Understanding the Complexity of Inflammatory Bowel Disease from a Multi-Omics Perspective
Hu, Shixian

DOI:
10.33612/diss.182834035

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 9

Discussion and future perspective
Inflammatory bowel disease (IBD) results from both genetics and environmental factors. Genome-wide association studies (GWAS) have led to the discovery of more than 240 IBD-associated genetic variants. The use of experimental animal studies has shed light on the molecular mechanisms for a few IBD-associated variants. But we are still lacking understanding of IBD genetics in human. In order to study how genetic risk factors are involved in the pathogenesis of IBD, a more holistic perspective is needed. In this context, this means we need to understand how the genetic variants regulate gene expressions, proteins, metabolites and gut microbiota and ultimately lead to IBD.

Therefore, this field has gradually switched from solely genotype-phenotype associations to the combination of omics data including genetics, transcriptomics, proteomics, metabolomics, gut microbiota and environment factors. During my PhD study, multi-omics data integration has been a powerful approach to study the complex mechanisms behind IBD in humans.

The main goal of my thesis was to add new insights on how genetic factors contribute to IBD by combining different molecular traits, including gene expression, protein level, metabolites, gut microbiota and a broad range of (disease)-phenotypes. The first part of this thesis focused on integrating transcriptomics and proteomics data to link genetic factors to disease from different aspects (chapter 2-3). Part two I put my effort at the host genetics-microbiota interactions as well as fecal metabolomics in both IBD and general populations (chapter 4-8). In the last chapter, I discussed the conclusions from my PhD thesis and future perspectives.

**Discussion**

**QTL effect is context-specific**

Studies have shown that quantitative trait loci (QTL) is context-specific, which means the genetic regulation effect on molecular traits, e.g. gene expression, protein level, metabolites, gut microbiota and a broad range of (disease)-phenotypes. The first part of this thesis focused on integrating transcriptomics and proteomics data to link genetic factors to disease from different aspects (chapter 2-3). Part two I put my effort at the host genetics-microbiota interactions as well as fecal metabolomics in both IBD and general populations (chapter 4-8). In the last chapter, I discussed the conclusions from my PhD thesis and future perspectives.

**Potential determining factors for molecular biomarkers**

Accurate and effective biomarkers are used for disease diagnosis and monitoring. Therefore, various biomarkers, such as blood C-reactive protein (CRP) and fecal calprotectin (FCP), have been widely used in IBD clinical practice. But neither of them is perfect because many factors (e.g. bacterial infections) can cause false positives. Alternatively, the levels of blood cytokines and growth factors have also been suggested as potential biomarkers.

In chapter 5, we identified mbQTL effects that were only present in patients with IBD but not in healthy controls. Variants in the IBD-associated gene *BTNL2*, which regulates T cell proliferation, were associated with an increased abundance of *Bacteroides cellulosilyticus* while this observation was not present in healthy people. In the other way around, the association of *LCT*- *Bifidobacterium* mentioned above was not observed in IBD. These phenomena might be due to the dysbiosis in disease which hides the host-microbiota interactions or makes them more evident. In chapter 4, mbQTL effects were also dependent on other conditions. The abundance of genus *Collinsella* and microbial lactose degradation pathway were associated with host *ABO* locus, but dependent on *FUT2* genotype (the gene that determines the antigen secretor status). We only observed the *ABO* mbQTL effects at secretors which suggested these associations are likely to be driven by the exposure of A/B antigens to gut bacteria.
Host and environmental factors of blood cardiovascular biomarkers have been studied in general populations\(^{20,21}\) which demonstrated strong genetic regulation effects. However, investigations on IBD-related blood biomarkers are still lacking. In chapter 3 we focused on inflammation biomarkers and firstly assessed the influence of both genetics and phenotype on protein levels. 69 out of 86 proteins were affected by age, sex, surgery, the use of medication, genetics, etc. For example, VEGFA, a vascular endothelial growth factor protein, which has been reported as a biomarker for structuring CD\(^{12,13}\), was significantly influenced by genomic variants nearby the corresponding protein-coding gene. This result is consistent with “human plasma pQTL was significantly influenced by genomic variants nearby the corresponding protein-coding gene.” Compared with mbQTLs associated with the impaired mucosal barrier function in mice studies\(^{27,28}\), the latter one also observed mbQTLs surrounding the histo-blood group ABO system transferase (ABO) gene. In chapter 6, we conducted an mbQTL analysis in the largest population with metagenomics sequencing (Dutch Microbiome Project) so far. We added new insights with higher resolution that multiple microbial signatures were associated with LCT and ABO loci. For example, we observed the Bilidobacterium bifidum and Collinsella aerofaciens were associated with ABO locus which have not been reported in previous 16S study. We also estimated that our current sample size was still not enough to identify genetic effect on low prevalent bacteria. As I mentioned above, studying mbQTLs in IBD might help to understand the role of host-microbiota interactions in disease. Sokol et.al have identified NOD2 genetic variants associated with lower abundance of Roseburia and Faecalibacterium genus in IBD patients\(^{24}\) in chapter 5, by using whole exome sequencing (WES) and metagenomics sequencing, we identified novel associations between the host protein-coding variants, copy number variants, rare mutations and microbial taxa and metabolic potential. Several genes were involved in IBD and immune diseases, for example, IL17REL, MYRF, and CD160, which suggested the functional mutated variants of immune-related genes play key roles in dysbiosis in IBD. Although the sample size is relatively small, we first revealed that the difference between host-microbiota interactions in disease and healthy controls. Still, we need to consider more sub-phenotypes like disease activity and replication in multiple IBD cohorts.

**Host genetics-gut microbe interactions**

Both host genetics and gut microbiota play important roles in IBD. Around ten years ago, people found evidence in twins that the composition of gut microbiota was partly heritable\(^{18}\). Later studies have showed a clear association between the functional mutated LCT (a lactase coding gene) genotype and higher abundance of Bifidobacterium\(^{15,20}\). This association indicates the potential of bacteria in treating lactose intolerance symptoms. Therefore, studying host genetics-microbiota interactions helps to understand the determining factors of human gut microbiota and how it is relevant to health.

In chapter 4, I introduced the current knowledge of host genetics-microbiota interactions based on populations, including tens of genome-wide significant microbial quantitative trait loci (mbQTL)\(^{21,22,23}\). However, the results across studies are poorly replicated between each other. The main reason is that the variation of individual gut microbiota is huge. Compared with environmental factors, the host genetics effect is relatively moderate and thereby much larger sample sizes are needed. Other reasons include the differences of cohorts, statistical methods and data generation techniques.

During my PhD study, there have been two large-scale mbQTL studies using multiple population-based cohorts and 16S sequencing technology. The MiBioGen study included 24 different cohorts with over 18,000 individuals\(^{24}\). The other study from Germany consisted of 8,956 individuals from four cohorts\(^{25}\). Both of these two studies replicated the association of LCT-Bifidobacterium. The gut microbiota has formed a mutual beneficial balance with the host during the long evolution, however, the balance can be disrupted in a disease situation like IBD. People have observed dysbiotic gut microbiota was associated with the impaired mucosal barrier function in mice studies\(^{27,28,29}\). Nowadays, studies have identified patients with IBD to be characterized by wide-spread changes in gut microbiota compared with healthy individuals. For example, short chain fatty acid (SCFA)-producing bacteria like Faecalibacterium and Roseburia showed lower abundance in patients with IBD\(^{10,31,32,33,34}\), which indicates the role of gut microbiota in disease pathology.
Some studies have used metagenomics sequencing to infer the microbial metabolic potential based on genes abundance but it cannot reflect the actual metabolic activity in vivo. Fecal metabolites contain small molecules involved in a wide range of pathways which are important mediators of host-microbiota interactions. Therefore, in chapter 7, we measured 1684 untargeted fecal metabolites and showed a broad metabolic difference between IBD and controls. We identified the decreased ratio from primary bile acids to secondary bile acids in CD patients indicated a disrupted microbial deconjugation activity which is consistent with HMP2 and PRISM studies. Recently, Hang et al. and Song et al. have also revealed that the bile acid metabolites modulate host T cell homeostasis which was important for controlling inflammation. Beyond these, our analysis has identified metabolic ratios presented better IBD prediction potential compared with using individual metabolite or microbial data. Moreover, other studies have revealed the large influence from diet on blood metabolites. However, we identified the dominant effect on fecal metabolites was the gut microbiota which further suggested the fecal metabolites acted as a functional readout of the gut microbiota. This study showed the disrupted gut microbial activities in IBD which is important complementary evidence for current metagenomics and 16S based studies.

The gut microbiota alteration has been observed in many disease conditions and even used to predict human mortality risk. In chapter 8, we observed largely changed gut microbial composition and functions in post organ (liver and renal) transplant patients which could be resulted from the use of immunosuppressive medications and surgery effects. Moreover, we identified the alpha diversity was significantly associated to an increased mortality in patients after one-year transplantation which showed a promising microbiota-based biomarker for clinical prediction.

**Multi-omics data and clinical care**

The main purpose of personalized medicine is to use individual factors like genetics and gut microbiota, to improve IBD clinical treatment. Rapid development of multi-omics data from IBD cohorts has investigated a wide range of molecular traits and identified the close relevance to the disease outcome.

Several IBD biobanks have been established during my PhD study. The Integrative Human Microbiome Project (HMP2) in the US enrolled 132 adult patients with IBD and followed them one year collecting biomaterials and generating metagenomics, proteomics, metabolomics, genetics and transcriptomics data. This study has identified that periods of IBD disease activity was characterized by temporal shift of microbial community. Another American cohort (RISK) collected fecal, biopsies and blood samples from more than 400 treatment-naïve and new-onset CD patients. Other large-scale IBD cohorts are also mainly in the North American and European countries, but most of them are still under the establishment for multi-omics data. These studies have discovered direct evidence of interactions between host and microbiota in IBD, for example, the increased expression of antimicrobial gene DUOX2 in gut mucosa was correlated with expansion of some facultative anaerobes. They provided a valuable resource for microbiota- or molecule-based therapeutic strategies. However, due to the large disease heterogeneity and individual difference, some findings are only observed in a particular set of patients, for example, gut microbial taxa or community have not been consistently associated with factors including cohort-specificity, disease subtypes, disease progress, or medication response, which suggests that larger sample size, more personalized factors should be considered.

The University Medical Center Groningen has established the largest cross-sectional IBD cohort with omics-data so far (1000IBD), with detailed prospective follow up of clinical and personal records. During my thesis writing, based on 1000IBD, Vich Vila et al. have shown hints that gut microbiota metabolizes drug compounds, and thereby potentially influences the medication response. Bolte et al. have revealed personal specific nutrients associated with pro- or anti-inflammatory species which provided the reference for diet intervention studies. In my thesis, by using host genomics and fecal metagenomics data, Chapter 3 and chapter 4 added further exploration on interactions between genetic factors and gut microbiota, pinpointing important mechanisms behind the disease risk. Chapter 5 inferred the mucosal cell type enrichment using transcriptomics data, highlighted the role of cell types in intestinal inflammation and eQTL effects. There is also an increasing number of studies that have suggested novel drug target potential at mucosal cells in patients. Chapter 6 addressed the genetic and phenotypic factors influence at blood biomarkers in IBD based on genomics and proteomics data, which emphasized the importance of patient stratification when analyzing the potential of blood-based biomarkers. Chapter 7 showed good IBD diagnosis potential of fecal metabolites ratios through metabolomics data. In chapter 8, we observed the gut microbiota dysbiosis was linked to higher mortality in solid organ transplantation using metagenomics data, which is another example to use high-throughput data to improve clinical prediction.
To summarize, I focused on different aspects in different chapters and showed the great health care potential from multi-omics data. Particularly, I have linked genetics to downstream traits such as gene expressions, proteins and gut microbiota. This is important for us to explain why and how genetics contributes to disease. However, as a cross-sectional cohort, 1000IBD consists of mainly well-established disease patients which not be suitable to explore the disease pathology from early onset stage. Therefore, we are still on the way to devote efforts to the cross-cohort collaboration and launch longitudinal studies for dynamic omics-data research.

Future perspective

New opportunities to expand the genetic architecture of IBD

Array-based GWAS have linked hundreds of common genomic loci to IBD risk, however, most of the variants are located in non-coding regions which is a challenge to infer the disease causal variants or genes.

Protein truncating variants (PTVs) have deleterious impact on downstream molecular functions and thereby link the causality with disease onset. One example is the splice SNV in CARD9 gene and it is discovered to be protective in CD by sequencing candidate exonic targets. And this finding is validated in following mice models where this SNV disrupts the downstream protein-protein interaction with TRIM62, which leads to decreased level of the pro-inflammatory NF-kappaB pathway. However, PTVs are difficult to be captured in array imputation-based GWAS due to the rare allele frequency (MAF <0.05). Therefore, using exome sequencing provides the opportunities to get the full understanding of IBD genetic architecture. During my PhD study, big world-wide collaborations have been started in IBD WES study. The Helmsley IBD Exome Sequencing Program (HIESP) aims to sequence more than 15,000 IBD whole exomes, covering a range of Europeans, Americans and Asians. In my opinion, this will provide a great resource of candidate IBD causal genes and therefore, more research can be performed for functional follow-up and identify new drug targets based on IBD-associated PTVs. For example, some PTVs are loss-of-function variants which are similar to “knockout” humans. Thus, drugs like molecular inhibitors can be predicted and developed.

IBD patients with long time inflammation duration have a high risk to colorectal cancer (CRC) which reminds us the intestinal somatic mutations might attribute to this progress. Somatic mutations can occur as a consequence of exposure to certain environment and enable the cells with autonomous self-renewal ability in tumorigenesis. However, somatic mutations have been ignored in most IBD genetics studies. At the time of writing this thesis, several studies firstly described the somatic mutations findings in the inflamed intestinal epithelium from IBD patients. They identified inflammation leads to cell genetic evolution. However, there are still lots of questions remain to answer. How are these somatic mutations related to IBD disease activity? What is the consequence to the downstream molecular traits and mucosal ecosystem? Can we use somatic mutations to monitor IBD progress to CRC? Can we find out mutated cells and consider them as novel drug targets? By comparing mutations from WES or RNAseq data between blood and gut mucosa, we can assess the landscape of somatic mutations in vivo. I believe studying the intestinal somatic mutations will be an important complement to IBD genomics.

Gut microbiota, an extension of human genome

We have obtained much knowledge about human genetic architecture during last decades. Here I suggest that gut microbiota is an extension of human genome, because their functional activity is closely related to our daily life, e.g. food digestion, drug metabolizing and molecules absorption. Thus, there is a need to explore not only gut microbial taxa, but also their genetic architecture, e.g. SNPs and structural variations, to see how it leads to the microbial metabolic differences and how it is related to our health.
Gut microbial strains have different functions\textsuperscript{61}. Therefore, one important question is can we find the right IBD-associated strains that have specific genomic characteristics and lead to pathological states? High-throughput metagenomics sequencing allows us to compare the bacterial genomic content based on reference genomes or identify novel strains by \textit{de novo} assembly. Hall et al. have identified two distinct strains of \textit{Ruminococcus gnavus} using metagenomics sequencing in IBD patients but only one is associated with increased disease activity. This strain is enriched with genes related to oxidative stress tolerance and it might be the real disease-causing pathogens\textsuperscript{62}. Strikingly, \textit{culturomics} approach, which uses multiple culture conditions for microbes, has been developed rapidly and isolated hundreds of novel human gut bacteria strains\textsuperscript{63}. This method can provide opportunities for \textit{in vitro} experiments validation. Therefore, future studies should use IBD patients with precise phenotype records, combine metagenomics sequencing and \textit{culturomics} approaches to characterize which is the likely disease-associated bacteria, what is the identified bacteria doing and how can we manipulate the microbiota for treatment.

Another question is how microbial genetics influence human phenotypes in populations. Metagenomics sequencing provides great potential to explore gut microbial structural variations (SVs) and mobile genetic elements\textsuperscript{64,65}. Therefore, we are able to link microbial genetics with host phenotypes. Chen et al. have discovered potential associations between microbial SVs with human metabolic diseases\textsuperscript{66}. Similar with human genetic association studies, we can perform microbial genetics-host phenotype associations in large-scale populations. This is important for microbiota-based treatment, for example, manipulating gut bacteria genetics and monitoring fecal microbiota transplantation (FMT) at higher resolution. Meanwhile, we can also explore the dynamic ecology between the microbiota and host. Using a longitudinal cohort, closely tracking how the gut microbiota and the host phenotypes change, we can reveal how the bacterial genomic evolution from people birth to adult, or from healthy status to disease onset.

**Study genetic effect in a context-specific manner**

It has been recognized that the genetic effect on molecular traits is specific to certain context like tissue, cell and environment. This might be the key answers to the question: why do we observe different phenotypes from the same genomic variants?

IBD is characterized by a relapsing and remitting course. There is little known that if IBD genetic effect can be temporal-specific and how it is involved in IBD repeated flares. During this thesis writing, Gutierrez-Arcelus et al. have observed \textit{HLA-DQB1} alleles have distinct eQTL effects in blood T cell dynamic physiological states\textsuperscript{68,69}. But these studies are from \textit{in vitro} experiments. The dynamic intestinal ecosystem in IBD patients is more complex as it is a combination of host factors (e.g. genetics and immune cells) and gut microbiota. Therefore, long-term tracking patients and intense sampling of relevant biopsies will bring us dynamic information. For example, can we identify different eQTL effects before, during and after flares? If yes, do these temporal-specific genetic effects influence the immune-related cell activities, e.g. secretion of signaling molecules? If yes, how is this relevant to the disruptions of gut microbial community? The ultimate aim here is to find the cellular pathways which could increase the remission period while suppress the time of flares in IBD patients. Thus, these findings would help the disease prevention and intervention.
**Time to establish causation**

The determination that if a correlation is causal is critical for IBD clinical applications because it enables us to take action to the true hazardous factors. However, we cannot address the causality from purely correlation in cross-sectional studies. Here I summarized multiple ways to estimate causality.

Mendelian Randomization (MR) is a statistic method to infer the causal relationship between exposure and outcome, using genomic variants as instrumental variables. Recently, the SCALLOP consortium has used blood pQTLs and diseases GWAS summaries and identified that higher circulating Suppression Of Tumorigenicity 2 (ST2) level causes higher risk of IBD. With the increasing sample sizes of IBD GWAS and omics-data summaries (eQTLs, pQTLs, mQTLs, mbQTLs) from populations, we have gained more power to infer the causality between different molecular traits with disease onset. Moreover, MR has been a promising approach to prioritize drug targets.

The strongest causal evidence is from experiments. We can observe the direct effect after manipulating the mouse models. For example, NOD2 deficiency triggers intestinal inflammation in mice; colonization with specific microbiota in germ-free mice induces colitis and T cell increase. Measuring the ability of bacteria metabolize hundreds of drugs provides causal evidence of microbial gene content and metabolic activity. This facilitates our understanding the role of gut microbiota in drug metabolizing. Recently, organ-on-a-chip shows great potential to mimic human intestinal ecosystem, which allows us to combine microbiota and other metabolites to stimulate “gut tissue”. Therefore, it is possible to explore multiple risk factors of IBD in vitro.

A longitudinal study in patients will also provide important information. IBD flare is unpredictable and the causality remains unknown. Until now, we still cannot distinguish between the change of IBD activity and gut microbiota, which is the first and which comes later. One possible reason might be the limited sampling timeframe. Therefore, densely sampling is necessary to guarantee a high resolution to find which factors proceed disease activity.

**Clinical omics in IBD**

Although the emerging omics-data accelerates our understanding of the molecular pathogenesis, we still have a long way to go to translate all the knowledge to IBD clinical applications. I listed several promising directions and big challenges so far.

**IBD diagnosis** To distinguish IBD from controls or other diseases, multiple layers of data provide a huge pool of biomarker candidates, including serum proteins, fecal microbiota, fecal metabolites and circulating non-coding RNAs. Further, biomarkers are also developed for IBD sub-classification, for example, stricturing behavior, penetrating behavior and fibrostenosing behavior. But there are several main considerations. One is as I discussed in chapter 6 and chapter 7, each biomarker should be carefully studied on its potential confounding factors. If one biomarker is strongly influenced by genetics or food intake, then we might need patients’ stratification to improve the sensitivity. The second one is the specificity. We might need combination of multiple biomarkers to have a better classification of disease sub phenotype or different status. The third one is the balance between omics-data cost and benefit. Clinicians should make a selection of the biomarkers, keeping the high diagnosis efficiency but with relatively low cost.

**IBD treatment** There are two main aspects. Firstly, the potential to prioritize novel drug targets. Omics-data reveals disease relevant pathways in different molecular levels (e.g. DNA, RNA and proteins), and thereby presents many putative targets. Secondly, repurposing of the existing drugs. As showed in chapter 5, we identified some drug targeted genes are largely dependent on genetics which might lead to low-response rates in a sub group patients based on their genetic background. Therefore, omics-data provides opportunities to predict and re-evaluate the current drug efficiency.

**IBD prevention** Multiple risk factors of IBD have been studied using omics-data, which allows the disease risk assessment to be made in human early life. For example, if a person has less healthy-related gut microbiota, clinicians could give suggestions to manipulating his or her daily diet and lifestyle towards a healthier status. Thus, intervention actions could be taken in an early stage to reduce the exposure time to disease risk factors.

Despite these advantages, we are facing some challenges. Firstly, omics-data sharing and collaboration. IBD patients have large individual variation (e.g. gut microbiota), which leads to some inconsistent findings between different IBD cohorts. For replication and meta-analysis purpose, I encourage that a world-wide IBD omics consortium should be established, similar with the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). Moreover, this could also benefit other developing countries in the world to help them with IBD research. Secondly, we need hypothesis-driven study.
Many omics-data studies in cross-sectional cohorts are hypothesis-free and data-driven, which can obtain plenty of candidate disease mechanisms. Due to the fact that these studies in general cannot establish the causality, there is still a big gap from findings to applications. Therefore, start from the cross-sectional study findings, generate with clear hypotheses, and conduct a smaller longitudinal or intervention validation, would definitely benefit translational medicine. Finally, methods development. Omics-data generation needs standard and uniformed protocols which guarantees the discoveries reliability and reproducibility. Researchers should put more efforts to get consent on the data production. In addition, omics-data brings challenge to data integration. Large number of statistic methods and bioinformatics tools has been created nowadays, for example, construct complex networks to integrate different layers. Therefore, analytic methods development could facilitate the “puzzle” completion. On the other hand, we should also keep in mind that proper biological and translatable interpretations are the key to reveal the “black box”.

Take home messages from this thesis

1. Genetics affects IBD onset through complex downstream pathways, including gene expressions, proteins, metabolites and gut microbiota.
2. Protein-coding and copy number variants potentially regulate gut microbiota in IBD. These variants are enriched in immune-related genes.
3. Current microbial genome-wide association studies are limited by sample size and much larger number of samples is needed.
4. Genomic variants regulate gene expressions in inflammation-dependent and cell enrichment-dependent manners in IBD intestinal tissue.
5. Plasma proteins in IBD are affected largely by genetics and clinical histories.
6. Fecal metabolic ratios are promising biomarkers for IBD diagnosis.
7. Gut microbiota can be used to predict post-organ-transplantation survival.

References

46. Collij, V. Gut microbiome and proteome changes as biomarker of response to vedolizumab in patients with inflammatory bowel disease.
Chapter 9 - Discussion and future perspective


