Gut mucosal gene expression in inflammatory bowel disease

Uniken Venema, Werna

DOI:
10.33612/diss.178571637

chapter EIGHT

Discussion

In the past five years, the treatment of IBD has rapidly evolved. The implementation of a variety of biological drugs has lowered the use of corticosteroids and the need for surgery[1,2]. Moreover, an ever-ongoing research effort is allocated to develop new pharmacologic, and recently also cellular, therapies. However, a decrease in likelihood of patients to progress from a mild to a more complicated disease course, in particular for CD[3], is yet to be achieved. The 'holy grail' therapeutic to cure our patients with IBD, or at least induce sustained remission, has yet to be discovered. Nevertheless, advancements in our understanding of IBD disease pathology and improvements in the application of the currently available therapies have been made. This chapter discusses the findings as presented in this thesis in the light of current knowledge.

Cell type-resolution biology of gut mucoal tissue through scRNAseq

Gene expression provides insight into the functioning of cells and tissues under various conditions. Previously, cell-specific gene expression was almost exclusively studied on human blood cells, given their wide availability and accessibility. Meanwhile, tissues other than peripheral blood were generally assessed as a whole, at first through low-resolution microarrays and later with bulk RNA sequencing techniques. It was with the rise of single-cell technologies that we and others could study solid gut tissue at high, single-cell resolution, both in the context of health and disease (**chapters 3,5,6**).

The main advantage of scRNAseq is that it allows to map cell functions and cell-cell interactions, as well as infer cellular development and differentiation trajectories. This has revolutionized human biology research, particularly the field of immunology[4]. With the shift from cell type definition by surface markers to using transcriptomics and a combination of both, the large heterogeneity of immune cells has been exposed. In **chapter 4**, we show that immune cells from CD gut mucosal tissue exert different expression patterns as compared to their equivalents derived from blood, indicating that the cells have tissue-specific functions. Expression of genetic risk genes for CD was mainly seen in some gut mucosal T cells such as Th17 cells, and less prevalent in the blood. This knowledge can be used to target cells in a tissue-specific way.

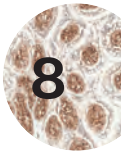
Cell type definition is not black and white in scRNAseq: we noticed that $CD8^+$ and $CD4^+$ cells intermix when clustered based on their transcription profiles. It suggests that although the surface marker expression of these T cell subsets differs, their transcriptomic compartments may be similar. This shows that in single-cell sequencing, cell types are defined based on function and also on developmental state. With the increased number of large multi-center studies, various novel cell subtypes (and developmental states) have been identified. Following these developments, global collaborative networks are

necessarily integrating the generated single-cell sequencing data from many tissues into online cell atlases such as the Human Cell Atlas[5], empowering exchange of knowledge within and between fields, and enabling fellow researchers to explore the available data.

For years, T cells were assumed to be the driving cell types in IBD pathogenesis. Recent scRNAseq studies, including those described in **chapter 5** of this thesis, revealed the biological roles of previously understudied cell types, such as fibroblasts and tuft cells, and novel cell types, such as *BEST4* epithelial cells, in IBD pathology[6–8].
Chapter 5 summarizes the scRNAseq-mediated discoveries that changed our view of human gut biology and gastrointestinal diseases. It was demonstrated that disease-specific variations in cell type composition exist between healthy and diseased gut tissues. Inflammatory fibroblasts, for example, which are only present in small numbers in healthy tissue, were found to be highly abundant in inflamed tissue in IBD. In the pre-scRNAseq era, this particularity was obscure. Whether these compositional variations delineate truly distinct cell types, or rather different acquired states of the same cell type, remains an unanswered question.

Technological developments at single cell level

Single-cell technologies went through a rapid technological development and optimization, which has improved results but also posed new challenges. In the early days of scRNAseq, cells were individually sorted in plates, making a simultaneous analysis of hundreds of cells possible (SMART-seq[9], CELL-seq[10]). Subsequently, methods have evolved towards fluidics-based high-throughput methods (Drop-seq[11]), enabling library generation of thousands of cells at a time, while at the same time lowering the costs per cell sequenced. Whereas in the beginning non-standardized, lab-specific protocols existed, now companies (e.g. Fluidigm C1, Takara ICELL8 and the 10X Genomics Chromium) have evolved, rendering rapid improvements in efficiency, standardization, accessibility and further applications of the method.



Various tools for single-cell measurements have been developed, such as those on the level of (surface)protein (CyTOF), DNA variations, repressive chromatin state (ChIPseq) and active chromatin state (e.g. ATACseq, DNase-seq), RNA levels and splice forms, protein levels (CITEseq, REAP-seq) and cell clonality (Single Cell V(D)J). Interactions between cells in an environment can now be modeled through bioinformatic approaches (e.g. CellPhoneDB[12], CellChat[13]). Novel methods for demultiplexing, such as cell hashing (CITEseq[14]) and variant-based demultiplexing (Souporcell[15]), enable pooling of multiple samples for sequencing, thereby reducing batch effects and costs.

Nevertheless, we are only at the start of the single-cell technological revolution, as it faces a number of issues. First, applying single-cell sequencing to solid tissues poses a challenge, due to the currently unavoidable step of sample dissociation from solid tissue to single cells. Gut mucosal tissue dissociation is a labor-intensive and time-consuming procedure, which should be approached with caution since scRNAseq is sensitive to noise and perturbations. A point of concern is that a variety of tissue dissociation protocols are used across the IBD research groups. In order to map their effect on scRNAseq data, we performed a comparison of three gut mucosal tissue dissociation protocols in **chapter 6**. We showed that a one-step collagenase protocol has an advantage over currently used multi-step protocols as it eliminates step-induced batch effects. However, depending on the research question and interest in specific cell subtypes, other dissociation protocols may be preferred. The choice of a certain dissociation protocol has a large impact on the measured gene expression, potentially hindering meta-analysis of previously presented results. Second, the most used protocols are not feasible for high-throughput studies, which are necessary to draw robust conclusions based on biological findings. We showed that tissue cryopreservation is a valuable tool for increasing the throughput while reducing potential batch effects in studies that utilize single-cell sequencing (**chapter 6**). In **chapter** 7, we used the preferred protocol for a study comparing PSC-IBD to UC.

Although cryopreservation is used in our research center on a daily basis, it has yet to be implemented across endoscopy centers and labs in the field to enable robust, comparable studies all over the world. Third, the bioinformatic tools for scRNAseq and in particular for paired and longitudinal downstream analyses, are continuously being renewed and further developed, rendering them non-standardized. Standardization of both lab and computational protocols is a key step for improving the quality and comparability between studies, but more importantly, for drawing solid conclusions that can be translated to effective clinical solutions. Large collaborative institutions such as the Human Cell Project could play a main role in standardizing protocols.

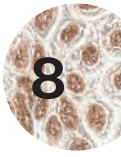
The promise of multi-omics

The exploration of only one data layer is often insufficient for understanding the complexity of human diseases such as IBD. For instance, grasping the heterogeneity of clinical presentations of IBD requires detailed knowledge of the phenotype of each participant in a cohort. Moreover, it has been shown that genetics and gene expression are only two pieces of the puzzle. Therefore, cohorts should include additional layers of information, such as the study of proteins, microbiota, exposome and diet among others. Only integration of this data with its clinical (*e.g.* medication use, disease behaviour) context produces meaningful insights in disease pathology. Unfortunately, studies that integrate multiple data layers are still scarce, thus there are large steps still to be taken.

An integrative analysis of the gene expression data, genotype data and inflammation status of IBD patients' gut mucosa is performed in **chapter 3** We showed that the effect of genetic variants on gut mucosal gene expression depends on the inflammation status of the tissue samples. Following that observation, we found 190 inflammation-dependent interactions between the genotype and mucosal gene expression in patients with IBD (eQTLs). A clear example of such inflammation-dependent eQTL is formed by *IL26*, encoding a key gene in antimicrobial defense. This gene shows an eQTL effect in inflamed tissue, whereas in non-inflamed tissue there is no genotype-dependent expression. These findings suggest that the effectiveness of the antimicrobial defense mechanism upon inflammation is affected in genetically susceptible individuals. Next, we found that the inflammation-dependent eQTLs are, in part, associated with gut mucosal cell type composition. For example, the eQTL of SNP rs76748970 with immunoglobulin heavy chain-encoding transcript *IGHV4-4*, was shown to be more pronounced in mucosa enriched for plasma cells, raising the question whether the gene expression is a cause or a consequence to the cell type enrichment. To find whether these eQTLs are cell type specific, single-cell eQTL studies need to be pursued. These inflammation-eQTLs support a personalized approach to treating gut inflammation.

Although the influence of SNPs on gene expression was previously thought to be a tissue-dependent (*i.e.* GTEx), static process, we showed that differences in phenotype (*e.g.* inflammation status) influence the relative context between genetic variants and gene expression. Others, for example, described that smoking induces an eQTL in *FPR1*, a gene involved in the inhibition of lung epithelial wound repair[16]. These findings have changed the way we look at penetrance of genetic variants. It may be that additional factors, such as appendectomy or growing up in the countryside, may influence eQTLs in the same way, providing possibly an explanation for their association with IBD. This complexity once again underlines the need for developing the methods to study disease pathology by multi-layer data integration.

Thus far, interactions between layers and (disease) networks have been revealed through integration of various data layers. This holds true for: gene expression and microbiome data[17], clinical phenotypes with microbiome data and single-cell expression data[18,19], as well as for protein interactions and effects of dietary components[20,21]. An example of a finding that resulted from the integration of multiple data layers, is the increasing number of reactive bacterial species along the colon (*i.e.* from proximal to distal parts) and its association with the accumulation of certain B cell clones[18]. This provides an insight into the interaction between the host and the microbiota, and causes or consequences to microbial dysbiosis in IBD. Even though the research efforts to integrate multiple data modalities and to study biological changes over the course of multiple time-points are



being made, (3TR[22], immuniverse[23], (sc)eQTLgen[24], human microbiome project[25]), the applied approach still needs to evolve before the day comes when we are able to explain the full complexity of IBD.

To enable better integration of multiple data layers, the following is required: (1) collaborative medical staff and participants; (2) an appropriate research infrastructure with dedicated people who collect, store and process samples in a correct, uniform way; and (3) bioinformatic approaches that ensure a systematic and statistically correct integration of such big data. In our research institute, the 1000IBD project is designed to enable the so-called 'systems biology' approach, as it aims for a combinatory analysis of many different data modalities, including clinical data, and is and will be a valuable source for future research in the IBD field[26] (**Figure 1**).

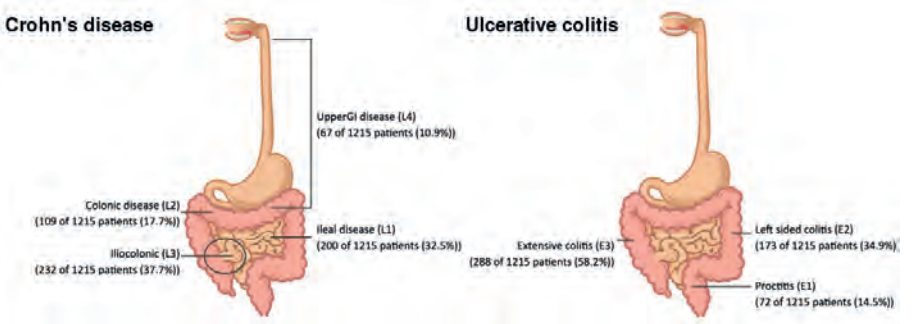


Figure 1. The 1000IBD project. This project consists of data from IBD patients in multiple data modalities[26]

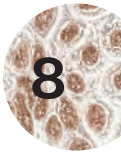
Primary sclerosing cholangitis (PSC) and IBD

Our understanding of the roles played by various gut cells in the pathophysiology of IBD, and also other gastrointestinal diseases, has expanded tremendously with the arrival of scRNAseq technologies (**chapter 5**). However, much less knowledge is available for the clinical cases when a patient suffers from two concomitant diseases, such as primary sclerosing cholangitis (PSC) and IBD. One of the theories behind PSC-IBD pathogenesis is that it arises through a defect in the blood-gut barrier. This defect would enable microbial trafficking from a leaky gut, via the portal venous system, to the bile ducts. There, it is thought to cause inflammatory reactions and subsequent sclerosis in the liver[27]. The genes that are known to be PSC risk genes are largely located in the human leukocyte antigen (HLA) region, indicating defective innate immune reactions. PSC-IBD, although not subject to large-scale global research, is an intriguing niche to study.

By using scRNAseq of gut mucosa from PSC-IBD and UC patients in **chapter 7**, we aimed to identify both distinct and overlapping patterns between PSC-IBD and UC. We show that the cell type composition is similar for PSC-IBD and UC, but that the function of those cell types of the same sort is different between the IBDs. The plasma cells in PSC show an upregulation of stress pathways upon inflammation, indicating that plasma cell stress or exhaustion may play a role in the disease. Supporting the 'leaky gut hypothesis', we see a downregulation of Claudin-3 in various epithelial cell types in PSC inflamed tissue, whereas only immature enterocytes downregulate this gene in UC inflammation. The gene may be an interesting target for treating PSC-IBD. To confirm this hypothesis, we need to study both liver and gut from PSC patients, preferably in a multi-organ disease model. In the near future, we will assess if we can explain differences in mucosal composition between inflamed and non-inflamed tissues using cell-cell interaction modeling. Furthermore, we will try to isolate the *DUOX2* enterocyte to analyze its functions. Strengths of this study are the use of paired samples to limit noise by genetic and environmental background and the fact that it has a relatively large sample size for the study of such rare disease phenotype as PSC-IBD. A weak point is, that paired inflamed and non-inflamed samples are obtained naturally from different parts of the colon, which are known to differ in cell type constitution and gene expression.

Drug repurposing: a chance for a fast-track IBD therapy

Drugs developed for diseases other than IBD may be subjected to drug repurposing[28]. As many auto-immune diseases share at least part of their pathophysiologic mechanisms (e.g. continuously overactivated state of immune cells), treatments also overlap: biologicals are used for treatment of rheumatoid arthritis as well as IBD. For example, methotrexate and adalimumab have a shared indication for IBD, rheumatoid arthritis and psoriasis. In contrast to the development of new therapeutics for IBD, drug repurposing is an economical and fast way to introduce new drugs to IBD patients who do not respond to the current treatment options. In **chapter 4 and 7**, we predicted currently prescribed medication, which is not in use (yet) for IBD, to target key cells in IBD pathology. For example, the PDE4D-inhibitor apremilast may target Th17 mucosal cells. A recent study followed up on this lead, and demonstrated that a PDE-inhibitor improves colitis in the clinic[29].



REFERENCES

1. Jeuring SFG, Biemans VBC, Heuvel TRA van den, et al. Corticosteroid sparing in inflammatory bowel disease is more often achieved in the immunomodulator and biological era—results from the Dutch population-based IBDSL cohort. Am J Gastroenterol 2018;113:384–395.
2. Lowe SC, Sauk JS, Limketkai BN, et al. Declining Rates of Surgery for Inflammatory Bowel Disease in the Era of Biologic Therapy. J Gastrointest Surg 2021;25:211–219.
3. Jeuring SFG, Heuvel TRA Van Den, Liu LYL, et al. Improvements in the Long-Term Outcome of Crohn's Disease over the Past Two Decades and the Relation to Changes in Medical Management: Results from the Population-Based IBDSL Cohort. Am J Gastroenterol 2017;112:325–336.
4. Chen H, Ye F, Guo G. Revolutionizing immunology with single-cell RNA sequencing. Cell Mol Immunol 2019;16:242–249.
5. Atlasy N, Bujko A, Brazda PB, et al. Single cell transcriptome atlas of immune cells in human small intestine and in celiac disease. bioRxiv 2019:721258.
6. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell 2019;178:714-730.e22.
7. Parikh K, Antanaviciute A, Fawkner-Corbett D, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature 2019;567:49–55.
8. Kinchen J, Chen HH, Parikh K, et al. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. Cell 2018;175:372-386.e17.
9. Picelli S, Björklund ÅK, Faridani OR, et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 2013;10:1096–1100.
10. Hashimshony T, Wagner F, Sher N, et al. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. Cell Rep 2012;2:666–673.
11. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 2015;161:1202–1214.
12. Efremova M, Vento-Tormo M, Teichmann SA, et al. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc 2020;15:1484–1506.
13. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun 2021;12:1088.
14. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 2017;14:865–868.
15. Heaton H, Talman AM, Knights A, et al. souporcell: Robust clustering of single cell RNAseq by genotype and ambient RNA inference without reference genotypes. bioRxiv 2019:699637.
16. Pouwels SD, Wiersma VR, Fokkema IE, et al. Acute cigarette smoke-induced eQTL affects formyl peptide receptor expression and lung function. Respirology 2020.
17. Hu S, Vich Vila A, Gacesa R, et al. Whole exome sequencing analyses reveal gene-microbiota interactions in the context of IBD. Gut 2021;70:285–296.
18. James KR, Gomes T, Elmentaite R, et al. Distinct microbial and immune niches of the human colon. Nat Immunol 2020;21:343–353.
19. Yu S, Balasubramanian I, Laubitz D, et al. Paneth Cell-Derived Lysozyme Defines the Composition of Mucolytic Microbiota and the Inflammatory Tone of the Intestine. Immunity 2020;53:398-416.e8.
20. Khan S, Waliullah S, Godfrey V, et al. Dietary simple sugars alter microbial ecology in the gut and promote colitis in mice. Sci Transl Med 2020;12.
21. Nielsen OH, Hansen TI, Gubatan JM, et al. Managing Vitamin D deficiency in inflammatory bowel disease. Frontline Gastroenterol 2019;10:394–400.

22. Anon. Home | 3TR. Available at: https://3tr-imi.eu/ [Accessed February 11, 2021].

23. Anon. ImmUniverse | IMI Innovative Medicines Initiative. Available at: https://www.imi.europa.eu/projects-results/project-factsheets/immuniverse [Accessed February 11, 2021].

24. eQTLGen. Available at: https://www.eqtlgen.org/ [Accessed February 11, 2021].

25. Anon. NIH Human Microbiome Project - Home. Available at: https://www.hmpdacc.org/ [Accessed February 11, 2021].

26. Souza HSP De, Fiocchi C, Iliopoulos D. The IBD interactome: An integrated view of aetiology, pathogenesis and therapy. Nat Rev Gastroenterol Hepatol 2017;14:739–749.

27. Dean G, Hanauer S, Levitsky J. The Role of the Intestine in the Pathogenesis of Primary Sclerosing Cholangitis: Evidence and Therapeutic Implications. Hepatology 2020;72:1127–1138.

28. Collij V, Festen EAM, Alberts R, et al. Drug Repositioning in Inflammatory Bowel Disease Based on Genetic Information. Inflamm Bowel Dis 2016;22:2562–2570.

29. Huang B, Chen Z, Geng L, et al. Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. Cell 2019;179:1160-1176.e24.