., HERFARTH, H. H., SARTOR, R. B., MCGOVERN, D. P. B., RAHBAR, R., SADIQ, T. S., KORUDA, M. J., FUREY, T. S. & SHEIKH, S. Z. 2018. Molecular classification of Crohn's disease reveals two clinically relevant subtypes. *Gut*, 67, 36-42.

- WELTER, D., MACARTHUR, J., MORALES, J., BURDETT, T., HALL, P., JUNKINS, H., KLEMM, A., FLICEK, P., MANOLIO, T. & HINDORFF, L. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*, 42, D1001-D1006.
- WILKINS, M. R., PASQUALI, C., APPEL, R. D., OU, K., GOLAZ, O., SANCHEZ, J.-C., YAN, J. X., GOOLEY, A. A., HUGHES, G. & HUMPHERY-SMITH, I. 1996. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/technology*, 14, 61-65.
- WILLIAMS, H. R., WILLSMORE, J. D., COX, I. J., WALKER, D. G., COBBOLD, J. F., TAYLOR-ROBINSON, S. D. & ORCHARD, T. R. 2012. Serum metabolic profiling in inflammatory bowel disease. *Digestive diseases and sciences*, 57, 2157-2165.
- XAVIER, R. & PODOLSKY, D. 2007. Unravelling the pathogenesis of inflammatory bowel disease. *Nature*, 448, 427-434.
- ZENG, Z., MUKHERJEE, A. & ZHANG, H. 2019. From Genetics to Epigenetics, Roles of Epigenetics in Inflammatory Bowel Disease. *Front. Genet.*, 10, 1017.
- ZHENG, J. J., ZHU, X. S., HUANGFU, Z., GAO, Z. X., GUO, Z. R. & WANG, Z. 2005. Crohn's disease in mainland China: a systematic analysis of 50 years of research. *Chinese journal of digestive diseases*, 6, 175-181.

List of publications

Crohn's Disease Mucosal Transcriptomes Highlight JAK/STAT Pathway Involvement in Postoperative Recurrence. (submitted)

Marjolaine Ngollo, **Kevin Perez**, Nassim Hammoudi, Yuri Gorelik, Marc Delord, Claire Auzolle, Hugo Bottois, Dominique Cazals-Hatem, Madeleine Bezault, Stéphane Nancey, Benjamin Pariente, Xavier Treton, Mathurin Fumery, Antony Buisson, Nicolas Barnich, Philippe Seksik, Shai S Shen-Orr, Lionel Le Bourhis, Matthieu Allez

T cell clonal expansions in ileal Crohn's disease are associated with smoking behaviour and postoperative recurrence, Gut 2019

Matthieu Allez, Claire Auzolle, Marjolaine Ngollo, Hugo Bottois, Victor Chardiny, Ana Maria Corraliza, Azucena Salas, **Kevin Perez**, Carmen Stefanescu, Stéphane Nancey, Anthony Buisson, Benjamin Pariente, Mathurin Fumery, Harry Sokol, Xavier Tréton, Nicolas Barnich, Philippe Seksik, Lionel Le Bourhis

Monitoring of mucosal T cells subsets under biotherapies in IBD: responders exhibit a phenotype similar to healthy controls (in preparation)

N Hammoudi, **K Perez**, D Hassid, ML Tran Minh, JM Gornet, C Baudry, N Lourenco, B Gergaud, V Chardiny, J Bonnet, L Le Bourhis, M Allez

Meta-analysis of IBD gut samples gene expression identifies specific markers of ileal and colonic diseases (submitted)

Kevin Perez, Marjolaine Ngollo, Nassim Hamoudi, Philippe Seksik, Lionel Le Bourhis, Ramnik J. Xavier, Mark J. Daly, Matthieu Allez

Conferences attended

United European Gastroenterology Week (UEGW) 2018, **UEG Top abstract Prize 2018.**

European Crohn's and Colitis Organisation (ECCO) 2017, Journées Francophones d'Hépatogastroentérologie et d'Oncologie Digestive (JFHOD) 2017

Appendix

A. 1. The ELYP study

Immune phenotyping and disease course of IBD patients undergoing biotherapies

N. Hammoudi, K. Perez et al.

Background

A dysregulated T cell response is involved in the pathophysiology of Inflammatory Bowel Diseases (IBD). T cells expressing the NKG2D activation receptor are implicated. Here, we investigated changes in intestinal T lymphocyte subsets overtime in patients treated by biotherapy for active disease.

Methods

In this prospective single-center study, patients with IBD who have started a biotherapy (Infliximab, Adalimumab, Golimumab, Vedolizumab, Ustekinumab) for active disease were included. Mucosal biopsies in inflamed segments were taken during endoscopies before treatment initiation (W0) and in the same location 14 weeks (W14) and 52 weeks (W52) after inclusion. Control biopsies were taken from 16 patients without IBD (7 ileal and 9 colonic). Immunophenotyping of isolated lymphocytes was performed by flow cytometry at each timepoint. Correlation with endoscopic response was evaluated. For Crohn's Disease (CD), response was defined by a 50% decrease of the CDEIS score. For Ulcerative Colitis (UC), response was defined by a 2 points decrease of the UCEIS

Results

Between February 2016 and November 2018, 85 patients were included. Forty-four were male (52%). Forty had active CD with the most inflamed segment located for 28 in the colon and 12 in the ileum. Thirty-nine patients had active UC and 6 pouchitis. An anti-TNF was started for 51 patients (60%), Vedolizumab for 16 patients (19%) and Ustekinumab for 18 (21%). Lymphocytes

phenotyping was performed at inclusion for the whole cohort, at W14 for 67 patients (79%) and at W52 for 49 patients (58%). Data concerning endoscopic response was available for all patients. At W0, IBD patients displayed higher rates of CD4 T cells and lower rates of CD8 T cells than controls (medians: 31% vs 48% for CD4, $p < 0.001$ and 52% vs 37% for CD8, $p < 0.001$ respectively). Endoscopic responders experienced a normalization of these mucosal populations as compared to non-responders (W14: 58% vs 45% for CD4, $p = 0.005$ and 40% vs 28% for CD8, $p = 0.03$; W52 56% vs 43% for CD4, $p = 0.01$ and 27 vs 43% for CD8, $p = 0.003$). Results remained significant when splitting the cohort in CD and UC patients.

Before treatment initiation, IBD patients displayed lower expression of NKG2D on CD8 T cells (78% vs 94%, $p < 0.001$). At W52, endoscopic responders presented significantly higher expression of this marker on CD8 T cells compared to non-responders (85% vs 68%, $p < 0.001$). Results remained significant in UC and CD patients.

Conclusion

The intestinal T cells of IBD patients with active disease present an increase of the CD4 to CD8 T cell ratio and lower expression of NKG2D on CD8 T cells reflecting an engagement of this receptor in inflammation. These T cell parameters are normalized in endoscopic responders but not in non-responders

