THE UNIVERSITY OF CHICAGO

THE CONTRIBUTION OF THE APPENDICEAL ADAPTIVE IMMUNE COMPARTMENT DURING INFLAMMATION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON IMMUNOLOGY

BY

EMMA C. STEWART

CHICAGO, ILLINOIS

MARCH 2025

ABSTRACT

The appendix is a secondary lymphoid organ that sits at the junction between the small and large intestine. Housing both dense lymphoid tissue and its own microbiome, it is uniquely poised to play a role in gut immune homeostasis, yet it is largely ignored. The purpose of this work was to investigate the appendiceal immune compartment across states of inflammation and to begin to interrogate the antigen reactivity of lymphocytes residing in the appendix that may contribute to various pathologies. The first portion of this thesis profiles lymphocyte phenotypes and receptor repertoires in appendicitis to try and understand if they contribute to appendicitis. Through spatial transcriptomics and single-cell sequencing, IgG-secreting plasma cells were found to be increasingly activated in appendicitis compared to controls and had greater mutational burden than other plasma cell subsets. These plasma cells are likely generated through a T cell-dependent process, via CD4+ T follicular helper cells. This study proposes the dysbiotic appendiceal microbiome during appendicitis as a source of antigen and used metagenomic sequencing and computational prediction methods to identify possible MHC II restricted bacterial peptides that could be recognized by CD4+ T cells. The remaining results described in this thesis focused on the relationship between the appendix and inflammatory bowel diseases. There is a long-standing epidemiological observation that appendectomy yields a reduced risk of ulcerative colitis, although this relationship does not exist for other intestinal pathologies such as Crohn's disease or colorectal cancer. Due to the known contribution of T cells in colitis and prior observations of T cell activation in the appendix of inflammatory bowel disease patients, the main hypothesis of this study was that the appendix was a site containing unique colitogenic T cells that were able to contribute to disease pathology in the colon. While T cells largely lacked evidence of antigenic drive in appendicitis, CD8+ T cells had marked clonal

expansion in the appendix of ulcerative colitis patients, with a subset of these cells predicted to be reactive to viral antigen. These viral-reactive CD8+ T cells in ulcerative colitis increasingly took on a terminal effector phenotype suggestive of prior antigenic stimulation in blood compared to healthy controls and were enriched in *GZMK*-expressing scRNA-seq clusters. Immunofluorescence staining showed granzyme K positive CD8+ T cells were increased in the colon of ulcerative colitis patients with intact appendices compared to patients who had undergone prior appendectomy, and epithelial cells from intact patients showed increased inflammatory gene expression by spatial transcriptomics. Overall, this project provides an in-depth report of the appendiceal lymphocyte populations, with a particular focus on T cell phenotypes and antigen reactivity, and beings to provide some mechanistic insight behind the observed association between the appendix and ulcerative colitis development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, Roshni Roy Chowdhury, for her support during my time in her lab, and her encouragement to pursue a unique thesis project investigating the immune cell compartment of the appendix. I would also like to thank Carmen, our senior technician in the lab who always lent a helping hand in lab especially during long tissue collection days, acted as an impromptu lab manager, and was always around for a gossip session or a coffee run.

I would also like to thank my committee members Eric Pamer, Marcus Clark, Marisa Alegre, and Samantha Riesenfeld for their insightful feedback during committee meetings which helped to steer the direction of this project. I would also like to thank the larger COI community and my cohort. I'd also like to thank the friends I've made in graduate school, especially Wiola, Joash, and Sophia. It's been awesome having some other people going through this process alongside me, and I can't wait to see where we all end up in the future!

Finally, I would like to thank my friends and family. To my parents, without your love and support I never would have gotten to this point. To my sister Megan, and my aunt and uncle who are like a second set of parents to me. To Kara, we've been able to maintain a friendship since first grade and I've so appreciated our weekly phone calls when we are across the country from one another, and I can't wait to celebrate on our friendship anniversary trip! To my college friends Ellie, Anna, Cassie, Sadie and a special shout out to Amy, Cindy and Nicole who flew out for my defense, I'm so grateful you were able to come out here in person. It's been quite a ride getting through grad school, and I am so thankful to each and every one of you for supporting me through it. I'm so excited to see where the next chapter brings for all of us.

CHAPTER 1: INTRODUCTION

1.1.The appendix

The appendix is an organ approximately 6-10 centimeters in length that sits at the junction between the small and large intestine. Looking at a tissue cross section, the layers consist of an outer serosa, followed by the muscularis propria, submucosa, and mucosa prior to reaching the lumen[1]. The appendix is a secondary lymphoid organ (SLO), and as such contains both diffuse lymphocytes as well as lymphoid follicles distributed throughout the mucosal and submucosal tissue. Relative to the colon, the appendix has a greater density of follicles, and increased plasma cell density[1]. Follicular structures in the appendix are comparable to Peyer's patches in the small intestine, with the exception of the "mixed-cell zone" that resides between the epithelium and mantle zone containing a mixture of macrophages, B, and T cells[1]. The appendix also contains M cells which function to translocate luminal antigens across the epithelium[2]. These cells likely help initiate immune responses to the native commensals and other intestinal antigens residing within the appendiceal lumen.

The first description of an appendiceal microbiome was in 2007 when biofilm was described within the lumen[3]. It is now appreciated that the appendix has its own microbiome and is a site of dense biofilm formation. It has been hypothesized that the appendix acts as a microbial reservoir capable of re-seeding the colon in times of dysregulation or infection since the organ is relatively isolated from the fecal stream. In support this theory, patients who underwent incidental appendectomy experienced long-term dysbiosis following bariatric surgery compared to patients who retained their appendices[4]. There is also evidence that prior appendectomy increases the risk for recurrent *Clostridium difficile* infections[5].

The interplay between the immune compartment and resident microbes in the appendix likely contributes to gut homeostasis. Colonization of the appendix happens in early life and coincides with formation of lymphoid follicles, which continue to increase in number through adolescence[6]. The appendix is also a major producer of IgA, which is known to support mucosal immunity. Patients who underwent appendectomy have approximately half the secretory IgA of those with an intact appendix in their serum[7]. IgA deficiency is typically compensated for in humans by IgM and IgG, so no immunodeficiency is expected, although selective IgA deficient patients can exhibit dysbiosis[8,9].

1.2.Appendicitis

1.2.1. Epidemiology and etiology

Appendicitis is one of the most common abdominal emergencies, with an estimated prevalence of 8.7% globally[10]. Appendicitis is more common in childhood through early adulthood, with 28 being the mean age of presentation[11]. Obstruction of the appendiceal lumen via fecalith, tumor, or lymphoid hyperplasia is the primary cause of disease[12,13]. However, fecalith and tumoral obstruction is relatively rare, and the root cause of lymphoid hyperplasia is undetermined, although host genetics, environmental changes, dysbiosis or infection have all been proposed as contributing factors[12]. To date, the precise etiology of appendicitis and the molecular basis of the inflammatory response are not well understood.

1.2.2. Pathophysiology

Luminal obstruction of the appendix leads to inflammation, thrombosis of small vessels, and lymphatic stasis[13]. As appendiceal wall pressure increases, localized ischemia leads to necrosis of the tissue[13]. Continued pressure puts patients at risk of appendiceal perforation and potential peritonitis[11]. Appendicitis can be categorized as uncomplicated or complicated. Uncomplicated appendicitis presents as phlegmonous tissue with neutrophil infiltration that has not perforated[14]. Complicated appendicitis on the other hand, presents as necrotic tissue that has either already perforated or is at risk of perforation[14]. This is clinically relevant, as it has been proposed that uncomplicated and complicated appendicitis are distinct physiological processes and may be managed differently.

1.2.3. The immune response in appendicitis

Thus far, descriptions of the immune response in appendicitis have primarily focused on innate immunity in the literature. Transmural neutrophil infiltration is the main pathological feature of acute appendicitis, and cytokines, such as IL-6, IL-8, IL-10, and TNF-α have been associated with disease[15,16]. However, early histological studies found increased infiltration of T cells and plasma cells in the inflamed appendiceal tissue compared to controls, implicating the adaptive response in appendicitis[1]. A recent report also identified a positive association between B cell infiltration and abundance of certain microbial genera in appendicitis[17]. Additionally, studies on genetic risk variants for appendicitis have identified T cell related genes such as *CTLA4*, *CD44*, *IL17*, and *IL13*[18]. Yet, the contribution of B and T cells to appendiceal inflammation remains unclear, both in terms of the key inflammatory subsets present during

appendicitis, as well as possible antigens and the downstream antigen-specific responses taking place.

1.2.4. The appendiceal microbiome in appendicitis

The appendiceal microbiome has been implicated in the pathogenesis of appendicitis. This proposal is reasonable as the appendix houses dense microbial biofilms, and multiple groups have observed a dysbiotic state during appendicitis[3,19–21]. The degree of dysbiosis appears to correlate with disease severity. Oral taxa such as *Fusobacterium* and opportunistic pathogens such as *Parvimonas micra* are more often enriched in complicated appendicitis, whereas uncomplicated cases have a lesser dysbiosis composed of typical gut commensals[22–24]. The APPAC randomized clinical trial, which demonstrated that antibiotic therapy alone can be effective in appendicitis treatment, further implicates the microbiome in disease etiology[25]. However, use of antibiotics alone has been shown to have a 25-40% failure rate in long term follow-up studies, with inflammation recurring up to years after the initial incident[26,27]. This may be due to differences in antibiotic-responsiveness between uncomplicated and complicated appendicitis, as a recent review found diagnostic imaging suggestive of complicated disease indicates an increased risk of antibiotic failure[13].

1.2.5. Current management

Appendicitis is diagnosed by a combination of physical exam, lab testing, and imaging. Common symptoms include peri-umbilical and right-lower quadrant pain, nausea and fever. Patients often present with elevated white blood counts and additional tests can rule out other conditions. Computed tomography (CT) scans are commonly used in suspected appendicitis for

its high sensitivity and specificity in diagnosis[13]. Laparoscopic appendectomy remains the gold-standard treatment along with a single dose of preoperative broad spectrum antibiotics, although antibiotic therapy alone can be adequate in some patients as discussed above[13]. Treatment decision for surgery versus an antibiotics first approach is determined based on clinical findings and patient preference.

1.3. Inflammatory bowel disease

1.3.1. Epidemiology

Inflammatory bowel disease (IBD) is a significant burden globally, with the highest prevalence in North America and Western Europe[28,29]. In 2017 there were an estimated 6.8 million cases of IBD worldwide[28]. The U.S. accounts for approximately 2-3 million cases (0.7-1.3% of the adult population) depending on the report[28,30–32]. While western countries currently shoulder the largest disease burden, cases are on the rise in many countries across South America, Africa and Asia[28,29]. Diagnosis typically occurs in early adulthood, with incidence reaching a peak in the third decade of life and leveling off afterwards[30]. This results in patients living with a progressive chronic disease for the majority of their adult life. Disease occurrence is highest in non-Hispanic white individuals, although any ethnic background can be affected[28,30]. Reports of sex-dependent differences in the literature are conflicting, ranging from slight female preference to no difference depending on the study[30,33,34].

There are two major subsets of IBD: ulcerative colitis (UC) and Crohn's disease (CD). Prevalence of UC tends to be higher in epidemiological studies that analyze the subsets separately[29,30]. The diseases are similar in that they are both chronic idiopathic conditions

resulting in inflammation of the gut. Both diseases are thought to have multifactorial etiologies and are managed with similar medication regimens. However, they differ in presentation of inflammation and disease complications.

UC and CD have distinct inflammatory patterns within the bowels. In CD, inflammation can occur anywhere along the GI tract and is discontinuous, interspersed by normal-appearing sections of mucosa[33,35]. Inflammation is also transmural, with repeated flares causing deep tissue damage resulting in strictures or fistulas requiring surgical intervention[33,35]. UC, on the other hand, is restricted to the colon. Inflammation is limited to the mucosal layer, typically beginning in the rectum and moving up and along the large intestine as disease progresses[34,36]. In severe cases UC patients also require surgical intervention, albeit at lower rates than CD patients. Both IBD subsets are at an increased risk for colorectal cancer (CRC) due to the effects of long-term chronic inflammation in the intestinal tissue[37]. However, over time cancer rates have declined possibly due to increased CRC screening and better disease management leading to tissue healing in this patient population[37].

1.3.2. Etiology

While there is no single cause for development of IBD, multiple factors can increase risk including the gut microbiome, environment, genetics, and immune dysfunction. Gut dysbiosis is described in both CD and UC and is possibly associated with dietary changes as people moved away from high-fiber low fat diets to more processed foods[35,36]. This relationship is disputed because diets are known to have transient effects on the gut microbiome, and also because it is not known whether dysbiosis is a cause or effect of IBD. However, a move towards processed foods typically coincides with industrialization, which in turn coincides with the epidemiology of

IBD incidence. Childhood antibiotic use is also associated with increased risk for both CD and UC, further implicating the gut microbiome and immune regulation[35,36]. Smoking has been shown to have a significant positive association with CD development, and surprisingly, quitting smoking is a risk factor for UC, although a mechanism to explain these observations is not known[35,36].

There is a degree of heritability in IBD as assessed by twin studies and genome-wide association studies (GWAS). A study with 512 pairs of twins from the German IBD twin registry found increased concordance rates, meaning both twins were diagnosed with IBD, in monozygotic versus dizygotic twins (UC: 17% vs. 2.4%, CD: 34.6% vs. 4%)[38]. Additional twin studies have also found a higher degree of concordance in CD up to 50%, compared to UC[39–41]. Of the discordant twins, those diagnosed with IBD were more likely to have history of GI infections and antibiotic use prior to diagnosis[38,41]. Family history impacts both CD and UC, with having one or more first-degree relatives affected being the greatest known risk factor for development of disease[42]. Over 200 genetic loci have been associated with IBD following identification of *NOD2* in 2001 as a susceptibility gene for CD[43–45]. Many of the identified genes play a role in immune responses and also show positive association with other inflammatory diseases. Genetic studies have identified loci shared between UC and CD (*IL23R, IL12B, NKX2-3, MST1, CARD9, PRDM1, PTPN22*) as well as loci unique to CD (*NOD2, ATG16L1, IRGM*) and UC (*IL10R, RNF186, HLA*) whose mutations are associated with disease risk[43]. In general, these genes play a role in pathways regulating inflammatory responses, autophagy, microbial sensing and barrier function, all of which are known to contribute to IBD pathology. These genetic studies support a role for the interplay between luminal microbes and the immune response in IBD pathology. However, identified risk alleles are estimated to explain

only 13.1% of disease variance in CD, and 8.2% in UC[46]. The limited effect of genetic variance in IBD limits usefulness in the clinic, and as a result genetic sequencing is not used to assist in diagnosis.

1.3.3. Pathophysiology

Despite differences in disease presentation, loss of epithelial barrier integrity in the gut is a hallmark of both UC and CD. In healthy tissue goblet cells produce mucus which provides a protective layer between the host tissue and luminal microbes. Reduced goblet cell numbers, loss of intercellular tight junctions, or lamina propria inflammation leads to the breakdown of this barrier[34,35]. Luminal microbes are then able to breach the epithelium and enter tissue, creating a chronic cycle of inflammation and epithelial breakdown characteristic of IBD. This barrier breakdown is thought to contribute to dysregulation of immune responses downstream.

Damage- and pathogen-associated signals are sensed by the innate immune system, which drive early stages of inflammation. Neutrophils, the first responders, follow cytokine and chemokine gradients from circulation to inflamed gut tissue, where they release proteases and inflammatory cytokines[47]. Inflammatory signals lead to upregulation of MAdCAM-1 on endothelial cells, which in turn promotes entry of circulating immune cells into the tissue via binding of α4β7 integrin[34]. Interrupting entry of immune cells into the colon can have significant impact on IBD pathology, as evidenced by the success of vedolizumab therapy[48,49]. Pro-inflammatory macrophages secreting TNF-α, IL-6, and IL-1β are increased in IBD intestinal tissue, as are colitogenic CD103+ dendritic cells[34,36,47]. This leads to a shift away from regulatory T cell generation and increased effector T cells, which will be discussed in more detail in Section 1.5.

The role of the humoral response in IBD remains an open question, and it is unclear whether changes in this compartment are protective, pathogenic, or a result of bystander activation via T cell responses[50]. Serum antibodies binding self or bacterial antigens have been reported in IBD, but likely represent a bystander effect of more severe disease[50]. One report found increased IgG-secreting plasma cells in the colon of UC patients with reduced diversity compared to control samples, and proposed that these cells were reactive to self and microbial antigens[51]. IgA-coated bacteria isolated from IBD patients were capable of inducing colitis when transferred into germ-free mice, suggesting humoral responses may play a protective role in anti-microbial responses[52].

1.3.4. Microbial dysbiosis in IBD

The gut microbiome is a critical factor in the development of IBD. The gut microbiome is required for development of colitis in genetically susceptible mice, highlighting its integral role in gut homeostasis and disease[53–55]. In both CD and UC patients there is an observed dysbiosis of gut commensals and reduced species diversity, although it is unknown whether this is a cause or effect of disease[56]. In the human colon, 4 phyla are the most abundant: *Firmicutes, Bacteroidetes, Proteobacteria,* and *Actinobacteria*[57]. There is great inter-individual variation in the gut microbiota, even amongst healthy people, and factors such as diet, lifestyle, host genetics or immune responses can contribute to microbiome changes. One longitudinal study of the IBD microbiome showed increased fluctuations in communities over time compared to healthy controls, with a more extreme phenotype in CD than UC[58]. This is an interesting finding in the context of a recent description of transient versus persistent microbial communities and suggests that IBD patients suffer from transient microbial destabilization more frequently, or fail to

recover in the same manner as healthy donors[59]. Common findings from gut microbiome studies in IBD include reduced diversity, decreased *Firmicutes,* and increased *Proteobacteria*[56,57,60–62]. In UC, multiple studies have identified decreases in known butyrate-producing bacteria such as *Roseburia, Eubacterium rectale* and *Faecalibacterium prausnitzii* and increases in epithelial-adherent strains like *Escherichia coli*[60,61]. UC is also associated with an increase in facultative anaerobes, which are better able to survive in environments under oxidative stress[57,63]. Metagenomic analysis of stool samples collected longitudinally from an IBD cohort identified transient increases of *Ruminococcus gnavus*, a member of the *Firmicutes* phylum, which often correlated with disease activity[63]. Despite correlative findings, no microbial species have been identified as causative in IBD pathology[56]. Administration of probiotics or fecal transplantation has been attempted in IBD with some clinical efficacy, although long-term impact is unknown[61,64,65]. This shows that the role of the microbiome is more complex than changes in community composition alone and supports the existing thought that genetic susceptibility and immune regulation are also critical factors in IBD inflammation.

1.3.5. Current management and unmet need

IBD is typically diagnosed via a combination of patient symptoms, endoscopy, and histopathology of intestinal biopsies. Common UC symptoms include chronic bloody stool and increased bowel movement frequency, and diseased tissue can be characterized by presence of erythema, edema, mucosal friability or ulceration depending on severity[34,36]. By histology, changes in colonic crypt morphology, immune cell infiltration and epithelial metaplasia are often observed within the mucosal tissue[36]. With CD the most common symptoms are diarrhea and abdominal pain, and over 50% of patients describe extraintestinal symptoms[33,35]. Unlike UC, CD

patients can develop fistulas and/or strictures as a result of their disease requiring surgical intervention[35]. Transmural inflammation is the hallmark of CD by histology, as is presence of epithelioid granulomas, although these are not observed in all patient specimens[33]. Both UC and CD are inflammatory diseases characterized by periods of symptom remission and recurrence, with disease progressing over time leading to intestinal damage. The gold-standard treatment outcome for both UC and CD is endoscopic healing.

	Approved for:		
Mechanism / medication class	***UC***	***CD***	***Both***
5-aminosalycyclic-acid derivatives	Mesalamine Sulfasalazine Balsalazide Olsalazine		
Thiopurines / Immunomodulators	Cyclosporine	Methotrexate	Azathioprine Mercaptopurine Tacrolimus
Anti-α4β7 integrin			Vedolizumab
Anti-α4 (also inhibits α4β1)		Natalizumab	
Anti-TNF-α	Golimumab	Certolizumab	Adalimumab Infliximab
Anti-IL-23	Mirikizumab		Risankizumab
Anti-IL-12/IL-23			Ustekinumab
JAK inhibitors	Tofacitinib		Upadacitinib
Glucocorticoids			Budesonide Methylprednisolone Prednisolone Prednisone
S1PR modulators	Etrasimod Ozanimod		
Antibiotics		Ciprofloxacin Metronidazole	

Table 1.1. Summary of FDA-approved medications for IBD

Early cases of IBD had mortality rates greater than 50% due to lack of effective treatments. Today patients can expect to have a normal life expectancy thanks to a vastly different treatment landscape. There is significant overlap in treatment strategies between CD and UC, with cytokine blockade, inhibition of lymphocyte trafficking and immunomodulation emerging as common themes (**Table 1.1**). Mild to moderate UC can typically be treated with 5-

aminosalycyclic-acid derivatives and corticosteroids, while moderate to severe patients are often maintained on thiopurines, biologics (e.g. anti-TNF) or targeted small molecules (e.g. JAK, S1PR inhibitors) [34,36]. Despite availability of multiple treatment options, efficacy remains limited, with clinical remission rates in the range of 30-40% for many of these agents[66–69]. Patients with severe disease or who fail multiple drug classes will ultimately require surgical intervention. While surgical rates have decreased with the introduction of biologics, an estimated 10-30% of UC patients, and 50-80% of CD patients will require a bowel resection over their lifetime[70]. Medication failure and surgical rates in IBD indicate a continued unmet need and justifies further research into factors contributing to disease pathology.

1.4. IBD and the appendix

1.4.1. Appendectomy protects against colitis development

Epidemiological studies have consistently found that appendectomy, especially early in life, results in reduced likelihood of a UC diagnosis later on[71,72]. This is not the case for CD, where appendectomy may lead to an increased risk of a future CD diagnosis but a less severe disease course[73,74]. Differential effect of appendectomy in IBD may be due to disease presentation. While UC typically presents as continuous distal inflammation beginning in the rectum, CD can occur anywhere along the GI tract, most commonly in the terminal ileum. It is possible the appendix is an early site of CD-related inflammation in some patients and misdiagnosed as appendicitis. In this instance the patient will be later diagnosed correctly with CD as disease manifests elsewhere. However, in UC, the appendix may act as a priming site for immune cells which traffic to the colon and drive inflammation.

Mouse models have validated the relationship between the appendix and colitis development. Cecal patch removal, the murine equivalent of appendectomy, resulted in decreased colitis incidence in two spontaneous colitis models[75,76]. Prior appendectomy also led to reduced disease severity in an induced model of colitis[77]. In the CD62L+CD4+ transfer model of colitis, labelled cells preferentially migrated to the appendix of mice, supporting the theory of this tissue as a priming site for colitogenic immune cells[78]. Interestingly, appendectomy appears to reduce immunosurveillance into the colon, resulting in increased cancer rates in mice[76,79]. Although, it is inconclusive as to whether this effect extends to humans[80–82].

1.4.2. The appendix as a skip lesion

The appendiceal orifice has long been observed to be a "skip lesion" in UC, meaning there is inflammation present that is discontinuous with diseased colon tissue. First reported as a case study in 1974, many clinicians have since observed peri-appendiceal inflammation in UC patients[83–86]. Inflammation is present in both pan-colitis and left sided colitis cases at greater rates compared to CRC cases, and both CD and UC cases show evidence of inflammation, supporting the appendix as a potential disease site in CD[84]. Notably, histological features in the inflamed appendices were consistent with UC, and distinct from appendicitis[84,85]. Appendiceal inflammation preceded UC development in a subset of cases in a 20-month follow-up period[87]. This observation further supports the appendix as a disease priming site in UC.

1.4.3. Appendectomy may ameliorate disease in established colitis

The impact of appendectomy on disease course in UC patients has also been studied. A meta-analysis failed to identify an association between appendectomy and colectomy risk,

although there appears to be variation depending on whether appendectomy occurs before or after diagnosis[88,89]. However, there is some evidence that appendectomy leads to slower disease course, with patients being diagnosed at later ages on average[89–91]. Appendectomy has also been investigated as a treatment alternative to colectomy. Case reports have found appendectomy leads to long-term remission and mucosal healing[92,93]. A prospective case series of 30 patients reported a 40% remission rate with patients discontinuing all therapy[94]. The PASSION study, a prospective cohort of medication refractory patients, found appendectomy led to significant improvements in severity scores at short (3 and 12 months) and long term follow-up (median 3.7 years)[95,96]. While not all patients benefitted, these small studies provide evidence that a subset of UC patients exist where appendectomy may be an appropriate treatment course and provide rationale for larger clinical studies. The recently completed ACCURE and COSTA trials should provide greater insight into efficacy of appendectomy in treating UC[97–99].

1.5. T cells in IBD

1.5.1. Colitis is T cell dependent

Murine models have demonstrated both CD4+ and CD8+ T cells can induce colitis. Transfer of naïve $CD45RB^{high}$ CD4+ T cells into lymphopenic ($Rag1^{-/-}$) hosts is commonly used to model colitis pathology[100]. Notably, this model relies on exclusion of $CD45RB^{low}$ CD4+ T regulatory (Treg) cells, and transfer of CD25+CD4+ Tregs can ameliorate colitis[101]. In a reversible T cell transfer model, CD4 T depletion led to resolution of inflammation within days while administration of anti-TNF therapy had a minimal effect[102]. While CD8+ T cell

dependency in colitis is less well established, induction of colitis using DNBS was shown to be CD8+ T dependent, with CD8+ T depletion inhibiting inflammation[103].

1.5.2. Increased T cell infiltration and activation in the IBD appendix

While data looking at immune cell populations within the appendix is relatively limited, there is evidence of increased T cell infiltration and activation within this tissue in UC patients. . Clinical studies of patients undergoing appendectomy to treat UC have also reported increased CD4+ T infiltration[92,95]. CD4+ T infiltration overall was increased in the appendix of patients with active colitis, as was the proportion of CD69+ CD4+ T cells[104]. Appendiceal T cell activation is also a feature of murine colitis models. Naïve CD62L+CD4+ T cells transferred to induce colitis not only preferentially entered the appendix, but also had increased CD154 expression compared to cells in the colon[78]. Looking at the site of disease, the colon, CD3+ T cells overall and CD8+ T cells were reduced following appendectomy in both mice and humans in a colitis-associated cancer setting, leading to poor immunosurveillance and increased tumor burden[79]. These observations support that T cells can be activated in the appendix and may migrate to the colon where they contribute to inflammation during colitis. While trafficking of lymphocytes between the appendix and colon is not well described in humans, studies in mice discovered that IgA+ B cells produced in the Peyer's patches preferentially traffic to the small intestine, while cells produced in the cecal patch are able to migrate to the colon[105]. Trafficking molecules such as integrin α4β7, CCR9, and CCR10 are known to mediate gut homing in mice and humans, and GPR15 expression on T cells has been implicated in increased homing to the colon in colitis[106–109].

1.5.3. CD4+ T cells in IBD

Alterations in the CD4+ T compartment during IBD leads to imbalances in regulatory versus effector subsets[36]. Antigen presenting cells (APCs) secreting IL-12, IL-18, and IL-23 activate Th1 and Th17 cells, two major pathogenic subsets in IBD[35]. At homeostasis, Th17 cells function independently of IL-23, secreting cytokines that promote barrier function such as IL-17 and IL-22[50]. Thus, while IL-17 levels have been shown to be increased in the intestines of IBD patients, use of anti-IL-17 therapy exacerbates existing disease, and in some cases has caused new-onset IBD in rheumatoid and psoriatic arthritis patients[110,111]. However, IL-23 induces a pathogenic phenotype in Th17 cells, leading to production of GM-CSF and IFNγ which promotes intestinal inflammation[50]. Targeting the shared IL-12/IL23 p40 subunit and IL-23 p19 subunit, in contrast to anti-IL-17, has had significant success in the clinic, highlighting the key role of this pathway in IBD pathology[35,67]. Traditionally, CD was thought to be a Th1 mediated disease, while UC was thought to have a Th2 component due to increased IL-13 expression in the colon. However, anti-IL-13 therapy failed to provide a benefit in a phase II clinical trial, and the contribution of these cells in disease pathology remains inconclusive. T regulatory cells (Tregs) play a crucial role in inhibiting intestinal inflammation via secretion of IL-10 and TGF-β, preferential binding of IL-2, and suppression of APC co-stimulation via inhibitory receptors such as CTLA-4[50,112]. Overall Treg numbers in the intestines are often reduced in IBD, which may be due to reduced infiltration of Tregs, or extrinsic factors inhibiting peripheral Treg induction, as intrinsic defaults in Treg populations have not been identified. One such extrinsic mechanism is Smad7 expression in effector CD4+ T cells which confers resistance to Treg suppression[113]. Additionally CD103+ dendritic cells, typically critical for Treg differentiation, are altered during intestinal inflammation, favoring induction of IFN-γ effector CD4+ T cells[114].

1.5.4. CD8+ T cells in IBD

Colitis in known to be a T-cell driven disease with a growing role for CD8+ T cells being described in the literature. Mouse models have demonstrated a role for IFN-γ producing antigen-specific CD8+ T cells in intestinal inflammation[103,115]. Additionally, CD8+ T cells from IBD patients upregulate transcriptional profiles related to antigen-dependent activation and exhaustion which correlates with clinical outcomes, further suggesting these cells contribute to disease pathology[116,117]. Single-cell profiling in UC patients has described increased *TNF* and *IL-17* expressing Tc17 cells in inflamed tissue[118]. Both pediatric and adult cohorts have identified *EOMES* and *GZMK* expressing subsets, and Eomes overexpression in CD8+ T cells was shown to exacerbate colitis in a DSS murine model[119–121]. Eomes is a master transcriptional regulator of effector and memory differentiation in CD8+ T cells, and is known to promote a cytotoxic effector profile leading to terminally differentiated cells[122–127]. While this function has been reported in established models of chronic antigen exposure such as anti-tumor responses and chronic viral infections, it is less well described in chronic inflammatory conditions such as UC.

1.5.5. Antigen specificity of T cells in IBD

A successful T cell response relies on both antigen specificity via T cell receptor-MHC contact and functional differentiation facilitated through environmental cues. The intestinal microbiome is a top candidate as a disease-relevant antigen source in IBD due to presence of dense luminal microbes and loss of barrier integrity at the site of disease. Crucially, the microbiome is known to impact T cell differentiation and function, at least in mice. Perhaps the most well-known example is segmented filamentous bacteria (SFB), which induces Th17 cells in the small intestine in an antigen-specific manner[128,129]. Additional studies employing mono-

colonization of germ-free mice have found *Bifidobacterium adoloscentis* and adherent strains of *Escherichia coli* induce Th17 cells as well[130–132]. *Bacteroides fragilis* colonization induces IL-10+ Tregs, and a conserved *Bacteroidetes* antigen is sufficient to induce anti-inflammatory CD4+ intra-epithelial lymphocytes in mice[133,134]. Certain clusters of *Clostridium* are able to induce Tregs, and oral administration of these species into adult mice attenuated disease in an inducible colitis model[135,136]. Mono-colonization experiments into germ-free mice found several microbial species were able to induce colonic RORγ^{+}Helios^{-} Tregs[137]. While induction of CD8+ T cells by the microbiome is less well studied, an 11-species commensal consortium induced IFN-γ production in CD8+ T cells. leading to increased control of disease in both a *Listeria* infection and tumor model[138].

Several microbial species have shown context-dependent effects of T cell differentiation. *Helicobacter* has been shown to induce Tregs under homeostatic conditions, but induces Th17 cells in murine colitis models[139,140]. In a gnotobiotic setting, *Akkermansia muciniphila* specific CD4+ T cells become follicular helper cells, while in conventionally housed mice, these T cells take on a variety of fates[141]. Interestingly, a screen of microbiota-reactive CD4+ T cells found there was significant cross-reactivity across a strain consortium. The group identified a conserved *Firmicutes* antigen which functions as part of an ATP-binding cassette transport system[142]. This finding highlights that while common microbial antigens may be present in both health and disease, the context they are recognized in is likely crucial to T cell function.

Knowledge of microbiota-reactive CD4+ T cell function in humans is relatively limited and relies on *in vitro* stimulation of TCR clones or primary cells isolated from blood or gut tissue. The first of such reports comes from Duchmann and colleagues, where they found TCR clones isolated from the inflamed lamina propria of IBD patients which proliferated following

co-culture with bacterial lysates in an MHC-II dependent manner[143]. Additional publications from the same group show that bacteria-reactive T clones were more abundant in inflamed intestine as compared to uninflamed intestine or peripheral blood, and that T clones which proliferate in the presence of *Bifidobacterium* and *Bacteroides* often showed cross-reactivity to enterobacteria[144,145]. In a more recent report, CD4+ T cells isolated from the blood of Crohn's patients showed an increased proportion of IL-17+ and IL-17+/IFNy+ expressing subsets after exposure to bacterial antigens[146]. Human CD4+ T PBMCs and gut lamina propria cells showed reactivity to bacterial lysates *in vitro* in both healthy and UC individuals, but reactivity was increased in UC in the gut, and CD4+ T cells showed increased IL-17 and IFNy production following stimulation as compared to healthy controls[147].

1.6. Summary of project rationale

The objective of this thesis is to (1) define the immune compartment of the human appendix during different inflammatory states, including appendicitis, CRC, and IBD in order to understand how lymphocyte phenotypes and repertoires are altered across disease states. The second goal of this project is to (2) investigate the role of appendiceal immune compartment in driving IBD pathology, in particular UC pathology, due to the observed association between appendectomy and reduced UC risk. Due to the known role of T cells in IBD pathology, the analysis was focused on the T cell profiles and receptor repertoires within patient samples. The working hypothesis for this project is that appendiceal inflammation present in UC is distinct from other conditions, including appendicitis, and that as a SLO the appendix acts as a unique priming site for pro-colitogenic T cells which contribute to disease in the colon.

CHAPTER 2: MATERIALS & METHODS

2.1. Experimental methods

2.1.1 Collection of human appendix, colon, and lymph node tissue

Potential tissue donors were identified daily by UChicago Medicine Anatomic Pathology in accordance with IRB #21-1241. Appendiceal tissue collected from appendicitis patients had severity graded as either grossly perforated or unperforated at the time of collection. Patients undergoing incidental appendectomy, primarily due to colorectal cancer or diverticular disease, were used as the control group in these studies. For inclusion appendix tissue was found to be uninvolved in disease processes by pathology. Appendiceal, colon, and lymph node tissue was collected from IBD patients undergoing surgery primary to their IBD diagnosis when available. Appendix and colon tissue was stored in PBS (Corning, #21-040-CV) and transported on ice to lab for downstream processing. Lymph nodes were transported on ice in RPMI-1640 (Cytiva, #SH30027.1).

2.1.2. Immune cell isolation from the appendix

Upon receipt in lab, appendiceal cross-sections were stored in MACS Tissue Storage Solution (Miltenyi Biotec, #130-100-008) at 4°C until processing began. Tissue pieces were then placed in a petri dish containing RPMI-1640 supplemented with 10% FBS (Corning, #MT35015CV) and 10ug/mL pan-caspase inhibitor (InvivoGen, #tlrl-vad). Appendiceal sections were cut open along the orifice to reveal the lumen, and the outer serosal layer was peeled away using forceps and discarded. The mucosal layer was gently peeled or scraped into the media

using forceps. Removal of the mucosal layer from the tissue was confirmed using a stereo microscope. The dish containing mucosal cells was then stored on ice while the remaining submucosal tissue was placed in a conical tube containing pre-warmed RPMI-1640 media supplemented with digestion enzymes from the human Tumor Dissociation Kit (Miltenyi Biotec, #130-095-929) according to the manufacturer's protocol. The tissue was chopped into small pieces with dissection scissors and incubated at 37°C in a water bath with shaking. Following enzymatic digestion, the reaction was quenched using RPMI-1640 with EDTA at a final concentration of 5mM. Both mucosal and submucosal cell suspensions were passed through 100µm and 70µm filters sequentially and washed with fresh media for cell counting. After counting, cells were washed once in enrichment buffer composed of Milli-Q water containing 1X BSA (Miltenyi Biotec #130-091-376) and 1X Annexin V Binding Buffer (BD Biosciences, #556454). Cells were resuspended at 100uL per 10 million cells with staining cocktail containing 1:100 APC anti-CD235a (clone HI264, BioLegend, #349114), 1:200 APC anti-EPCAM (clone 9C4, BD Pharmingen, #566842), and 1:100 APC Annexin V (Invitrogen, #BMS306APC-100) in enrichment buffer. After a 20-minute incubation on ice, cells were washed and resuspended in 1:10 dilution of anti-APC Microbeads (Miltenyi Biotec, #130-090-855) in enrichment buffer for another 20 minutes on ice. Finally, cells were washed and passed through an LS column (Miltenyi Biotec, #130-042-401) on the QuadroMACS Separator (Miltenyi Biotec, #130-090-976) to remove red blood cells, epithelial cells, and dying cells in order to enrich for live CD45+ immune cells.

2.1.3. Immune cell isolation from full-thickness colon

Processing of full-thickness colon tissue was adapted from Jørgensen and colleagues to isolate mucosal immune cells[148]. Colon tissue was placed mucosal-layer down in a petri dish containing RPMI-1640 supplemented with 10% FBS and pan-caspase inhibitor. Fat and the outer serosal and muscle layer were trimmed away and discarded. The remaining tissue was placed into pre-warmed RPMI-1640 containing 4mM DTT (Thermo Fisher, #426380100) and incubated at 37°C in a water bath with shaking for 10-20 minutes. After incubation the tissue was rinsed and placed in a petri dish with fresh media where the submucosa was trimmed away using dissection scissors. Then, the mucosal layer was carefully peeled away using forceps. This layer was placed into pre-warmed media containing enzymes from the human Tumor Dissociation Kit at half concentration. Tissue was chopped into small pieces and placed in 37°C water bath for up to 30 minutes or until single cell suspension was achieved. Cells were double filtered and washed. An RBC lysis step was included by adding 2mL of RBC Lysis for two minutes before quenching with media. After cell counts, cells were enriched in the same manner as appendiceal immune cells.

2.1.4. Immune cell isolation from lymph nodes

Lymph nodes were placed in petri dishes containing RPMI-1640 and perfused with a syringe to release lymphocytes. Perfusion was repeated until the media ran clear, and the resulting cell suspension was washed once prior to counting.

2.1.5. Preparation and sequencing of 10X single-cell libraries

Following immune cell enrichment, cells were washed once more with 0.05% UltraPure BSA (Thermo Fisher Scientific, #AM2616) and counted in preparation for the single-cell workflow. Cells were resuspended to the recommended cell density in the 10X 5' v2 protocol. GEM generation was performed according to the manufacturer's recommendations using the 10X Chromium controller. Single cell libraries were generated using the 5' v2 reagent kits and associated protocols. Library QC was performed using the BioAnalyzer High Sensitivity DNA Kit (Agilent, #5067-4626) and run on the 2100 BioAnalyzer Instrument. Samples were pooled and submitted to the UChicago Functional Genomics Core, where they were sequenced using the Illumina NovaSEQ-6000 or NovaSeqX.

2.1.6. DNA extraction and HLA typing

After enrichment approximately 1 million cells were collected for DNA extraction for HLA typing. DNA was extracted using the QIAmp DNA Micro Kit (Qiagen cat. #56304) and associated protocol, with subsequent elutions into 65µL and 35µL buffer AE. Both aliquots were quantified by Nanodrop and stored at -20°C until submitted for HLA typing. DNA extracts containing at least 2µg of material at a 10ng/µL concentration were shipped overnight to Histogenetics (300 Executive Blvd. Ossining, NY 10562). Results of HLA typing for MHC I alleles are reported in Section 2.3.

2.1.7. Tetramer staining of PBMCs

HLA-A*02:01 typed peripheral blood mononuclear cells (PBMCs) from healthy controls were purchased from STEMCELL Technologies (Vancouver, CA) and PBMCs from UC patients

were banked from a previous study and generously gifted by Joel Pekow at UChicago[149]. All samples were transported on dry ice and stored frozen until use. Cells were thawed in a water bath and washed in pre-warmed RPMI-1640 for counting. Cells were stained on ice for 20 minutes with 1:300 Live Dead Aqua (Thermo Fisher Scientific, #L34957), followed by pooled tetramer stains at 3.33µg/mL concentration for each tetramer for 30 minutes on ice, and finally surface stained for 30 minutes on ice (**Table 2.1**). Cells were resuspended in FACS buffer and run on the Cytek Aurora. During analysis one sample from each study group was identified as being an outlier based on tetramer-positive CD8+ T cell proportions by Grubbs' test and was excluded from the analysis due to suspected active viral infection at time of blood draw.

Tetramer Reagents		
Fluorophore	***Reagent***	***Source***
PE	HLA-A*02:01 | Influenza A M1 58-66 | GILGFVFTL	NIH Tetramer Core Facility
PE	HLA-A*02:01 | SARS-CoV-2 N 222-230 | LLLDRLNQL	
PE	HLA-A*02:01 | EBV BMLF1 259-267 | GLCTLVAML	
PE	HLA-A*02:01 | CMV pp65 495-503 | NLVPMVATV	
Surface Antibodies		
Dilution	***Reagent***	***Vendor***
1:200	BUV496 anti-human CD8 (clone RPA-T8, #612942)	BD Biosciences
1:100	BV605 anti-human CD3 (clone UCHT1, #300460)	BioLegend
1:200	Spark NIR 685 anti-human CD4 (clone SK3, #344658)	BioLegend
1:200	PE/Dazzle 594 anti-human CD45RA (clone HI100, #304146)	BioLegend
1:200	AF700 anti-human CD27 (clone 0323, #302814)	BioLegend
1:200	APC-Cy7 anti-human CD19 (clone HIB19, #302218)	BioLegend
1:200	APC-Cy7 anti-human CD16 (clone 3G8, #302018)	BioLegend
1:200	APC-Cy7 anti-human CD14 (clone 63D3, #367108)	BioLegend

Table 2.1. Tetramer reagents and surface antibodies for PBMC experiment

2.1.8. Immunofluorescence staining of FFPE colon sections

Archived formalin-fixed paraffin-embedded (FFPE) colon tissue was identified from UC patients who underwent colectomy with either an intact appendix or prior appendectomy at the time of surgery. Fresh sections were cut by the Human Tissue Resource Center core facility at

the University of Chicago. Severity scoring was performed by a pathologist at UChicago Medicine. To process sections, paraffin was removed from slides with a double wash in xylenes (Fisher Scientific #X3P-1GAL), then subsequent decreasing ethanol (Fisher Scientific, #04-355-233) concentrations into a final water wash. Antigen retrieval was performed using Tris-EDTA buffer (Abcam, #ab93684) at 50°C for 40 minutes. Tissue was permeabilized using 10% DMSO (Sigma #D2438) and 0.1% Triton X-100 (VWR #EM-9400) prior to rinsing in TBS (Fisher Scientific, #BP24711). Tissue sections were circled with wax pen and blocked in 10% donkey serum (Sigma-Aldrich, #s26-100ML) for 1 hour. Primary stains were incubated overnight in a humidified slide box in the fridge with 5% donkey serum, 1:100 mouse anti-pan Cytokeratin (clone AE-1/AE-3, BioLegend, #914204) and 1:100 rabbit anti-Granzyme K (Invitrogen, #PA5-50980). Slides were washed and incubated for one hour at room temperature with previously spun down secondary antibodies AF555 donkey anti-mouse IgG (Invitrogen, #A-31570) and AF593 donkey anti-rabbit IgG (Invitrogen, #A-21207) at 1:500 dilution. Slides were incubated with 300nm DAPI (BioLegend, #422801) for 5 minutes, washed, and mounted with ProLong Gold Antifade Mountant (Invitrogen, #P10144) with a #1.5 coverslip and sealed with clear nail polish. Whole scan slides were collected at 40X magnification on an Olympus VS200 Slideview scanner through the University of Chicago Microscopy Core.

After imaging, slides were bleached with 4.5% H_2O_2 and 25mM NaOH with 1 hour direct light in accordance with the CycIF protocol[150]. The staining protocol was repeated with 20 minutes of Tris-EDTA antigen retrieval at 90°C, staining with 1:100 mouse anti-CD3 (clone OKT3, Invitrogen) and AF555 donkey anti-mouse IgG (Invitrogen, #A-31570), followed by staining with 1:100 AF647 anti-human CD8 (clone C8/144B, BioLegend, #372906). These slides were also imaged at 40X magnification on an Olympus VS200 Slideview scanner.

2.1.9. Spatial transcriptomics using the Nanostring GeoMx platform

For the appendicitis study FFPE tissue blocks were identified for appendicitis samples and age-matched controls who had undergone incidental appendectomy. Slides stained with anti-CD20, anti-CD3 (clone F7.2.38, abcam, #ab17143), AF532 anti-pan-Cytokeratin (clone AE-1/AE-3, Novus Biologicals, #NBP2-33200AF532), and SYTO 13 dye (Invitrogen, #S7575) for region of interest (ROI) selection.

For the IBD study, tissue sections from UC patients with and without appendices were selected for spatial transcriptomics. Sections were stained with anti-pan-Cytokeratin (clone AE-1/AE-3), anti-CD3 (clone F7.2.38, abcam, #ab17143), anti-CD8 (clone C8/144B, BioLegend, #372906) and SYTO 13 dye (Invitrogen, #S7575) to identify ROIs.

For all slides fresh sections were cut by the Human Tissue Resource Center and stored in the fridge prior to use. Downstream processing was completed according to the Nanostring GeoMx workflow and manufacturer's instructions. ROIs were profiled for spatially indexed transcriptomics using the Human Whole Transcriptome Atlas and sequenced in the UChicago Functional Genomics Core using the NovaSeq6000 platform.

2.1.10. Collection of appendiceal swabs and DNA extraction

Brushes (Cook Medical, #CCB-7-240-3-S) were inserted into the appendiceal lumen upon tissue arrival to Anatomic Pathology, spun five times, then placed in a sterile tube. A second brush was inserted, spun, and placed in a sterile tube containing 20% glycerol. Both brush samples were immediately placed on dry ice and stored at -80°C. Brush samples were transferred to the Microbiome Metagenomics Facility (MMF) at UChicago for DNA extraction

and sequencing. DNA extraction was performed using the QIAmp PowerFecal Pro DNA Kit (Qiagen, #51804).

2.1.11. 16S rRNA amplicon sequencing

Following DNA extraction, the V4-V5 region of the 16S rRNA genes were PCR-amplified using custom barcoded dual-index primers. Illumina compatible libraries were generated using the Qiagen Q1ASeq 1-step amplicon kit (Qiagen, #180419), and sequencing was performed on the Illumina MiSeq platform using 2x250 paired end reads with a goal of generating 5,000 – 10,000 reads per sample. Raw V4-V5 16S rRNA gene sequence data was demultiplexed by the MMF Core and fastq files were shared for downstream processing.

2.1.12. Shotgun metagenomic sequencing

Following DNA extraction, Illumina compatible libraries were generated using the QIAseq FX Library Kit (Qiagen, #180477). Sequencing runs were performed on the Illumina NextSeq1000 platform in the MMF using the 2x150 paired ends reads cassette. Shallow shotgun sequencing was performed with a goal of generating 2-3 million reads per sample.

2.2. Data & statistical analysis

2.2.1. Appendicitis spatial transcriptomics analysis

Raw count data was processed and normalized using the R package GeoMx Tools (v3.2.0)[151]. Filtering thresholds (<1,000 raw reads, <75% aligned, <80% trimmed, stitched sequencing reads) were used to remove poor quality reads. Additional thresholds (<50%

sequencing saturation, negative <1, no template control > 9000) were used to exclude poor quality segments. A probe was removed if (geometric mean of probe count from all segments / geometric mean of all probe counts from all segments) < 0.1 or if the probe was an outlier according to Grubb's test in at least 20% of the segments. Segments with less than 5% of genes detected were removed, and genes detected in fewer than 10% of segments were removed. Counts were then normalized by the Q3 normalization.

Differential expression for each subset (FB, FT, EPI) was modeled using linear models with experimental factors as the predictor using R package Limma[152]. The main factor of disease group (i.e. Appendicitis versus Normal) was used as a covariate. The topTable function was applied to calculate test statistics including the log fold-change and adjusted p-value for all genes. An adjusted p-value cutoff of ≤ 0.05 was used as a significance threshold for differential expression.

Gene set enrichment analysis (GSEA) was performed using total gene lists from each compartment using the PreRanked function within the GSEA software (v4.3.2) using default settings and the C5 Gene Ontology Biological Processes gene sets[153,154]. T test values were used for rankings. GSEA pathway significance was determined using an FWER p-value cutoff of ≤ 0.05.

2.2.2. Appendicitis single cell library QC and GEX analysis

Initial processing of 10X single-cell RNA seq data was performed by the CRI Bioinformatics Core at UChicago using Cell Ranger from 10X Genomics and the scRICA package[155]. Using scRICA, cells with mtRNA content >20% were removed, as were doublets identified by DoubletDecon[156]. Samples were integrated following standard workflows with

Seurat (v4.3.0)[157]. Integrated data were projected into a uniform manifold approximation and projection (UMAP) space, and SNN clustering was performed at resolution 0.5. The top markers defining each of the clusters were identified with the *FindMarkers* function. Clusters with fewer than 150 cells or top genes consisting primarily of rRNA were excluded from downstream analysis. UMAP clusters were annotated manually based on cluster gene expression with the help of the PanglaoDB human datasets and the CellTypist immune cell encyclopedia as references[158,159].

2.2.3. Appendicitis BCR repertoire analysis

Cell Ranger output BCR libraries were exported for downstream analysis. C gene and clonotype classification was based on the enclone output as part of the Cell Ranger workflow. Violin plots, bar graphs, and heat maps were generated in GraphPad Prism (v9.5.1) along with associated significance testing. Chord diagram for C-gene usage and cluster allocation was generated using TCR_Explore (v1.0)[160]. Clonotypes of size $n \geq 4$ were analyzed for C gene usage and B cell compartment based on cluster localization in the UMAP. Plasma (C8, C20), germinal center (GC) (C6, C9, C11, C14) or other (C0, C1, C3, C17, C19, C22) groupings were assigned. Filtered BCR contig.fasta files from the Cell Ranger output were used for V(D)J assignment using IMGT/HighV-QUEST[161,162]. Germline reconstruction was performed using the CreateGermlines function within the Dowser package (v2.1.0)[163]. Mutational frequencies were calculated using the observedMutations function as part of the SHazaM package (v1.2.0)[164].

2.2.4. Appendicitis TCR repertoire analysis

Cell Ranger output of TCR libraries was exported for downstream analysis. TCRs were annotated into CD8+ versus CD4+ based on their expression of *CD8A/B* or *CD4*. In the absence of co-receptor gene expression, TCRs were annotated on the basis of the UMAP cluster they mapped to. TCRs that were unable to be annotated were excluded at this stage. TCRs missing alpha chain sequences or CDR3β sequences were also removed. Final TCR repertoires were composed of paired TCR sequences with CD4 or CD8 annotation. Clonal counts were defined as TCRs with identical Vα-Jα-CDR3α and Vβ-Jβ-CDR3β pairs based on amino acid sequence. If expanded clones were expressed in multiple clusters, they were assigned to their most popular cluster. Pie charts and bar graphs, and associated significance testing were completed using GraphPad Prism (v9.5.1). Chord diagrams were generated using TCR_Explore (v1.0)[160]. Motif alignment was also performed in TCR_Explore using MUSCLE (v3.34.0) and viewed using motifStack (v1.36.1)[165,166]. GLIPH2.0 was used to cluster TCRs of putative similar specificity using default settings and version 2.0 of the reference dataset[167]. Violin plots for specified CD8+ T cell cluster markers were generated using the *FindMarkers* function in Seurat and significance threshold was based on adjusted p-value cutoff ≤ 0.05.

2.2.5. Appendicitis 16S taxonomy assignment and downstream analysis

Quality control for demultiplexed files was performed using dada2 (v1.18.0)[168]. Forward and reverse reads were truncated at 180bp, merged amplicon sequences between 300-360bp were retained, and chimeras were removed using the default consensus method. The resulting amplicon sequence variants (ASVs) were used for taxonomy assignment using the Silva reference database (v138.1) to identify bacteria at the genus level[169]. Additional species level

information was added when available on the basis of identical matches between the input ASV and the Silva reference sequence. Downstream analysis was performed using MicrobiomeAnalyst 2.0[170]. Sequencing saturation was assessed using rarefaction curves, and samples were rarefied to the minimum library size of 1,856 reads. Low count reads were filtered out using count ≥ 4, percent prevalence across samples ≥10% to remove low abundance features. This strategy was used for analysis of relative abundance profiles for the top 20 genera, along with alpha and beta diversity. For edgeR and sparCC correlation network analysis, samples were not rarefied, rather total sum scaling (TSS) normalization was applied prior to downstream analysis. EdgeR analysis resulted in 57 significant taxa comparing appendicitis versus control samples using an FDR cutoff ≤ 0.05[171]. For correlation network analysis, the sparCC method was employed, using correlation statistic cutoff of ≥ ±0.3 and p-value of ≤0.05 for inclusion[172].

2.2.6. Appendicitis microbial-derived MHC II binding peptide analysis

Raw reads had adaptors trimmed followed by quality control analysis by Trimmomatic (v.0.39)[173]. Host genome was identified and removed using KneadData (v0.7.10), and contigs were assembled using MEGAHIT (v1.2.9)[174,175]. Gene calling was performed using Prodigal (v2.6.3)[176]. Functional annotation was performed with eggNOG-mapper (v2.1.12)[177,178]. Results were annotated as either gram positive or gram negative based on taxonomy predicted from eggNOG-mapper output followed by input into PSORTb (v3.0.3) to predict cellular localization[179]. Sequences with predicted extracellular, cell wall, or outer membrane expression based on PSORTb analysis were prioritized for peptide identification. Prioritized sequences were used as input for NetMHCIIpan (v4.3)[180]. To select HLA alleles for inclusion, The Allele Frequency Net Database was referenced, where the top three most frequent alleles for each locus

based on U.S. Caucasian and African-American populations were included to better represent our patient population[181].

2.2.7. IBD analysis of CD8+ T cells from the scRNA-seq cohort

Initial processing of 10X single-cell RNA sequencing data was performed by the CRI Bioinformatics Core at UChicago using Cell Ranger from 10X Genomics and the scRICA package[155]. Using scRICA, cells with mtRNA content >20% were removed, as were doublets identified by DoubletDecon[156]. At this point cell barcodes were filtered for expression of *CD3* genes, *TR* genes, or TCR sequence expression for inclusion in the initial integration with the Seurat package (v4.1.1)[182]. Integrated data were projected into a uniform manifold approximation and projection (UMAP) space, and SNN clustering was performed at resolution 0.5. The top markers defining each of the clusters were identified with the *FindMarkers* function. Major cell types were annotated manually to identify T cell clusters. Final T cell barcodes were identified by negative *CD19* expression and localization in a T cell cluster. Cells single positive for a *CD8* gene or *CD4* were annotated accordingly, and double negative cells were retained if they clustered with a CD8 T cluster as defined by the initial UMAP. CD8+ T cells were integrated and projected onto a separate UMAP and SNN clustering was performed at resolution 0.5. Top markers for each cluster were once again identified using the *FindMarkers* function, and T subset clusters were manually annotated. Violin plots were generated using the *VlnPlot* function. Cells with positive *EOMES* or *GZMK* expression were plotted as a density map over the existing UMAP. Pseudotime analysis of CD8+ T cell subsets was performed using Slingshot (v2.6.0), where a minimum spanning tree was first generated within principle component analysis clustering, with final branches overlaid onto the CD8 T UMAP space[183]. Differential

expression analysis between C0 and C4 was performed using the *FindMarkers* function. A correlation network was generated from genes that were positively expressed along with *EOMES* and *GZMK* to identify a core gene signature shared amongst these cells.

2.2.8. IBD CD8+ T repertoire analysis in the scRNA-seq cohort

Barcodes with TCR sequences annotated as CD8+ T cells were included for downstream analysis. Mucosal-associated invariant T cells (MAIT) TCRs were excluded on the basis of *TRAV1-2* and *TRAJ12/20/33* pairing. Sequences with only alpha chains were excluded. All beta chains were used for viral matching using the Immune Epitope Database (IEDB) TCRMatch tool[184]. Matches were initially filtered to contain MHC-I restricted, human TCR, and viral epitope results only. CDR3β sequences with an exact database match were retained, then patient MHC I expression was cross-referenced to the database entry, at which point TCRs were considered "viral-reactive". Paired TCRs were retained for clonality analysis and generation of clone size counts. To validate repertoire sampling depth, an adapted cumulative distribution function for each donor was generated measuring total paired TCRs versus unique TCRs present in the paired CD8+ T repertoire. As the extent of repertoire diversity is approached for each donor, the plot will plateau off. Repertoire samples needed to demonstrate a degree of plateauing for inclusion in the CD8+ T cohort. GLIPH2.0 was used to cluster viral-reactive TCRs of putative similar specificity using default settings and version 2.0 of the reference dataset[167].

Viral reactivity for expanded clones was classified by virus, with "multiple" indicating the TCR was predicted to react to more than one virus. For chi-squared analysis, TCRs were classified as "latent" (HCMV, EBV), "respiratory" (Influenza A, SARS-CoV2), or "shared" (had

matches to both viral types). Repertoire bubble plots were generated in R (v4.2.1) using a custom script.

2.2.9. CD8+ TCR analysis from an independent cohort

Rectum TCR data, GEX data, and cluster identification was all pulled from the Gene Expression Omnibus accession number GSE125527 that was referenced in the original *Boland et. al*. publication[121]. This yielded an initial dataset of 56,462 TCRs that underwent a similar filtering and viral matching process to our dataset. Barcodes were attributed to CD8+ T cells if negative for *CD4* and *CD19* and mapped to a CD8+ T cell cluster from the original publication (T1, T2, T10, or T14), or expressed a *CD8* gene and mapped to any T cell cluster. MAIT receptors were again removed, as were unpaired sequences or barcodes that had multiple α or β chain results. Cumulative distribution functions were also generated here to evaluate TCR repertoire sequencing depth from mucosal biopsies, which resulted in exclusion of 3 healthy control samples for failure to reach a curve. The final dataset contained 3,117 paired TCRs from 5 healthy and 7 UC biopsies. TCR β chains were used for viral matching, again retaining only exact CDR3β matches that were confirmed MHC I restricted, human TCRs reacting to viral epitopes. As the Boland dataset does not contain HLA data, receptor matches were not able to be matched to allele information. To evaluate functional differences, TCRs with a cluster assignment from the original publication were used to compare the viral-reactive cells between healthy and UC samples. Differences in cluster assignment composition were assessed by Chi-squared testing.

2.2.10. Image analysis of GZMK+ CD8+ T cells

Whole slide scans were imported to QuPath (v0.5.1.) and sequential scans were overlaid to generate a single image using the ImageCombinerWarpy extension[185]. Regions of interest (ROIs) were selected to obtain representative sections of T cell infiltration within the mucosa for all sections. Automated cell detection was performed based on DAPI staining in QuPath. ROI area and nuclei counts were collected for all annotations to ensure fairness of sampling. To determine cutoffs for positive values for CD3, CD8, and Granzyme K (GZMK) staining, auto-thresholding using the Triangle method was performed with auto-down-sampling of the image and a smoothing sigma value of 0 on the full tissue for each fluorescent channel. These thresholds were used to define positivity for CD3, CD8, and GZMK. These values were then used to classify cells as T cells (CD3+), CD8+ T cells (CD3+CD8+), and GZMK+ CD8+ T cells (CD3+CD8+GZMK+) for quantitative analysis. Uninflamed sections used underwent an additional step where a classifier was trained to identify epithelium versus lamina propria cells in QuPath as mucosal tissue was too thin to select ROIs of comparable size with only immune cells. Lamina propria cells were retained for analysis and underwent an identical auto-thresholding and quantification process as inflamed sections.

2.2.11. IBD analysis of epithelial ROIs from spatial transcriptomics

Sequencing data was processed using the GeoMx tools package (v3.10.0)[151]. Following NanoString's recommended workflow, segments were checked for the number of reads, the fraction trimmed, the fraction stitched, the fraction aligned, sequence saturation, the number of negative control counts, the number of nuclei, and area. All segments passed standard GeoMX QC checks, which yielded 11,462 genes present in at least 10% of the segments, with 470

differentially expressed genes using an adjusted p-value cutoff of <0.05. Principle component analysis was performed to compare +AP to -AP segments. Gene sets from a prior publication of epithelial cells in UC were used as a reference gene list to narrow the search space of relevant genes[118]. 2,033 genes were shared between the two datasets. Differential gene expression between the +AP and -AP segments was performed using DEseq2, yielding 109 significant results based on an adjusted p-value cutoff of <0.05[186].

2.2.12. Flow cytometry

All flow cytometry data was analyzed in FlowJo (v10.9.0). Representative plots and population statistics were exported for figures or downstream statistical analysis.

2.2.13. Statistical analysis

Unless otherwise specified, graphs and statistical tests were generated in GraphPad Prism (v9.5.1). Bar plots are shown with individual donors as dots with bars denoting the mean ± SEM. Normality testing was performed using the Shapiro-Wilk test for sample distributions. Two-group comparisons were performed with unpaired T-tests for parametric, and Mann-Whitney for nonparametric. Three-group comparisons were performed using one-way ANOVA for parametric or Kruskal-Wallis for nonparametric with multiple comparison testing. Two-way ANOVA testing with multiple comparison correction was performed for compositional data with multiple values. Chi-squared testing was performed to evaluate compositional changes within the TCR repertoire. Significant p-values (< 0.05) are denoted in the figures.

2.3. Human sample cohorts and demographic information

Donor ID	*Study Group*	*Severity*	*Diagnosis/Indication*	*Surgery*	*Age*	*Sex*
N1	Normal	NA	Ovarian adenocarcinoma	Cytoreduction	30	F
N2	Normal	NA	Benign ovarian cyst	Appendectomy	33	F
N3	Normal	NA	Ovarian cystadenoma	Salpingo-oophorectomy, omentectomy, appendectomy	38	F
N4	Normal	NA	Gunshot wound	Appendectomy	21	M
A1	Appendicitis	UP	Acute appendicitis	Appendectomy	32	F
A2	Appendicitis	UP	Acute appendicitis	Appendectomy	32	F
A3	Appendicitis	UP	Acute appendicitis	Appendectomy	41	M

Table 2.2. Appendicitis spatial transcriptomics cohort. **Abbreviations:** NA = not applicable, UP = unperforated

Donor ID	*Study Group*	*Severity*	*Diagnosis/Indication*	*Surgery*	*Age*	*Sex*
ID50	Appendicitis	UP	Acute appendicitis	Appendectomy	13	F
ID52	Appendicitis	UP	Acute appendicitis	Appendectomy	15	F
ID63	Appendicitis	UP	Acute appendicitis	Appendectomy	17	M
ID32	Appendicitis	P	Acute appendicitis	Appendectomy	8	M
ID94	Appendicitis	P	Interval appendicitis	Appendectomy	7	M
ID24	Control	NA	Cecal adenocarcinoma	Right hemicolectomy	56	M
ID39	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	75	M
ID42	Control	NA	Diverticulitis	Total abdominal colectomy	48	F
ID58	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	61	F
ID64	Control	NA	Tubular adenoma, concern for malignancy	Right hemicolectomy	68	M
ID65	Control	NA	Neuroendocrine tumor	Right hemicolectomy	45	F
ID72	Control	NA	Colonic adenocarcinoma	Total abdominal colectomy	37	M
ID77	Control	NA	Colonic adenocarcinoma	Right extended hemicolectomy	73	F
ID89	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	66	F
ID93	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	52	M

Table 2.3. Appendicitis BCR repertoire control and appendicitis cohort. **Abbreviations:** NA = not applicable, UP = unperforated, P = perforated

Donor ID	*Study Group*	*Severity*	*Diagnosis/Indication*	*Surgery*	*Age*	*Sex*
ID50	Appendicitis	UP	Acute appendicitis	Appendectomy	13	F
ID52	Appendicitis	UP	Acute appendicitis	Appendectomy	15	F
ID63	Appendicitis	UP	Acute appendicitis	Appendectomy	17	M
ID32	Appendicitis	P	Acute appendicitis	Appendectomy	8	M
ID94	Appendicitis	P	Interval appendicitis	Appendectomy	7	M

Table 2.4. Appendicitis scRNA-seq cohort. **Abbreviations:** UP = unperforated, P = perforated

Donor ID	*Study Group*	*Severity*	*Diagnosis/Indication*	*Surgery*	*Age*	*Sex*
ID12	Control	NA	Colonic adenocarcinoma	Cytoreduction	61	F
ID18	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	80	F
ID24	Control	NA	Cecal adenocarcinoma	Right hemicolectomy	56	M
ID26	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	72	F
ID45	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	78	F
ID5	Control	NA	Neuroendocrine tumor	Right hemicolectomy	56	M
ID61	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	78	F
ID64A.G	Control	NA	Tubular adenomas	Right hemicolectomy	68	M
ID72A	Control	NA	Colonic adenocarcinoma	Total colectomy	37	M
ID9	Control	NA	Colonic adenocarcinoma	Right hemicolectomy	65	M
ID14	Appendicitis	UP	Acute appendicitis	Appendectomy	13	F
ID36	Appendicitis	UP	Acute appendicitis	Appendectomy	32	M
ID46	Appendicitis	UP	Interval appendicitis	Appendectomy	13	F
ID83A	Appendicitis	UP	Acute appendicitis	Appendectomy	23	M
ID91A	Appendicitis	UP	Acute appendicitis	Appendectomy	11	M
ID25	Appendicitis	P	Acute appendicitis	Appendectomy	2	M
ID32	Appendicitis	P	Acute appendicitis	Appendectomy	8	M
ID57	Appendicitis	P	Acute appendicitis	Appendectomy	5	F
ID59	Appendicitis	P	Acute appendicitis	Appendectomy	16	F
ID92A	Appendicitis	P	Acute appendicitis	Appendectomy	15	F
ID94A	Appendicitis	P	Interval appendicitis	Appendectomy	7	M
ID98A	Appendicitis	P	Acute appendicitis	Appendectomy	13	F

Table 2.5. Appedicitis16S microbiome cohort demographic and clinical characteristics. **Abbreviations:** NA = not applicable, UP = unperforated, P = perforated

Donor ID	*Study Group*	*Severity*	*Diagnosis/Indication*	*Surgery*	*Age*	*Sex*
ID36	Appendicitis	UP	Acute appendicitis	Appendectomy	32	M
ID46	Appendicitis	UP	Interval appendicitis	Appendectomy	13	F
ID63	Appendicitis	UP	Acute appendicitis	Appendectomy	17	M
ID25	Appendicitis	P	Acute appendicitis	Appendectomy	2	M
ID32	Appendicitis	P	Acute appendicitis	Appendectomy	8	M
ID57	Appendicitis	P	Acute appendicitis	Appendectomy	5	F
ID59	Appendicitis	P	Acute appendicitis	Appendectomy	16	F
ID60	Appendicitis	P	Interval appendicitis	Appendectomy	11	F

Table 2.6. Appendicitis metagenomic analysis cohort. **Abbreviations:** UP = unperforated, P = perforated

Sample ID						
AP	***COL***	***LN***	***Age***	***Sex***	***Race***	***Surgery***
CRC1			56	M	Black or African American	Right hemicolectomy
CRC2			48	F	Black or African American	Total colectomy
CRC3			61	F	Black or African American	Right hemicolectomy
CRC4			68	M	Black or African American	Right hemicolectomy
CRC5			46	F	White	Right hemicolectomy
CRC6			67	F	Black or African American	Ileocecectomy
CRC7			52	M	Unknown	Right hemicolectomy
CD1			28	M	Unknown	Ileocecectomy
CD2			41	F	White	Subtotal colectomy
CD3			28	F	White	Ileocecectomy
CD4			72	M	White	Total proctocolectomy
CD5			26	F	White	Ileocecectomy
CD6			46	F	White	Ileocecectomy
CD7			26	M	White	Ileocecectomy
UC1			30	F	White	Total colectomy
UC2			76	M	White	Total proctocolectomy
UC3			59	F	Black or African American	Total proctocolectomy
UC4	COL1	LN1	30	M	Unknown	Total colectomy
UC5			20	M	White	Total colectomy
	COL2	LN2	31	F	Unknown	Proctocolectomy
	COL3		48	M	White	Ileocolectomy
	COL4	LN3	18	M	White	Total colectomy
	COL5		30	M	Unknown	Total colectomy
	COL6		43	M	White	Total colectomy

Table 2.7. Patient demographics CD8+ T scRNA-seq cohort

Sample ID			HLA-A		HLA-B		HLA-C	
AP	***COL***	***LN***	***Allele 1***	***Allele 2***	***Allele 1***	***Allele 2***	***Allele 1***	***Allele 2***
CRC1			*02:01	*74:11	*15:03	*49:01	*02:10	*07:01
CRC2			*01:01	*30:01	*42:01	*53:01	*06:02	*17:01
CRC3			*02:01	*33:03	*07:05	*41:02	*08:02	*17:01
CRC4			*23:01	*74:01	*15:03	*49:01	*02:10	*07:01
CRC5			*01:01	*11:01	*08:01	*40:02	*02:02	*07:01
CRC6			*03:01	*30:01	*07:02	*42:01	*07:02	*17:01
CRC7			*02:01	*03:01	*18:01	*41:01	*07:01	*17:01
CD1			*01:01	*02:01	*08:01	*56:01	*01:02	*07:01
CD2			*01:01	*01:01	*08:01	*27:05	*01:02	*07:01
CD3			*29:02	*30:01	*13:02	*44:03	*06:02	*16:01
CD4			*01:01	*02:01	*07:02	*50:01	*06:02	*07:02
CD5			*02:01	*03:01	*07:04	*07:02	*07:02	*07:02
CD6			*03:01	*03:01	*07:02	*15:01	*03:03	*07:02
CD7			*02:01	*02:01	*44:02	*51:01	*05:01	*14:02
UC1			*02:01	*02:01	*15:01	*40:01	*03:04	*03:04
UC2			*02:01	*24:02	*07:02	*15:01	*03:03	*07:02
UC3			*02:05	*34:02	*44:03	*58:02	0	0
UC4	COL1	LN1	*02:06	*11:01	*35:01	*40:02	*03:04	*04:01
UC5			*02:01	*30:01	*13:02	*27:05	*02:02	*06:02
	COL2	LN2	*02:02	*25:01	*18:01	*41:01	*12:03	*17:01
	COL3		*02:01	*24:02	*27:05	*51:09	*01:02	*01:02
	COL4	LN3	*24:02	*24:02	*14:02	*51:01	*02:02	*05:01
	COL5		*01:01	*32:01	*08:01	*14:01	*07:01	*08:02
	COL6		*01:01	*24:02	*08:01	*40:02	*02:02	*07:01

Table 2.8. MHC I alleles CD8+ T scRNA-seq cohort. Zero values indicate alleles that were not determined during MHC I typing

Sample ID	***Age***	***Sex***	***Race***
HC1	21	M	White
HC2	24	F	Black or African American
HC3	30	M	White
HC4	53	M	White
HC5	65	F	White
UC1	37	M	White
UC2	25	M	White
UC3	23	M	White
UC4	58	F	Black or African American
UC5	23	F	Black or African American

Table 2.9. Patient demographics for the CD8+ T PBMC cohort

Sample ID	*Age*	*Sex*	*Diagnosis*
APP1*	72	M	Severely active ulcerative colitis with extensive ulceration
APP2*	51	M	Moderately active ulcerative colitis involving distal 90 cm of the colon
APP3	20	M	Moderately active ulcerative colitis involving the distal 36 cm of colon
APP4	75	F	Severely active ulcerative colitis with ulceration, inflammatory polyps, and marked regenerative changes
APP5	21	F	Mild to severely active ulcerative colitis
APN1	28	M	Severely active ulcerative pancolitis with villiform changes and reactive atypia
APN2*	70	F	Severely active ulcerative colitis involving the distal colon, also with diverticular disease and a cecal tubular adenoma
APN3	56	M	Moderately active ulcerative colitis involving distal 30 cm of the colon
APN4	73	M	Moderately and severely active ulcerative colitis involving the entire colon and rectum
APN5	68	M	Diffuse severely active ulcerative colitis with polypoid low-grade dysplasia
APN6*	43	M	Moderately active ulcerative colitis involving the distal 50 of colon

Table 2.10. Patient demographics for the CD8+ T imaging and spatial transcriptomics cohort. *indicates inflamed tissue block was used for spatial transcriptomics

CHAPTER 3: B CELL CLASS SWITCHING AND BYSTANDER T CELL ACTIVATION ORCHESTRATE APPENDICITIS PATHOLOGY

3.1. Summary

Appendicitis is one of the most common abdominal emergencies worldwide. Neutrophil infiltration induced by IL-8 is the best-known hallmark of appendicitis pathology. However, studies have identified increased T and B cell infiltration during appendicitis, supporting a possible role for antigen-driven adaptive immune response in disease pathogenesis. Using spatial transcriptomics and scRNA-sequencing with lymphocyte receptor repertoire analysis of the human appendix this study attempts to answer the questions (1) Which B and T cell subsets are the key inflammatory drivers of disease in appendicitis, and (2) Are these responses antigen-specific, and if so, what are possible antigenic targets? This study identifies increased germinal center class switching of B cells to IgG, supporting an antigen-specific response. In contrast T cells mediate an antigen-independent response characterized by elevated expression of IL-32, a potent IL-8 inducer. Epithelial barrier dysfunction identified via spatial transcriptomics coupled with microbial dysbiosis represented by outgrowth of *Parvimonas* and *Fretibacterium* motivated identification of microbial epitopes via computational prediction that can act as a disease-relevant antigen source. This project offers new insights into the potential role of antigen-driven adaptive immune responses in appendicitis.

3.2. Results

3.2.1. Upregulation of B cell activation pathways in the appendicitis follicle

To explore the immune response patterns within their native spatial context, we collected spatially resolved transcriptomic data on archival clinical FFPE sections of appendix tissue from three patients with appendicitis and four age-matched normal controls (**Table 2.4**). Tissue sections were stained to highlight B cells (CD20), T cells (CD3), and epithelial cells (pan-cytokeratin) to identify regions of interest (ROIs) using the NanoString GeoMx platform. ROIs containing greater than 300 estimated cells were collected across donors for 2 compartments: (1) follicular B cells (FB) and (2) follicle-adjacent T cells (FT) (**Figure 3.1a**). After data pre-processing and filtering, 8,805 genes were retained for downstream analysis.

To explore changes taking place in the follicles during appendicitis, differential expression analysis was performed between appendicitis FB ROIs (n = 8) and normal FB ROIs (n = 12) (**Figure 3.1b**). This analysis identified 314 genes that were significantly upregulated and 979 genes that were downregulated in appendicitis (**Figures 3.1c-d**). Genes related to B cell lineage (*PAX5*), activation (*CD40, CD79A, CD22, LYN*), and isotype switching (*SWAP70, IGHM, IGHG2, IGHG4*) were among the top 30 most upregulated. All four IgG subclass genes had increased fold change in appendicitis, with *IGHG1* and *IGHG3* just shy of the *p*-value cutoff (**Figure 3.1e**). The gene encoding follicular dendritic cell secreted protein, *FDCSP*, was also upregulated in FB segments (**Figures 3.1d-e**). This molecule is produced by follicular dendritic cells, is known to bind preferentially to B cells activated by a T cell-dependent anti-CD40 mechanism and was shown to regulate T cell-dependent antigen responses in the germinal center in a transgenic mouse model[187,188]. Gene set enrichment analysis (GSEA) of FB segments found

enrichment in B cell activation as well as antigen processing and presentation pathways (**Figures 3.1f, 3.2a**). Taken together, these findings show there is increased B cell activation in the follicle during appendicitis and provide support for antigen-driven germinal center responses during disease.

3.2.2. Upregulation of T cell activation pathways in the appendicitis follicle

An analogous differential expression analysis was performed between appendicitis FT ROIs (n = 6) and normal FT ROIs (n = 3) to explore changes in the T cell compartment (**Figure 3.1b**). From this analysis, 53 genes were upregulated and 12 were downregulated in appendicitis (**Figure 3.1g**). Genes related to lymphocyte activation (*CD69*, *JAK3, NAMPT, IL2RA, IFNGR1*) were upregulated in appendicitis, as were genes related to complement activation (*CTSL*) and apoptosis regulation and oxidative stress (*CEBPD, CTSB, SOD2*) (**Figure 3.1h**). Notably, there was enhanced expression of *IL32*, which encodes a pro-inflammatory cytokine that induces the production of TNF-α, IL-6, and IL-8, suggesting a key role for T cells in initiating the inflammatory cascade that drives neutrophil recruitment. Consistently, GSEA analysis revealed enrichment in T cell activation and differentiation pathways (**Figure 3.1i**). Additional pathways related to antigen presentation, innate immune activation, and immunoglobulin-mediated immunity were also enriched (**Figure 3.2b**).

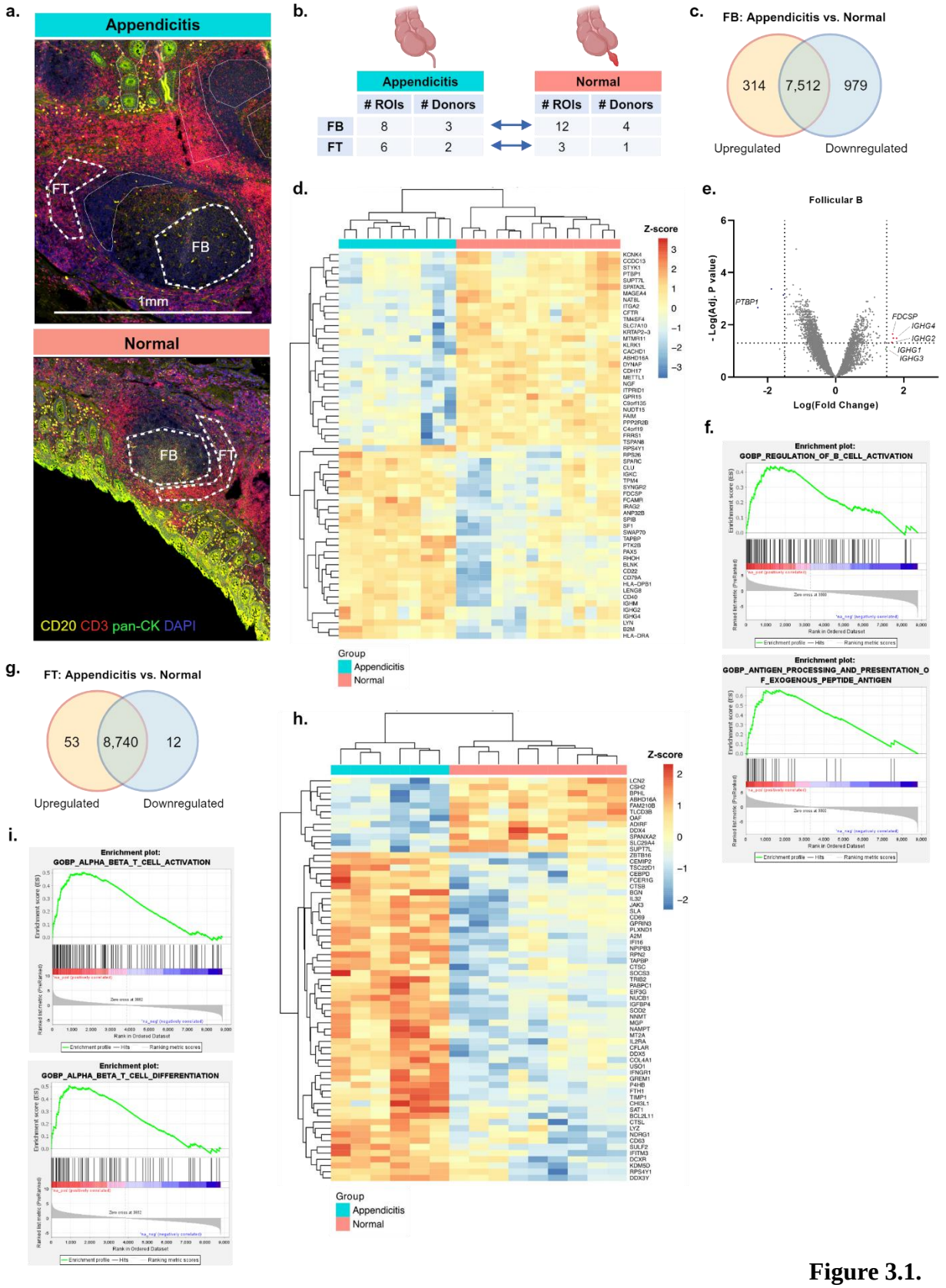

a.
Appendicitis
FT
FB
1mm
Normal
FB
FT
CD20 CD3 pan-CK DAPI
b.
Appendicitis
ROIs
Donors
Normal
ROIs
Donors
FB
8
3
12
4
FT
6
2
3
1
c.
FB: Appendicitis vs. Normal
314
7,512
979
Upregulated
Downregulated
d.
Z-score
Group
Appendicitis
Normal
e.
Follicular B
- Log(Adj. P value)
Log(Fold Change)
PTBP1
FDCSP
IGHG4
IGHG2
IGHG1
IGHG3
f.
Enrichment plot:
GOBP_REGULATION_OF_B_CELL_ACTIVATION
Enrichment plot:
GOBP_ANTIGEN_PROCESSING_AND_PRESENTATION_OF_EXOGENOUS_PEPTIDE_ANTIGEN
g.
FT: Appendicitis vs. Normal
53
8,740
12
Upregulated
Downregulated
h.
Z-score
Group
Appendicitis
Normal
i.
Enrichment plot:
GOBP_ALPHA_BETA_T_CELL_ACTIVATION
Enrichment plot:
GOBP_ALPHA_BETA_T_CELL_DIFFERENTIATION

Figure 3.1.

Figure 3.1. FB and FT regions show upregulation of lymphocyte activation pathways, supporting a role for an antigen-driven adaptive response in appendicitis. (a) Representative ROIs from appendiceal tissue sections. ROIs are outlined by white dashed line and labelled FB or FT accordingly. Sections are colored for CD20 (yellow), CD3 (red), pan-cytokeratin (green), DAPI (blue) and labelled with 1mm scale bar. **(b)** Workflow schematic for ROI analysis **(c)** Summary of DE genes from FB analysis (adjusted p-value cutoff ≤ 0.05) **(d)** Heatmap showing the top 30 upregulated and downregulated DE genes from FB segments, colored by z-score of gene expression. **(e)** Volcano plot for DE genes from FB segments, horizontal dashed line indicates adjusted p-value cutoff ≤ 0.05, while vertical dashed line indicates log(FC) ≥ ± 1.5. Genes upregulated in appendicitis are colored red, downregulated genes are colored blue, Ig genes that were near significant colored green. Relevant significant genes are annotated. **(f)** Representative GSEA plots from FB segment analysis. GSEA pathway significance was determined using an FWER p-value cutoff ≤ 0.05. **(g)** Summary of DE genes from FT analysis (adjusted p-value cutoff ≤ 0.05) **(h)** Heatmap showing all significant DE genes from FT segments, colored by Z-score of gene expression **(i)** Representative GSEA plots from FT segment analysis. GSEA pathway significance was determined using an FWER p-value cutoff ≤ 0.05. **Abbreviations**: DE = differentially expressed, FC = fold change, FWER = family wise error rate

Figure 3.2. Significant GSEA pathways from spatial transcriptomic analysis of Follicular B and T cell regions in appendicitis. (a) Significant GSEA pathways from follicular B cell (FB) segments as determined using an FWER p-value cutoff of ≤0.05. Pathways are ordered by normalized enrichment score (NES). Pathways relevant to disease pathology are highlighted according to the figure legend. **(b)** Significant GSEA pathways from follicle-adjacent T cell (FT) segments as determined using an FWER p-value cutoff of ≤0.05. Pathways are ordered by normalized enrichment score (NES). Pathways relevant to disease pathology are highlighted according to the figure legend.

3.2.3. scRNA-seq profiles of lymphocytes in appendicitis

In order to validate the observation that IgG responses were increased in appendicitis, we performed B cell receptor (BCR) repertoire analysis of appendiceal B cells from 10 control specimens and 5 appendicitis specimens using the 5' 10X Genomics VDJ library workflow (**Table 2.5**). IgG overall was increased in appendicitis compared to controls in the single cell BCR repertoire data as well (**Figure 3.3a**). To obtain a higher resolution investigation of the immune compartment in appendicitis, gene expression (GEX) and T cell receptor (TCR) scRNA-seq libraries were generated for immune cells isolated from the appendices of the five appendicitis patients using the 10X Genomics workflow (**Table 2.6, Figure 3.3b**). After initial processing and quality control using the Cell Ranger pipeline, a UMAP projection was created which resulted in 33 initial clusters. Clusters of less than 150 cells were excluded from all downstream analysis, as well as one cluster where top genes were rRNA genes, resulting in 21 final UMAP clusters containing 47,179 cell profiles (**Figure 3.3c**). All clusters contained cells from each of the 5 donors. Major cell types included T cells, B cells, and one macrophage cluster, which were represented relatively equal in proportion across donors (**Figure 3.3e**).
Of the 12 B cell clusters, 2 were plasma cells, 4 were germinal center cells, and the remaining 6 were memory and innate-like populations (**Figure 3.3c**). Plasma clusters (C8, C20), which were the least abundant, were identified on the basis of *MZB1, XBP1, CCR10,* and *JCHAIN* expression (**Figure 3.3d**). Germinal center clusters (C6, C9, C11, C14) expressed *BCL6* and *CD38*. Actively cycling clusters (C6, C11) expressed cell cycle genes such as *MKI67, PCNA, TOP2A*. Dark zone clusters (C6, C9) were differentiated from light zone (C11, C14) by *AICDA* expression.
Additional B cell clusters included resting B cells (C0, C1) (*IGHM, IGHD, MS4A1, CD19, CD74*) and marginal zone B cells (C3) (*NOTCH2, CR1, CR2*). Smaller clusters also included

activated marginal zone B cells (C17), plasmablasts (C19) (*IGHV-genes*, *MZB1*, *IGHA1*, *IGHG1/3)*, and atypical memory B cells (C22) (*FCRL4*, *FCRL5*, *ITGAX ENTPD1)* (**Figure 3.3d, 3.3f, 3.4b**).

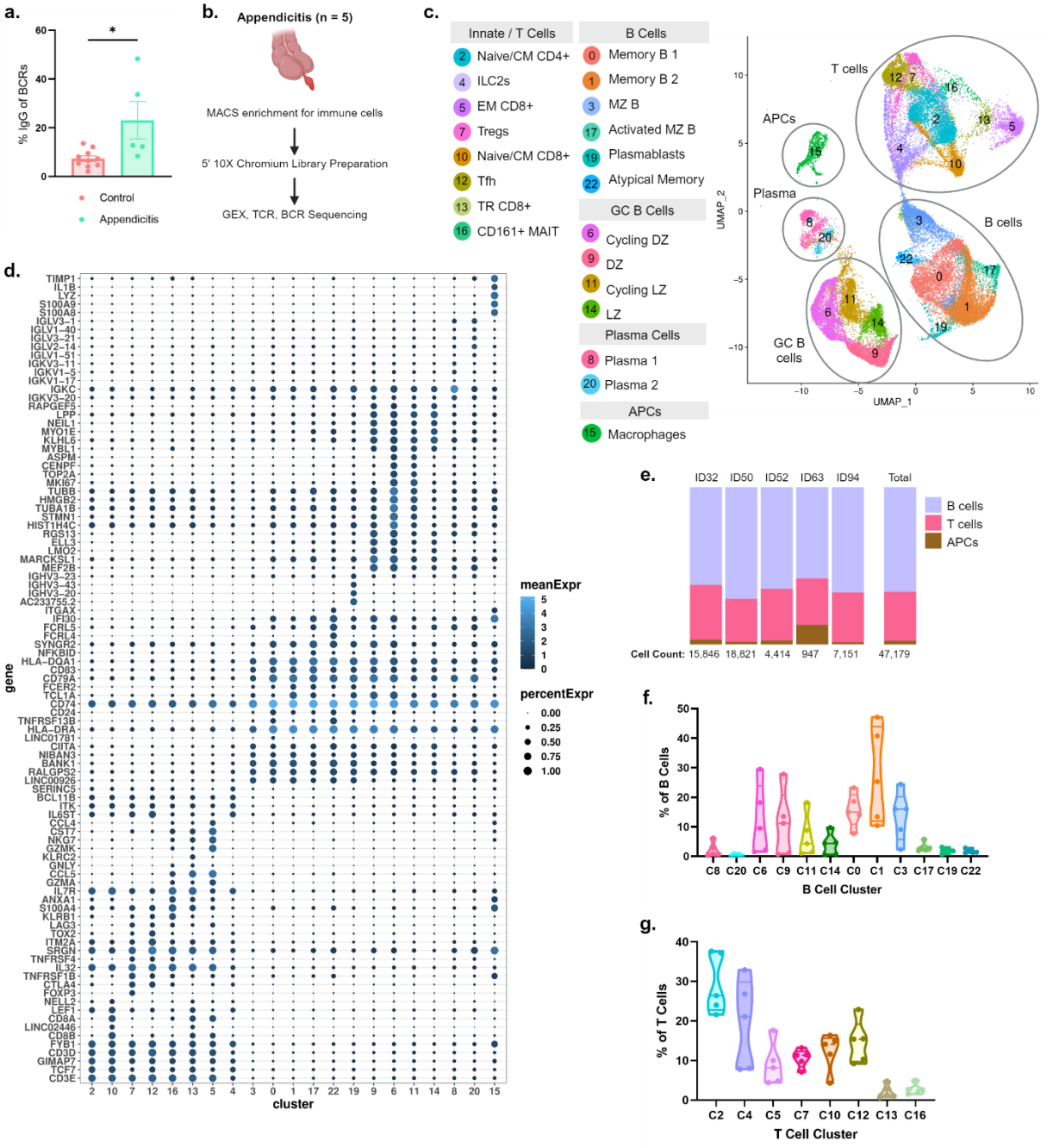


Figure 3.3. scRNA-seq of appendiceal immune cells in appendicitis show diverse activated phenotypes in T and B cell compartments. (a) Bar plot of IgG C gene usage as a percentage of BCR heavy chains from scRNA-seq data. Each dot represents one donor, bars indicate mean ±SEM. Significance testing by Mann-Whitney (p value < 0.05 = *). **(b)** Workflow schematic for 10X scRNA-seq sample collection **(c)** UMAP

identifying immune cell subsets from appendicitis donors. Major cell types are circled, with cluster annotations at left. **(d)** Bubble plot showing expression of key genes across clusters. Genes are shown on the y axis, with clusters on the x axis. Bubble size corresponds to percent of cells within the cluster expressing that gene and are colored by mean gene expression. **(e)** Proportions of major cell types by donor. Total cells per donor are listed beneath the respective bar. **(f)** Truncated violin plot for frequency of each B cell cluster as a percentage of all B cells, each dot represents one donor. Horizontal lines represent quartiles, along with minimum and maximum values from the dataset. **(g)** Truncated violin plot for frequency of each T cell cluster as a percentage of all T cells, each dot represents one donor. Horizontal lines represent quartiles, along with minimum and maximum values from the dataset. **Abbreviations:** GC = germinal center, APC = antigen presenting cell, CM = central memory, ILC = innate lymphoid cell, EM = effector memory, TR = tissue resident, MAIT = mucosal associated invariant T cell, MZ = marginal zone, DZ = dark zone, LZ = light zone

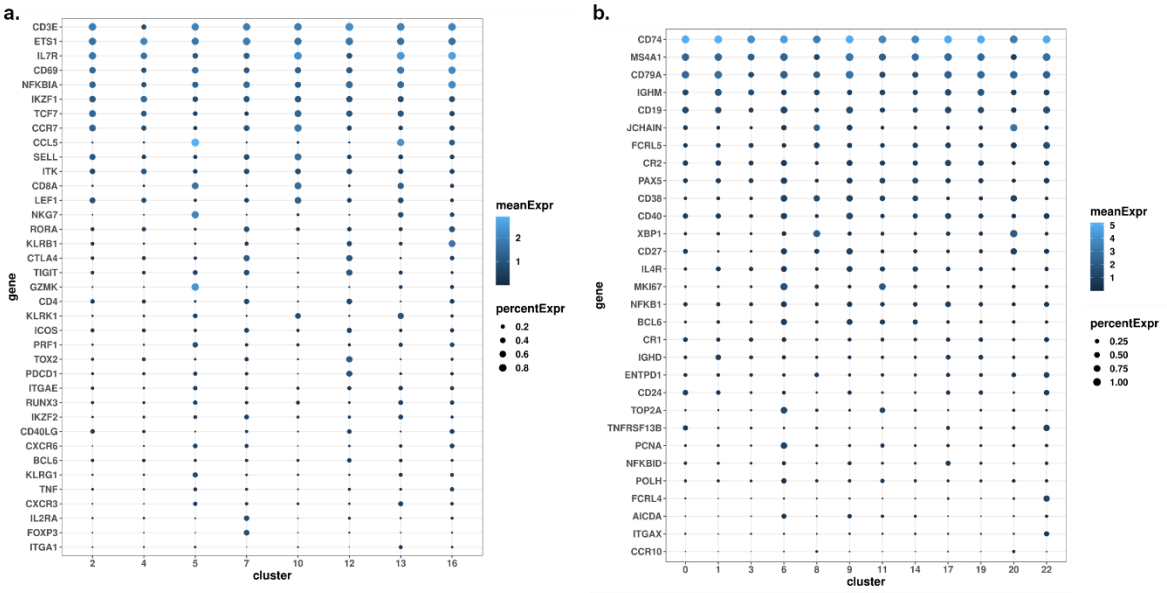


Figure 3.4. Summary bubble plots for additional phenotypic genes in T and B single cell clusters. (a) Bubble plot showing expression of key genes across T cell clusters. Genes are shown the on the y axis, with clusters shown on the x axis. Bubble size corresponds to percent of cells within the cluster expressing that gene and are colored by mean gene expression. **(b)** Bubble plot showing expression of key genes across B cell clusters. Genes are shown on the y axis, with clusters shown on the x axis. Bubble size corresponds to the percentage of cells within the cluster expressing that gene and are colored by mean gene expression.

Eight of the clusters were grouped into the T cell portion of the UMAP. Of these, three clusters were the major CD8+ populations: resting naïve/CM CD8+ T cells (C10) (*TCF7*, *IL7R*, *CCR7*, *SELL*), cytotoxic effector CD8+ T cells (C5) (*CCL5*, *GZMK*, *KLRG1*, *PRF1*), and Trm CD8+ T cells (C13) (*ITGA1*, *ITGAE*, *RUNX3*, *CXCR6*). Another three clusters comprised the major CD4+ T cell clusters: resting naïve/CM CD4+ T cells (C2) (*TCF7*, *IL7R*, *LEF1*, *CCR7*), which was the most abundant of all T cell clusters, T regulatory cells (Tregs) (C7) (*FOXP3*,

CTLA4, IKZF2, IL2RA), and T follicular helper cells (Tfh) (C12) (*PDCD1, ICOS, BCL6*) (**Figure 3.3c-d, 3.3g, 3.4a**). Cluster 16 (C16) was a CD161+ (*KLRB1*) T cell cluster containing MAIT cell alpha chains. This cluster was tissue-resident (*ITGAE, CD69*) and expressed activation genes such as *NFKBIA, CD40LG*, and *TNF* (**Figure 3.3d, 3.4a**). Cluster 4 (C4), while it had some T cell receptor (TCR) calling, was predominantly an innate lymphoid cell (ILC) cluster. Several transcription factors known to play a role in maintaining the ILC2 lineage were expressed (*TCF7, IKZF1, ETS1, RORA, ITK*) (**Figure 3.3d, 3.4a**). Except for the ILC2s, all T cell clusters expressed high levels of *IL32* (**Figure 3.3d**), which is consistent with the spatial gene expression analysis.

3.2.4. BCR repertoire analysis in appendicitis

Since both the spatial and single cell transcriptomics datasets demonstrated an increase in IgG in appendicitis, the next step was to evaluate where in the B cell compartment IgG+ B cells were present. IgG+ BCRs predominantly mapped to germinal center clusters C6 and C9, along with the C8 plasma cell cluster (**Figure 3.5a**). Two donors had a majority of their IgG+ BCRs map to a resting B cluster (C0) (**Figure 3.5a**). Interestingly, these two donors had limited clonal expansion compared to the other three donors (**Figure 3.5c**). This may be due to (1) heterogeneity in immune responses during appendicitis or (2) the possibility these patients underwent appendectomy at an earlier stage in disease course where a substantial adaptive immune response had not yet formed.

To gain a global understanding of which BCR isotypes mapped to each cluster, chord diagrams were created to visualize movement of isotypes between clusters. From this diagram it is obvious that memory cluster C1 is predominantly IgM+. The two largest germinal center

clusters C6 and C9 contained about half IgA+ BCRs, a quarter IgM+, and a quarter IgG+. C8, the largest plasma cluster, was roughly evenly split between IgA+ and IgG+ (**Figure 3.5b**). By plotting percentage of IgG+ BCRs within each cluster compared to the overall frequency of IgG

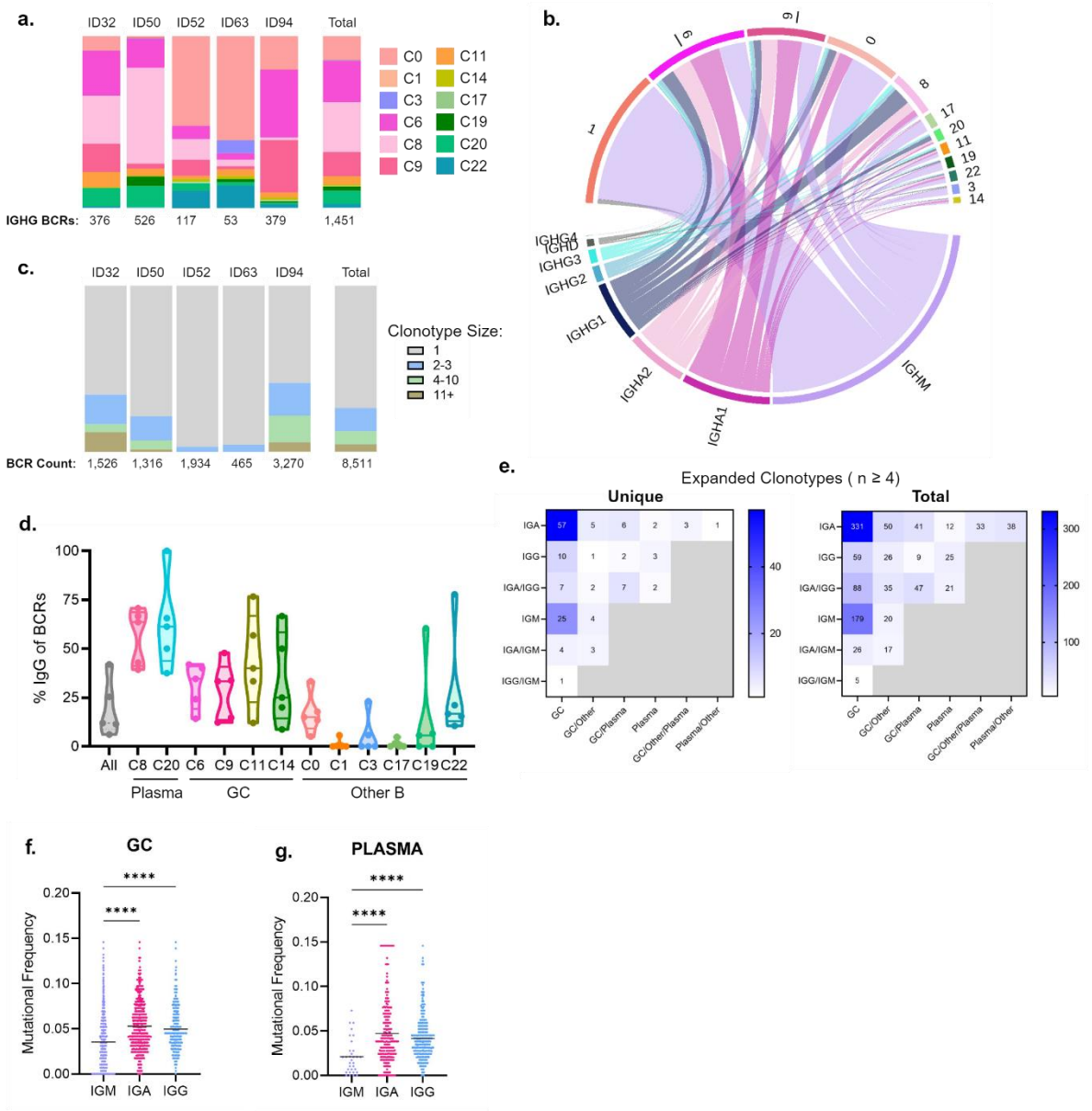


Figure 3.5. BCR repertoire analysis in appendicitis reveals expanded clonotypes in germinal center and plasma cell clusters with increased IgG usage. (a) Relative frequencies of B cell clusters as a proportion of IgG+ heavy chains. Each bar plot represents one donor with IgG BCR heavy chain counts listed below. **(b)** Chord diagram showing distribution of IgC genes across B cell clusters. Heavy chain constant regions are listed along the bottom, cluster numbers are distributed across the top. Graph shows total BCR heavy chains from the appendicitis cohort. **(c)** Bar plots showing relative frequencies of clonal expansion across donors. Each bar represents one donor with BCR heavy chain counts labelled below. **(d)** Truncated violin plots showing the percentage of IgG+ BCRs across B cell clusters. Each dot represents one donor, horizontal lines

represent minimum, maximum, and quartiles. Overall IgG versus cluster comparison was performed using Mann-Whitney testing with multiple comparison correction (all comparisons ns) **(e)** Heatmaps of BCR clonotypes with size ≥ 4, showing distribution of clones across BCR cluster group (x axis), and IgC gene (y axis). Left panel shows unique clonotype counts, right panel shows total BCR heavy chain counts. **(f)** Mutational frequencies plotted by C gene usage in the germinal center and **(g)** plasma cell compartment. Significance testing by Kruskal-Wallis with multiple comparison testing (p value ≤ 0.0001 = ****).

in the BCR repertoire, we found IgG was enriched in both plasma cell clusters, and trending towards an increase in the germinal center clusters, although this was not significant (**Figure 3.5d**).

BCRs likely descended from the same lineage were annotated into clonotypes using the *enclone* module within the Cell Ranger pipeline. All five donors had some clonal expansion observed (n > 2), indicating B cell activation, with three of the donors having larger clone sizes (n ≥ 4) (**Figure 3.5c**). Expanded clones of size ≥ 4 were categorized by their isotype and UMAP localization (either plasma, germinal center, or other). Clonotypes with multiple C genes were jointly labelled. From the resulting heatmap it is apparent the germinal center holds the largest count of unique BCR clonotypes, with a smaller fraction present in the plasma compartment (**Figure 3.5e**). IgA was the most common isotype for expanded clones, with joint IgA/IgG and IgG clones were also present in the plasma compartment. When looking at mutational frequencies between isotypes, both IgA and IgG BCRs had greater mutation rates than IgM BCRs in the germinal center and plasma cell compartments (**Figure 3.5f-g**). These increased mutation rates indicate somatic hypermutation is taking place in the germinal center, leading to likely higher affinity antibodies in the resulting plasma cell population.

3.2.5. CD8+ T cells in appendicitis show evidence of bystander activation

To investigate the presence of antigen-dependent processes in the T cell compartment, we performed TCR repertoire analysis using the single cell dataset. After initial quality control,

TCRs were annotated as CD4+ or CD8+ on the basis of co-receptor gene expression or based on UMAP cluster assignment. Afterwards, unpaired sequences were removed, resulting in 2,176 CD8+ TCRs (**Figure 3.6a**). All 5 donors had TCRs present in the three major CD8+ clusters (C5, C10, C13). As expected, clonal expansion in resting cells (C10) was minimal, while the effector subset had relatively higher expansion, both in terms of larger clone sizes and overall percentage of expanded clones, although this was non-significant (**Figure 3.6b-c**).

Next, we compared the C5 effector repertoire (n = 514) versus the C10 resting repertoire (n = 991) to look for evidence of antigenic drive in the effector subset. However, there were no significant differences in Vβ gene usage or CDR3β sequence length, nor was there enrichment for specific Vβ-Jβ or Vβ-Vα gene pairings in effector cells (**Figure 3.7a-d**). The CDR3β region between CD8 subsets were analyzed for sequence enrichment as this region of the TCR is critical for antigen recognition. Expanded clones from the effector population with CDR3β length of 12-17 amino acids (n = 44) and a random sample of resting sequences of the same length range (n = 44) underwent multiple sequence alignment. Visualization did now show any obvious sequence enrichment, with effector and resting populations looking highly similar, indicating the effector group also consisted of random TCRs (**Figure 3.6d**). Next, we employed the GLIPH2.0 algorithm, which uses CDR3β sequences and Vβ gene usage to predict TCRs with likely similar reactivity. However, inputting the effector CD8 TCRs resulted in no clusters with multiple donors, and single donor clusters consisted of a single shared CDR3β sequence, further indicating a lack of antigenic drive in this population.

The lack of TCR sequence enrichment in the effector population supports TCR-independent bystander activation, a mechanism that has previously been described for multiple microbial infections in both mice and humans[189,190]. Expression of *PRF1, IFNG,* and *CCL5,*

which encode known effectors in bystander activation, were all significantly increased in the effector population compared to resting CD8+ T cells (**Figure 3.6e**). With this mechanism, pro-inflammatory cytokines such as IL-15 are sufficient to activate memory CD8+ T cells, leading to

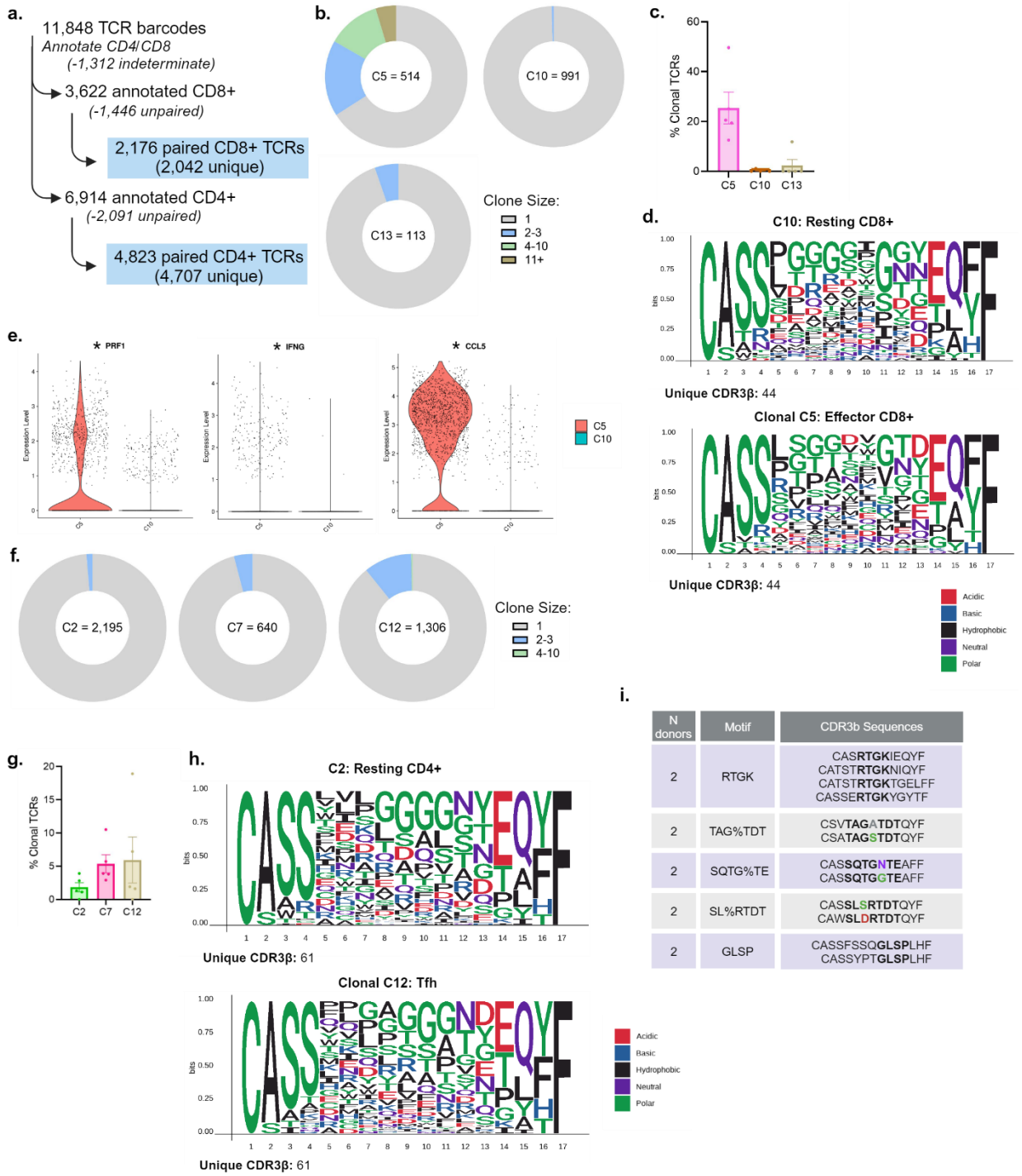


Figure 3.6.

Figure 3.6. CD8+ effectors show evidence of bystander activation, while CD4+ T cells show some evidence of antigen-dependent activation in appendicitis. (a) Workflow schematic to generate the TCR repertoire dataset **(b)** Pie charts showing clonal expansion of total TCRs for CD8+ T cell clusters (C5, C10, and C13). **(c)** Bar graph showing percent of expanded clones across CD8+ T clusters, each dot represents one donor, bars denote mean ± SEM. Significance testing by Kruskal-Wallis with multiple comparison testing (ns). **(d)** Multiple sequence alignment of CDR3β sequences, relative amino acid usage is shown for resting CD8s in the top panel and expanded effector CD8s in the bottom panel. **(e)** Violin plots showing gene expression by CD8+ T cell cluster for selected genes involved in bystander activation. Significance denoted by * symbol, determined by the *FindMarkers* function (adjusted p value: $PRF1 = 1.4x10^{-103}$, $IFNG = 9.1x10^{-70}$, $CCL5 = 1.0x10^{-296}$). **(f)** Pie charts showing clonal expansion of total TCRs for CD4+ T cell clusters (C2, C7, and C12). **(g)** Bar graph showing percent of expanded clones in the CD4+ T clusters, each dot represents one donor, bars denote mean ±SEM. Significance by Kruskal-Wallis with multiple comparison testing (ns). **(h)** Multiple sequence alignment of CDR3β sequences, relative amino acid usage is shown for resting CD4s in the top panel and expanded Tfh CD4s in the bottom panel. **(i)** Summary of GLIPH2.0 results using all CD4+ Tfh TCRs as input. Clusters with 2 or more donors were included here, sequence motifs are bolded with distinct residues highlighted.

perforin and granzyme production. In the single cell dataset, a small cluster of epithelial cells was positive for *IL15* expression while the effector CD8+ cluster (C5) expressed IL2RB (CD122), further supporting the possibility of bystander activation in appendicitis.

3.3.6. CD4+ Tfh cells in appendicitis may show evidence of antigenic drive

To investigate the possibility of antigen-driven activation in the CD4+ T cell compartment, the same workflow as previously described was performed, resulting in 4,823 total paired CD4+ TCRs for downstream analysis (**Figure 3.6a**). Again, the first step was to determine the extent of clonal expansion in the CD4+ compartment. The overall CD4+ T cell expansion was lower compared to CD8+ T cells, with the Tfh cluster (C12) showing the greatest degree of expansion (10.8% of total TCRs) (**Figure 3.6f-g**). Because Tfh cells play a role in supporting B cell activation in the germinal center, we compared the Tfh TCRs (n = 1,306) and resting CD4+ TCRs (n = 2,195) for evidence of antigenic drive. Once again, there was no observed enrichment of Vβ gene usage, Vβ-Jβ pairing, Vβ-Vα pairing, or differences in CDR3β sequence length between the two clusters (**Figure 3.8a-d**). Motif sequence alignment was

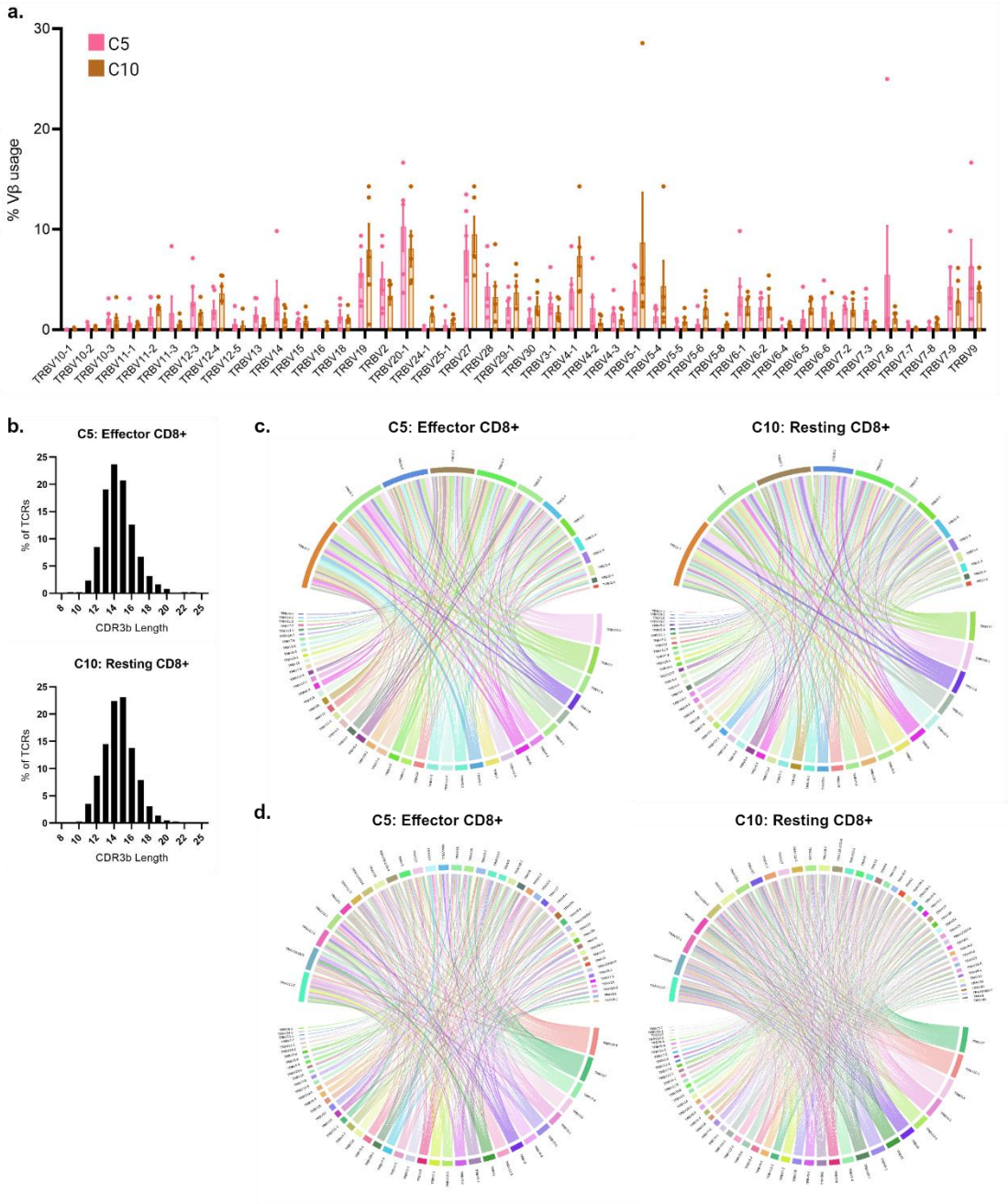


Figure 3.7. Comparison of Resting CD8+ versus Effector CD8+ TCR repertoires in appendicitis. (a) Frequency of Vβ gene usage in resting CD8+ (C10, brown) versus effector CD8+ (C5, pink) appendicitis TCRs. Each dot represents one donor. All comparisons were non-significant by multiple Wilcoxon matched-pairs signed rank test with FDR cutoff <1% for multiple comparison correction. **(b)** CDR3β length comparison between C5 and C10 CD8+ T cell populations. Bars show percentage of all TCRs with that length. Differences in length distribution were non-significant by Mann-Whitney test. **(c)** Chord diagrams showing Vβ – Jβ gene pairings in CD8+ T cell populations. **(d)** Chord diagrams showing Vα – Vβ gene pairings in CD8+ T cell populations.

performed using CDR3β sequences of length 12-17 amino acids from clonally expanded CD4+ TCRs from the Tfh cluster (n = 61) and a random selection from the resting CD4 cluster (n = 61). Within the Tfh TCRs there was still substantial diversity in relative amino acid use, with some variation in positions 5-8 compared to the resting cohort, along with enrichment of aspartic acid at position 13 (**Figure 3.6h**). All Tfh TCRs were also input into GLIPH2.0, with multiple output clusters shared between donors. These included both local and global sequence motifs, suggesting the possibility of shared epitope specificity within this cluster (**Figure 3.6i**). These results support antigen-driven activation in the Tfh compartment of the CD4+ T cells.

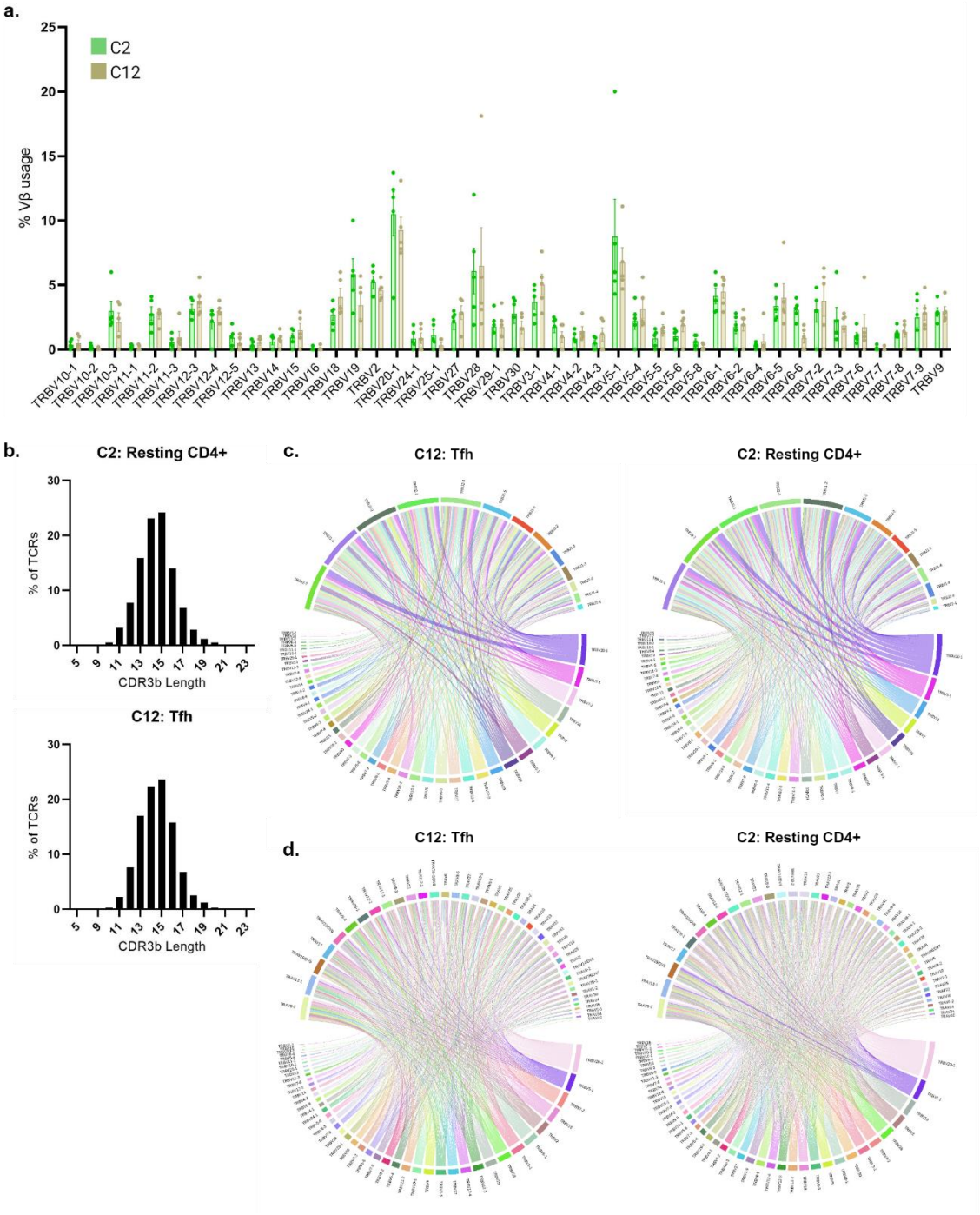


Figure 3.8. Comparison of Resting CD4+ versus Tfh CD4+ TCR repertoires in appendicitis. (a) Frequency of Vβ gene usage in resting CD4+ (C2, green) versus Tfh CD4+ (C12, gold) appendicitis TCRs. Each dot represents one donor. All comparisons were non-significant by multiple Wilcoxon matched-pairs signed rank test with FDR cutoff <1% for multiple comparison correction. **(b)** CDR3β length comparison between C2 and C12 CD4+ T cell populations. Bars show percentage of all TCRs with that length. Differences in length distribution were non-significant by Mann-Whitney test. **(c)** Chord diagrams showing Vβ – Jβ gene pairings in CD4+ T cell populations. **(d)** Chord diagrams showing Vα – Vβ gene pairings in CD4+ T cell populations.

3.2.7. Appendicitis and outgrowth of opportunistic pathogens

Because the appendix houses its own microbiome, and there are several reports of dysbiosis during appendicitis, we sought to characterize the appendiceal microbiome during disease that may contribute to the inflammatory response. Appendiceal swabs were collected for 16S rRNA sequencing from 10 controls who underwent incidental appendectomy and 12 appendicitis donors, which were further separated into perforated (n = 7) and unperforated (n = 5) subgroups (**Table 2.7**). When comparing relative frequencies for the top 20 most abundant genera in the dataset, it was readily apparent that the appendicitis group, in particular perforated appendicitis, had an outgrowth of *Parvimonas* (**Figure 3.9a**). In line with existing literature, there was no significant difference in alpha diversity between the three study groups as measured by the Shannon diversity index (**Figure 3.9b**). However, beta diversity using the Bray-Curtis index was significantly different between study groups, and pairwise comparison found a significant alteration in community composition between perforated appendicitis samples and controls (**Figure 3.9c**). These results indicate that while the appendicitis microbiome is not less diverse than that of controls, the community composition is altered, especially in the perforated cases. To identify which taxa were altered in appendicitis versus controls, differential expression analysis was performed using edgeR, which yielded 57 significant results. Opportunistic pathogen genera such as *Parvimonas* and *Peptostreptococcus* were increased in appendicitis (**Figure 3.9d**). A previous report found increased *Parvimonas* in complicated appendicitis that was associated with oral bacteria such as *Fusobacterium,* a genera that has repeatedly been associated with appendicitis in the literature[20,191–193]. While *Fusobacterium* trended towards an increase in appendicitis, it was non-significant (**Figure 3.9d**). However, another oral genus, *Fretibacterium*, was significantly increased (**Figure 3.9d**). In order to define a "signature" of

microbial taxa that positively correlated with one another, SparCC network correlation analysis was applied to compare appendicitis and control samples. Within the resulting network, *Parvimonas, Peptostreptococcus, Fusobacterium,* and *Fretibacterium* were all positively correlated with one another (**Figure 3.9e**). Microbes from these taxa have been isolated from subgingival plaque biofilms and are associated with periodontal disease. Additionally, *Fusobacterium nucleatum* has been reported to invade the tissue during appendicitis[194]. The association of these taxa with inflammation provides a rationale for further studies looking at their mechanistic role in appendicitis.

3.2.8. Prediction of MHC II restricted peptides from the appendicitis microbiome

After identifying possible antigen-dependent activation in the B cell compartments as well as a dysbiotic state of the appendiceal microbiome in appendicitis, we hypothesized that the appendiceal microbiome acted as an antigen source during appendicitis. Because modeling of BCR epitope binding remains technically challenging, we pursued identification of CD4+ Tfh antigens. Computational prediction of MHC-II restricted microbial-derived epitopes has been described in the literature, demonstrating feasibility of this technique[195]. To identify MHC II-restricted epitopes from the appendicitis microbiome that could potentially be recognized by CD4+ Tfh cells, a subset of appendiceal swabs were submitted for metagenomic sequencing. This cohort consisted of 3 unperforated and 5 perforated samples (**Table 2.8**). A similar workflow was employed as previously described to prioritize contigs encoding genes with successful functional annotation that were predicted to localize to the membrane or extracellular space (**Figure 3.9f**). The resulting 323 contigs were input into NetMHCIIPan, along with the most frequent MHC II alleles at each locus based on representative U.S. populations. This

method identified 4,187 total 15-mer peptides predicted to be strong binders (SB) from the metagenomic data. we further restricted the search to 130 contigs derived from the *Peptoniphilaceae* family, of which *P. micra* is a member. This filtering step yielded 1,932 SB results likely derived from *P. micra*. The majority of SB peptides were predicted to bind alleles at the DP locus (**Figure 3.9g**). To understand which proteins were yielding the most SB results, contigs were sorted based on their protein assignment, and discovered that by far the most common protein yielding SB results were ABC transporters (**Figure 3.9h**). These transporters are a conserved *Firmicutes* protein and are highly abundant in the microbiome, and have previously been described as a T cell antigen[142]. This exploratory analysis demonstrates a high potential for microbial-derived antigens to contribute to adaptive immune responses in appendicitis and provides rationale for future studies into the role of bacterial epitopes in this disease.

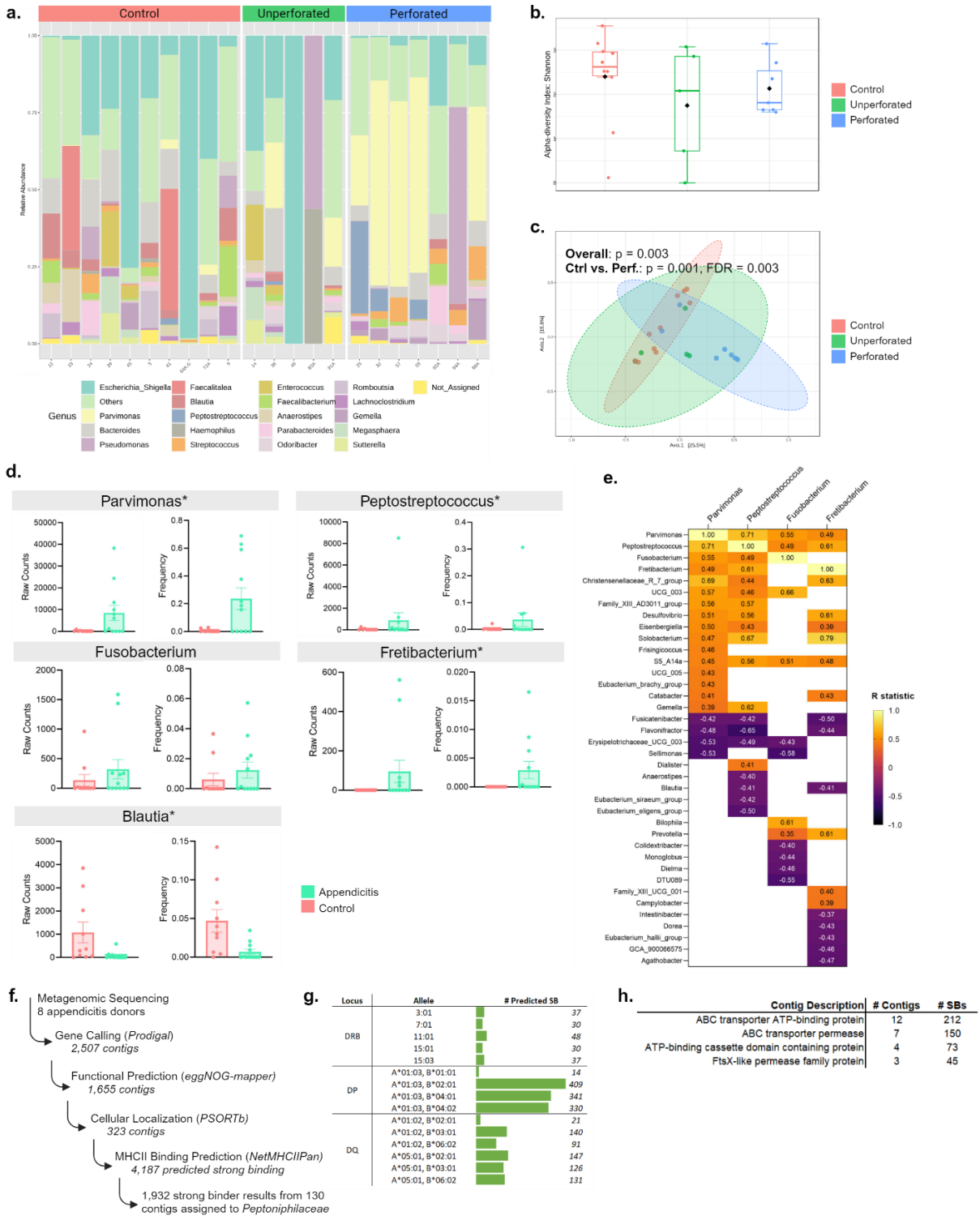


Figure 3.9. Appendicitis is characterized by outgrowth of opportunistic pathogens which are positively correlated with oral biofilm-forming taxa. (a) Relative abundance for the top 20 genera by donor **(b)** Alpha diversity by Shannon diversity index. Significance by ANOVA with multiple comparison testing (ns) **(c)** Beta diversity using the Bray-Curtis index and PCoA visualization. Significance with PERMANOVA and pair-wise comparison testing (Benjamin Hochberg adjustment, FDR cutoff ≤ 0.05) **(d)** Raw counts and frequencies are plotted for selected genera, bars indicate mean ± SEM, each dot represents one donor. Labels with * were

significant by *edgeR* analysis, significance determined by FDR cutoff ≤ 0.05. **(e)** Heatmap results of *sparCC* correlation network analysis for select taxa (cutoffs R ≥ 0.3, p-value ≤ 0.05 for inclusion). Heatmap is colored by R value, which is noted in each cell. **(f)** Workflow schematic for metagenomic analysis to identify microbially derived *Peptoniphilaceae* peptides predicted to bind MHC II molecules. **(g)** Table summarizing number of predicted strong binding sequences by MHC allele. **(h)** Summary of most frequent contigs present in the strong binder results. **Abbreviations:** FDR = false discovery rate

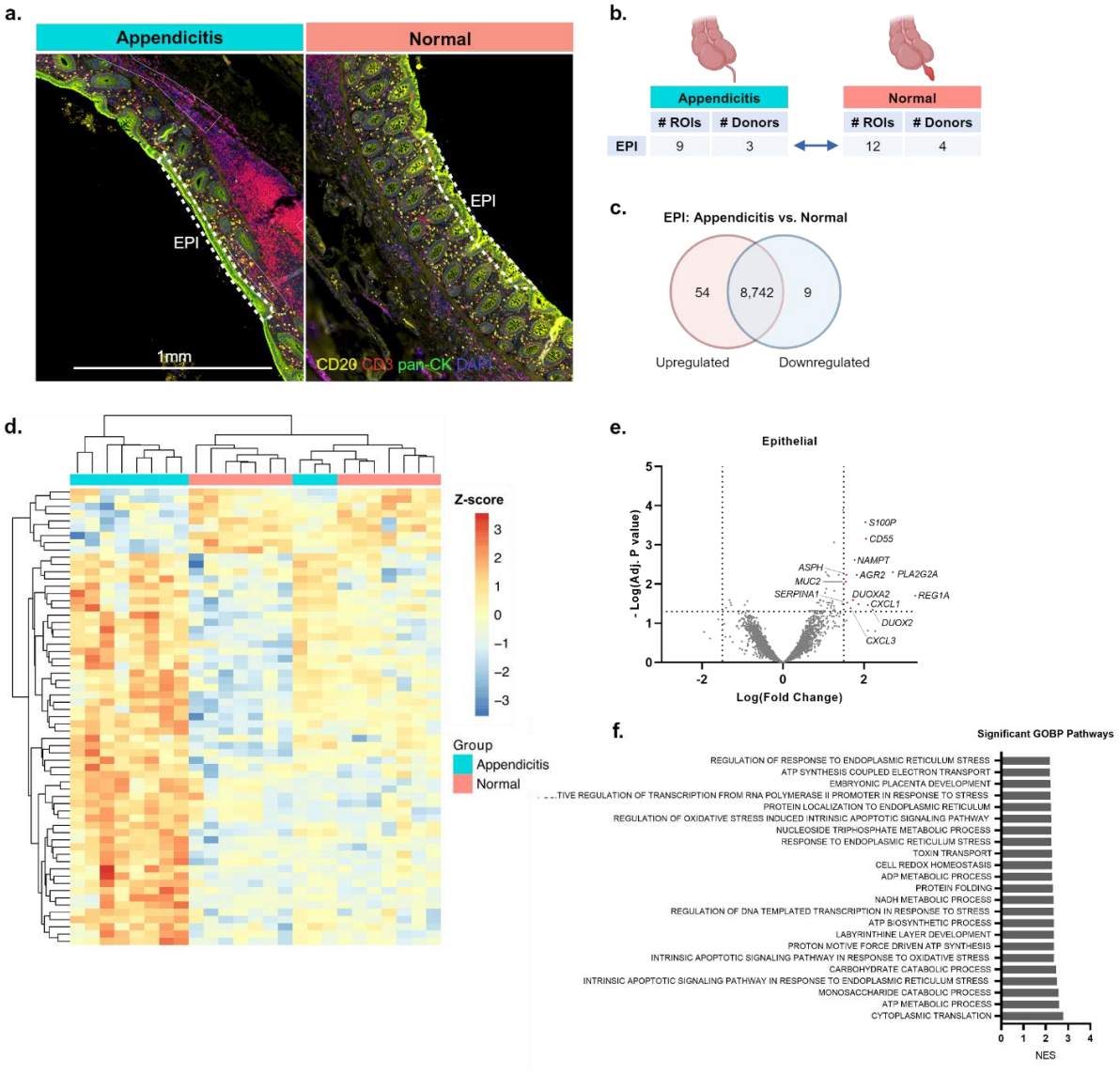


Figure 3.10. Disruption of epithelial barrier integrity in appendicitis identifies a likely route for tissue invasion by microbes. (A) Representative ROIs for EPI segments from appendiceal tissue. ROIs are outlined by dashed white line. Sections are colored for CD20 (yellow), CD3 (red), pan-cytokeratin (green), DAPI (blue) and labelled with 1mm scale bar. **(B)** Workflow schematic for ROI analysis. **(C)** Summary of DE genes from EPI analysis (adjusted p value ≤ 0.05) **(D)** Heat map showing of significant DE genes, colored by z-score of gene expression. **(E)** Volcano plot for DE genes from EPI segments, horizontal dashed line indicates adjusted p-value cutoff ≤ 0.05, while vertical dashed line indicates log(FC) ≥ ± 1.5. Genes upregulated in appendicitis are colored red. Relevant significant genes are annotated. **(F)** Significant GSEA pathways ordered by normalized enrichment score. Significance determined using and FWER p-value cutoff ≤ 0.05.

3.2.9. Epithelial barrier disruption in appendicitis identifies a likely route for tissue invasion by microbes

Microbial entry at mucosal sites is tightly regulated by mucus production, the epithelial barrier, and the host immune system. However, when epithelial barrier integrity is compromised, microbes can invade the tissue, leading to damage and inflammation. we analyzed spatial transcriptomic data from epithelial cell (EPI) ROIs to determine whether there was evidence of epithelial barrier disruption in appendicitis (**Figure 3.10a**). Differential expression analysis between appendicitis ROIs (n = 9) and normal ROIs (n = 12) resulted in 54 upregulated genes and 9 downregulated genes in appendicitis (**Figure 3.10b-d**). Both *DUOX2* and *DUOXA2* were significantly increased in the epithelium of appendicitis tissue (**Figure 3.10e**). *DUOX2,* along with its regulatory counterpart *DUOXA2*, encodes a protein which produces hydrogen peroxide in a microbiota-dependent fashion[196]. This has been shown in mice with increased activation of DUOX2, where increased bacterial translocation led to increased pro-inflammatory cytokine signaling[197]. Increased expression of this gene has been found in patients with inflammatory bowel disease, and conditional knockout of *Duox2* in intestinal epithelium protected mice against DSS-induced colitis[198]. *REG1A* was also increased in the epithelium of appendicitis patients in the spatial dataset (**Figure 3.10e**). This gene is also upregulated in inflammatory bowel disease, where it is expressed in response to IL-6/IL-22 mediated JAK/STAT3 signaling and acts as a protective factor to inhibit inflammatory responses[199]. GSEA analysis of the epithelial segments also found multiple pathways related to cell stress and apoptosis (**Figure 3.10f**). These changes in gene expression in the epithelium of appendicitis samples supports a loss of epithelial barrier integrity and the potential for microbes to invade the nearby tissue and contribute to disease pathology.

CHAPTER FOUR: APPENDICEAL CD8+ T CELLS CONTRIBUTE TO ULCERATIVE COLITIS PATHOLOGY

4.1. Summary

Ulcerative colitis (UC) is a chronic and progressive inflammatory condition mediated largely by T cells. There has been a long-standing clinical observation that removal of the appendix protects against later development of UC, and appendectomy is shown to decrease CD8+ T cell infiltration into the colon. However, the role of appendiceal T cells and their possible antigen reactivity in UC has not yet been explored. Through collection of a human scRNA-seq dataset, we identified a subset of clonally expanded CD8+ T cells predicted to be viral reactive. These cells were enriched in an effector *GZMK*+ subset in the appendix that was also present in the colon of two independent scRNA-seq datasets. Using immunofluorescence, granzyme K+ CD8+ T cells were found to be increased in the colon of UC patients with appendices compared to patients who underwent prior appendectomy, suggesting the appendix is a priming site for these cells. Epithelial transcriptional signatures from UC patients with an intact appendix show upregulation of inflammatory genes compared to UC cases without an appendix, tying this tissue to pathology in UC. This study provides the first in-depth report of the appendiceal CD8+ T repertoire in UC and begins to shed light on the mechanistic basis for the relationship between appendectomy and UC.

4.2. Results

Human tissue was sourced from patients undergoing surgery for inflammatory bowel disease (IBD) or incidental appendectomy where the appendix was uninvolved in disease processes. Immune cells were isolated and processed using the 10X Genomics 5' scRNA-seq workflow to generate gene expression (GEX) and T cell receptor (TCR) data. HLA-typing for all donors was also performed. Quality control steps resulted in 241,259 cells overall, which after filtering led to a final cohort of 27,490 CD8+ T cells and 14,189 paired TCRs from 28 samples across three compartments: appendix, colon, and cecum-draining lymph node (**Figures 4.1., 4.2, Tables 2.9, 2.10**).

4.2.1. Predicted viral-reactive CD8+ T cells are increasingly clonally expanded in the appendix of UC donors compared to CD or CRC samples

Paired appendiceal CD8+ TCRs were analyzed from 7 colorectal cancer (CRC, n =1,940), 7 Crohn's disease (CD, n = 2,585) and 5 UC (n = 5,008) patient samples to assess clonal expansion and potential antigen reactivity (**Figure 4.3a-b**). CD8+ T cells were increased in the UC samples both by count and relative abundance in the scRNA-seq data, although this was not correlated with clonal expansion or quality of repertoire sampling (**Figures 4.1h, 4.2g-h, 4.3c**). In the overall repertoire there was a trend toward increased clonal expansion and significant increases in clonal dominance in the UC group, indicating increased CD8+T activation compared to CD or CRC (**Figure 4.3b,d**). As CD8+ T cells are the canonical anti-viral responders, and IBD patients are known to have increased susceptibility to viral infections, we wanted to investigate the viral reactivity of the appendiceal CD8+ TCRs[200]. In order to predict antigen reactivity, our TCR sequences were used as input into the TCRMatch tool within the

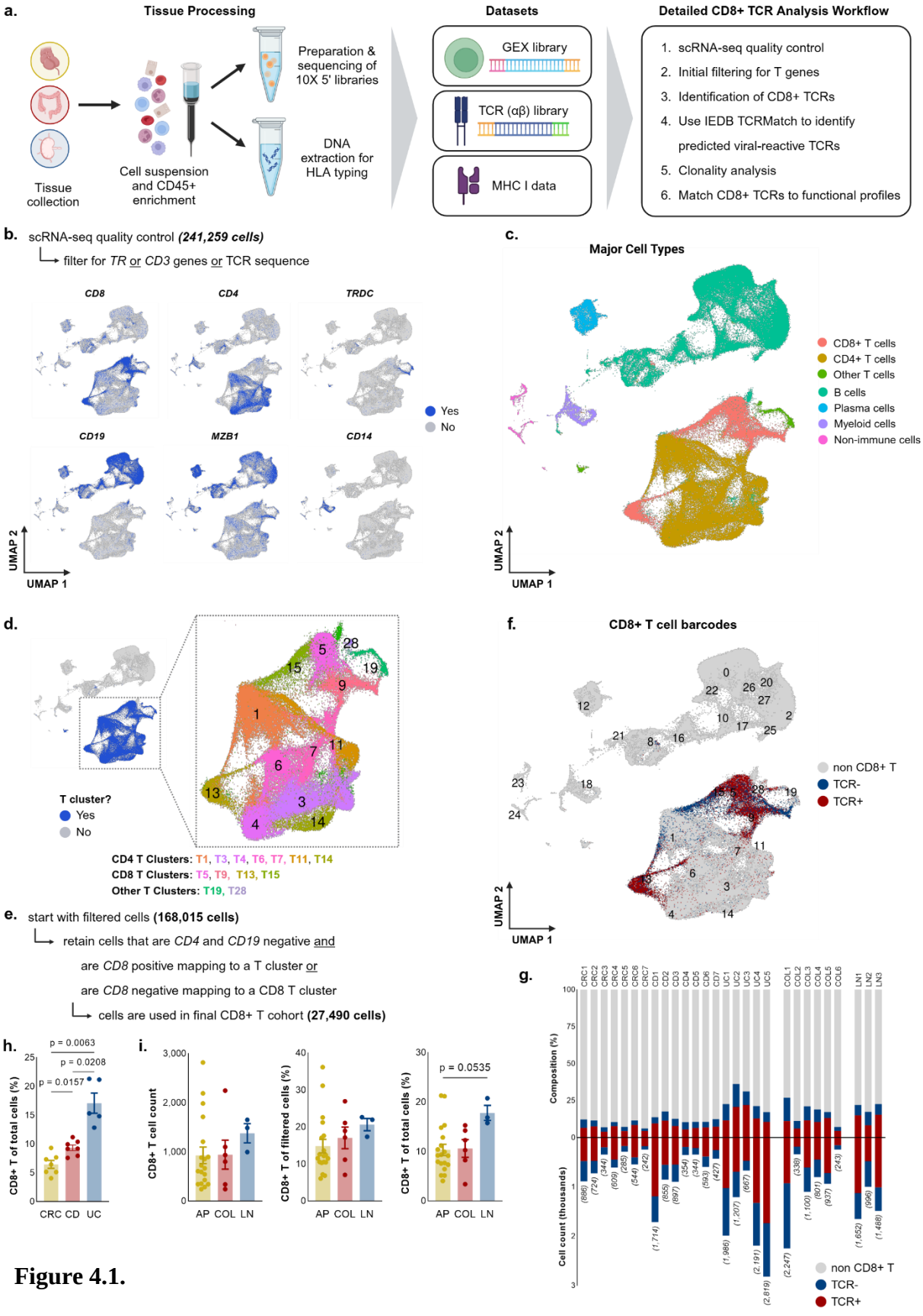

a.
Tissue Processing
Tissue collection
Cell suspension and CD45+ enrichment
Preparation & sequencing of 10X 5' libraries
DNA extraction for HLA typing
Datasets
GEX library
TCR (αβ) library
MHC I data
Detailed CD8+ TCR Analysis Workflow
1. scRNA-seq quality control
2. Initial filtering for T genes
3. Identification of CD8+ TCRs
4. Use IEDB TCRMatch to identify predicted viral-reactive TCRs
5. Clonality analysis
6. Match CD8+ TCRs to functional profiles
b.
scRNA-seq quality control (241,259 cells)
filter for TR or CD3 genes or TCR sequence
CD8
CD4
TRDC
CD19
MZB1
CD14
Yes
No
UMAP 2
UMAP 1
c.
Major Cell Types
CD8+ T cells
CD4+ T cells
Other T cells
B cells
Plasma cells
Myeloid cells
Non-immune cells
UMAP 2
UMAP 1
d.
T cluster?
Yes
No
CD4 T Clusters: T1, T3, T4, T6, T7, T11, T14
CD8 T Clusters: T5, T9, T13, T15
Other T Clusters: T19, T28
e.
start with filtered cells (168,015 cells)
retain cells that are CD4 and CD19 negative and
are CD8 positive mapping to a T cluster or
are CD8 negative mapping to a CD8 T cluster
cells are used in final CD8+ T cohort (27,490 cells)
f.
CD8+ T cell barcodes
non CD8+ T
TCR-
TCR+
UMAP 2
UMAP 1
g.
Composition (%)
Cell count (thousands)
non CD8+ T
TCR-
TCR+
h.
p = 0.0063
p = 0.0208
p = 0.0157
CD8+ T of total cells (%)
CRC CD UC
i.
CD8+ T cell count
CD8+ T of filtered cells (%)
CD8+ T of total cells (%)
p = 0.0535
AP COL LN

Figure 4.1.

Figure 4.1. Identification of CD8+ T cells from the scRNA-seq cohort. (a) Workflow for identification and analysis of CD8+ TCRs from the scRNA-seq dataset **(b)** initial filtering steps with UMAPs colored by selected marker genes for major cell types **(c)** UMAP colored by major cell types with annotations **(d)** T cell clusters from the initial integration **(e)** logic workflow to identify CD8+ T cells **(f)** final barcodes defined as CD8+ T cells colored by presence of TCR sequence (red) or not (blue) **(g)** Composition (top) and cell counts (bottom) of CD8+ T cells by presence or lack of a TCR sequence, total CD8+ T cell counts added to each column **(h)** percentage of CD8+ T cells of total cells by appendiceal study group **(i)** counts and proportions of CD8+ T cells by compartment in the scRNA-seq cohort. Bar plots represent the mean ± SEM for each group with individual dots as donors. Normality testing was performed using the Shapiro-Wilk test, followed by Kruskal-Wallis with Dunn's multiple comparison correction for significance testing. Parametric distributions were compared by one-way Welch's ANOVA with Dunnett's multiple comparison correction. P values < 0.05 were considered significant.

publicly available Immune Epitope Database (IEDB)[184]. This resulted in 600 matches based on exact CDR3β sequences and HLA allele expression (CRC = 94, CD = 141, UC = 365) (**Figure 4.3e**). The viral-reactive repertoire accounted for approximately 5% of the TCRs and were significantly clonally expanded in the UC appendix compared to the other groups, indicating these cells are seeing and reacting to antigen (**Figure 4.3e-f**). To evaluate TCR similarity, the viral-reactive sequences were input into GLIPH2.0[167]. Clusters containing expanded viral-reactive TCRs (clone size >1) from multiple IBD donors were present, indicating these TCRs are more likely to be responding to the same or similar antigens. One example showed nearly identical β chain sequences mapping to epitopes from influenza $M1_{58-66}$ and SARS-CoV-2 $NCAP_{222-230}$ antigens (**Figure 4.3g**). This is consistent with recent findings of cross-reactivity between SARS-CoV-2 epitopes and the immunodominant M1 epitope from influenza[201]. These findings establish a subset of CD8+ T cells in the appendix that are predicted to react to viral antigens and are uniquely clonally expanded in UC compared to CD or CRC.

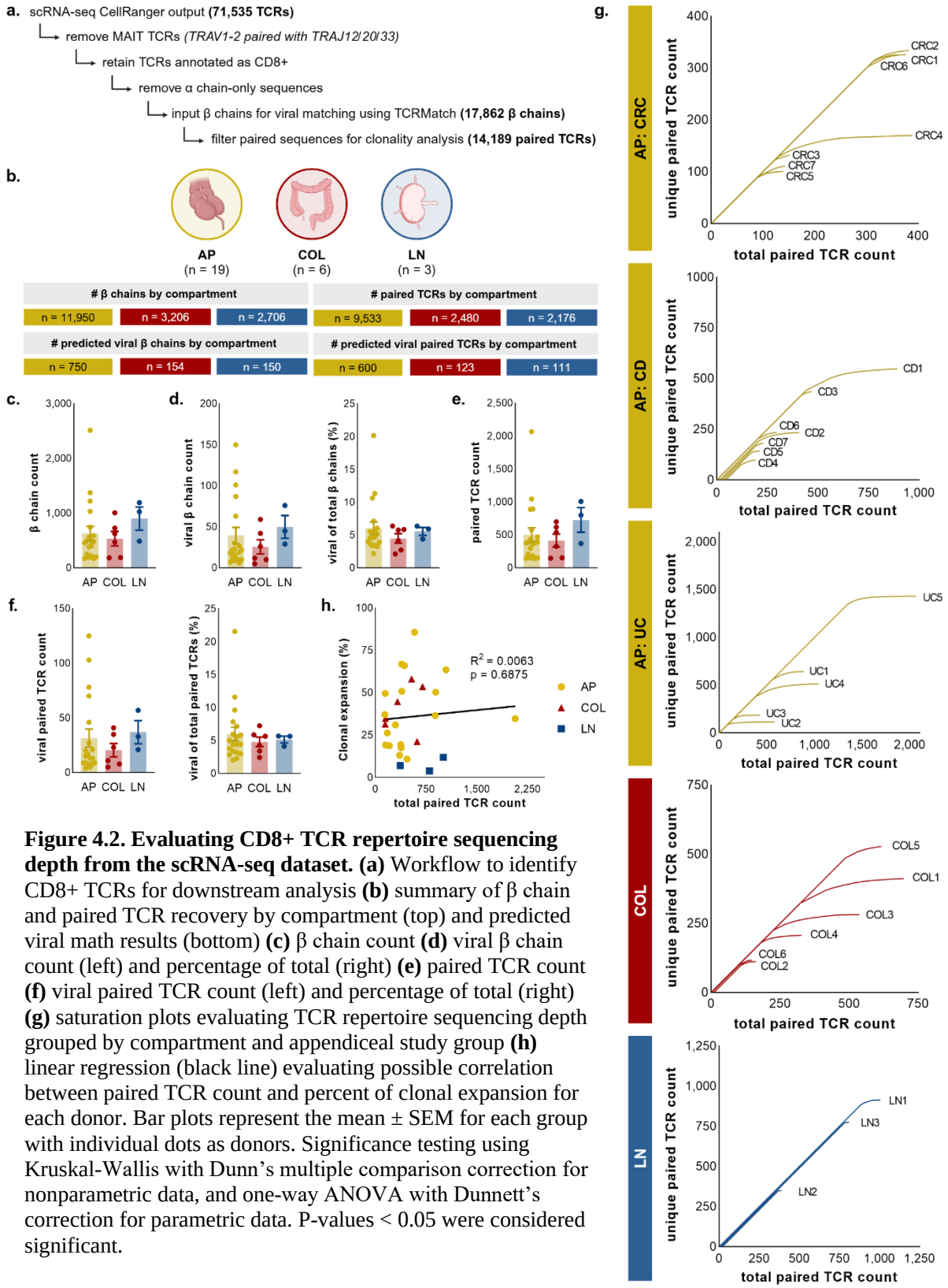
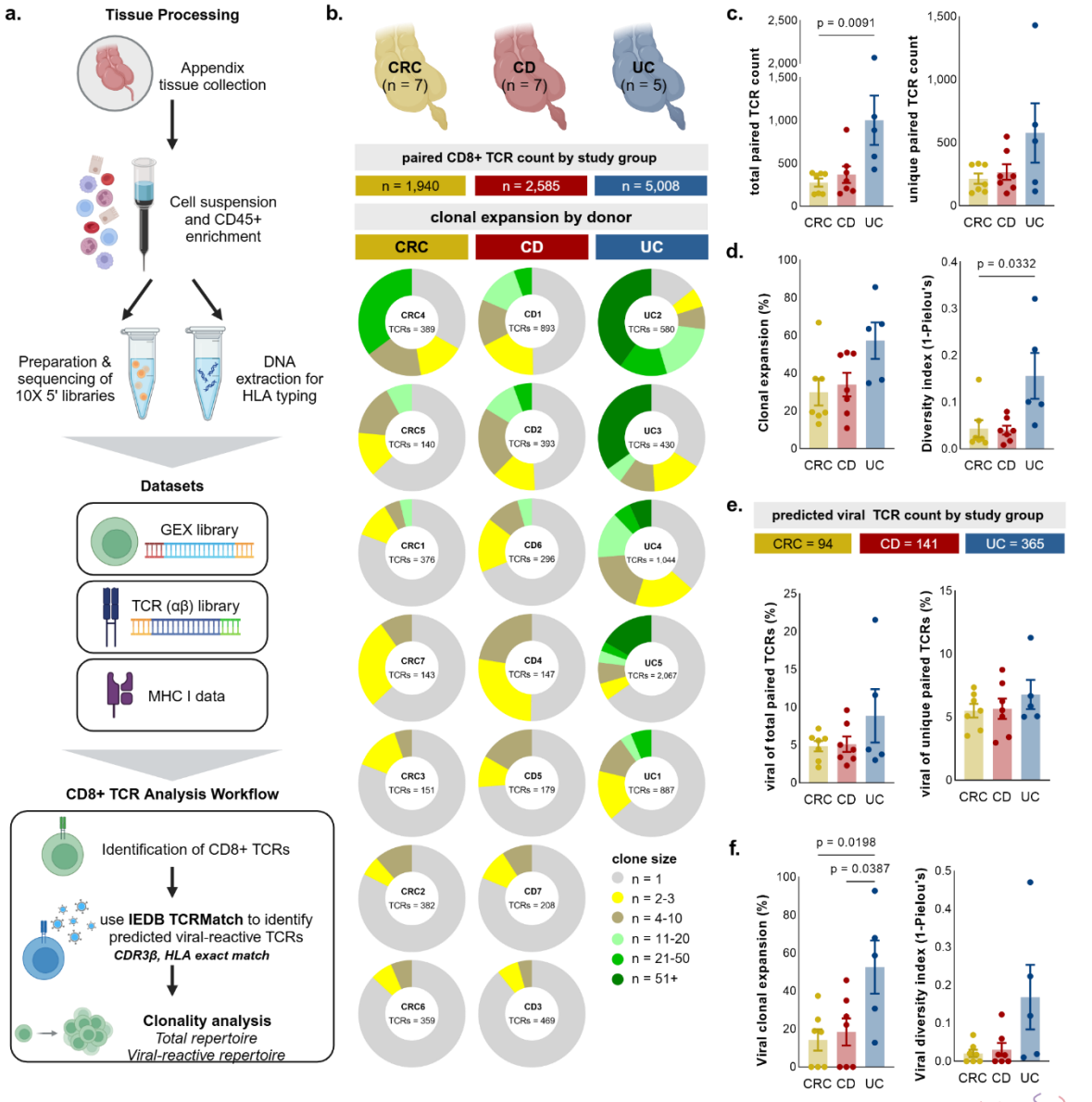


Figure 4.2. Evaluating CD8+ TCR repertoire sequencing depth from the scRNA-seq dataset. (a) Workflow to identify CD8+ TCRs for downstream analysis **(b)** summary of β chain and paired TCR recovery by compartment (top) and predicted viral math results (bottom) **(c)** β chain count **(d)** viral β chain count (left) and percentage of total (right) **(e)** paired TCR count **(f)** viral paired TCR count (left) and percentage of total (right) **(g)** saturation plots evaluating TCR repertoire sequencing depth grouped by compartment and appendiceal study group **(h)** linear regression (black line) evaluating possible correlation between paired TCR count and percent of clonal expansion for each donor. Bar plots represent the mean ± SEM for each group with individual dots as donors. Significance testing using Kruskal-Wallis with Dunn's multiple comparison correction for nonparametric data, and one-way ANOVA with Dunnett's correction for parametric data. P-values < 0.05 were considered significant.



g.

	α chain V	α chain (CDR3)	α chain J	β chain V	β chain (CDR3)	β chain J	TCR Count	Predicted epitope(s)
[UC1] SI%SGSYNE	TRAV10	CVVS**P**GREYGNKLVF	TRAJ47	TRBV19	CASSI**S**SGSYNEQFF	TRBJ2-1	30	GILGFVFTL
	TRAV10	CVVS**A**GREYGNKLVF	TRAJ47	TRBV19	CASSI**T**SGSYNEQFF	TRBJ2-1	2	LLLDRLNQL
	TRAV12-2	CAVNRGR**ETSGSR**LTF	TRAJ58	TRBV19	CASSI**T**SGSYNEQFF	TRBJ2-1	5	LLLDRLNQL
	TRAV12-2	CAVNRGR**DSGGSNYK**LTF	TRAJ58	TRBV19	CASSI**T**SGSYNEQFF	TRBJ2-1	17	LLLDRLNQL
[UC5] SLT%T	TRAV34	CGAA**F**LNYGGSQGNLIF	TRAJ42	TRBV7-9	CASSLT**E**TQYF	TRBJ2-5	2	YVLDHLIVV
	TRAV34	CGAA**LL**NYGGSQGNLIF	TRAJ42	TRBV7-9	CASSLT**D**TQYF	TRBJ2-3	2	SARS-CoV2 pool
[CD1] S%GTGE	TRAV13-1	CAA**GG**GNKL**T**F	TRAJ10	TRBV12-4	CASS**Q**GTGELFF	TRBJ2-2	2	VQELYSPIFLIV
	TRAV13-1	CAA**SD**GNKL**V**F	TRAJ47	TRBV12-4	CASS**Q**GTGELFF	TRBJ2-2	4	FIAGLIAIV
	TRAV13-1	CAA**SE**GNKL**V**F	TRAJ47	TRBV12-4	CASS**S**GTGELFF	TRBJ2-2	7	IMLIIFWFSL
	TRAV13-1	CAA**GE**GNKL**V**F	TRAJ47	TRBV12-4	CASS**L**GTGELFF	TRBJ2-2	2	SARS-CoV2 pool

Figure 4.3.

Figure 4.3. CD8+ TCRs predicted to be viral-reactive are uniquely clonally expanded in the appendix of UC, but not CRC or CD patients. (a) workflow schematic for tissue processing and clonality analysis of CD8+ TCRs **(b)** human appendix cohort summary and CD8+ TCR repertoire by donor grouped by clone size, TCR counts included with each pie chart **(c)** total (left) and unique right) paired TCR counts **(d)** clonal expansion (left) and diversity index (right) by study group in the overall paired TCR repertoire **(e)** summary counts of viral-matched TCRs and percentage matched of total (left) and unique (right) TCRs **(f)** clonal expansion (left) and diversity index (right) by study group in the viral reactive repertoire **(g)** selected results from GLIPH analysis highlighting clusters with global CDR3β similarity that contain clonally expanded (size > 1) viral-reactive TCRs with epitope predictions. Bar plots represent the mean ± SEM for each group with individual dots as donors. Normality testing was performed using the Shapiro-Wilk test. Significance testing was performed using Kruskal-Wallis test with Dunn's multiple comparison correction for nonparametric data, or by ANOVA testing with Tukey's multiple comparison correction for parametric data. P-values <0.05 were considered significant.

4.2.2. Expanded clones in the appendix have distinct viral reactivities in UC compared to CD and CRC

We profiled the expanded viral (clone size > 1) CD8+ TCRs in the appendix to understand whether viral reactivity differed across study groups, as these sequences were more likely to be disease relevant compared to the singlet repertoire. Expanded clones in UC were predominantly predicted to react to respiratory viruses such as Influenza A and SARS-CoV2, while CRC and CD had a more even split between respiratory and latent viruses such as Epstein-Barr (EBV) and human cytomegalovirus (HCMV) (**Figure 4.4a-c**). Common predicted antigens included pp65 in HCMV, BMLF1 in EBV, and matrix protein 1 in flu (**Figure 4.5a-b**). The top three SARS-CoV2 antigens were surface glycoprotein, ORF3a, and nucleoprotein (**Figure 4.5b**). Interestingly, persistence of SARS-CoV2 antigens in the gut of IBD patients has been reported, providing some rationale for the expansion of SARS-CoV2 reactive T cells in the absence of active infection, as patients would not have been admitted to surgery with an active infection[202]. All tissue samples were collected post-COVID-19 pandemic and we anticipate equivalent rates of vaccination across study groups. All UC donors were also negative for CMV colitis.

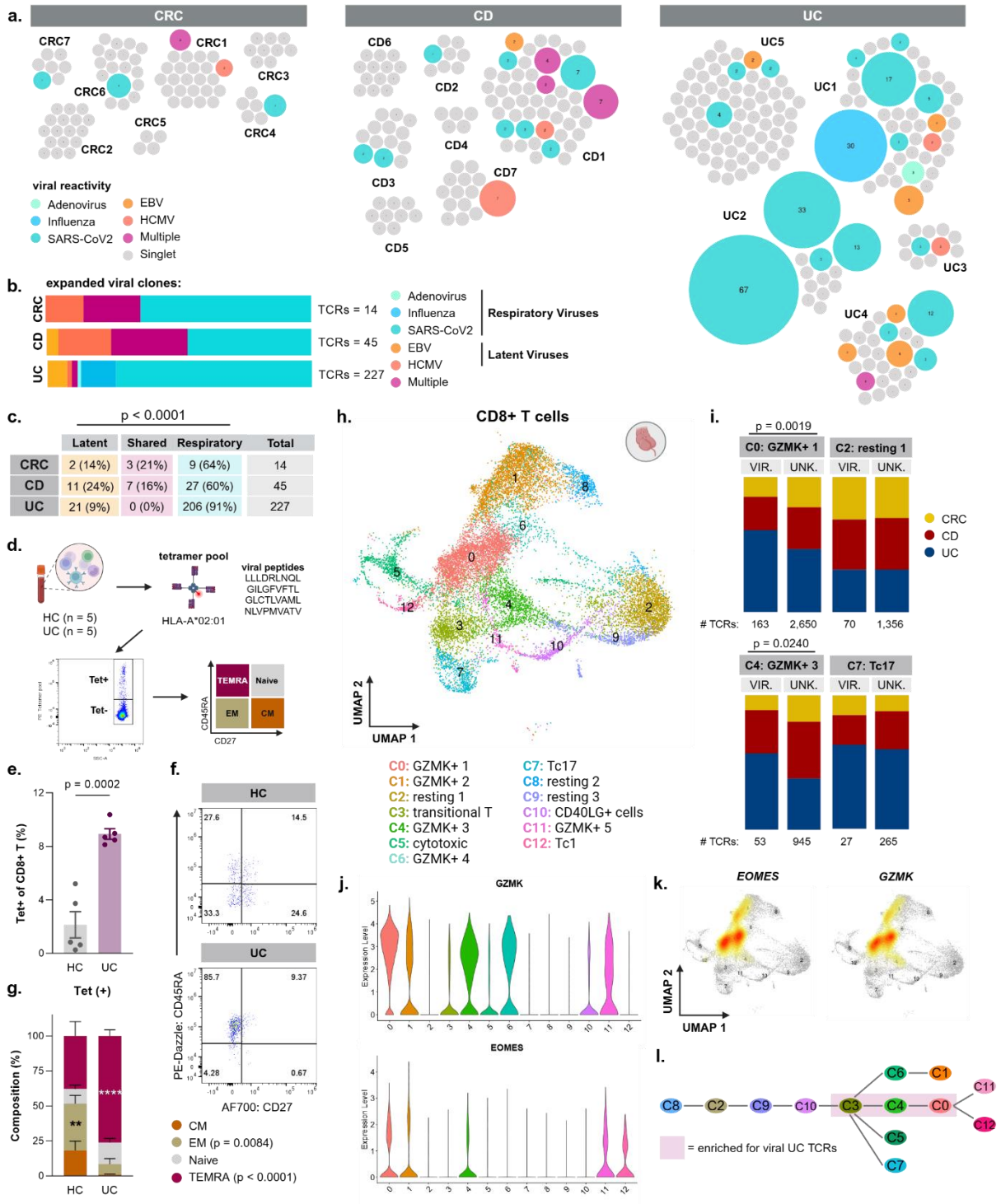


Figure 4.4. Viral-reactive CD8+ T cells in UC have distinct reactivities and functional phenotypes. (a) packed bubble plots of predicted viral reactivities of expanded clones in the scRNA-seq cohort by donor **(b)** relative abundance of expanded viral-reactive TCRs by study group **(c)** contingency table used for Chi-squared testing showing TCR counts and relative abundance for each subset **(d)** workflow schematic for viral tetramer staining of PBMCs **(e)** percentage of Tet+ cells within the CD8+ T population by flow **(f)** representative flow plots of Tet+ cells from PBMCs **(g)** composition of Tet+ CD8+ T cells between study groups **(h)** CD8+ T

UMAP colored by cluster with annotations **(i)** cluster composition of CD8+ T cells with viral reactivity versus unknown reactivity broken down by study group for selected clusters, significant Chi-squared results are shown **(j)** violin plots of *EOMES* (top) and *GZMK* (bottom) expression across clusters **(k)** *EOMES* (left) and *GZMK* (right) positive cells overlaid as density plots onto the CD8 T UMAP **(l)** flowchart showing results from pseudotime analysis. Compositional data was analyzed by two-way ANOVA with Sidak's multiple comparison correction in (g) and by Chi-squared testing in (c) and (i). P-values <0.05 were considered significant.

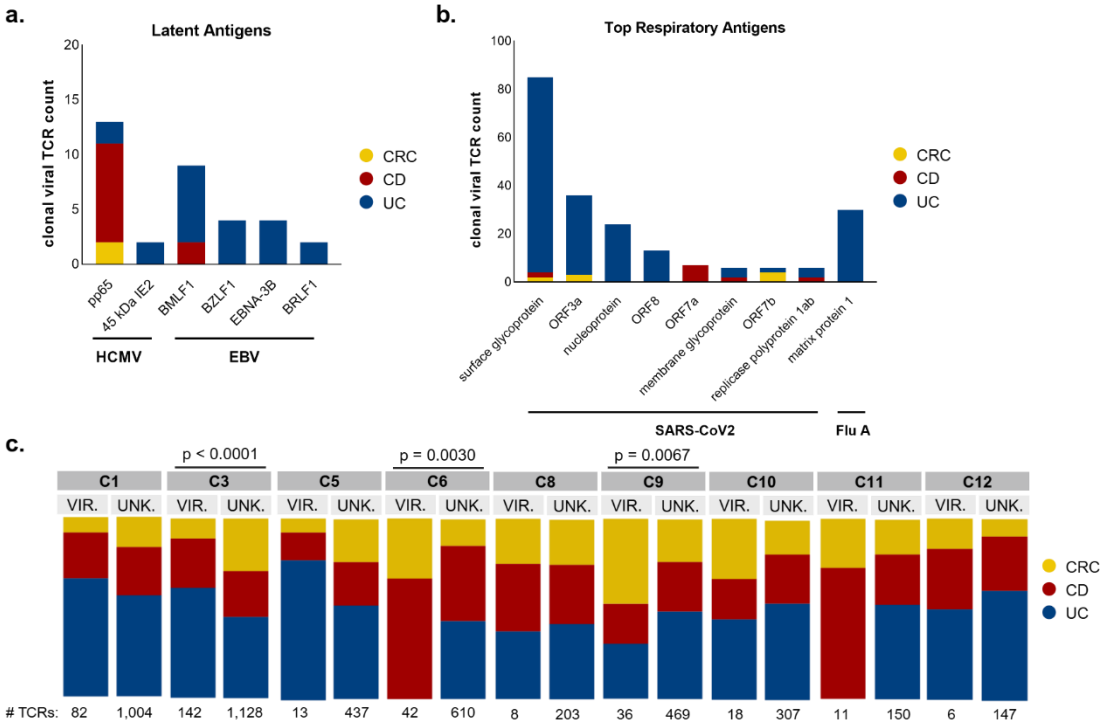


Figure 4.5. Reactivity and function of viral CD8+ TCRs in the appendix. Count of expanded viral TCRs broken down by predicted antigen reactivity and study group for **(a)** latent and **(b)** respiratory viral antigens. **(c)** cluster composition for predicted viral and unknown TCRs by study group for remaining CD8+ T clusters. Significant Chi-squared results are shown.

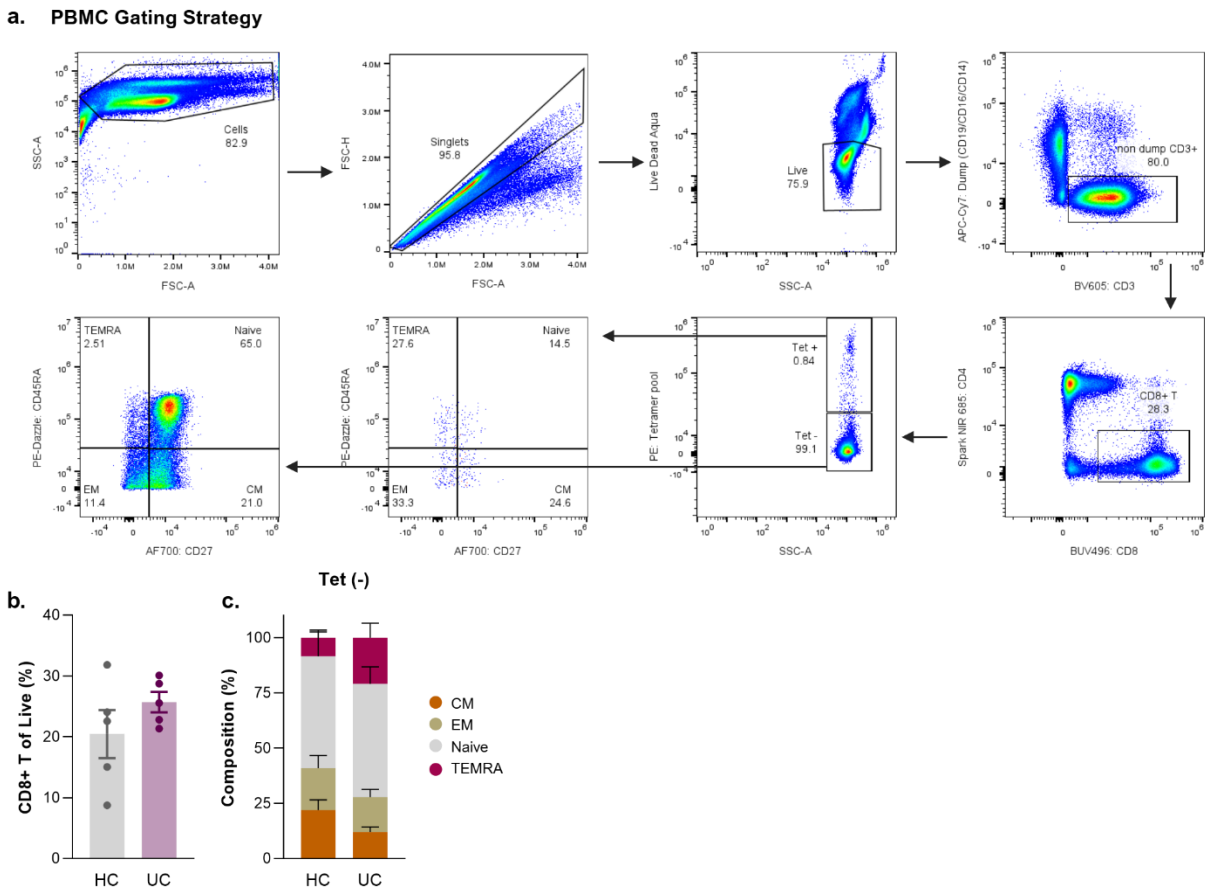


Figure 4.6. PBMC viral tetramer stains. (a) gating strategy for the PBMC tetramer stain experiment **(b)** percentage of CD8+ T cells of live cells **(c)** composition of Tet- CD8+ T cells by study group. Bar plots represent the mean ± SEM for each group with individual dots as donors. Significance testing was performed with Mann-Whitney for nonparametric data. Compositional data was analyzed by two-way ANOVA with Sidak's multiple comparison correction, error bars indicate the SEM for each group. P-values <0.05 were considered significant.

4.2.3. Viral-reactive CD8+ T cells in UC have a terminal effector phenotype in the blood

In order to validate the presence of increased viral-reactive CD8+ T cells in UC, we collected HLA-typed peripheral blood mononuclear cells (PBMCs) from 5 healthy controls (HC) and 5 UC donors, due to ease of collection and evidence there is substantial repertoire overlap between the blood and colon of UC patients (**Table 2.11**)[203]. Samples were stained with HLA-A*02:01 tetramer pools containing viral peptides LLLDRLNQL (SARS-CoV-2 $NCAP_{222-230}$), GILGFVFTL (Influenza $M1_{58-66}$), GLCTLVAML (EBV $BMLF1_{280-288}$), and NLVPMVATV

(CMV pp65$_{495\text{-}503}$) (**Figure 4.4d**). CD8+ T cells were identified by flow cytometry and separated into tetramer-positive and -negative populations (**Figure 4.6a**). While CD8+ T cell proportions were equivalent between HC and UC samples, tetramer-positive cells were significantly increased in UC (**Figures 4.6b, 4.4e**). The viral-reactive cells took on a predominant terminally differentiated (TEMRA) phenotype in UC, while HC samples had an effector-memory phenotype (**Figure 4.4f-g**). As expected, tetramer-negative CD8+ T cells were majority naïve in both HC and UC blood (**Figure 4.6c**). This finding suggests viral-reactive CD8+ T cells have a dysfunctional effector state in UC compared to healthy controls, potentially as a result of chronic antigen stimulation in the gut.

4.2.4. Viral-reactive CD8+ T cells in the UC appendix are enriched in EOMES+GZMK+ clusters

To evaluate CD8+ T functional phenotypes in the gut, we integrated the TCR and GEX data from the scRNA-seq cohort and visualized clusters by uniform manifold approximation and projection (UMAP). This resulted 16,792 appendiceal cells represented across 13 clusters (**Figures 4.4h, 4.7c-d**). To understand whether certain CD8+ T phenotypes were enriched in the viral-reactive repertoire, we compared the cluster composition by study group between the viral and unknown repertoires. This resulted in an observed enrichment of viral-reactive TCRs from UC samples in C0, C3, and C4 (**Figures 4.4i, 4.5c**). This was not a finding common across clusters, as the resting C2 cluster was evenly composed of CRC, CD, and UC cells. It was also not due to overall pathologic changes, as Tc17 cells in C7 were composed of a majority of UC cells, but there was no viral-specific enrichment (**Figure 2i**).

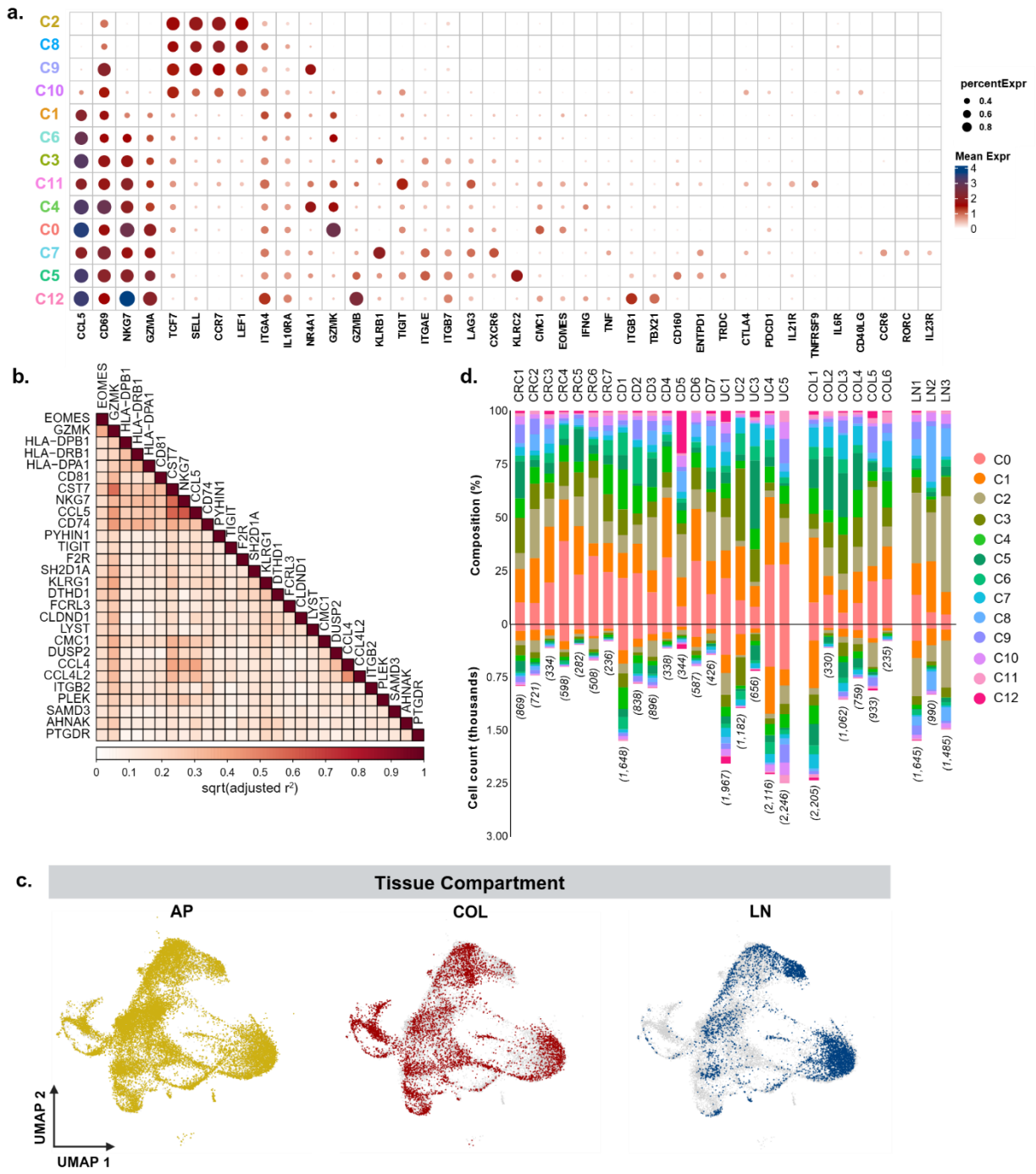


Figure 4.7. Transcriptional signatures and composition of the CD8+ T UMAP. (a) key gene bubble plot for the CD8+ T UMAP, bubble size corresponds to percent of cells expressing a marker and colored corresponds to mean expression level, markers expressed in <20% of cells were excluded **(b)** core signature of genes positively correlated with *EOMES* and *GZMK* expression **(c)** UMAP colored by compartment **(d)** Composition (top) and cell counts (bottom) for clusters broken down by sample, total cell counts added to each column

When looking at these clusters in more depth, C3 expressed tissue residency genes (*ITGAE, ITGB7, KLRB1*) and migratory markers (*CCL5, CD69*) (**Figure 4.7a**). C0 and C4 were two of five clusters with high *GZMK* expression (**Figure 4.4j**). When looking at the dataset, *GZMK* and *EOMES* were often co-expressed and were part of a core gene set including *CMC1, CCL5,* and *NKG7* that was often upregulated together (**Figures 4.4j-k, 4.7b**). Both C0 and C4 appeared to be effector subsets, with C4 expressing increased TCR activation genes (*JUN, FOS, NR4A1*) and cytokine *IFNG,* and C0 expressing increased cytotoxic genes (*GZMA, NKG7*) and inhibitory receptor genes (*TIGIT, LAG3*) (**Figure 4.7a**).

4.2.5. The appendix as a priming site for diverse CD8+ T cell effector phenotypes

To understand how CD8+ T cell states in the appendix are related to one another, we performed trajectory analysis using SlingShot[183]. Unsurprisingly the earliest pseudotime states passed through the three resting clusters C8, C2, and C9 (*LEF1, SELL, TCF7, CCR7*) before moving to C10 which expressed markers of early TCR activation (*CD40LG, IL6R, CD28*) (**Figure 4.4l**). C3, the transitional tissue resident cluster, appeared to be the major bifurcation point, with Tc17 cells in C7 (*RORC, IL23R, TNF, IL18R1*) and cytotoxic cells in C5 (*CD160, TNFRSF18, GZMB, PRF1*) splitting off here (**Figure 4.4l**). The remaining two trajectories contained *GZMK* and *EOMES* expressing clusters. The first had cells in C6, which expressed fewer genes than the rest of the clusters, moving into C1, which did not express classical exhaustion markers but did express genes known to impact T cell dysfunction (*MALAT1, NEAT1, HNRNPA2B1, IKZF3, RNF19A*), indicating a potential terminal cell state (**Figure 4.4l**). The final trajectory began in C4 before moving to C0, which was in line with our expectations based on cytotoxic and inhibitor gene expression, before reaching a final cell state in C11 or C12

(**Figure 4.4l**). C12 contained Tc1 cells (*TBX21*, *GZMB*, *PRF1*, *CX3CR1*) (**Figure 4.7a**). C11 expressed CD137 (*TNFRSF9*) along with inhibitory receptors (*TIGIT*, *LAG3*, *CTLA4*, *PDCD1*) and exhaustion-associated transcription factors (*TOX*, *IRF4*) (**Figure 4.7a**).

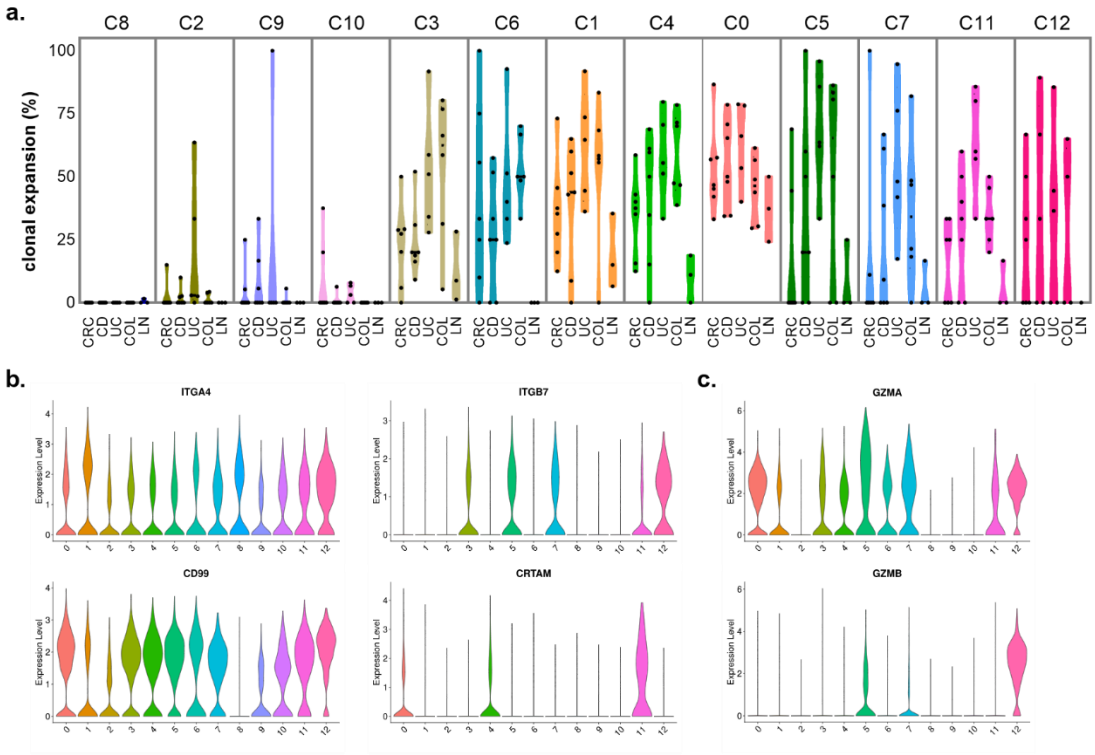


Figure 4.8. Extended measures from the CD8+ T UMAP. (a) truncated violin plots showing percent of clonal expansion across clusters and study groups, each dot represents one sample. Selected gene expression across clusters for **(b)** adhesion and migration markers and **(c)** granzymes.

The trajectory analysis aligned with the percentage of clonal expansion across clusters, with resting clusters having minimal expansion that increased in later cell stages (**Figure 4.8a**). Interestingly, viral-reactive cells were enriched in relatively early cell states in the UC appendix (C0, C3, C4), suggesting that these cells are primed in the appendix, and have the capacity to take on both effector and exhausted phenotypes downstream. This analysis also shows that the highly cytotoxic C5 and Tc17 cells in C7 were independent of this *GZMK* signature. Looking at known gut adhesion and migration markers, *CCR10* and *GPR15* were minimally expressed (data

not shown), while *ITGA4*, *ITGB7*, *ITGAE*, *CXCR6*, and *CD99* were expressed in both positive and negative *GZMK* clusters (data not shown, **Figure 4.8b**). Interestingly, *CRTAM* was expressed only in C0, C4, and C11, suggesting this could be a marker unique to this *GZMK* cell trajectory (**Figure 4.8b**). Inhibitory markers were generally upregulated in later cell stages, with the greatest expression seen in C11. While *GZMA* was expressed almost ubiquitously in effector populations, *GZMB* was expressed only in C5, C7, and C12, indicating this gene is mutually exclusive with *GZMK* (**Figure 4.8c**).

While we focused on viral-reactive cells so far, these were not the only TCRs that could be contributing to immune responses in the appendix. When evaluating the overall CD8+ TCR repertoire, it is apparent that all three study groups contained large clones of unknown reactivity (**Figure 4.9a**). In order to understand the functional potential of expanded clones overall, we looked at cluster mapping for major clones (defined as clone size ≥ 10). Inter-patient variability in TCR function appeared greater than intra-patient variability, with most major clones within a single donor showing similar breakdown in functional profiles (**Figure 4.9b**). Shared functionalities within clones aligned with findings from the trajectory analysis. Clones often mapped to both C4 and C0, with a larger fraction mapping to C0 reflecting a later stage of differentiation. Clones with shared C3 and C7 functional mapping were also commonly observed, and rarely contained cells mapping to C0 which was predicted to be an independent trajectory (**Figure 4.9b**).

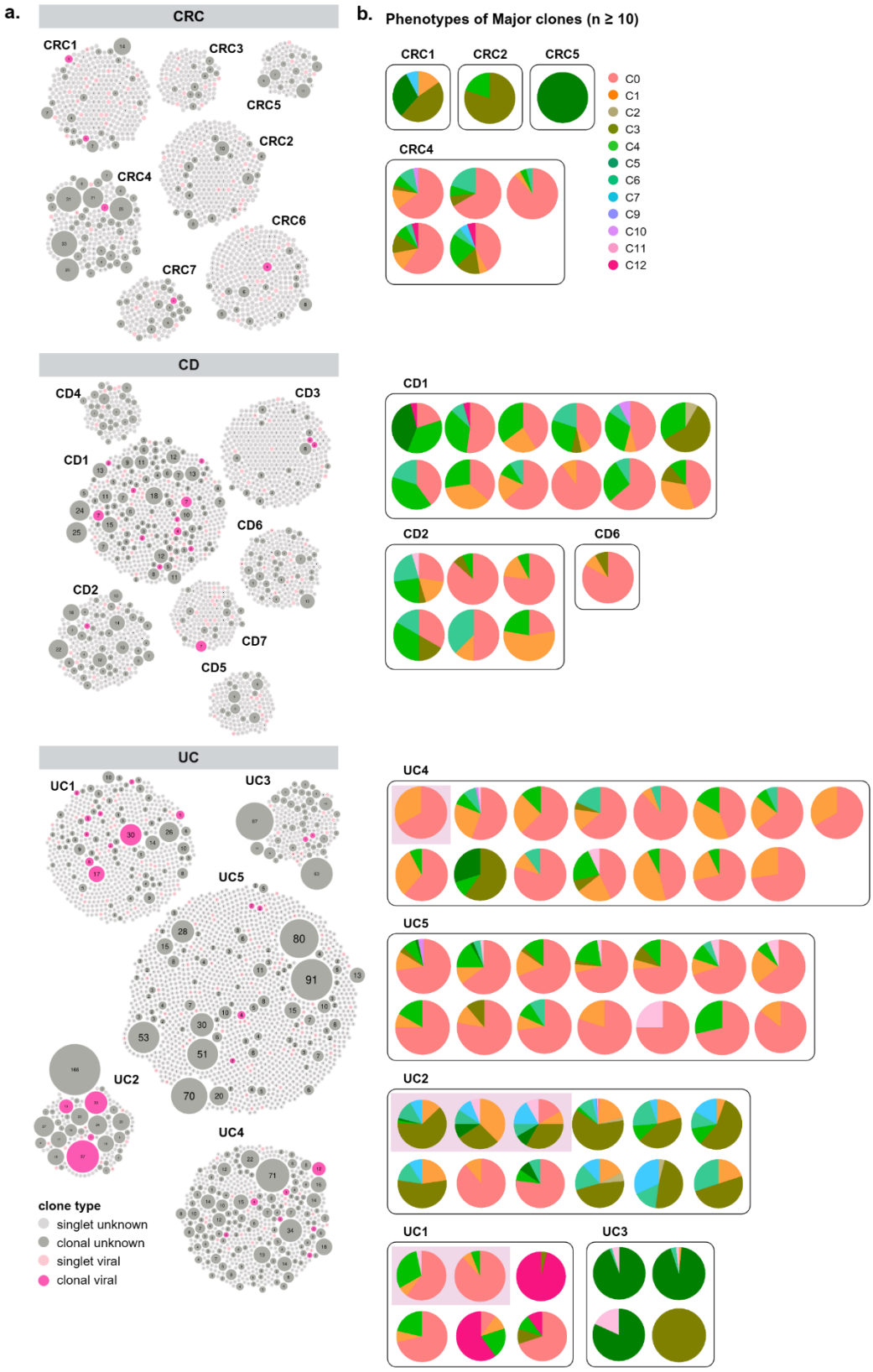

a.
CRC
CRC1
CRC3
CRC5
CRC2
CRC4
CRC6
CRC7
CD
CD4
CD3
CD1
CD6
CD2
CD7
CD5
UC
UC1
UC3
UC5
UC2
UC4
clone type
singlet unknown
clonal unknown
singlet viral
clonal viral
b.
Phenotypes of Major clones (n ≥ 10)
CRC1
CRC2
CRC5
CRC4
C0
C1
C2
C3
C4
C5
C6
C7
C9
C10
C11
C12
CD1
CD2
CD6
UC4
UC5
UC2
UC1
UC3

Figure 4.9.

Figure 4.9. Shared functions across major expanded clones in the appendix. (a) packed bubble plots visualizing reactivities and extent of clonal expansion across all donors in the appendix cohort **(b)** major clones (size ≥ 10) by donor mapped to their functional profile

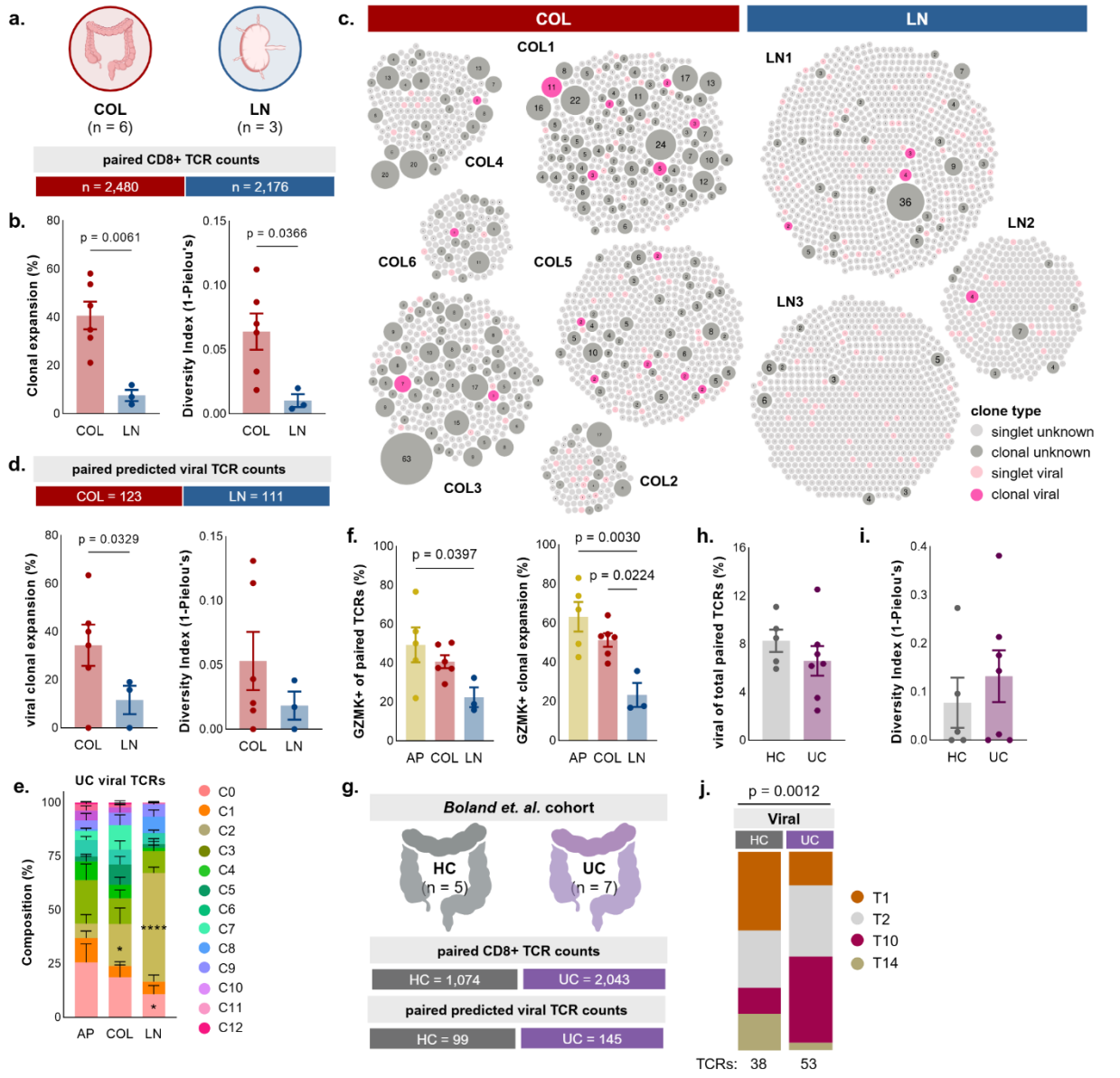


Figure 4.10. *GZMK*+ CD8+ T cells are present in the colon of UC patients in two scRNA-seq cohorts. (a) Summary of CD8+ TCR recovery from the colon and lymph node **(b)** percent of clonal expansion (left) and diversity index (right) from the overall TCR repertoire **(c)** packed bubble plots showing the clone size and reactivity of TCRs **(d)** summary of viral-reactive TCR counts, as well as clonal expansion (left) and diversity index (right) in the viral-reactive repertoire **(e)** composition of viral-reactive TCRs across compartments from UC donors **(f)** percentage of TCRs mapping to a *GZMK*+ cluster (left) and percent of clonal expansion within these *GZMK*+ cells (right) across UC compartments **(g)** summary of CD8+ TCR recovery and viral-matching from an independent scRNA-seq cohort **(h)** percentage of viral matched of total TCRs **(i)** diversity index of viral TCRs **(j)** cluster composition of viral-reactive CD8+ T cells using barcodes that mapped to originally published phenotypes, Chi-squared results are shown. Bar plots represent the mean ± SEM for each group with individual dots as donors. Normality testing was performed using the Shapiro-Wilk test. For nonparametric data, significance testing was performed by Mann-Whitney (2-group) or Kruskal-Wallis test (3-group) with

Dunn's multiple comparison correction. For parametric data, significance testing was performed by T-test (2-group) or ANOVA (3-group) with Tukey's multiple comparison correction. Compositional data in (e) was analyzed by two-way ANOVA with Sidak's multiple comparison correction P-values <0.05 were considered significant.

4.2.6. Viral-reactive CD8+ T cells are clonally expanded in the colon but not the cecum-draining lymph node of UC patients

Our next question was whether viral-reactive cells were also clonally expanded at the site of disease in UC, the colon. We collected tissue and identified CD8+ TCRs from 6 colon pieces (COL, n = 2,480) and 3 cecum draining lymph nodes (LN, n = 2,176) from UC patients (**Figures 4.1-4.2, 4.10a**). While CD8+ TCR counts were increased in the lymph node, clonal expansion and clonal dominance was significantly increased in the colon overall (**Figures 4.10b-c, 4.11a-b**). Viral-reactive CD8+ TCRs were identified (COL = 123, LN = 111), which again accounted for approximately 5% of the repertoire (**Figures 4.10d, 4.11c**). Viral-reactive CD8+ T cells were significantly clonally expanded, with a trend towards increased clonal dominance in the colon compared to the lymph node of UC donors (**Figure 4.10d**). Expanded viral-reactive clones in the colon were also predicted to react to predominantly respiratory viruses (**Figure 4.11d**).

4.2.7. GZMK+ CD8+ T cells are present in the colon of UC patients in two scRNA-seq cohorts

To understand the function of these cells, colon and lymph node CD8+ T cells were integrated into the UMAP. When looking at the viral CD8+ T repertoire, the lymph node had a significant decrease in C0 cells, with the predominant phenotype being resting cells compared to the other compartments (**Figure 4.10e**). The appendix and colon looked similar to one another in terms of cell composition, further supporting the appendix as a distinct niche compared to the

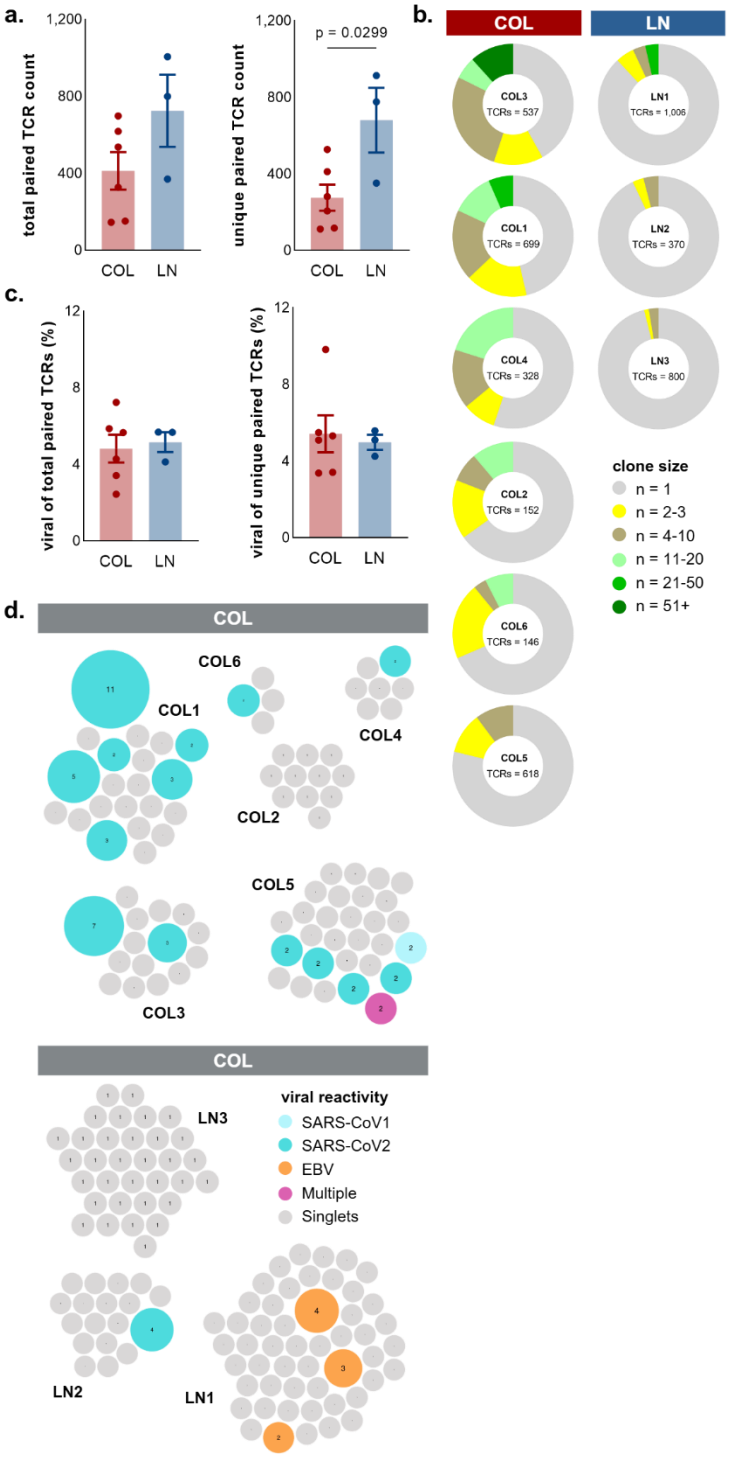


Figure 4.11. The CD8+ TCR repertoire in colon and lymph node of UC patients. (a) total (left) and unique (right) paired TCR counts **(b)** pie charts showing clonal expansion by donor, TCRs grouped by clone size, total TCR counts are included within each chart **(c)** percentage of viral matches by total (left) and unique (right) TCRs **(d)** packed bubble plots of viral-reactive repertoires by donor, expanded clones are colored for predicted reactivity. Bar plots represent the mean ± SEM for each group with individual dots as donors. Significance testing was performed by T-test for parametric, or Mann-Whitney test for nonparametric data. P-values < 0.05 were considered significant.

nearby SLO of the cecum-draining lymph node (**Figure 4.10e**). When looking at the overall CD8+ T repertoire, the percentage of cells belonging to a *GZMK*+ cluster (C0, C1, C4, C6, C11)

was significantly higher in the appendix than the lymph node, and clonal expansion of these cells was increased in both the appendix and colon compared to the lymph node (**Figure 4.10f**). This suggests these cells are sourced from the appendix and travel to the colon, rather than being activated in the nearby lymph node.

To validate our scRNA-seq cohort findings, we obtained TCR and GEX data from a previously published scRNA-seq cohort of rectal mucosal biopsies[121]. A similar filtering process led to identification of CD8+ TCRs from 5 HC (n = 1,074) and 7 UC (n = 2,043) donors (**Figures 4.10g, 4.12a-b**). TCR counts and clonal expansion measures were similar between study groups (**Figure 4.12c-e**).Viral matching was performed similarly identifying 244 sequences (HC = 99, UC = 145), although HLA matching was not possible for this cohort, which may explain the slight increase in predicted viral reactive repertoire (**Figures 4.10g-h, 4.12f**). Expansion of viral-reactive TCRs was highly variable yet equivalent between study groups, although there was a trend towards increased clonal dominance in UC (**Figures 4.10i, 4.12g-h**). Interestingly, when TCRs were mapped back to their functional clusters from the original publication, viral-reactive TCRs from UC patients increasing mapped to a *GZMK*-expressing cluster (T10 in the original publication) (**Figure 4.10j**). T10 had similar gene expression to the *GZMK*+ clusters we described (*EOMES, IFNG, CRTAM, KLRG1, TIGIT*) and showed a high degree of clonal expansion in line with our earlier observations[121]. By contrast, virus-specific cells from the healthy controls largely exhibited a TRM-like phenotype that lacked *EOMES* expression but was characterized by ICOS and IL7R expression (designated T1 in the original publication)[121].

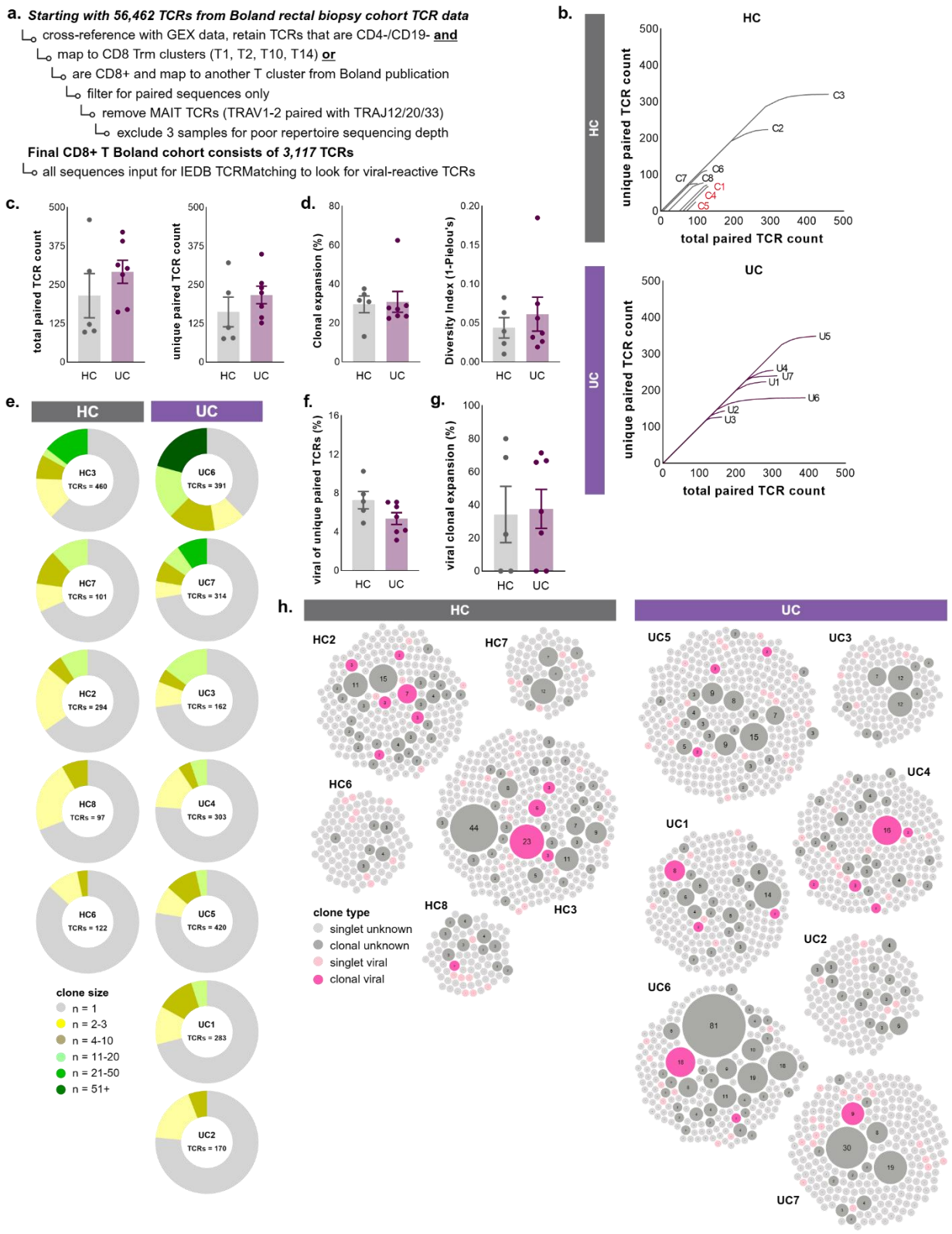


Figure 4.12. Evaluating viral reactivity in the *Boland et. al.* CD8+ T repertoire. (a) workflow summarizing filtering and viral matching of scRNA-seq barcodes from an independent cohort of rectum mucosal biopsies

(b) saturation plots evaluating TCR repertoire sequencing depth by study group, samples excluded for poor depth are highlighted in red **(c)** total (left) and unique (right) paired TCR counts for the overall repertoire **(d)** clonal expansion (left) and diversity index (right) for the overall repertoire **(e)** pie charts showing clonal expansion by donor, TCRs are colored by size with total counts within each chart **(f)** percent of viral matches of unique TCRs **(g)** clonal expansion within the viral repertoire **(h)** packed bubble plots of the CD8+ T repertoire colored by clone size and predicted reactivity broken down by donor and study group. Bar plots represent the mean ± SEM for each compartment with individual dots as donors. Significance testing was performed by Mann-Whitney for nonparametric, and T-test for parametric data. P-values <0.05 were considered significant.

4.2.8. Granzyme K+ CD8+ T cells are increased in the inflamed colon of UC patients when the appendix is intact

In order to understand whether the appendix acts as a source for Granzyme K+ CD8 T cells, we identified archival tissue sections from 5 UC patients with an appendix (+AP) and 6 UC patients who had previously undergone appendectomy (-AP) at the time of colectomy (**Table 2.12**). Sections were cut and stained for immunofluorescence staining and imaging, followed by selection of representative regions of interest (ROIs) of CD8+ T infiltration in the mucosa (-AP = 60, +AP = 50 ROIs) (**Figure 4.13a**). Patient tissues had similar severity scores both by IBD-DCA and an overall pathology score (**Figure 4.13b**). ROIs were equivalent by area and nuclei count (**Figure 4.13c-d**). CD8+ T cell infiltration was reduced in donors lacking an appendix, and within CD8+ T cells, there was also a reduction in the percentage of granzyme K positive cells (**Figure 4.13e-f**). This finding supports the appendix as a source for granzyme K+ CD8 T cells, and in the absence of the organ equivalent cells cannot infiltrate or be activated within the colon itself.

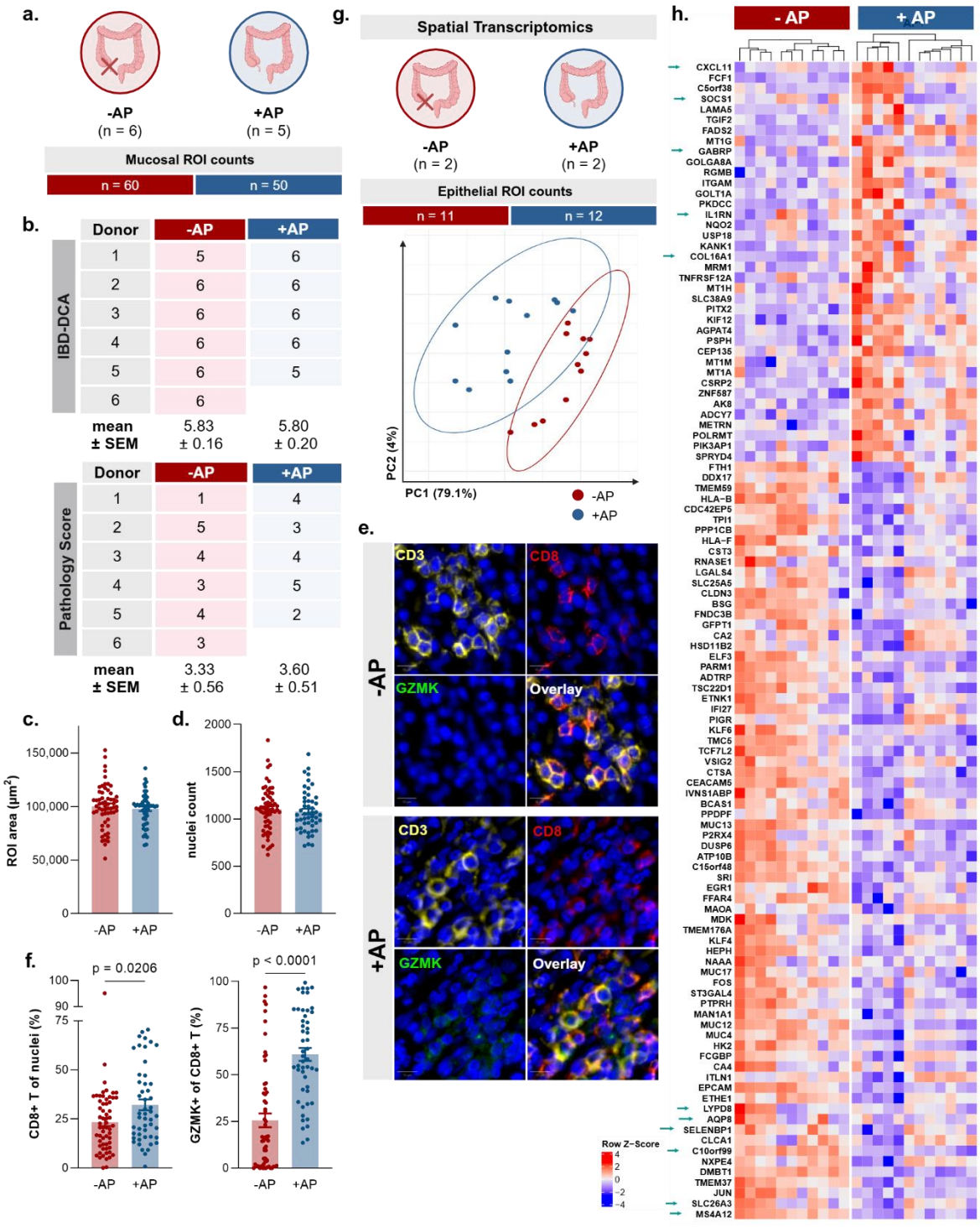


Figure 4.13. Granzyme K+ CD8+ T cells are increased in the colon of UC patients with appendices compared to UC patients with prior appendectomy. (a) summary of ROI collection for immunofluorescence study in UC colon tissue sections **(b)** severity measures by IBD-DCA score (top) and overall pathology score (bottom) of inflamed tissue sections **(c)** ROI area and **(d)** nuclei count per ROI **(e)** representative images of CD8+ T infiltrated colonic lamina propria **(f)** percent of CD8+ T of nuclei (left) and GZMK+ of CD8+ T (right) **(g)** summary of epithelial segment ROI collection for spatial transcriptomics (top) and PCA plot of epithelial ROIs colored by study group with ellipses (bottom) **(h)** heatmap of differentially

expressed genes (adjusted p value < 0.05), genes called out in the text are annotated with an arrow. Bar plots represent the mean ± SEM for each group with individual dots as donors. Significance testing was performed by T-test for parametric, or Mann-Whitney for nonparametric data. P-values < 0.05 were considered significant.

4.2.9. Colonic epithelium from UC donors with intact appendices further upregulates inflammatory genes

To further evaluate the impact of the appendiceal immune compartment on UC pathology, a subset of tissue sections was used for spatial transcriptomics using the Nanostring GeoMx platform (**Table 2.12**). Epithelial ROIs were collected from -AP (n = 11) and +AP (n = 12) tissues and processed according to the manufacturer's workflow to evaluate changes in the epithelial layer (**Figure 4.13g**). Genes detected from the GeoMx probes were cross-referenced to a reference dataset of 2,389 genes previously identified in scRNA-seq analysis of colonic epithelial cell [118]. Epithelial segments separated well by PCA plot, and DEseq2 identified 109 differentially expressed genes based on an adjusted p-value cutoff of 0.05 (**Figure 4.13g-h**). Many of the genes upregulated in the +AP segments are upregulated in inflamed colonic segments from the reference dataset (e.g., *CXCL11, SOCS1, GABRP, IL1RN,* and *COL16A1*). *CXCL11* and *SOCS1* have known roles in promoting epithelial-mesenchymal transition in cancer, which could have implications for the increased CRC risk in UC[204,205]. Conversely, genes that are upregulated in the -AP specimens tend to be upregulated in uninflamed or healthy reference segments of the colon (e.g., *MS4A12, SLC26A3, C10orf99,* and *LYPD8*). Specifically, genes associated with normal absorptive colonocyte and goblet cell functions, like extracellular redox homeostasis (e.g., *SELENBP1*), epithelial barrier integrity and anti-microbial response (e.g., *SLC26A3, LYPD8),* and normal nutrient absorption (e.g., *AQP8*) are upregulated in -AP specimens (**Fig. 4h**). Reduced expression of *AQP8* in UC has been associated with the disruption of epithelial polarity and barrier dysfunction[206]. This is in line with a previous report that showed

GZMK alone is sufficient to induce epithelial tissue remodeling, suggesting that in the absence of $GZMK^+$ effector $CD8^+$ T cells, distinct (appendix-independent) mechanisms drive disease pathology[207,208].

CHAPTER 5: DISCUSSION

5.1 The appendix and appendicitis

This study provides a high-resolution characterization of the appendix during appendicitis, a common yet poorly understood disease. Spatial transcriptomic analysis of the lymphoid follicles found increased IgG genes and B cell activation, indicating an antigen-driven response, as IgG is canonically produced in a T cell- and antigen-dependent fashion. BCR repertoire analysis confirmed this observation and showed enrichment of IgG in the germinal center and plasma cell compartments, further supporting an antigen-driven mechanism. In the T cell compartment, I propose that CD8+ T cells are driven by bystander activation, as there was no clear evidence of antigenic drive based on the TCR repertoire analysis. However, some evidence of antigenic drive was seen in the Tfh CD4+ subset based on motif enrichment by GLIPH2.0 analysis. I hypothesized that the microbiome is a source of antigen during appendicitis. To support this idea, I observed loss of epithelial barrier integrity in the spatial transcriptomics dataset, as well as dysbiosis of the appendiceal microbiome during disease. This led us to perform exploratory analysis of potential microbial derived-MHC II restricted peptides using a subset of metagenomic sequences from appendicitis samples. As BCR epitopes remain technically difficult to predict, this gives us a hint as to which peptide sequences the CD4+ T-follicular helper cells may be recognizing to assist in germinal center reactions. I identified predictive epitopes from microbial groups that have been reported to increase in richness in appendicitis, providing important foundational evidence that warrants further investigations into antigen-specific B cell response in appendicitis disease pathogenesis.

I classified appendicitis cases as perforated versus unperforated upon gross examination as a proxy for complicated versus uncomplicated disease. Complicated appendicitis is typically defined as gangrenous or perforated tissue, although some micro-perforations may require histological analysis. Uncomplicated appendicitis cases show evidence of inflammation but no perforation. This classification is clinically relevant, as a recent review found diagnostic imaging suggestive of a complicated disease course are indicative of increased risk of antibiotic failure[13]. Persisting challenges in identifying patients where antibiotic therapy is appropriate provides a rationale for further research into the disease mechanisms of appendicitis.

It is up for debate whether appendicitis is a progressive disease, where left untreated all cases become complicated, or if these classifications represent two distinct diseases with unique inflammatory mechanisms. In support of appendicitis as a progressive disease, retrospective studies have found that delaying surgery by greater than 24 hours increases the risk of perforation and complications[209]. Additionally, while overall incidence of appendicitis was decreased, perforation rates increased during the COVID-19 pandemic, presumptively due to treatment delays while patients and providers were trying to limit contact[210]. While the progressive versus two-disease debate is an open question, I found increased B cell activation and IgG production in both perforated and unperforated samples in the transcriptomic datasets, supporting the role of the humoral response in both “types” of appendicitis.

Appendicitis is typically a pathology with sudden onset on the scale of hours, with surgical or antibiotic intervention happening relatively soon afterwards. However, germinal center development and B cell maturation within the follicle takes place over the course of 1-3 weeks generally, which seemingly invalidates the possibility of B cells playing a role in disease here. However, while appendicitis itself is ischemia created by obstruction of the appendiceal

lumen, which is responsible for the sudden onset of pain, lymphoid hyperplasia is a common cause of this obstruction[12,13]. Lymphoid hyperplasia is observed commonly in response to infection, and often subsides on its own as the infection is resolved. This pathology can go undetected for weeks as follicles become enlarged and begin responding to antigen, allowing for the eventual class switching and somatic hypermutation in the B cell compartment described in this study. In the absence of obstruction by fecalith or tumor, it is reasonable to consider lymphoid hyperplasia as a cause of appendiceal obstruction. This patient subset may also be more likely to respond to non-surgical antibiotic intervention if there is low risk of perforation as determined by a physician.

The appendix has largely been neglected in scientific literature due to the assumption it is a vestigial organ. While non-essential, the appendix houses dense lymphoid tissue and a resident microbiota, which are intimately linked[1]. Follicle development coincides with bacterial colonization of the appendix in early human life[6]. The appendix is a major producer of IgA, which helps to maintain microbial homeostasis at mucosal surfaces[1,6,7,105]. Due to its narrow lumen and relative isolation from the fecal stream, the appendix has been hypothesized to act as a microbial reservoir that can re-seed the gut following infection or other upset. In support of this idea, bariatric surgery patients who underwent prophylactic appendectomy had reduced fecal microbial diversity compared to patients with intact appendices at follow-up[4]. The appendix has been associated with multiple diseases, including colorectal cancer, ulcerative colitis, and Parkinson's disease, via mechanisms unknown[71,211–214]. Additional research of the appendiceal environment may reveal new insights into both organ function as well as the pathophysiology of multiple chronic diseases.

A limitation of this study was difficulty in obtaining appropriate control samples. Control samples used in the IgG+ BCR repertoire comparison as well as the 16S microbiome analysis were primarily from colorectal cancer patients with appendices uninvolved by disease. In general, these control patients were much older than the appendicitis patients and as a result, age may be a confounding factor. However, I identified similar patterns in IgG in the spatial cohort, which did have age-matched controls. Additionally, I identified similar microbial features in this cohort as previous literature reports, supporting the validity of these results.

The presence of subgingival plaque taxa is a common finding in appendicitis, and in line with this, I identified *Fusobacterium* and *Fretibacterium* in this dataset. Subgingival plaque is a biofilm, and the appendix also contains dense biofilms[3]. The network correlation analysis found that *Fusobacterium, Fretibacterium, Parvimonas,* and *Peptostreptococcus* were all positively correlated with one another in appendicitis. Interestingly, *Parvimonas micra* and *F. nucleatum* have been shown to have synergistic biofilm formation, and *Peptostreptococcus micros* has been shown to co-aggregate with *F. nucleatum*[215,216]. Additional species assignment using the Silva taxonomy database for the 16S data found that the majority of *Fusobacterium* sequences belonged to *F. nucleatum,* with the remainder unassigned. *P. micra* and *Peptostreptococcus stomatis* were the predominant species in their respective genera. These three species, *F. nucleatum, P. micra,* and *P. stomatis* have previously been associated with colorectal cancer (CRC), and were proposed as part of a biomarker panel for identifying CRC patients[217,218]. These species have been implicated in multiple inflammatory conditions, including appendicitis, periodontitis, and cancer, and could be an area of future study looking at the role of biofilm formation during disease states.

It is possible that the aforementioned genera form appendiceal biofilms that are better able to survive in the inflammatory conditions of appendicitis, leading to the relative outgrowth observed in the 16S cohort. Periodontal disease has previously been described as having inflammophilic bacterial communities, meaning the species present thrive under inflammatory conditions, allowing them to outcompete other species[219,220]. Gut dysbiosis has been linked to increased IL-32 production, and both *Parvimonas* and *Fusobacterium* have been shown to induce inflammatory cytokines like IL-6 and IL-8, suggesting that IL-32 mediated inflammatory network might play a pivotal role in appendicitis pathogenesis[220,221]. *Fusobacterium* was also shown to invade the epithelial barrier in appendicitis[194]. In a model of periodontitis, this invasion was mediated via adhesin FadA, and the authors found negative effects of bacterial invasion could be attenuated via treatment with piperlongumine or fisetin[221]. These studies highlight the role bacterial invasion plays in mucosal inflammation and raise the possibility of drug repurposing to optimize therapeutic strategies in appendicitis.

5.2. CD8+ T cells of the appendix and IBD

A summary of findings as well as remaining open questions from the CD8+ T IBD study are outlined in **Figure 5.1** and discussed in more detail in this section. This study identifies a population of clonally expanded predicted viral-reactive CD8+ T cells in the appendix of UC patients. It remains an open question how CD8+ T cells in UC are being exposed to viral antigens, or if this is a case of molecular mimicry. It is well established that IBD patients have increased susceptibility to viruses due to use of immunomodulatory agents, and both latent viruses and SARS-CoV2 antigens have been observed in the colonic tissue[200,202,222,223]. However, our donor tissue was negative for CMV-colitis, and immunomodulatory agents are prescribed in

both CD and UC, yet expansion of these cells in the appendix was almost unique to UC samples, which supports an underlying biological difference.

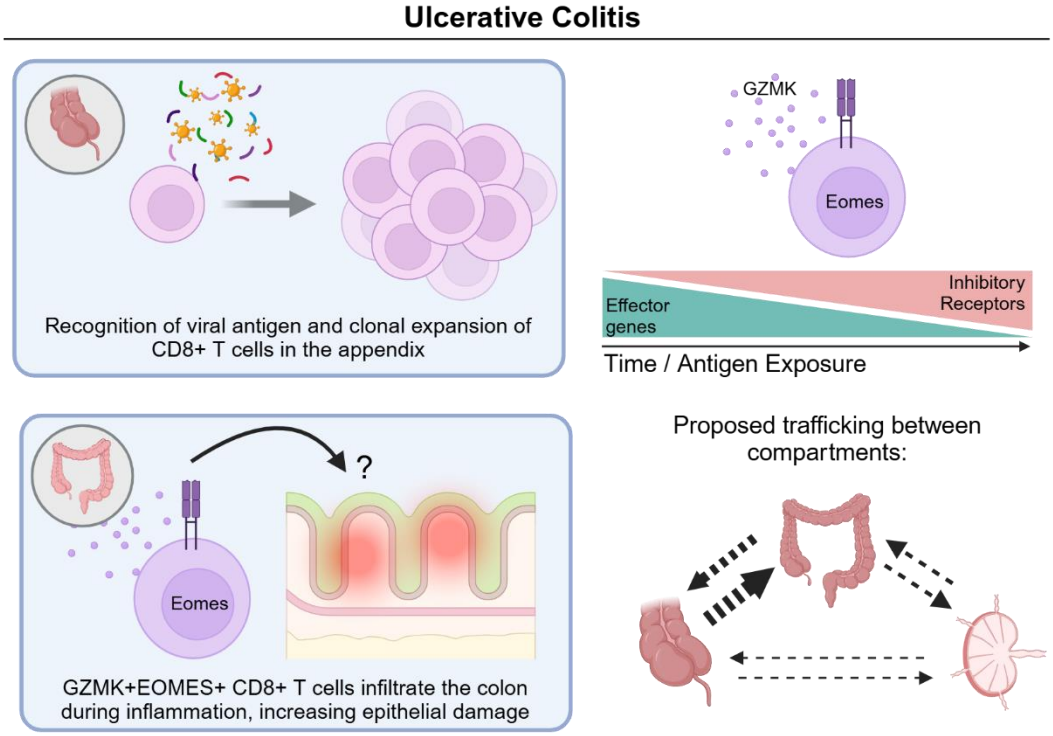


Figure 5.1. Summary of findings. The model here describes the increased clonal expansion of viral-reactive CD8+ T clones found in the appendix that are enriched for GZMK+EOMES+ phenotype (top left). In the colon, these cells infiltrate the mucosa in inflamed tissue during UC of patients with an intact appendix, resulting in increased epithelial inflammation through a yet unknown mechanism (bottom left). GZMK+EOMES+ CD8+ T cells appear to take on an increased exhaustion phenotype over time, although this still needs to be verified experimentally (top right). Finally, overlaps in receptor repertoires and cell phenotypes suggests increased trafficking between the appendix and colon relative to the lymph node and colon, with minimal trafficking occurring between the appendix and lymph nodes. Mechanisms of cell movement between the appendix and other tissues as well as extent of trafficking is yet unknown (bottom right)

In addition to the predominance of predicted SARS-CoV2 reactive CD8+ T cells in UC, there was also a relative increase in the proportion of EBV reactive TCRs compared to the other study groups (**Figure 4.4b**). EBV infected B cells have previously been observed in the colon of UC patients, and in some cases show evidence of viral replication[224]. EBV-infected cells were more frequently observed in UC compared to CD or control cases, suggesting the immune response is uniquely failing to control EBV infection in this setting, allowing for increased viral

antigen in infected individuals. In line with this report, preliminary analysis of our scRNA-seq dataset found EBV genes specifically in UC donor samples. While this is not a conclusive finding, as our analysis was not designed to measure viral transcripts, and is limited to detection of DNA viruses, it supports the increased presence of EBV antigens in the UC samples compared to other study groups.

It is possible that EBV-infected B cells densely packed into the appendix act as a source of antigen, leading to recognition by CD8+ T cells. EBV infection led to induction of CD8+ T cells with a tissue resident phenotype in a humanized mouse model[225]. These cells showed cytotoxic potential *ex vivo*, yet failed to control EBV infection in the mucosa, while CD8+ T cells in the spleen showed greater effector function and infection control. Our CD8+ T subset of interest also showed expression of tissue residency markers, and showed cytotoxic potential based on mRNA expression of GZMK and GZMA. However, functional capacity was not evaluated in our T cell population so it has not yet been verified whether these cells have the ability to control viral infection or produce cytokines. Based on the upregulation of inhibitory receptors in the GZMK+ CD8+ T cells over the course of our trajectory analysis, it is likely that these cells lose effector potential over time as they are chronically exposed to low levels of viral antigen, leading to the eventual dysfunctional state seen in both C11 in the scRNA-seq data as well as in the tetramer-positive PBMCs of UC patients.

Because the appendix is a site of dense B cell infiltration, and EBV-infected B cells were increasingly found in the colon of UC patients in the literature, this provides a plausible explanation for increased viral antigen burden in the appendix of UC patient specifically, leading to increased clonal expansion of EBV-reactive CD8+ TCRs. For viruses such as SARS-CoV2 which infect epithelial cells rather than B cells, for example, it is less clear how CD8+ T cells

increasingly respond to viral antigen in UC but not in CD or CRC. SARS-CoV2 antigen has been shown to persist in the intestines of IBD patients for months following infection, likely facilitated in part due to the epithelial barrier disruption that is a hallmark of both diseases[202]. Differences in SARS-CoV2 reactive CD8+ T cells observed in our data may be due in part to the compartments analyzed. Viral-reactive CD8+ T cells may be present in greater proportion in the ileum and Peyer's patches of the small intestine in CD, as this is the most common site of disease. Differences in cell and antigen trafficking between the colon and appendix, and appendix and small intestine could possibly explain the differences seen in expanded viral-reactive clones between UC and CD in our dataset.

Another possibility is that these CD8+ T cells don't necessarily recognize viral epitopes, but similar sequences from another source. NCBI Blast searches reveal that while viral epitopes show minimal overlap with self-proteins, they do show overlap with bacterial epitopes, including species that are known to inhabit the human gut. The microbiome is known to be altered in UC, and epithelial barrier disruption paves the way for potential entry by viral and/or bacterial antigens into the tissue. The observation of viral-reactive CD8+ T cells with exhausted-like phenotypes in the blood and scRNA-seq datasets of UC donors suggests these cells are being chronically stimulated by antigen over long periods of time, leading to eventual dysfunction.

We propose that the *GZMK*+ CD8+ T phenotype is primed in the appendix, where they are able to traffic to the colon. This is supported by the observation that relatively earlier cell stages (C3, C4, C0) had enrichment of viral-reactive CD8+ T cells in UC, which was not present in later stage effectors like Tc1 or Tc17 cells. These cells were also most present and the most clonally expanded in the appendix, followed by the colon, in both the overall CD8+ T repertoire and viral-reactive CD8+ T repertoire. The lymph node had a significant reduction in these cells,

and was primarily composed of resting singlets, suggesting this SLO is not the source for *GZMK*+ CD8+ T cells. Finally, image analysis showed that granzyme K+ CD8+ T cells were decreased in UC patients who underwent prior appendectomy, demonstrating that in the absence of the appendix, the colon does not compensate by inducing more of these cells. Spatial transcriptomic analysis of the epithelium from these tissue sections found upregulation of pro-inflammatory markers, providing evidence that these cells contribute to colitis pathology in the colon.

While we believe Granzyme K+ T cells are trafficking from the appendix to the colon, it is unknown whether there are specific mechanisms controlling migration between these two tissues. While past studies have defined a role for CCR10 and GPR15 for movement of immune cells to the colon, these genes were minimally expressed in our dataset[105,107]. Components of the $\alpha_4\beta_7$ integrin (*ITGA4*, *ITGB7*) were widely expressed across clusters, in line with the known efficacy of vedolizumab. CRTAM, which was present only in *GZMK*-expressing clusters, is known to support tissue residency of T cells in the small intestine and promotes CD8+ T infiltration into tumors[226,227]. The expression pattern in this dataset suggests this marker could be used to selectively target pathogenic Granzyme K+ cells sourced from the appendix and may be a viable target molecule for novel IBD therapeutic strategies.

Granzyme K+ CD8+ T cells have been described previously in both healthy controls and inflamed tissue samples from patients with autoimmune diseases[228–230]. Blood and synovial fluid from rheumatoid arthritis patients contained IFNγ-producing GZMK+CD8+ (similar to our C4) that were distinct from highly cytotoxic GrB+CD8+ T cells (similar to our C5)[229,230]. Notably, GZMK+GZMB+ CD8+ T cells responded to both TCR and cytokine stimuli while single positive GZMK+ CD8+ T cells only responded to TCR stimuli[229]. This supports antigen-

dependent activation within the *GZMK*+ CD8+ T populations reported here, which largely lacked *GZMB* expression in the scRNA-seq dataset (**Figure 4.8c**). A virology study found that infection history had a profound impact on CD8+ T cell marker expression in the blood, with EBV inducing GZMK+ cells and HCMV inducing GZMKB+ cells[231]. In this study samples came from latently infected but otherwise healthy adults and found that viral-reactive cells took on largely effector and central memory phenotypes, in line with our findings in healthy controls. This demonstrates a role for viral infection producing differential CD8+ T cell responses based on epitope recognition and inflammatory context, an interesting observation in light of increased viral-reactive terminal-effector CD8+ T cells seen in the blood of UC patients.

This study proposes that the GZMK+ CD8+ phenotype is primed in the appendix. Unlike GZMB, which cleaves caspases 3 and 7 to induce apoptosis of a target cell, GZMK does not directly activate caspase cascades. Rather, it is known to induce apoptosis indirectly, or to interact with extracellular receptors PAR-1 and PAR-2, leading to induction of IL-6, IL-8, and CCL2 by fibroblasts and endothelial cells *in vitro*[232]. However, appendix-dependent mechanisms are not the only contributors to IBD inflammation. While analyzing the CD8+ T cell data it became apparent that *TNF* was almost exclusively expressed by Tc17 cells, a trajectory that was independent of the *GZMK* signature enriched in the UC appendix. GSEA analysis of the epithelial spatial transcriptomics segments found enrichment of a TNF responsiveness pathway in the -AP group, along with increased expression of SLC26A3, which protects against barrier disruption in response to TNFα[233]. TNF is a known driver of IBD pathology, and observation of its involvement in multiple transcriptomics datasets suggests that this is an appendix-independent pathway during intestinal inflammation.

The results described here provide the first detailed description of the immune compartment of the human appendix across multiple inflammatory conditions and proposes that the appendix is a unique niche within the gut capable of priming GZMK+ CD8+ T cells which contribute to inflammation in UC in an antigen-specific manner. In a broader sense, this project suggests that the appendix has a distinct role as a SLO compared to gut LNs, and more closely mirrors the immune populations in the colon, while LNs are composed of predominantly naïve singlet T cell clones. While we show an enrichment for GZMK+ CD8+ T cells in the appendix and colon in UC, these cells are also present in CRC and CD appendix. It remains an open question what role these cells have, if any, in these conditions considering the lack of clinical association between appendectomy and CD or CRC. In CD specifically, GZMK+ T cells have been observed in the ileum, but have not been as widely reported in diseased tissue as GZMK+ T cells in UC[229]. To our knowledge an increase in GZMK+ CD8+ T cells in CD relative to healthy controls has not been reported. One possible hypothesis for the differential impact of these cells in CD versus UC may be the difference in trafficking of these cells to other intestinal tissue, or the contribution of Peyer's patches and immune cells from the small intestine-draining lymph nodes. A prior report found that Peyer's patch cells traffic almost exclusively to the small intestine, while appendiceal cells traffic to both the small and large intestine[105]. It is also appreciated that SLOs across the intestine induce different adaptive responses, with LNs in the small intestine inducing more tolerogenic responses, while large intestinal LNs are more inflammatory[234]. It may be an interesting avenue to compare the appendix to the Peyer's patches in the small intestine, due to their similar morphology and immune-priming ability, to determine if cells from both sites have a differential impact on inflammation in the ileum in CD versus the colon in UC.

Potential limitations of this study are largely related to patient cohorts and availability of public TCR sequences. The IBD patients used in our scRNA-seq cohort had severe medication refractory disease, so all observations may not be generalizable to less severe patients. To offset this, blood samples were collected from UC patients during scheduled clinic visits when disease was often controlled or minimal and we still observed changes in the viral-reactive CD8+ T population. In a similar vein the Boland et. al. cohort was taken from biopsy samples which can have a larger range in disease severity and displayed a similar *GZMK*+ population. Samples may also have been impacted by medications used to treat their disease, although we believe our findings to be distinct to UC, as CD tissue was also collected and did not show the same findings, and we were able to replicate our observations across multiple UC cohorts. With regards to viral-reactivity we were limited to searching known TCR sequences, so it is highly possible that additional sequences that react to viral epitopes were not identified. Public databases were also highly impacted by SARS-CoV2 research, and a large number of entries are for SARS-CoV2 epitopes. While our scRNA-seq samples were all collected post-pandemic, the Boland cohort was not, and we still observed some predicted SARS-CoV2 TCRs. This may be due to cross-reactivity to other coronaviruses, or possibly other antigen sources as discussed above.

5.3. Future directions

5.3.1. Additional models to evaluate the GZMK+ CD8+ T cell mechanism in UC

While human studies are beneficial in the sense that observations here are true representations of human disease, an inherent limitation of human sample research is the

inability to establish causation in the same way that is possible using animal models. While we have an observed presence of GZMK+ CD8+ T cell in the UC appendix, and that these cells are associated with inflammation in the colon in UC patients who still have an appendix, it remains unknown whether this cell subset is truly driving disease. This question could be addressed through the use of murine models, each with their own set of challenges.

Humanized mouse models would be advantageous here to study the impact of the observed CD8+ T cell subset on colitis. Transfer of CD34+ hematopoietic stem cells (HSCs) or PBMCs into immunodeficient NSG mice followed by induction of colitis using a 7-day DSS protocol are established preclinical models to study UC pathology. A prior group has even used PBMCs from IBD patients, demonstrating increased disease pathology compared to healthy control blood following induction of colitis[235]. A similar model could be used for future experiments, sourcing immune cells from both healthy donors and UC patients, as we suspect this colitogenic profile is unique to UC. However, it is unknown if the GZMK+EOMES+ CD8+ T subset of interest is present in the blood. While we observed increased viral-reactive CD8+ T cells with a terminal effector phenotype, additional phenotyping of these cells would be needed to confirm a similar mRNA phenotype in the blood. If the phenotype is confirmed by additional scRNA-seq of PBMCs, transfer experiments could be performed with blood samples. However, the protocol could also be adapted to test transfer of appendiceal CD45+ immune cells into NSG mice, to determine whether these cells can be successfully grafted. Using appendiceal immune cells would be preferred if possible, given our hypothesis that the appendiceal niche is a critical source of this CD8+ T subset.

The first step here would be establishing the initial model of colitis induction, using either bulk PBMCs or appendiceal immune cells (possibly from both healthy and UC samples) ,

and confirming that (1) the CD8+ T cell subset is present in the transferred cells and (2) colitis induction is possible using this transfer model coupled with DSS. Afterwards another experiment using CD8 depleting antibodies can be used to deplete CD8 T cells prior to colitis induction. If CD8 T cells overall are contributing to inflammation, mice with CD8 depletion will experience reduced disease severity compared to mice who were administered and isotype control antibody. Additionally, to evaluate if CD8+ T cells alone can drive colitis, bulk CD8+ T cells from either PBMCs or the appendix can be transferred prior to DSS colitis induction. Disease severity would be anticipated to be worse than mice with CD8+ T transfer where colitis was not induced, and either intermediate or on par with DSS exposed mice who experienced bulk immune cell transfer.

So far these experiments would establish a colitis-inducing effect of appendiceal immune cells, and bulk appendiceal CD8+ T cells, yet does not address the specific subset identified in our human samples. To evaluate this, the GZMK+EOMES+ CD8+ T subset of interest would need to be reliably identifiable by FACS. I propose CRTAM here as a unique surface marker of this subset that was identified in the scRNA-seq data that could be identified by FACS for sorting of these cells (following confirmation of this marker at the protein level). CRTAM+ and CRTAM- CD8+ T cells from the appendix could be sorted and transferred into humanized mice as before, followed by DSS induction of colitis. Prior FACS experiments have yielded very low viability of appendiceal cells post-spin down in our hands, in which case it may be necessary to perform magnetic enrichment for the equivalent cell populations and verifying purity by flow cytometry prior to transfer.

It remains an open question whether the colitogenic function of the *GZMK+ EOMES+* CD8+ T cells is the direct effect of GZMK, or if it is due to the larger cell program downstream

of EOMES. CD8+ T cells could be modified to express an empty GFP lentiviral vector, or one with either GZMK or EOMES. These cells could once again be transferred, followed by induction of colitis using DSS. A similar experiment was previously conducted in $Rag1^{-/-}$ mice and found EOMES-overexpressing CD8+ T cells did in fact exacerbate colitis severity[121]. I suspect it is the larger EOMES-driven cell program, rather than the effects of GZMK itself that are responsible for exacerbating colitis. In part based on the results of this single mouse study, but also due to the known role of EOMES in driving memory and terminal effector programs in other disease models, as well as the reduced cytotoxic potential of GZMK compared to other molecules such as GZMB. It may be worthwhile to pursue future ATAC-seq experiments in CD8+ T cells to better understand the transcriptional regulation downstream of EOMES. Our study has identified genes that are commonly upregulated along with EOMES (GZMK, CCL5, CMC1, NKG7, CRTAM) although it remains unclear what the exact functional implications these cells have in the tissue, and how these genes are regulated.

It is possible that establishing a humanized mouse model will fail, either due to (1) lack of this CD8+ T population in blood or (2) failure to establish colitis following appendiceal immune cell transfer. If this is the case, standardized mouse models may need to be used. Appendectomy followed by DSS induction of colitis has been shown to reduce disease severity, and lead to reduced infiltration of CD8+ T cells to the colon in mice[77,79]. Once again, it is not known if an equivalent CD8+ T population exists in the mouse appendix, which would need to be evaluated to establish validity of the model. Our study also hypothesizes that these cells arise in the setting of viral antigen presentation, possibly with the background of chronic intestinal inflammation. Induction of these cells may be possible using a chronic LCMV infection model to with or without chronic UC modeled using multiple rounds of DSS exposure. This method, if

successful, would require significant optimization but could potentially lead to induction of an analogous CD8+ T cell population in mice, allowing for manipulation of this system to investigate the colitogenic potential of this T cell subset. An analogous set of experiments could be performed using CD8+ T depletion antibodies, T cell transfers into RAG knockout mice, and appendectomy to evaluate both CD8+ T cells in driving colitis as well as the appendix as a priming site or source of these cells.

Another route to support a causal mechanism between the GZMK+EOMES+ CD8+ T cells and ulcerative colitis pathology would be implementing human organoid models. Colonic organoids derived biopsy samples or stem cells can recapitulate the epithelial compartment *in vitro*[236,237]. If colitis cannot be recapitulated in either humanized or standard mouse models, this method could still provide valuable insight. The CD8+ T cell subset of interest could be sorted or enriched from the appendiceal immune compartment and transferred into cultures, as could transduced primary cells expressing GFP-GZMK or GFP-EOMES. This would allow a reductionist approach to evaluate the impact of these T cells on the intestinal epithelium. Gene expression changes in the epithelial cells could be evaluated for signs of inflammation and damage.

5.3.2. Targeting the causative CD8+ T subset in UC

The CD8+ T cell subset we observed were clonally expanded in medically refractory UC patients, suggesting that these cells are resistant to existing treatment modalities. If additional models as described above support these cells as being causative in colitis, it could be worth pursuing targeting these cells using a novel therapeutic. While the CD8+ T subset is defined in the results as being *GZMK+EOMES+*, neither of these are attractive targets, GZMK because it is

non-specific, and EOMES because it is a transcription factor, and also plays a major role in T cell responses. The ideal target would be specific to this subset, and minimally expressed in other cell types and tissues. Here I am using CRTAM as an example target candidate, as it was exclusively expressed in the scRNA-seq data in CD8+ T cell clusters of interest, although there are likely additional molecules to pursue. Additional analysis of the scRNA-seq data could provide more candidates with specific expression in the population of interest. Expression of the target would have to be confirmed at the protein level using either immunofluorescence, flow cytometry or Western blot for example.

Another factor in evaluating target candidates is expression across other cell types and tissues. For example, CRTAM is highly expressed in the cerebellum at the RNA level, with unknown protein expression according to the Human Protein Atlas. Additional IHC staining for CRTAM across human tissues could help elucidate unintended targeting of a biologic. CRTAM is also expressed in NK cells and T cells and helps control intestinal retention of T cells in the mucosa. Interestingly, *CRTAM* also appears to be expressed on different clusters than the combined *ITGA4* and *ITGB7* that make up the α4β7 integrin in our dataset (Figure 4.8b). Targeting trafficking of lymphocytes to the intestines is a successful therapeutic strategy as evidenced by vedolizumab, and this data suggests that CRTAM targeting would impact a different subset of CD8+ T cells in the intestines, possibly providing a unique therapeutic benefit.

After target selection, antibody candidates could be generated and evaluated in preclinical models. CRTAM and other candidate target antibodies could be screened using phage display libraries or through immunization of transgenic mice to generate fully human antibodies[238]. Fully human antibodies are less immunogenic than chimeric or humanized antibody products.

Additional screening would occur to increase the affinity of candidates for their target, evaluate specificity of binding, and to optimize the constant region as well as glycosylation patterns, which can impact downstream effector properties[239]. Antibody candidates can be evaluated for efficacy in ameliorating colitis via administration prior to induction of colitis in an established mouse model as described earlier. If efficacious, I would expect the candidate antibody to have reduced colitis severity compared to an isotope control or DSS alone. This initial experiment would provide proof of concept data and provide rationale for further preclinical experimentation targeting CRTAM or another marker specific to this CD8+ T cell subset. Existing therapies such as anti-TNF and anti-integrin agents could be used to perform head-to-head comparisons between existing biologics and the novel candidate(s).

5.3.3. Rationale to investigate additional immune cell compartments

These findings also provide rationale for future studies looking at changes in other appendiceal immune cell compartments during intestinal disease. While not discussed in detail here, the appendix is a major site of IgA production, and also houses dense B cell and plasma cell populations[1]. Considering the known role of IgA in regulating the intestinal microbiome, it is possible that the appendix contributes to maintenance of homeostasis in the gut. Dysbiosis is an appreciated feature of chronic intestinal diseases like IBD and CRC, and it is reasonable to consider that the appendiceal B cell response is altered during disease, contributing to some of the observed microbial changes. In the appendicitis dataset both IgG and IgA showed evidence of clonal expansion, suggesting that IgA cells also respond to inflammatory stimuli in the appendix.

Investigating the appendiceal CD4+ T cell compartment is also a reasonable future direction. In IBD CD4+ T cells have a known role in pathology, and previous reports have identified increased activation and infiltration of these cells into the appendix in IBD patient tissues. Preliminary data from our lab has observed that in scRNA-seq samples, while expansion of CD4+ T cells is limited overall compared to CD8+ T cells, there is increased clonal expansion in the UC appendix compared to CD or CRC. Considering the dysregulation of Treg versus effector T cells in IBD, it is also of interest to better understand the phenotype of CD4+ T cells in the appendix, and whether this SLO imparts a regulatory or pro-inflammatory effect on these cells, and how that may change at steady state versus during disease. In mice, a handful of microbial antigens have been determined to impact CD4+ T cell differentiation and activation in the gut, as discussed in more detail in Section 1.5.5. However, studying antigen-specific responses in human CD4+ T cells remains a significant challenge, in part due to the lack of publicly available sequences and limitations in manipulating human systems compared to murine models. Lack of viable methodologies to interrogate CD4+ T cell antigen specificity in human samples is a major limitation of probing T cell repertoires in diseased states, particularly when the antigens are not clearly defined as is the case in IBD.

5.3.4. Computational challenges in predicting TCR recognition

Understanding the rules that govern TCR recognition of peptide-MHC complexes, and building accurate tools to predict successful binding, is a major challenge in the field. Solving this problem would have major implications for every antigen-dependent T cell mediated disease process, including anti-tumor responses, autoimmunity, and infection. However, the expansive

variability in the T cell receptor repertoire, and the relative paucity of known receptor-antigen binding pairs limits our ability to predict TCR recognition of antigen.

TCR recognition occurs in an MHC-restricted manner. This means that antigens need to be presented on the surface of a cell by an MHC molecule. Within humans, there is significant variability in MHC allele expression, with humans expressing 6 MHC I alleles, and 6 to 8 MHC II alleles. Frequencies of MHC molecules will vary by ethnic background and geographic region, so sampling diverse populations is critical when building models for TCR recognition of peptide-MHC (pMHC). MHC molecules are highly polymorphic, which allows for presentation of the vast peptide pool they must encounter from potential pathogens, but also introduces complexity when trying to understand TCR binding. Once an MHC molecule is loaded with a peptide sequence and presented on the surface of the cell, there are specific regions on the TCR that come into direct contact with the molecule that are thought to be especially important in (pMHC) recognition. These complementarity determining regions (CDRs) are present on both the α and β chain of the TCR. CDR1 and CDR2 contact the MHC molecule directly, and are germline encoded within the V-gene region. The CDR3 sequence, however, contacts the peptide directly and is hypervariable, allowing for recognition of highly diverse peptide sequences an individual may encounter throughout their lifetime[240]. The CDR3 sequence is thought to be a critical portion in conferring specificity of antigen recognition, and as such is used to predict binding in existing computational models.

The theoretical diversity of TCR sequences is vast – on the scale of 10^{15} sequences prior to thymic selection[241]. While the real diversity within a single individual is much lower, restricted by the number of T cells within the human body, along with thymic selection and MHC haplotype, there is still huge variability in the population level diversity of the TCR

repertoire. When considering TCR diversity, antigen peptide sequences, and MHC allele diversity all combining to specific TCR-pMHC interactions, the likelihood of predicting any specific binding pair is exceedingly low. By generating datasets of experimentally validated TCR-pMHC pairs and using this data to understand the "rules" that govern binding, predictive models could be generated to more accurately identify TCRs that bind to specific antigens.

Developing computational models is dependent on the quality of existing TCR-pMHC datasets. Advances in library screening offer higher-throughput methods compared to bulk or single-cell TCR sequencing, and may help accelerate dataset generation in the future[242,243]. However, a recent publication reported that less than 1 million TCR sequences are listed in publicly available datasets, a minority of which provide paired α and β chain information[243]. These results are also biased to include more frequent HLA alleles, meaning the diversity of TCR-pMHC binding is not sufficiently captured in the existing datasets available today.

Some computational models exist to predict TCR recognition of peptides, although accuracy is limited. Existing methods such as TCRMatch and GLIPH that were used in the results section rely largely on the CDR3 sequence similarity, and only take the β chain into account when calculating sequence similarity, ignoring the α chain altogether[167,184]. The TCRdist algorithm uses both the α and β CDRs to provide a score that can be used to cluster TCRs based on sequence similarity[244]. However, this requires some knowledge of the antigen specificity for at least a fraction of TCRs in the clustered dataset, and still requires additional experimentation to evaluate binding of similarly-clustered TCRs.

More recent models rely on deep neural networks to try and predict TCR binding. Using the area under the receiver-operating characteristic curve (ROC-AUC) as a metric of accuracy, where 50% ROC-AUC indicates a random guess, supervised predictive models appear to show

some increased accuracy for a limited number of epitopes, showing an ROC-AUC score in the range of 60-80%[243]. This is true when an epitope is "seen" meaning the epitope and known TCRs were used in the training dataset to identify additional TCRs that are likely to bind. This increased accuracy is often lost for "unseen" epitopes that were not included in the training dataset, again limiting the usefulness of these algorithms models. Computational models are only as good as the training data used to develop them, and the limited accuracy in existing models today highlights the need for additional efforts to generate datasets reflecting the true diversity of TCR-pMHC binding.

REFERENCES

1. Kooij, I. A., Sahami, S., Meijer, S. L., Buskens, C. J. & te Velde, A. A. The immunology of the vermiform appendix: a review of the literature. *Clinical and Experimental Immunology* **186**, 1–9 (2016).
2. Corr, S. C., Gahan, C. C. G. M. & Hill, C. M-cells: origin, morphology and role in mucosal immunity and microbial pathogenesis. *FEMS Immunology & Medical Microbiology* **52**, 2–12 (2008).
3. Bollinger, R., Barbas, A. S., Bush, E. L., Lin, S. S. & Parker, W. Biofilms in the large bowel suggest an apparent function of the human vermiform appendix. *Journal of Theoretical Biology* **249**, 826–831 (2007).
4. Sánchez-Alcoholado, L. *et al.* Incidental Prophylactic Appendectomy Is Associated with a Profound Microbial Dysbiosis in the Long-Term. *Microorganisms* **8**, 609 (2020).
5. Im, G. Y. *et al.* The appendix may protect against Clostridium difficile recurrence. *Clin Gastroenterol Hepatol* **9**, 1072–1077 (2011).
6. Gebbers, J.-O. & Laissue, J.-A. Bacterial Translocation in the Normal Human Appendix Parallels the Development of the Local Immune System. *Annals of the New York Academy of Sciences* **1029**, 337–343 (2004).
7. Andreu-Ballester, J. C. *et al.* Secretory immunoglobulin A (sIgA) deficiency inserum of patients with GALTectomy (appendectomy and tonsillectomy). *Clinical Immunology* **123**, 289–297 (2007).
8. Eriksen, C. *et al.* IgG and IgM cooperate in coating of intestinal bacteria in IgA deficiency. *Nat Commun* **14**, 8124 (2023).

9. Catanzaro, J. R. *et al.* IgA-deficient humans exhibit gut microbiota dysbiosis despite secretion of compensatory IgM. *Sci Rep* **9**, 13574 (2019).
10. Guan, L. *et al.* The global, regional, and national burden of appendicitis in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *BMC Gastroenterol* **23**, 44 (2023).
11. Lotfollahzadeh, S., Lopez, R. A. & Deppen, J. G. Appendicitis. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024).
12. Bhangu, A., Søreide, K., Di Saverio, S., Assarsson, J. H. & Drake, F. T. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *The Lancet* **386**, 1278–1287 (2015).
13. Moris, D., Paulson, E. K. & Pappas, T. N. Diagnosis and Management of Acute Appendicitis in Adults: A Review. *JAMA* **326**, 2299 (2021).
14. Bom, W. J., Scheijmans, J. C. G., Salminen, P. & Boermeester, M. A. Diagnosis of Uncomplicated and Complicated Appendicitis in Adults. *Scand J Surg* **110**, 170–179 (2021).
15. Rivera-Chavez, F. A. *et al.* Innate Immunity Genes Influence the Severity of Acute Appendicitis. *Annals of Surgery* **240**, 269–277 (2004).
16. Peeters, T. *et al.* An observational study of innate immune responses in patients with acute appendicitis. *Sci Rep* **10**, 17352 (2020).
17. Bi, Y. *et al.* The Gut Microbiota and Inflammatory Factors in Pediatric Appendicitis. *Disease Markers* **2022**, 1–11 (2022).

18. Dimberg, J., Rubér, M., Skarstedt, M., Andersson, M. & Andersson, R. E. Genetic polymorphism patterns suggest a genetic driven inflammatory response as pathogenesis in appendicitis. *Int J Colorectal Dis* **35**, 277–284 (2020).
19. Zhong, D. Acute appendicitis in children is associated with an abundance of bacteria from the phylum Fusobacteria. *Journal of Pediatric Surgery* 6 (2014) doi:10.1016/j.jpedsurg.2013.06.026.
20. Rogers, M. B., Brower-Sinning, R., Firek, B., Zhong, D. & Morowitz, M. J. Acute Appendicitis in Children Is Associated With a Local Expansion of Fusobacteria. *Clin Infect Dis.* **63**, 71–78 (2016).
21. Salö, M. *et al.* Evaluation of the microbiome in children's appendicitis. *Int J Colorectal Dis* **32**, 19–28 (2017).
22. Vanhatalo, S. *et al.* Appendiceal microbiome in uncomplicated and complicated acute appendicitis: A prospective cohort study. *PLoS ONE* **17**, e0276007 (2022).
23. Blohs, M. *et al. Acute Appendicitis Manifests as Two Microbiome State Types with Oral Pathogens Influencing Severity*. http://biorxiv.org/lookup/doi/10.1101/2022.04.13.488268 (2022) doi:10.1101/2022.04.13.488268.
24. Jackson, H. T. *et al.* Culture-Independent Evaluation of the Appendix and Rectum Microbiomes in Children with and without Appendicitis. *PLoS ONE* **9**, e95414 (2014).
25. Paulina Salminen, MD, PhD; Hannu Paajanen, MD, PhD; Tero Rautio,MD, PhD; Pia Nordström, MD, PhD; Markku Aarnio,MD, PhD;, Tuomo Rantanen, MD, PhD; Risto Tuominen, MPH, PhD; Saija Hurme, MSc; Johanna Virtanen, MD; Jukka-Pekka Mecklin, MD, PhD;, & Juhani Sand, MD, PhD; Airi Jartti, MD; Irina Rinta-Kiikka, MD, PhD; Juha M. Grönroos, MD, PhD. Antibiotic Therapy vs. Appendectomy for Treatment of

Uncomplicated Acute Appendicitis: The APPAC Randomized Clinical Trial. *JAMA* **313(23):2340–2348**, (2015).

26. Salminen, P. *et al.* Five-Year Follow-up of Antibiotic Therapy for Uncomplicated Acute Appendicitis in the APPAC Randomized Clinical Trial. *JAMA* **320**, 1259 (2018).
27. Varadhan, K. K., Neal, K. R. & Lobo, D. N. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. *BMJ* **344**, e2156–e2156 (2012).
28. Alatab, S. *et al.* 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. *The Lancet Gastroenterology & Hepatology* **5**, 17–30 (2020).
29. Ng, S. C. *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *The Lancet* **390**, 2769–2778 (2017).
30. Lewis, J. D. *et al.* Incidence, Prevalence, and Racial and Ethnic Distribution of Inflammatory Bowel Disease in the United States. *Gastroenterology* **165**, 1197-1205.e2 (2023).
31. Dahlhamer, J. M. Prevalence of Inflammatory Bowel Disease Among Adults Aged ≥18 Years — United States, 2015. *MMWR Morb Mortal Wkly Rep* **65**, (2016).
32. Weisman, M. H. *et al.* Inflammatory Bowel Disease Prevalence: Surveillance data from the U.S. National Health and Nutrition Examination Survey. *Prev Med Rep* **33**, 102173 (2023).
33. Torres, J., Mehandru, S., Colombel, J.-F. & Peyrin-Biroulet, L. Crohn's disease. *The Lancet* **389**, 1741–1755 (2017).

34. Berre, C. L., Honap, S. & Peyrin-Biroulet, L. Ulcerative colitis. *The Lancet* **402**, 571–584 (2023).

35. Roda, G. *et al.* Crohn's disease. *Nat Rev Dis Primers* **6**, 1–19 (2020).

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

37. Stidham, R. W. & Higgins, P. D. R. Colorectal Cancer in Inflammatory Bowel Disease. *Clinics in Colon and Rectal Surgery* **31**, 168–178 (2018).

38. Spehlmann, M. E. *et al.* Risk factors in German twins with inflammatory bowel disease: results of a questionnaire-based survey. *J Crohns Colitis* **6**, 29–42 (2012).

39. Tysk, C., Lindberg, E., Järnerot, G. & Flodérus-Myrhed, B. 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**, 990–996 (1988).

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

41. Ng, S. C., Woodrow, S., Patel, N., Subhani, J. & Harbord, M. Role of genetic and environmental factors in British twins with inflammatory bowel disease. *Inflamm Bowel Dis* **18**, 725–736 (2012).

42. Gordon, H., Trier Moller, F., Andersen, V. & Harbord, M. Heritability in Inflammatory Bowel Disease: From the First Twin Study to Genome-Wide Association Studies. *Inflammatory Bowel Diseases* 1 (2015) doi:10.1097/MIB.0000000000000393.

43. Uniken Venema, W. T., Voskuil, M. D., Dijkstra, G., Weersma, R. K. & Festen, E. A. The genetic background of inflammatory bowel disease: from correlation to causality. *The Journal of Pathology* **241**, 146–158 (2017).

44. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).

45. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).

46. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979–986 (2015).

47. Saez, A., Herrero-Fernandez, B., Gomez-Bris, R., Sánchez-Martinez, H. & Gonzalez-Granado, J. M. Pathophysiology of Inflammatory Bowel Disease: Innate Immune System. *Int J Mol Sci* **24**, 1526 (2023).

48. Feagan, B. G. *et al.* Vedolizumab as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* **369**, 699–710 (2013).

49. Sandborn, W. J. *et al.* Vedolizumab as Induction and Maintenance Therapy for Crohn's Disease. *New England Journal of Medicine* **369**, 711–721 (2013).

50. Uhlig, H. H. & Powrie, F. Translating Immunology into Therapeutic Concepts for Inflammatory Bowel Disease. *Annual Review of Immunology* **36**, 755–781 (2018).

51. Uzzan, M. *et al.* Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity. *Nat Med* **28**, 766–779 (2022).

52. Palm, N. W. *et al.* Immunoglobulin A Coating Identifies Colitogenic Bacteria in Inflammatory Bowel Disease. *Cell* **158**, 1000–1010 (2014).

53. Sellon, R. K. *et al.* Resident Enteric Bacteria Are Necessary for Development of Spontaneous Colitis and Immune System Activation in Interleukin-10-Deficient Mice. *Infect Immun* **66**, 5224–5231 (1998).

54. Qian, B.-F., Tonkonogy, S. L., Hoentjen, F., Dieleman, L. A. & Sartor, R. B. Dysregulated luminal bacterial antigen-specific T-cell responses and antigen-presenting cell function in HLA-B27 transgenic rats with chronic colitis. *Immunology* **116**, 112–121 (2005).

55. Madsen, K. L. *et al.* Antibiotic therapy attenuates colitis in interleukin 10 gene–deficient mice. *Gastroenterology* **118**, 1094–1105 (2000).

56. Shan, Y., Lee, M. & Chang, E. B. The Gut Microbiome and Inflammatory Bowel Diseases. *Annu. Rev. Med.* **73**, 455–468 (2022).

57. Glassner, K. L., Abraham, B. P. & Quigley, E. M. M. The microbiome and inflammatory bowel disease. *Journal of Allergy and Clinical Immunology* **145**, 16–27 (2020).

58. Halfvarson, J. *et al.* Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* **2**, 17004 (2017).

59. Lee, S. *et al.* Transient colonizing microbes promote gut dysbiosis and functional impairment. *npj Biofilms Microbiomes* **10**, 1–11 (2024).

60. Pittayanon, R. *et al.* Differences in Gut Microbiota in Patients With vs Without Inflammatory Bowel Diseases: A Systematic Review. *Gastroenterology* **158**, 930-946.e1 (2020).

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

62. Magro, D. O. *et al.* Remission in Crohn's disease is accompanied by alterations in the gut microbiota and mucins production. *Sci Rep* **9**, 13263 (2019).

63. Hall, A. B. *et al.* A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med* **9**, 103 (2017).

64. Feng, J. *et al.* Efficacy and safety of fecal microbiota transplantation in the treatment of ulcerative colitis: a systematic review and meta-analysis. *Sci Rep* **13**, 14494 (2023).

65. Boicean, A., Birlutiu, V., Ichim, C., Anderco, P. & Birsan, S. Fecal Microbiota Transplantation in Inflammatory Bowel Disease. *Biomedicines* **11**, 1016 (2023).

66. Sands, B. E. *et al.* Vedolizumab versus Adalimumab for Moderate-to-Severe Ulcerative Colitis. *New England Journal of Medicine* **381**, 1215–1226 (2019).

67. Sands, B. E. *et al.* Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. *New England Journal of Medicine* **381**, 1201–1214 (2019).

68. Sandborn, W. J. *et al.* Ozanimod as Induction and Maintenance Therapy for Ulcerative Colitis. *New England Journal of Medicine* **385**, 1280–1291 (2021).

69. Rutgeerts, P. *et al.* Infliximab for Induction and Maintenance Therapy for Ulcerative Colitis. *New England Journal of Medicine* **353**, 2462–2476 (2005).

70. Khoudari, G. *et al.* Rates of Intestinal Resection and Colectomy in Inflammatory Bowel Disease Patients After Initiation of Biologics: A Cohort Study. *Clinical Gastroenterology and Hepatology* **20**, e974–e983 (2022).

71. Sahami, S. *et al.* The Link between the Appendix and Ulcerative Colitis: Clinical Relevance and Potential Immunological Mechanisms. *American Journal of Gastroenterology* **111**, 163–169 (2016).

72. Andersson, R. E. & Ekbom, A. Appendectomy and Protection against Ulcerative Colitis. *The New England Journal of Medicine* 7 (2001).

73. Kaplan, G. G. *et al.* The risk of developing Crohn's disease after an appendectomy: a population-based cohort study in Sweden and Denmark. *Gut* **56**, 1387–1392 (2007).

74. Mark-Christensen, A., Kristiansen, E. B., Laurberg, S. & Erichsen, R. Prior Appendectomy Is Associated With a Milder Clinical Course in Crohn's Disease: A Nationwide Population-based Cohort Study. *Inflammatory Bowel Diseases* izae059 (2024) doi:10.1093/ibd/izae059.

75. Mizoguchi, A., Mizoguchi, E., Chiba, C. & Bhan, A. K. Role of appendix in the development of inflammatory bowel disease in TCR-alpha mutant mice. *Journal of Experimental Medicine* **184**, 707–715 (1996).

76. Harnoy, Y. *et al.* Effect of appendicectomy on colonic inflammation and neoplasia in experimental ulcerative colitis. *British Journal of Surgery* **103**, 1530–1538 (2016).

77. Krieglstein, C. F. *et al.* Role of Appendix and Spleen in Experimental Colitis. *Journal of Surgical Research* **101**, 166–175 (2001).

78. Farkas, S. A. *et al.* Preferential Migration of CD62L+ Cells into the Appendix in Mice with Experimental Chronic Colitis. *European Surgical Research* **37**, 115–122 (2005).

79. Collard, M. K. *et al.* The Appendix Orchestrates T-Cell Mediated Immunosurveillance in Colitis-Associated Cancer. *Cellular and Molecular Gastroenterology and Hepatology* **15**, 665–687 (2023).

80. Liu, Z., Ma, X., Zhu, C. & Fang, J.-Y. Risk of colorectal cancer after appendectomy: A systematic review and meta-analysis. *Journal of Gastroenterology and Hepatology* **38**, 350–358 (2023).

81. Rothwell, J. A. *et al.* Colorectal cancer risk following appendectomy: a pooled analysis of three large prospective cohort studies. *Cancer Commun (Lond)* **42**, 486–489 (2022).

82. Viennet, M. *et al.* Increased risk of colon cancer after acute appendicitis: a nationwide, population-based study. *eClinicalMedicine* **63**, (2023).

83. Cohen, T., Pfeffer, R. B. & Valensi, Q. ‘Ulcerative appendicitis’ occurring as a skip lesion in chronic ulcerative colitis; report of a case. *Am J Gastroenterol* **62**, 151–155 (1974).

84. Scott, I. S., Sheaff, M., Coumbe, A., Feakins, R. M. & Rampton, D. S. Appendiceal inflammation in ulcerative colitis. *Histopathology* **33**, 168–173 (1998).

85. Heuthorst, L. *et al.* High prevalence of ulcerative appendicitis in patients with ulcerative colitis. *UEG Journal* **9**, 1148–1156 (2021).

86. Park, S. H., Loftus, E. V. & Yang, S.-K. Appendiceal skip inflammation and ulcerative colitis. *Dig Dis Sci* **59**, 2050–2057 (2014).

87. Park, S. H. *et al.* Long term follow-up of appendiceal and distal right-sided colonic inflammation. *Endoscopy* **44**, 95–98 (2011).

88. Welsh, S., Sam, Z., Seenan, J. P. & Nicholson, G. A. The Role of Appendicectomy in Ulcerative Colitis: Systematic Review and Meta-Analysis. *Inflammatory Bowel Diseases* **29**, 633–646 (2023).

89. Parian, A. *et al.* Appendectomy does not decrease the risk of future colectomy in UC: results from a large cohort and meta-analysis. *Gut* **66**, 1390–1397 (2017).

90. Agrawal, M. *et al.* The appendix and ulcerative colitis — an unsolved connection. *Nat Rev Gastroenterol Hepatol* **20**, 615–624 (2023).

91. Cosnes, J. *et al.* Effects of appendicectomy on the course of ulcerative colitis. *Gut* **51**, 803–807 (2002).

92. Okazaki, K. *et al.* A patient with improvement of ulcerative colitis after appendectomy. *Gastroenterology* **119**, 502–506 (2000).

93. Noh, C. H. *et al.* Remission of Ulcerative Colitis after Appendectomy: A Case Report. *Korean J Gastroenterol* **56**, 201–204 (2010).

94. Bolin, T. D., Wong, S., Crouch, R., Engelman, J. L. & Riordan, S. M. Appendicectomy as a therapy for ulcerative proctitis. *Am J Gastroenterol* **104**, 2476–2482 (2009).

95. Sahami, S. *et al.* Appendectomy for Therapy-Refractory Ulcerative Colitis Results in Pathological Improvement of Colonic Inflammation: Short-Term Results of the PASSION Study. *Journal of Crohn's and Colitis* **13**, 165–171 (2019).

96. Stellingwerf, M. E. *et al.* Prospective cohort study of appendicectomy for treatment of therapy-refractory ulcerative colitis. *British Journal of Surgery* **106**, 1697–1704 (2019).

97. Gardenbroek, T. J. *et al.* The ACCURE-trial: the effect of appendectomy on the clinical course of ulcerative colitis, a randomised international multicenter trial (NTR2883) and the ACCURE-UK trial: a randomised external pilot trial (ISRCTN56523019). *BMC Surg* **15**, 30 (2015).

98. Visser, E. *et al.* Clinical statistical analysis plan for the ACCURE trial: the effect of appendectomy on the clinical course of ulcerative colitis, a randomised international multicentre trial. *Trials* **25**, 218 (2024).

99. Bemelman, W. A. *The Effect of Appendectomy in Ulcerative Colitis Patients With Active Disease: COlonic Salvage by Therapeutic Appendectomy*. https://clinicaltrials.gov/study/NCT03912714 (2024).

100. Ostanin, D. V. *et al.* T cell transfer model of chronic colitis: concepts, considerations, and tricks of the trade. *Am J Physiol Gastrointest Liver Physiol* **296**, G135–G146 (2009).

101. Mottet, C., Uhlig, H. H. & Powrie, F. Cutting edge: cure of colitis by CD4+CD25+ regulatory T cells. *J Immunol* **170**, 3939–3943 (2003).

102. Brasseit, J. *et al.* CD4 T cells are required for both development and maintenance of disease in a new mouse model of reversible colitis. *Mucosal Immunol* **9**, 689–701 (2016).

103. Nancey, S. *et al.* CD8+ Cytotoxic T Cells Induce Relapsing Colitis in Normal Mice. *Gastroenterology* **131**, 485–496 (2006).

104. Matsushita, M. Appendix is a priming site in the development of ulcerative colitis. *WJG* **11**, 4869 (2005).

105. Masahata, K. *et al.* Generation of colonic IgA-secreting cells in the caecal patch. *Nat Commun* **5**, 3704 (2014).

106. Adamczyk, A. *et al.* Differential expression of GPR15 on T cells during ulcerative colitis. *JCI Insight* **2**, e90585 (2017).

107. Kim, S. V. *et al.* GPR15-Mediated Homing Controls Immune Homeostasis in the Large Intestine Mucosa. *Science* **340**, 1456–1459 (2013).

108. Nguyen, L. P. *et al.* Role and species-specific expression of colon T cell homing receptor GPR15 in colitis. *Nat Immunol* **16**, 207–213 (2015).

109. Habtezion, A., Nguyen, L. P., Hadeiba, H. & Butcher, E. C. Leukocyte Trafficking to the Small Intestine and Colon. *Gastroenterology* **150**, 340–354 (2016).

110. Grümme, L., Dombret, S., Knösel, T., Skapenko, A. & Schulze-Koops, H. Colitis induced by IL-17A-inhibitors. *Clin J Gastroenterol* **17**, 263–270 (2024).

111. Ju, J. *et al.* Crohn's disease exacerbated by IL-17 inhibitors in patients with psoriasis: a case report. *BMC Gastroenterology* **20**, 340 (2020).

112. Clough, J. N., Omer, O. S., Tasker, S., Lord, G. M. & Irving, P. M. Regulatory T-cell therapy in Crohn's disease: challenges and advances. *Gut* **69**, 942–952 (2020).

113. Fantini, M. C. *et al.* Smad7 Controls Resistance of Colitogenic T Cells to Regulatory T Cell-Mediated Suppression. *Gastroenterology* **136**, 1308-1316.e3 (2009).

114. Laffont, S., Siddiqui, K. R. R. & Powrie, F. Intestinal inflammation abrogates the tolerogenic properties of MLN CD103+ dendritic cells. *European Journal of Immunology* **40**, 1877–1883 (2010).

115. Steinhoff, U. *et al.* Autoimmune Intestinal Pathology Induced by hsp60-Specific CD8 T Cells. *Immunity* **11**, 349–358 (1999).

116. McKinney, E. F., Lee, J. C., Jayne, D. R. W., Lyons, P. A. & Smith, K. G. C. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature* **523**, 612–616 (2015).

117. Lee, J. C. *et al.* Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J Clin Invest* **121**, 4170–4179 (2011).

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

119. Huang, B. *et al.* Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. *Cell* **179**, 1160-1176.e24 (2019).

120. Corridoni, D. *et al.* Single-cell atlas of colonic CD8+ T cells in ulcerative colitis. *Nat Med* **26**, 1480–1490 (2020).

121. Boland, B. S. *et al.* Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. *Sci. Immunol.* **5**, eabb4432 (2020).

122. Pearce, E. L. *et al.* Control of Effector CD8+ T Cell Function by the Transcription Factor Eomesodermin. *Science* **302**, 1041–1043 (2003).

123. Intlekofer, A. M. *et al.* Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nat Immunol* **6**, 1236–1244 (2005).

124. Banerjee, A. *et al.* Cutting Edge: The Transcription Factor Eomesodermin Enables CD8+ T Cells To Compete for the Memory Cell Niche. *The Journal of Immunology* **185**, 4988–4992 (2010).

125. Paley, M. A. *et al.* Progenitor and Terminal Subsets of CD8+ T Cells Cooperate to Contain Chronic Viral Infection. *Science* **338**, 1220–1225 (2012).

126. Simonetta, F. *et al.* High Eomesodermin Expression among CD57+ CD8+ T Cells Identifies a CD8+ T Cell Subset Associated with Viral Control during Chronic Human Immunodeficiency Virus Infection. *Journal of Virology* **88**, 11861–11871 (2014).

127. Llaó-Cid, L. *et al.* EOMES is essential for antitumor activity of CD8+ T cells in chronic lymphocytic leukemia. *Leukemia* **35**, 3152–3162 (2021).

128. Ivanov, I. I. *et al.* Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria. *Cell* **139**, 485–498 (2009).

129. Yang, Y. *et al.* Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature* **510**, 152–156 (2014).

130. Tan, T. G. *et al.* Identifying species of symbiont bacteria from the human gut that, alone, can induce intestinal Th17 cells in mice. *Proc. Natl. Acad. Sci. U.S.A.* **113**, (2016).

131. Viladomiu, M. *et al.* IgA-coated *E. coli* enriched in Crohn's disease spondyloarthritis promote T_H17-dependent inflammation. *Sci. Transl. Med.* **9**, eaaf9655 (2017).

132. Atarashi, K. *et al.* Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells. *Cell* **163**, 367–380 (2015).

133. Round, J. L. & Mazmanian, S. K. Inducible $Foxp3^+$ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12204–12209 (2010).

134. Bousbaine, D. *et al.* A conserved Bacteroidetes antigen induces anti-inflammatory intestinal T lymphocytes. *Science* **377**, 660–666 (2022).

135. Atarashi, K. *et al.* Induction of Colonic Regulatory T Cells by Indigenous *Clostridium* Species. *Science* **331**, 337–341 (2011).

136. Atarashi, K. *et al.* Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).

137. Sefik, E. *et al.* Individual intestinal symbionts induce a distinct population of RORγ+ regulatory T cells. *Science* **349**, 993–997 (2015).

138. Tanoue, T. *et al.* A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. *Nature* **565**, 600–605 (2019).

139. Chai, J. N. *et al. Helicobacter* species are potent drivers of colonic T cell responses in homeostasis and inflammation. *Sci. Immunol.* **2**, eaal5068 (2017).

140. Xu, M. *et al.* c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* **554**, 373–377 (2018).

141. Ansaldo, E. *et al. Akkermansia muciniphila* induces intestinal adaptive immune responses during homeostasis. *Science* **364**, 1179–1184 (2019).

142. Nagashima, K. *et al.* Mapping the T cell repertoire to a complex gut bacterial community. *Nature* **621**, 162–170 (2023).

143. Duchmann, R. *et al.* Tolerance exists towards resident intestinal flora but is broken in active inflammatory bowel disease (IBD). *Clinical and Experimental Immunology* **102**, 448–455 (1995).

144. Duchmann, R., Märker-Hermann, E. & Meyer Zum Büschenfelde, K. -H. Bacteria-Specific T-Cell Clones are Selective in their Reactivity Towards Different Enterobacteria or *H.*

pylori and Increased in Inflammatory Bowel Disease. *Scandinavian Journal of Immunology* **44**, 71–79 (1996).

145. Duchmann, R. *et al.* T cell specificity and cross reactivity towards enterobacteria, *Bacteroides* , *Bifidobacterium* , and antigens from resident intestinal flora in humans. *Gut* **44**, 812–818 (1999).
146. Calderón-Gómez, E. *et al.* Commensal-Specific CD4+ Cells From Patients With Crohn's Disease Have a T-Helper 17 Inflammatory Profile. *Gastroenterology* **151**, 489-500.e3 (2016).
147. Hegazy, A. N. *et al.* Circulating and Tissue-Resident CD4+ T Cells With Reactivity to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered During Inflammation. *Gastroenterology* **153**, 1320-1337.e16 (2017).
148. Jørgensen, P. B. *et al.* Identification, isolation and analysis of human gut-associated lymphoid tissues. *Nat Protoc* **16**, 2051–2067 (2021).
149. Antigen-driven colonic inflammation is associated with development of dysplasia in primary sclerosing cholangitis | Nature Medicine. https://www.nature.com/articles/s41591-023-02372-x.
150. Lin, J.-R., Fallahi-Sichani, M., Chen, J.-Y. & Sorger, P. K. Cyclic Immunofluorescence (CycIF), A Highly Multiplexed Method for Single-cell Imaging. *Current Protocols in Chemical Biology* **8**, 251–264 (2016).
151. Griswold M, Ortogero N, Yang Z, Vitancol R, Henderson D. GeomxTools: NanoString GeoMx Tools. (2022).
152. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47–e47 (2015).

153. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).

154. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267–273 (2003).

155. Li, Y. yan-cri/scRICA. (2023).

156. DePasquale, E. A. K. *et al.* DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data. *Cell Reports* **29**, 1718-1727.e8 (2019).

157. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

158. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**, (2019).

159. Domínguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).

160. Mullan, K. A. *et al.* TCR_Explore: A novel webtool for T cell receptor repertoire analysis. *Computational and Structural Biotechnology Journal* **21**, 1272–1282 (2023).

161. Brochet, X., Lefranc, M.-P. & Giudicelli, V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research* **36**, W503–W508 (2008).

162. Lefranc, M.-P. *et al.* IMGT(R), the international ImMunoGeneTics information system(R). *Nucleic Acids Research* **37**, D1006–D1012 (2009).

163. Hoehn, K. B., Pybus, O. G. & Kleinstein, S. H. Phylogenetic analysis of migration, differentiation, and class switching in B cells. *PLoS Comput Biol* **18**, e1009885 (2022).

164. Gupta, N. T. *et al.* Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

165. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).

166. Ou, J., Wolfe, S. A., Brodsky, M. H. & Zhu, L. J. motifStack for the analysis of transcription factor binding site evolution. *Nat Methods* **15**, 8–9 (2018).

167. Huang, H., Wang, C., Rubelt, F., Scriba, T. J. & Davis, M. M. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol* **38**, 1194–1202 (2020).

168. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).

169. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596 (2012).

170. Lu, Y. *et al.* MicrobiomeAnalyst 2.0: comprehensive statistical, functional and integrative analysis of microbiome data. *Nucleic Acids Research* **51**, W310–W318 (2023).

171. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

172. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput Biol* **8**, e1002687 (2012).

173. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

174. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* **31**, 1674–1676 (2015).

175. biobakery/kneaddata. bioBakery (2024).

176. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

177. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829 (2021).

178. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314 (2019).

179. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).

180. Nilsson, J. B. *et al.* Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. *Sci. Adv.* **9**, eadj6367 (2023).

181. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Research* gkz1029 (2019) doi:10.1093/nar/gkz1029.

182. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

183. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

184. Chronister, W. D. *et al.* TCRMatch: Predicting T-Cell Receptor Specificity Based on Sequence Similarity to Previously Characterized Receptors. *Front. Immunol.* **12**, (2021).

185. Chiaruttini, N. *et al.* An Open-Source Whole Slide Image Registration Workflow at Cellular Precision Using Fiji, QuPath and Elastix. *Front. Comput. Sci.* **3**, (2022).

186. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

187. Marshall, A. J. *et al.* FDC-SP, a Novel Secreted Protein Expressed by Follicular Dendritic Cells. *The Journal of Immunology* **169**, 2381–2389 (2002).

188. Al-Alwan, M. *et al.* Follicular Dendritic Cell Secreted Protein (FDC-SP) Regulates Germinal Center and Antibody Responses. *The Journal of Immunology* **178**, 7859–7867 (2007).

189. Kim, T.-S. & Shin, E.-C. The activation of bystander CD8+ T cells and their roles in viral infection. *Exp Mol Med* **51**, 1–9 (2019).

190. Lee, H., Jeong, S. & Shin, E.-C. Significance of bystander T cell activation in microbial infection. *Nat Immunol* **23**, 13–22 (2022).

191. Guinane, C. M. *et al.* Microbial Composition of Human Appendices from Patients following Appendectomy. *mBio* **4**, e00366-12 (2013).

192. Swidsinski, A. *et al.* Acute appendicitis is characterised by local invasion with Fusobacterium nucleatum/necrophorum. *Gut* **60**, 34–40 (2011).

193. Blohs, M. *et al.* Acute appendicitis manifests as two microbiome state types with oral pathogens influencing severity. *Gut Microbes* **15**, 2145845 (2023).

194. Swidsinski, A. *et al.* Acute appendicitis is characterised by local invasion with Fusobacterium nucleatum/necrophorum. *Gut* **60**, 34–40 (2011).

195. Pedersen, T. K. *et al.* The CD4+ T cell response to a commensal-derived epitope transitions from a tolerant to an inflammatory state in Crohn's disease. *Immunity* **55**, 1909-1923.e6 (2022).

196. Sommer, F. & Bäckhed, F. The gut microbiota engages different signaling pathways to induce Duox2 expression in the ileum and colon epithelium. *Mucosal Immunology* **8**, 372–379 (2015).

197. Hazime, H. *et al.* EPITHELIAL DUOX2 ACTIVATION INCREASES GUT PERMEABILITY AND BACTERIAL TRANSLOCATION THAT IS RESCUED WITH BUTYRATE SUPPLEMENTATION. *Gastroenterology* **162**, S49 (2022).

198. Castrillón-Betancur, J. C. *et al.* Epithelial Dual Oxidase 2 Shapes the Mucosal Microbiome and Contributes to Inflammatory Susceptibility. *Antioxidants* **12**, 1889 (2023).

199. Mao, H. *et al.* Protective and anti-inflammatory role of REG1A in inflammatory bowel disease induced by JAK/STAT3 signaling axis. *International Immunopharmacology* **92**, 107304 (2021).

200. Dehghani, T. *et al.* Association Between Inflammatory Bowel Disease and Viral Infections. *Curr Microbiol* **80**, 195 (2023).

201. Chaisawangwong, W. *et al.* Cross-reactivity of SARS-CoV-2– and influenza A–specific T cells in individuals exposed to SARS-CoV-2. *JCI Insight* **7**, (2022).

202. Zollner, A. *et al.* Postacute COVID-19 is Characterized by Gut Viral Antigen Persistence in Inflammatory Bowel Diseases. *Gastroenterology* **163**, 495-506.e8 (2022).

203. Williams, K. G. *et al.* T Cell Repertoire Homogeneity and Blood-Gut Overlap in Patients With Inflammatory Bowel Disease. *Cellular and Molecular Gastroenterology and Hepatology* **17**, 119–130 (2024).

204. Gao, Y. J. *et al.* Down-regulation of CXCL 11 inhibits colorectal cancer cell growth and epithelial-mesenchymal transition. *OTT* **11**, 7333–7343 (2018).

205. Berzaghi, R. *et al.* SOCS1 favors the epithelial-mesenchymal transition in melanoma, promotes tumor progression and prevents antitumor immunity by PD-L1 expression. *Sci Rep* **7**, 40585 (2017).

206. Ricanek, P. *et al.* Reduced expression of aquaporins in human intestinal mucosa in early stage inflammatory bowel disease. *CEG* **8**, 49–67 (2015).

207. Kaiserman, D. *et al.* Granzyme K initiates IL-6 and IL-8 release from epithelial cells by activating protease-activated receptor 2. *PLoS One* **17**, e0270584 (2022).

208. Cooper, D. M., Pechkovsky, D. V., Hackett, T. L., Knight, D. A. & Granville, D. J. Granzyme K Activates Protease-Activated Receptor-1. *PLOS ONE* **6**, e21484 (2011).

209. Omundsen, M. & Dennett, E. DELAY TO APPENDICECTOMY AND ASSOCIATED MORBIDITY: A RETROSPECTIVE REVIEW. *ANZ Journal of Surgery* **76**, 153–155 (2006).

210. Köhler, F. *et al.* Changes in appendicitis treatment during the COVID-19 pandemic – A systematic review and meta-analysis. *International Journal of Surgery* **95**, 106148 (2021).

211. Wu, S.-C. *et al.* Association between Appendectomy and Subsequent Colorectal Cancer Development: An Asian Population Study. *PLoS ONE* **10**, e0118411 (2015).

212. Park, Y. Y. *et al.* A link between appendectomy and gastrointestinal cancers: a large-scale population-based cohort study in Korea. *Sci Rep* **10**, 15670 (2020).

213. Li, P. *et al. Gut Microbiome Dysbiosis Is Associated with Elevated Toxic Bile Acids in Parkinson's Disease*. http://biorxiv.org/lookup/doi/10.1101/2020.09.26.279851 (2020) doi:10.1101/2020.09.26.279851.

214. Killinger, B. A. *et al.* The vermiform appendix impacts the risk of developing Parkinson's disease. *Sci. Transl. Med.* **10**, eaar5280 (2018).

215. Horiuchi, A., Kokubu, E., Warita, T. & Ishihara, K. Synergistic biofilm formation by Parvimonas micra and Fusobacterium nucleatum. *Anaerobe* **62**, 102100 (2020).

216. Kremer, B. H. A. & Steenbergen, T. J. M. *Peptostreptococcus micros* coaggregates with *Fusobacterium nucleatum* and non-encapsulated *Porphyromonas gingivalis*. *FEMS Microbiology Letters* **182**, 57–61 (2000).

217. Koliarakis, I. *et al.* Oral Bacteria and Intestinal Dysbiosis in Colorectal Cancer. *IJMS* **20**, 4146 (2019).

218. Osman, M. A. *et al.* Parvimonas micra, Peptostreptococcus stomatis, Fusobacterium nucleatum and Akkermansia muciniphila as a four-bacteria biomarker panel of colorectal cancer. *Sci Rep* **11**, 2925 (2021).

219. Hajishengallis, G. Periodontitis: from microbial immune subversion to systemic inflammation. *Nat Rev Immunol* **15**, 30–44 (2015).

220. Higashi, D. L. *et al.* Who is in the driver's seat? *PARVIMONAS MICRA* : An understudied pathobiont at the crossroads of dysbiotic disease and cancer. *Environ Microbiol Rep* **15**, 254–264 (2023).

221. Wang, Y. *et al.* Study of the inflammatory activating process in the early stage of Fusobacterium nucleatum infected PDLSCs. *Int J Oral Sci* **15**, 8 (2023).

222. Zhang, H., Zhao, S. & Cao, Z. Impact of Epstein–Barr virus infection in patients with inflammatory bowel disease. *Front. Immunol.* **13**, (2022).

223. Cytomegalovirus and Inflammatory Bowel Diseases (IBD) with a Special Focus on the Link with Ulcerative Colitis (UC). https://www.mdpi.com/2076-2607/8/7/1078.

224. Spieker, T. & Herbst, H. Distribution and Phenotype of Epstein-Barr Virus-Infected Cells in Inflammatory Bowel Disease. *The American Journal of Pathology* **157**, 51–57 (2000).

225. Kirchmeier, D. *et al.* Epstein-Barr virus infection induces tissue-resident memory T cells in mucosal lymphoid tissues. *JCI Insight* **9**, (2024).

226. Zheng, S. *et al.* CRTAM promotes antitumor immune response in triple negative breast cancer by enhancing CD8+ T cell infiltration. *International Immunopharmacology* **129**, 111625 (2024).

227. CRTAM controls residency of gut CD4+CD8+ T cells in the steady state and maintenance of gut CD4+ Th17 during parasitic infection | Journal of Experimental Medicine | Rockefeller University Press. https://rupress.org/jem/article/211/4/623/41436/CRTAM-controls-residency-of-gut-CD4-CD8-T-cells-in.

228. Xu, T. *et al.* Single-cell profiling reveals pathogenic role and differentiation trajectory of granzyme K^+CD8^+ T cells in primary Sjögren's syndrome. *JCI Insight* **8**, (2023).

229. Jonsson, A. H. *et al.* Granzyme K+ CD8 T cells form a core population in inflamed human tissue. *Science Translational Medicine* **14**, eabo0686 (2022).

230. Duquette, D. *et al.* Human Granzyme K Is a Feature of Innate T Cells in Blood, Tissues, and Tumors, Responding to Cytokines Rather than TCR Stimulation. *The Journal of Immunology* **211**, 633–647 (2023).

231. van Aalderen, M. C. *et al.* Infection History Determines the Differentiation State of Human CD8+ T Cells. *Journal of Virology* **89**, 5110–5123 (2015).

232. Jonsson, A. H. Granzyme K+ CD8 T cells in autoimmunity. *Best Practice & Research Clinical Rheumatology* **38**, 101930 (2024).

233. Ding, X. *et al.* SLC26A3 (DRA) prevents TNF-alpha-induced barrier dysfunction and dextran sulfate sodium-induced acute colitis. *Laboratory Investigation* **98**, 462–476 (2018).

234. Esterházy, D. *et al.* Compartmentalized gut lymph node drainage dictates adaptive immune responses. *Nature* **569**, 126–130 (2019).

235. Weß, V. *et al.* Humanized NSG Mouse Models as a Preclinical Tool for Translational Research in Inflammatory Bowel Diseases. *International Journal of Molecular Sciences* **24**, 12348 (2023).

236. Pleguezuelos-Manzano, C. *et al.* Establishment and Culture of Human Intestinal Organoids Derived from Adult Stem Cells. *Current Protocols in Immunology* **130**, e106 (2020).

237. Crespo, M. *et al.* Colonic organoids derived from human induced pluripotent stem cells for modeling colorectal cancer and drug testing. *Nat Med* **23**, 878–884 (2017).

238. Lu, R.-M. *et al.* Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science* **27**, 1 (2020).

239. Ducancel, F. & Muller, B. H. Molecular engineering of antibodies for therapeutic and diagnostic purposes. *mAbs* **4**, 445–457 (2012).

240. Attaf, M., Huseby, E. & Sewell, A. K. αβ T cell receptors as predictors of health and disease. *Cell Mol Immunol* **12**, 391–399 (2015).

241. Nikolich-Žugich, J., Slifka, M. K. & Messaoudi, I. The many important facets of T-cell repertoire diversity. *Nat Rev Immunol* **4**, 123–132 (2004).

242. Joglekar, A. V. & Li, G. T cell antigen discovery. *Nat Methods* **18**, 873–880 (2021).

243. Hudson, D., Fernandes, R. A., Basham, M., Ogg, G. & Koohy, H. Can we predict T cell specificity with digital biology and machine learning? *Nat Rev Immunol* **23**, 511–521 (2023).

244. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017).