REVIEW

Making sense of the cause of Crohn’s – a new look at an old disease [version 1; peer review: 1 approved, 1 approved with reservations]

Anthony W. Segal
University College London, London, WC1E 6BT, UK

Abstract
The cause of Crohn’s disease (CD) has posed a conundrum for at least a century. A large body of work coupled with recent technological advances in genome research have at last started to provide some of the answers. Initially this review seeks to explain and to differentiate between bowel inflammation in the primary immunodeficiencies that generally lead to very early onset diffuse bowel inflammation in humans and in animal models, and the real syndrome of CD. In the latter, a trigger, almost certainly enteric infection by one of a multitude of organisms, allows the faeces access to the tissues, at which stage the response of individuals predisposed to CD is abnormal. Direct investigation of patients’ inflammatory response together with genome-wide association studies (GWAS) and DNA sequencing indicate that in CD the failure of acute inflammation and the clearance of bacteria from the tissues, and from within cells, is defective. The retained faecal products result in the characteristic chronic granulomatous inflammation and adaptive immune response. In this review I will examine the contemporary evidence that has led to this understanding, and look for explanations for the recent dramatic increase in the incidence of this disease.

Keywords
Crohn’s, Inflammatory Bowel Disease, GWAS, Immunology, Infection, Bacteria, Gastroenteritis, Gene

This article is included in the University College London collection.
General introduction
The enigma that is the cause of Crohn’s disease (CD) has puzzled clinicians and scientists from time immemorial. It is generally accepted that CD results from an aberrant immune response to commensal microflora in genetically susceptible individuals1, however, the nature of the immune defects, the responsible microflora and the genetic susceptibility remain incompletely defined and actively debated. With advances in genomic technologies our understanding of this puzzling condition is evolving, and answers forthcoming. The purpose of this article is to undertake a holistic review of the aetopathogenesis of CD in which historical concepts are integrated with recent discoveries.

The bowel mucosa, an interface between faeces and the tissues
The distal ileum and colon contain >10^11 bacteria per gram of faecal material2, which pose an immediate threat to life if they penetrate into the underlying tissues. The bowel microflora are isolated by a thin film of mucus and a single layer of columnar epithelial cells with a surface area of approximately 32m^2. The requirement for the absorption of fluids and nutrients by the bowel mucosa means that the bowel lining cannot simply be a tough impermeable barrier, and as a consequence provision must be made to defend the vulnerable mucosal epithelial cell layer against its contents. Mucus secreted by goblet cells forms a continuous, weak, viscoelastic gel, lining, 5–500 µm thick3. In addition to acting as a physical barrier and lubricant, the mucus is the site of action of a variety of antimicrobial mechanisms including secretory IgA, antimicrobial enzymes and peptides4 and H2O2 generated by the DUOX electron transport chain5. Despite these barriers, the separation of the tissues from the gut microbiome is not absolute, and even in health the mucosa is constantly penetrated by relatively small numbers of enteric organisms and soluble microbial products that gain access into the tissues6–10. Scattered amongst the epithelial cells overlying lymph follicles are Microfold (M) cells11–13, a unique intestinal epithelial cell (IEC) subset that are highly specialized for the phagocytosis and transcytosis of gut lumen macromolecules, particulate antigens and commensal microorganisms, which they transfer across the epithelium to mucosal macrophages and dendritic cells. This slow, constant, transit is important for the development, priming and maintenance of a potent immune system in the submucosa14–15. The protective role of the bowel immune system must be combined with tolerance to ingested antigens and commensal organisms to maintain homeostasis in a healthy bowel.

The immune system in the bowel
The bowel is the interface between a dense population of microbes and the immune system. Although an in depth review of the immune system in the bowel is well beyond the scope of this review, it is important to briefly cover this subject because defects in innate immunity are central to the development of CD whereas aberrant adaptive immunity causes bowel inflammation of a very different type, and a range of largely inaccurate animal models of CD.

Adaptive immunity. Most of what is known of classical adaptive immunology relates to the immune system of the bowel, but there are in addition some specialised features unique to the intestinal mucosa16–19.

The mucosa and exocrine glands harbour the largest activated B-cell system of the body, amounting to some 80–90% of all immunoglobulin (Ig)-producing cells in humans15. The major product of these lymphocytes is polymeric (pIgA (mainly dimers) with associated J chain. Both pIgA and pentameric IgM contain a binding site for the polymeric Ig receptor (pIgR), or secretory component (SC), which is a requirement for their active external transport through secretory epithelia into the overlying mucus19.

M cells, and intestinal dendritic cells, that phagocytose bacteria interact with B and T cells in the Peyer’s patches, inducing B cells to produce IgA directed against intestinal bacteria20. IgA+ B cells home to the intestinal lamina propria and secrete IgA that is transcytosed across the epithelium and deposited on the apical surface. The transcytosed IgAs bind to luminal bacteria, preventing microbial translocation across the epithelial barrier21,22.

After the initiation of the immune response by antigen processing and presentation to B and T cells in Peyer’s patches, primed lymphocytes leave the mucosa via the thoracic duct. Finally they migrate back to the mucosa where they exert effector functions.

There has been considerable recent interest in IL-23 and IL-17 in relation to the aetiology of CD. IL-23 is secreted by macrophages and dendritic cells and transforms naive T cells into Th (T-helper) 17 cells which then produce IL-17, IL-21 and IL-2223. IL-17 induces numerous cell types including T-cells, mast cells, macrophages, neutrophils, keratinocytes, and natural killer cells to produce a raft of pro-inflammatory mediators including IL-1b, IL-6, IL-8, IL-11, Gro-α, G-CSF, GM-CSF, IL-4, IL-5, IL-13, IgE, and eotaxin24. An important outcome of this cytokine cascade appears to be the recruitment of neutrophils to inflammatory sites25. Despite the apparent importance of IL-17 for intestinal barrier function26 and for diverse pro-inflammatory activities, there must be considerable redundancy in the pro-inflammatory repertoire as defects in the IL-17 pathway are associated with a very narrow predisposition to disease in the form of mucocutaneous candidiasis27. None of the hundreds of patients with this condition had CD28.

IL-17 has been considered to be detrimental in CD as a consequence of its apparent pro-inflammatory actions. It is therefore of interest that a trial of the treatment of CD with monoclonal antibodies against IL-17 had to be stopped because of the deterioration of the patients’ condition29.

Adaptive immunity in the bowel protects against commensal organisms, or those previously encountered in infections that were successfully overcome. This is accomplished by the production of a barrier of secreted IgA that permeates the lining mucus layer and by the production of specific IgG and IgM that opsonise penetrating organisms for phagocytosis. Immunity to pathogenic bacteria like Salmonella, Shigella, Vibrio cholera and Escherichia coli is generally not very potent or long lasting, which, together with the propensity of bacteria to mutate, makes vaccines relatively ineffective30–33. A degree of immunity does develop as a result of repeated reinfection in endemic areas but because this is not permanent, it is gradually lost after emigration to cleaner environments, which might be an important factor in relation to the
subsequent triggering of CD by infection in individuals moving from regions of low to high prevalence of this condition.

**Innate immunity.** The submucosa of the bowel is particularly vulnerable to microbial invasion if the mucosal barrier is breached as large numbers of organisms can achieve rapid access and the conditions are conducive to microbial proliferation. There is inadequate time for adaptive immunity to take effect and reliance must be placed on the innate system to contain and eliminate potentially harmful stimuli. At its heart this means the rapid and florid release of pro-inflammatory cytokines from lamina propria macrophages\(^{32}\), recruited from blood monocytes\(^{33}\), mast cells\(^{34-35}\), eosinophils and innate lymphoid cells\(^{36-38}\) when activated by bowel contents. Paneth cells are specialised intraepithelial secretory epithelium of the small intestine that reside in small clusters at the base of crypts of Lieberkühn in the small intestine. Large secretory granules in these cells contain a wide variety of proteins, the most abundant of which are antimicrobials such as the alpha defensins that are discharged into the crypt lumen. These effector molecules also diffuse from the crypt and disseminate into the mucous layer that overlies the mucosal epithelium, where they contribute to the mucosal antimicrobial barrier\(^{39}\).

Pro-inflammatory cytokines induce changes in the microvasculature\(^{40-41}\) leading to the extravasation of plasma proteins and to the recruitment of neutrophils\(^{42}\). A critical concentration of neutrophils is required to eliminate invading bacteria\(^{43}\) and immediately after bacterial penetration of the mucosa there is direct competition between bacterial replication and neutrophil recruitment and bacterial phagocytosis and killing. In the absence of specific antibodies, uptake of the foreign material is enhanced by non-specific opsonins like pentraxins, collectins and complement\(^{44}\). The neutrophils then undergo apoptosis or necrosis and the purulent collection is most probably discharged into the bowel lumen, with the residual debris being phagocytosed and cleared by macrophages\(^{45}\).

**Bowel inflammation in very early onset Inflammatory Bowel Disease (IBD)**

Bowel homeostasis requires an intact mucosal barrier, itself requiring the integrated function of many different cell types and molecules, and the largest collection of immunological cells in the body to present an integrated defence against the intestinal microbiome. It is therefore not surprising that defects in genes coding for proteins required for the integrity of this barrier, or for normal immune surveillance, manifest as mucosal inflammation.

As might be expected, these conditions present very early in life, and because they affect the mucosa as a whole, they result in a diffuse, non-specific inflammation, predominantly in the large bowel where concentrations of bacteria are highest. Uhlig et al. (Figure 1)\(^{46}\) found that about 5% of their cases of IBD had infantile or very early onset disease. This will represent a much higher proportion of cases found that about 5% of their cases of IBD had infantile or very early onset disease. This will represent a much higher proportion of cases occurring in the general population, because most cases of IBD occur in adults and are handled in non-specialist facilities, whereas rare inherited diseases gravitate to specialist centres like those of Uhlig and his co-authors. The monogenic lesions identified provide important identifiers of the molecules required for bowel integrity and adaptive immunity.

Mutations in the barrier function genes such as COL7A1, FERMT1, TTC7A and ADAM17 generally result in infantile bowel dysfunction and inflammation.

The severe immunodeficiency syndromes such as atypical Severe Combined Immunodeficiency (SCID) and Immunodysregulation Polyendocrinopathy Enteropathy X-linked Syndrome (IPEX) also generally have a very early onset and also do not have intestinal changes characteristic of either CD or UC. In contradistinction, the neutrophil defects, exemplified by Chronic Granulomatous Disease (CGD), and including Hermansky-Pudlak, congenital neutropenia and leukocyte adhesion deficiency all have a CD phenotype clinically, endoscopically and histopathologically, with a lot of perianal disease and granulomata evident on biopsy\(^{47}\). The neutrophil defects generally present later than the abnormalities of mucosal barrier function, or the severe immunological diseases.

**Figure 1. Age of onset of IBD-like symptoms in patients with monogenic diseases.** Multiple genetic defects are summarized in the group of atypical Severe Combined Immunodeficiency (SCID), Hoyeraal-Hreidarsson syndrome, Chronic Granulomatous Disease (CGD), and Hermansky-Pudlak syndrome. By comparison, an unselected IBD population is presented (Oxford IBD cohort study; paediatric and adult referral-based IBD cohort, \(n = 1605\) patients comprising CD, Ulcerative Colitis (UC), and IBD unclassified [IBDU]). Symbols represent individual patients. Bars represent the age range of case series if individual data were not available. Reproduced from \(^{46}\) with permission from the publisher.
Mutations in the genes coding for IL-10 and IL-10 receptor both present very early in life, as seen in mucosal abnormalities and immunodeficiencies, and exhibit the bowel phenotype characteristic of defective neutrophil function. These observations would appear to be at variance with the prevailing view that IL-10 down-regulates macrophage function, and that the bowel inflammation in its absence is a manifestation of uncontrolled macrophage activation. IL-10 does appear to be required for normal intestinal development. Multisystem abnormalities were observed in the original description of the IL-10 knock-out mouse in which there was a general enterocolitis with greatest abnormalities in the duodenum and jejunum, not locations associated with a high burden of commensal bacteria. Under specific-pathogen free (SPF) conditions the bowel lesions persisted, but were limited to the proximal colon. In addition the mice exhibited a severe growth defect, and were severely anaemic with a paucity of erythroid precursors in the bone marrow that was filled with myeloid precursors. The anaemia was unrelated to the extent of bowel involvement. These phenotypic features indicate that IL-10 is required for the normal growth and development of the bowel mucosa and haemopoetic tissue, in addition to its suppressant effect on macrophages.

The observed phenotype of patients with IL-10 and IL-10 receptor deficiency is in keeping with that of defective neutrophil function. Almost all the patients showed evidence of bacterial infection in the form of folliculitis, and ear and respiratory tract infections. Most revealing was the almost universal occurrence of perianal disease with abscesses, fissures and fistulae that are highly characteristic of the neutrophil deficiency diseases like CGD, Hermansky-Pudlack and glycogen storage disease 1b. If the IL-10 deficient phenotype results from the impairment of normal cellular development and an immunodeficiency rather than an excessive, unregulated, macrophage response as proposed, then it might be expected that treatment with immunosuppressive therapy would be ineffective, which was indeed found to be the case.

It is also noteworthy that bowel inflammation is not a feature of the classical autoinflammatory diseases in which deregulated macrophage activation is a feature. These diseases include cryopyrin-associated periodic syndrome (CAPS) in which activating mutations in the NLRP3 gene result in increased excretion of IL-17 excretion and other pro-inflammatory cytokines, and the haemophagocytic syndrome, in which the uncontrolled activation of antigen-presenting cells (macrophages and histiocytes) and T cells produces an exaggerated inflammatory response and cytokine storm.

Of mice and men - mouse models of “IBD”

Mouse models of IBD have been extensively reviewed in the literature. These models are very important because they are depended upon by clinicians and scientist trying to understand the causes of these diseases, and by the pharmaceutical industry attempting to produce drugs with which to treat them.

Induced inflammation. In general, impairment of the innate immune system as in mouse models of CGD or Wiscott-Aldrich does not result in the spontaneous development of bowel inflammation, although they do exhibit an exaggerated response to insult.

Predisposition to bowel inflammation may be exposed by reducing intestinal barrier function, thereby allowing access of the contained microbiome to the underlying tissues. Barrier function can be compromised through the genetic manipulation of proteins required for the production of mucus or the maintenance of epithelial integrity or by the use of chemicals or infectious agents.

Chemical agents employed for this purpose fall into three main groups. Those that produce direct damage to the mucosa such as dextran sodium sulphate (DSS), acetic acid and carrageenan. The second group are those such as Haptens (e.g. 2, 4-Dinitrochlorobenzene (DNCB) or Dinitrobenzene sulphonic acid (DNBS)) that induce an immune response. Finally infection with bacteria such as Salmonella, E.coli or Citrobacter, or parasites may be utilised.

Genetic models. The advent of gene targeting technology has provided immunologists with powerful tools with which to explore the immune system. In the course of investigating its diverse components, hundreds of different genes have been knocked-out, some of which resulted in the spontaneous development of bowel inflammation. Because of this, these mice have been proposed as models of IBD.

Prominent examples of such mice include the IL-2, T cell receptor (TCR)αβ, and IL-10 knockout models. With the exception of IL-10-deficient mice, which possess some features of human CD, the majority of these models have diffuse colonic inflammation. A strain of mice (TnfAREmice) was developed in which elements of the tumour necrosis factor (TNF) gene that are required to restrict the overproduction of this cytokine have been removed. In their absence the mice exhibit sustained overproduction of TNF which results in a diffuse arthritis and terminal ileal and caecal inflammation. IL-17 and IL-22 deficiency exacerbate induced colitis. Spontaneous models. The C3H/HeJ/Bir model of colitis was discovered by chance when mice in breeding colonies developed loose bowel actions. The pathology is characterised by spontaneous and chronic focal inflammation localised to the right colon and caecal region although not involving the small intestine. The colitis occurs in young mice and tends to resolve with age, without recurrence. The genetic mechanisms underlying these abnormalities remain to be identified.

SAMP/Yit mice were developed from senescence-accelerated mice. They spontaneously developed ileitis and gastritis even under germ free conditions. The underlying aetiology is unknown but there is some evidence that the primary defect lies in the epithelial cell barrier and that B cells appear to play a role in the pathogenesis of inflammation at both sites.

Adoptive transfer. One of the most commonly cited models for the study of the role of T lymphocytes in bowel inflammation in mice (as a proposed model of CD) is the adoptive transfer model in which T cells are transfused into SCID mice, resulting in bowel inflammation. By observing the effects of varying the populations of cells infused, conclusions have been drawn as to the regulatory interaction of the various cell populations. It is important to understand that these host SCID mice have hardly any B or T
lymphocytes and, as their name suggests, are hypogammaglobulinaemic and severely immunocompromised. In the CD45RB\textsuperscript{high} transfer model, first described by Morrissey\textsuperscript{74} a subset of lymph node CD4\textsuperscript{+} T-cells, expressing high levels of the marker CD45RB (CD45RB\textsuperscript{hi}), were injected into SCID mice. The mice developed a wasting disease accompanied by massive hyperplasia of the intestinal mucosa with a dense infiltration of lymphocytes thought to be due to “an augmented, unregulated reaction towards higher levels of luminal-derived bacteria or bacterial products”. These changes were not seen when the animals were infused with unfractionated CD4\textsuperscript{+} or CD45RB\textsuperscript{lo} cells, indicating that the extreme reaction to bacterial products\textsuperscript{75} by the CD45RB\textsuperscript{hi} cells could be controlled by the CD45RB\textsuperscript{lo} cells. Soon after, similar experiments were conducted by Powrie and colleagues\textsuperscript{76,77} who observed the same pathological changes in the bowel which they equated to those changes found in “inflammatory bowel disease” in humans. They showed that these changes could be prevented by antibodies to interferon-gamma and by recombinant IL-10\textsuperscript{78}, and identified the cells in the CD45RB\textsuperscript{lo} population responsible for controlling CD45RB\textsuperscript{hi} induced inflammation as the population of suppressor T-cells called T-reg cells\textsuperscript{78}.

These observations and their extrapolation to human IBD sparked a large body of work into the role of regulatory T cells in the pathogenesis of IBD. Over the past decade, multiple groups have failed to find abnormalities in these cells in the intestines or blood of patients with IBD\textsuperscript{79}, not altogether surprising given the extreme artificiality of the animal model from which their presumed role in human disease was derived.

These mouse inflammation models are undoubtedly of great value in dissecting out immunological mechanisms and attributing roles to specific cellular populations and their associated cytokines. However, equating genetic mutations leading to bowel inflammation in mice with causal mechanisms of diseases in humans can have serious consequences as it may misdirect clinicians and scientists as to the underlying pathophysiology, and mislead pharmaceutical companies as to the relevant biological pathways against which to attempt to develop drugs. On the other hand, mouse models can be of great value when the problem is turned the other way around and they, and other animals like zebrafish, are used to inform as the population of suppressor T-cells called T-reg cells\textsuperscript{78}.

CD\textsuperscript{91} usually involves the terminal ileum, and the caecum and colon to a variable extent, where the lesions are patchy, known as “skip lesions”, and associated with strictures, and fistulae between the bowel and other loops of bowel, the skin, and pelvic organs like the bladder and vagina. Outside the bowel, at the sites of transmural inflammation, the mesenteric adipocytes hypertrophy, covering the exterior of the bowel with a layer of protective fat, a process known as fat wrapping. Anal disease affects about 40% of these patients\textsuperscript{83} exemplified by abscess, fistulation and skin tags. The inflammation is described as transmural, extending deep into the wall of the bowel, and contains diagnostic granulomata, collections of macrophages which represent a characteristic tissue response to retained foreign material. “The basic etiological factor in the case of all granulomas is probably the presence of a nidus of insoluble material which, if small enough is ingested by phagocytic cells, or, if too large, remains extracellular”\textsuperscript{84}. The central macrophages in these granulomata are surrounded by lymphocytes.

UC is very different in that it starts at the rectum and extends proximally, although occasionally, when it involves the whole large bowel there can also be involvement of the terminal ileum, a condition known as “backwash ileitis”. The inflammation in UC is superficial, being limited to the lamina propria, and the histological hallmarks are crypt abscesses and depletion of goblet cells that normally contain mucus.

Although a syndrome, the diagnostic features of classical CD are quite precise, and very different from the very rare cases of very early onset IBD, and the vast majority of genetically abnormal mice, both characterised by bowel inflammation rather than the clinical criteria used to diagnose CD or UC.

### The three phases of Crohn’s disease

A unifying model of CD pathogenesis has been proposed in which this condition develops in three temporally distinct phases\textsuperscript{85}:

- The trigger - gastrointestinal infection;
- A defective response to the consequences of this infection;
- A subsequent prolonged chronic inflammatory adaptive immune response.

### The trigger - Breeching the mucosa – The infectious environmental factor

**Epidemiology.** There is strong evidence for the role of an infectious environmental factor in the pathogenesis of CD. This is most obviously seen when populations or families emigrate from one country to another. A high proportion of family members have been documented as developing the disease after moving from Morocco to Belgium\textsuperscript{1}, from Albania to Greece\textsuperscript{17} and from India to Canada\textsuperscript{18}. After being imported into the household enteric infections can spread to family members\textsuperscript{19}. At a population level, an increased incidence of CD has been described in recent immigrants from Ethiopia to Israel\textsuperscript{19}, and from Eastern European and Iraq to Sweden\textsuperscript{19}.

The epidemiology of CD has been the subject of a large body of work and multiple reviews. Most pertinent to this paper are issues concerning environmental influences, several of which are clearly associated with CD as outlined below.

---

**Classification of IBD**

“Medicine is learned by the bedside and not in the classroom. Let not your conceptions of disease come from words heard in the lecture room or read from the book. See, and then reason and compare and control. But see first.”

*Sir William Osler*

IBD has referred to CD and UC because both can largely affect the colon and terminal ileum, however, although there may be overlap at the interface of these two conditions, their classical manifestations are quite different\textsuperscript{1,3}. They are both syndromes, rather than specific diseases, where common clinical pictures are united by a common set of diagnostic criteria produced by similar pathophysiological mechanisms.
Figure 2. The immunopathogenesis of CD occurs in three temporally distinct stages. Penetration of luminal contents into underlying tissues occurs in stage 1, which may be facilitated by environmental factors such as infection, or inherent defects in the mucosal barrier. In healthy individuals, resident macrophages secrete pro-inflammatory cytokines in response to this material, resulting in neutrophil accumulation, clearance of the material, and thereby resolution. In CD patients, defective secretion of pro-inflammatory cytokines by macrophages results in impaired neutrophil influx and clearance of foreign material (stage 2). Subsequently, chronic inflammatory responses (stage 3) will be triggered, giving rise to the characteristic features of the CD lesion. From Figure 1 (reproduced with permission from the publisher).
CD is generally more common in urban females of higher socioeconomic status\textsuperscript{95}, with a male to female ratio of about 1.5–2:1.

The difference in prevalence of CD by country could be partly explained by genetic factors; however, evidence from migration studies emphasise the importance of the environment. A limited number of studies investigating the incidence of CD among recent immigrants have been undertaken. The most informative of these assessed the risk of IBD in first- and second-generation immigrants to Sweden from many different countries\textsuperscript{92}. They found that overall risk of CD was lower in many groups of first-generation immigrants than in the native-born Swedish reference group but that in most groups of second-generation immigrants these decreased risks disappeared, and in some cases even exceeded those in the native Swedish population. First generation Middle Eastern immigrants to Australia developed CD at a much later age (~57 years) than the second-generation who developed it at about 28 years of age, roughly the standard age in Western society\textsuperscript{96}. CD is very rare in Ethiopia but emerged in Ethiopian Jews migrating to Israel after a median lag of about 12 years after arrival\textsuperscript{10}. CD is also more common in Bangladeshi immigrants to England\textsuperscript{99}. Combined, these studies imply that immigrants from underdeveloped countries initially have a resistance to CD that wanes over the subsequent decade or so.

**Infection.** Infection has long been considered to cause CD. Attempts were made to transmit a CD agent from gut or lymph node tissue of patients to wild-type or immunodeficient mice\textsuperscript{100}. More granulomata were found in the mice receiving CD tissue, but that could have been due to the fact that the inflamed tissue contained enteric organisms or inflammatory cytokines. In the first description by Danzel in 1913 of what was later to be called Crohn’s disease, the similarity between “chronic interstitial enteritis” and Johne’s disease in cattle\textsuperscript{101}, which is caused by infection with *Mycobacterium avium paratuberculosis*, was commented upon\textsuperscript{102}. Evidence that this agent was also responsible for human CD has been extensively sought\textsuperscript{103} but has not been forthcoming\textsuperscript{104,105}.

Several prospective studies have followed the course of patients after infections with enteric organisms and all have found an increased incidence of IBD as compared with uninfected control subjects\textsuperscript{106–108}. In one of these\textsuperscript{106} the risk was similar whether or not an infecting agent was identified, suggesting that it was the damage to the bowel rather than a specific infection that was important.

Enteric infections are most commonly caused by viruses, particularly *Norovirus*\textsuperscript{111}, and by *Campylobacter, Salmonellae, Shigella, Entamoeba histolytica, Cytomegalovirus* and *Yersinia*\textsuperscript{112,113}. Particular attention has been paid to an adherent-invasive subgroup of *E. coli*, that has been linked to the development of CD\textsuperscript{114–116}. The natural lesions produced by these organisms might provide some insight into those most likely to trigger CD. *Norovirus* mainly affects the proximal small intestine\textsuperscript{117,118} and *Amoebic* and *Salmonella* infections generally produce a diffuse colitis whereas the other infections result in lesions located in the terminal ileum and colon, with a patchy distribution, similar to those of the lesions of CD\textsuperscript{119–122}.

Whereas the incidence of most bacterial gastrointestinal infections is steady or falling, that induced by the commonest bacterial pathogen, *Campylobacter* is increasing in countries like North America, Europe, Scandinavia, Australia and New Zealand and Japan\textsuperscript{126} (Figure 4). This could be because this organism is a common contaminant of poultry, the consumption of which is increasing in these countries, however, broiler flocks are heavily contaminated with both *Campylobacter* and *Salmonellae*\textsuperscript{127} and the incidence of infection by the latter is steady or falling (Figure 4).

“Relapses” in cases of IBD have been reported to be associated with infections with various organisms including *Clostridium difficile,*
Shigella, Salmonella, Campylobacter, E. coli and Listeria\textsuperscript{[128]}. These subsequent infections might be inducing the development of a novel set of Crohn’s lesions in a pre-dosed bowel, rather than recrudescence of the original disease, as a result of infection by different organisms. These subsequent infections could be predisposed to by the immunosuppressive treatments commonly used in this condition including corticosteroids, cytotoxic and biological agents.

Most gastrointestinal infections do not generally produce homogeneous mucosal damage but lead to focal areas of ulceration\textsuperscript{[124,125]}, often in the ileocaecal region of the bowel. Because infection with invasive gastrointestinal pathogens is a stochastic process\textsuperscript{[129]}, the age at which this occurs is highly variable, as is the outcome after the infection, which will depend upon the severity of the infection, extent of ulceration, quantity of bowel contents gaining access to the tissues and to the effectiveness of the innate immune response.

The microbiome, prebiotics, probiotics and faecal transplants.

In the search for possible causal infectious agents, stool samples from CD patients have been extensively cultured and examined without a positive result (see for example\textsuperscript{[130]}). This is not entirely surprising because the average time from the onset of symptoms to diagnosis of CD is over six months\textsuperscript{[131]} by which time an infectious organism will have been eliminated if it was a triggering agent rather than the cause of a chronic infection. With the advent of next generation 16S rRNA gene sequencing the phylogeny and taxonomy of samples from complex microbiomes can be determined without the need for them to be viable or culturable. Dysbiosis of the faecal microbiome is well recognised in CD\textsuperscript{[132,133]}, with a decrease in the abundance and diversity of the \textit{Firmicutes} phylum and an increased abundance of \textit{Proteobacteria}, and alterations in the fungal composition\textsuperscript{[134]}. Differences were also found between the microbiotas of CD patients with ileal and with colonic disease\textsuperscript{[135]}. This could reflect an epiphenomenon secondary to the disease process. Major alterations in the microbiota are induced by diarrhoea\textsuperscript{[136]}, enteral nutrition\textsuperscript{[137]}, antibiotics\textsuperscript{[138]}, which most of these patients receive\textsuperscript{[139]}, and by iron therapy\textsuperscript{[140],} which is often prescribed because these patients are generally anaemic. In general, gut and mouth microbiomes display universal dynamics, unlike microbial communities associated with certain skin sites that are probably shaped by differences in host environment\textsuperscript{[141]}.

Because CD predominantly occurs in those regions of the bowel with a high bacterial count, and given the differences in the microbiotas in CD described above, attempts have been made to alter the intestinal microbiota in the treatment of this condition. Prebiotics are typically non-digestible, fibre rich materials, which stimulate the growth or activity of advantageous bacteria that colonize the large bowel, whereas probiotics are live microorganisms that are directly administered by mouth. Neither prebiotics nor probiotics have been shown to be beneficial in CD\textsuperscript{[142-144]}. An alternative means of directly altering the intestinal microbiota is by faecal microbiota transplantation, the transfer of faeces from a healthy donor, to restore the intestinal microbiota of a diseased individual. Whilst this is a logical treatment for \textit{Clostridium difficile} infection, which generally develops in a colon depleted of its natural microbiome by antibiotics, it has not been found to be effective in the treatment of CD\textsuperscript{[145,146]}.

The Hygiene Hypothesis.

The considerable increase in the incidence of CD in developed countries in recent decades\textsuperscript{[147]} has been attributed to immunological changes to alterations in the environment as outlined in the Hygiene Hypothesis\textsuperscript{[148]}\textsuperscript{[149]}. This hypothesis\textsuperscript{[149]} was first described by Strachan in 1989 who stated that “over the past century declining family size, improvements in household amenities, and higher standards of personal cleanliness have reduced the opportunity for cross infection in young families. This may have resulted in more widespread clinical expression of atopic disease”\textsuperscript{[150]}. Subsequently modern living conditions have been held responsible for the increasing incidence of a variety of so-called “auto-immune” diseases, including CD, which have been attributed to exposure a reduced load of microbes of decreased diversity. Certainly CD is less common in rural societies where there is exposure to animals, pets and soil, bedroom sharing is more common, and there is less access to hot water for ablutions\textsuperscript{[150]}.

According to this theory, standards of hygiene are lower in lower socioeconomic societies, leading to a greater abundance and variety
of gastrointestinal pathogens. This would lead to a high incidence of gastrointestinal infections in infancy and childhood, resulting in death or immunity. CD is very uncommon in underdeveloped societies in Asia, South America, China, and sub-Saharan Africa and the increase in its incidence is closely associated with the improvement in income and living standards. Enteric infections are endemic in these developing societies in which diarrhoea is a major cause of death in children less than 5 years of age. The population in underprivileged societies also host a large burden of gastrointestinal helminths and the low incidence of CD recorded in developing countries has been attributed to the high rates of gastrointestinal infections with these organisms. Helminthic infection was found not to be protective against CD in Denmark and the outcome of several trials of iatrogenic infection with helminths as therapy for CD are awaited, but current evidence does not suggest that they will be efficacious.

Immunity to enteric organisms is transient and may be strain specific, and would be boosted by frequent reinfection in less advanced countries. This herd immunity would be lost over time after immigration to socially advanced, cleaner, societies, which would accord with the later age of onset of CD in first generation immigrants. One could envisage a situation in which the population of more socially advanced countries are living under increasingly hygienic conditions and are exposed to a less diverse repertoire of the enteric microorganisms capable of producing gastrointestinal infection. With less frequent gastrointestinal infection, the bowel is uninfammed, with less primed macrophages, mast cells and dendritic cells in the lamina propria and adaptive immunity is more restricted, and relatively feeble, through the lack of repeated boosts by infection, making the bowel vulnerable to attack by a novel or virulent organism.

If we postulate that the trigger for CD is enteric infection, how can the fact that the incidence of food-borne gastroenteritis is fairly steady in most developed countries be reconciled with the rapidly increasing incidence of CD? Due to greater regulation and control of food production and distribution the incidence of foodborne outbreaks of disease have remained steady or have declined.

It is important to consider the age distribution at which patients present with Crohn’s disease. It rises to a peak at between 20 and 30 years of age after which it demonstrates a steady decline, a pattern that is remarkably consistent, and very different from that of UC, across the geographical spectrum.

The peak incidence, generally at a later age than puberty, coincides with a stage in life accompanied by major lifestyle changes. These include the movement of individuals out of the family home, in which the ambient microbiome is likely to be relatively stable, into environments in which the risks of exposure to infection are much greater. The main two ways in which young adults are exposed to infectious enteric organisms is through the ingestion of contaminated food or fluids, or by person to person contact, the risk of both being increased by travel to places where exposure to novel organisms is more likely.

### Sexual transmission

Although enteric infections are generally considered to be foodborne, only about one half are in fact transmitted in this way, most of the rest being transferred by person to person contact. Sexual transmission is worthy of consideration as a means of transmission of faecal organisms between individuals because, as might be expected, the peak age for the acquisition of sexually transmitted diseases is very similar to that of CD (Figure 5).

Epidemiological studies from developed countries have reported an increasing prevalence of invasive infections by *Entamoeba histolytica*, *Shigella*, *Cryptosporidia* and *Campylobacter*, among men who have sex with men (MSM), which is not surprising because of the increased risk of exposure to coliform organisms by oral, anal and oro-anal sexual practises. The ingestion of as few as 10 virulent *Shigella* organisms can confer full-blown dysentery. It is easier to establish the causality of infectious outbreaks in these groups of individuals as compared with the general population, because they fall into more readily identifiable groupings which facilitate the epidemiological studies. It would be important to establish the incidence of CD in MSM, but diagnosis in these individuals is complicated by the relatively small proportions of individuals attending gastroenterology facilities, the presence of compounding factors such as “gay bowel” and of sexually transmitted diseases like lymphogranuloma venerium that can masquerade as CD.

Given that oral and oro-anal sexual practises have been demonstrated to be responsible for the transmission of enteric infections in MSM, they must also pose a risk in other populations. Although rectal bacterial flora are present on the perineum of both sexes and in the vagina, it is unlikely that an increase in gastrointestinal infection would result from vaginal intercourse alone. Only about 5% of the sexually active individuals in countries like Britain and the United States are not heterosexual. In the heterosexual community the anal sex is practised by 30 – 40% of the population in England and North America, and fellatio and cunnilingus are almost universal. In England the participation in anal sex has almost doubled over the last three decades, a similar increase to that of the incidence of CD. In terms of absolute numbers, approximately seven times more women than homosexual men engage in unprotected receptive anal intercourse. In addition, the ratio of homosexual to bisexual men is about 3:1, and the latter can act as “bridgers”, transmitting infections from men who have sex with men into the heterosexual community.

Those countries with a high standard of living and low rate of CD like Japan, Taiwan, China, Korea, Saudi Arabia, and Malasia, appear to have low rates of heterosexual anal sex and in these countries the sex ratio of the disease, which is commoner in females than in males in Western countries, is reversed, implying that men are particularly vulnerable to infection in these places. This does not appear to be due to a reporting bias because the sex ratio of UC in these countries matches that of Europe and North America.
Figure 5. (A) The age distributions of Crohn's disease and ulcerative colitis in several European countries. (Reproduced from 172 with permission from the publisher). (B) Global age-sex distribution of new HIV infections (reproduced with permission).

Figure 6. Heterosexual sexual practices in Britain. Redrawn from the three National Surveys of Sexual Attitudes and Lifestyles. Each line connects values for the same birth cohort at different ages.
Monogamous heterosexual couples develop complementary microbiomes\textsuperscript{[197–198]} which would suggest that the highest risk to infection of either partner through sexual contact would be in the early stages of a relationship, and that the risk to an individual would be related to the numbers of sexual partners, some of whom might be asymptomatic carriers of pathogenic organisms\textsuperscript{[99–201]}.

Clearly the above arguments are conjectural and their validity will require validation through sound sociological, epidemiological and microbiological investigations.

Four other factors, smoking, antibiotics, appendectomy and invasive pneumococcal disease have positive correlations with the incidence of CD.

**Smoking**

Smoking of tobacco is the strongest environmental influence on CD, roughly doubling the incidence\textsuperscript{[202]} and relapse rate\textsuperscript{[203]}. Smoking and nicotine impair intestinal\textsuperscript{[204,205]} and gastric\textsuperscript{[206]} mucosal blood flow. Adequate blood flow is central to the development of an effective acute inflammatory response. Smoking also reduces levels of acute inflammatory cytokines in the bowel wall\textsuperscript{[207]} and lumen\textsuperscript{[208]} in patients with CD.

**Antibiotics and appendectomy**

There is an increased frequency of antibiotic use in CD prior to diagnosis\textsuperscript{[93,209]}. The increased frequency of antibiotic use may be explained by an increased number of childhood bacterial infections. Similarly, an increased frequency of tonsillectomies\textsuperscript{[209]} has been reported in CD and this may be an indication of recurrent pharyngitis. A further indication of a predisposition to infection in CD comes from the demonstration that these patients are more susceptible to invasive pneumococcal infection\textsuperscript{[210]}.

A history of a greater frequency of appendectomy in CD is also in keeping with an increased susceptibility to childhood infection in this condition\textsuperscript{[199]}.

**Phase two – a defective inflammatory response**

“Any infectious agent associated with Crohn’s disease is likely to be a widely distributed organism to which some people react abnormally - that is, the disease is unlikely to show the characteristic features of an infectious disease”\textsuperscript{[211,212]}.

As described above, there is very good reason to believe that the initiating lesion in CD is infection by one of a number of enteric pathogens. The key to comprehending how the pathological lesions of the disease then develop lies in understanding the response to that initial infection. The infection by the organisms described above is very unlikely to persist, or the causal connection would have been clearly established some time ago. This is also the reason that antibiotics are of only limited efficacy in the treatment of CD\textsuperscript{[212]}.

These patients are unlikely to be unduly susceptible to infection by these organisms, or else the onset would occur earlier, and systemic disease would be expected, as occurs with *Salmonella*ae when the interferon-gamma/IL-12 axis is disrupted\textsuperscript{[11]}.

**Immunoparesis of the acute inflammatory response is the underlying Crohn’s phenotype**

The underlying pathology in Crohn’s disease is the ineffective manner in which the faecal material entering the tissues through the damaged mucosa is dealt with. Infective damage to the mucosa followed by the entry of faecal material with a bacterial count of greater than 10\textsuperscript{11} bacteria per ml into the tissues poses an existential threat that must be dealt with vigorously. This is accomplished by the acute inflammatory response, a non-specific local reaction to tissue damage that recruits the innate immune system. It includes the secretion of inflammatory mediators from mast cells and macrophages, complement activation, markedly increased blood flow, capillary dilatation and permeability, the deposition of a fibrin network, and most importantly in the context of CD, a massive influx of neutrophil leukocytes, highly motile phagocytes that ingest and kill invading bacteria and fungi and digest foreign organic material.

The underlying, and unifying, predisposition to the development of CD is a systemic incompetence of this acute inflammatory response. I will deal with the evidence supporting this immunoparesis in some detail because these experiments have been performed on CD patients, and healthy control subjects, and in some cases patients with UC and represent a unique set of data that have not been repeated, possibly because of the invasive and uncomfortable investigations required to obtain them.

The delay in the recruitment of neutrophils to sites of trauma to the body by the innate immune response has been demonstrated in patients with CD in several different but complimentary ways. In 1976 I demonstrated that the accumulation of neutrophils in superficial abrasions on the arm called “skin windows”, was grossly deficient when compared with healthy subjects or patients with another chronic inflammatory condition, rheumatoid arthritis\textsuperscript{[214]}. It was observed that “This abnormality of neutrophil function in Crohn’s disease appears to be secondary to a defective acute inflammatory response as the neutrophils themselves were found to behave normally on in-vitro testing. A weak acute inflammatory response to particulate or antigenic material in the bowel wall could result in the chronic inflammation observed in this condition.”

The next in these series of experiments was conducted on the ileal and rectal mucosa, and again on the skin\textsuperscript{[11]} A small mucosal biopsy was taken from the ileum or rectum, and this was then followed 6 hours later by a further biopsy of the previous biopsy site. This was repeated in both regions of the bowel. In addition to healthy subjects, control individuals with UC were studied and their neutrophil recruitment was normal. Trauma to the skin reproduced the impaired neutrophil recruitment into skin windows, as well as reduced secretion of IL-8 and IL-1β from them.

The direct injection of heat killed *E. coli* into the subcutaneous tissues of the forearm of normal subjects was followed by profound rise in local blood flow. This was considerably impaired in CD, but not in UC. Blood flow is important in recruiting innate
immune cells to sites of inflammation and this already paltry vascular response in CD would be further compromised by smoking tobacco.

The third of these experiments directly measured the accumulation of neutrophils at the site at which E. coli had been injected subcutaneously, and the rate of clearance of these organisms. In this study peripheral blood neutrophils were purified from the individual under investigation, labelled with the gamma-ray emitting radioisotope Indium-111, and reinfected intravenously at the same time that unlabelled E. coli were injected subcutaneously into the forearms. The rate of accumulation of the radioactive neutrophils over the site of the injected bacteria was determined. A much smaller proportion of neutrophils were recruited to the injected bacteria in the CD subjects than in the HC or UC individuals (Figure 7).

The next step was to radiolabel the E. coli with Phosphorus-32 and to then determine the rate of clearance of the bacteria from the tissues. This was a two phase process in HC and UC subjects with a very rapid initial clearance lasting about 4 hours followed by a slower phase, with total clearance being achieved by 7 to 10 days. In the CD subjects, initial clearance was much less efficient and total clearance was markedly delayed and was predicted to last from several weeks to infinity. This study showed unequivocally that coliform bacteria are cleared less efficiently from the tissues than normal in CD. It might be considered that this delayed recruitment of neutrophils to bacteria in the tissues should predispose these individuals to an increased incidence of clinically evident infections, which is not an obvious manifestation of CD. The reason for this apparent discrepancy is that the numbers of bacteria injected into the tissues were required to reach a certain critical load before the clearance defect was unmasked (Figure 7e). In this study 10⁹ organism were cleared normally whereas 10⁶ were not, indicating that a significant bacterial load must enter the tissues before the clearance systems are overwhelmed. The bowel is the only location in the body where such a burden of microbes is readily available to enter the tissues.

Phase three - The consequences of the failure to clear intestinal contents from within the bowel wall

In the absence of an adequate acute inflammatory response and the complete clearance of the inciting agent by neutrophils, the retained foreign material produces a granulomatous inflammation. E. coli, Streptococci and Listeria have been demonstrated immunohistochemically in macrophages, giant cells and lymph nodes of CD patients, and E.coli DNA has been identified in Crohn’s granulomata isolated by laser capture microdissection. The retention of this faecal material within the bowel leads to an intense adaptive immune response and the tissues become infiltrated with large numbers of T-cells. It is not therefore surprising that when actively inflamed CD tissues are biopsied, any number and variety of adaptive immune cells can be identified and immune mechanisms evoked in the pathogenesis of the condition. The macrophages and adaptive immune cells, reacting to the foreign antigenic material, will produce cytokines such as IL-1β and TNFα that lead to local inflammation and systemic symptoms.

The clinical picture of an inflamed bowel containing large numbers of macrophages and T-cells has led to the erroneous belief that Crohn’s was an autoimmune disease. It is however clear that the cytokines produced by these inflammatory foci in their response to foreign faecal material contribute to the local and systemic inflammation, and failure of mucosal healing, as evidenced by the, often dramatic, responses to anti-TNF drugs. However, only about half the patients respond to this treatment, and in those that do the response is often partial and temporary.

Chronic inflammation leads to fibrosis, or scarring, which in a hollow muscular organ causes narrowing, or stricture formation. Under some circumstances the material in the bowel wall undergoes liquefaction, as may occur with tuberculosis. This material then tracks to adjoining organs, possibly driven by the osmotic pressure produced by the breakdown of the organic material within the abscess, and discharges into them. This can then produce fistulae between the these organs for example between bowel and bowel, bowel and skin, bladder or vagina. The perianal fistulae between the rectum and periureum are characteristic of CD and of immunodeficiencies of the innate immune system, particularly those of neutrophil function.

This failure to clear organic material from the tissues offers an explanation for the false positive Kveim tests observed in CD. The Kveim test was designed to diagnose sarcoidosis, another chronic granulomatous disease. The intradermal injection of a crude homogenate of an extract of sarcoid tissue, usually from lymph node, produced epithelioid cell granulomas in subjects with sarcoidosis, reproducing those diagnostic of this disorder. Initially it was thought that the injected material contained some sarcoid specific factor, such as an infectious agent or antigen, but it has been recognised more recently that it relates to an abnormal host response:

“The “immune paradox” (delayed type hypersensitivity anergy in a setting of exuberant systemic granulomatous response) resists explanation. Its relationship to the Kveim test is poorly understood. Immunological investigations generated the thesis that the characterizing systemic granuloma arise as a fall-back reaction to inefficient cellular immune processing, due most often to impaired myeloid dendritic cell function of unknown cause.”

This is precisely the nature of the pathogenic mechanism in CD and it is therefore not surprising that positive tests are found in both conditions and that both diseases occasionally coexist in the same individual.

On the location of the CD lesions

Symptomatic lesions are largely confined to the terminal ileum, caecum and colon, probably due to the combination of mucosal damage by enteric infection coupled with the ready presence of massive numbers of bacteria to penetrate into the wall of the bowel when this happens. However, it is becoming apparent that the gastrointestinal tract is generally diffusely, sub-clinically, abnormal.
Figure 7. Neutrophil accumulation and subsequent clearance of *E. coli* from the tissues is markedly delayed in a dose-dependent manner in CD. Reproduced from 218 with permission from the publisher. 111Indium-labeled autologous neutrophils were injected intravenously at the same time as killed *E. coli* were injected subcutaneously into each forearm. (a) Radioactivity measured over the injection sites showed a much smaller proportion of labelled cells accumulating in CD subjects. (b) γ-Camera image of a CD patient at 24 h after injection, demonstrating focal accumulations of radioactivity at bacterial injection sites (arrows) and confirming lack of bowel inflammation. (c) 32P-labeled killed *E. coli* were injected into the subcutaneous tissues of the forearm and radioactivity was measured at the skin surface. Clearance of radioactivity was much slower in CD than in HC or UC. Extrapolating these curves indicated that almost complete removal (99%) would take 10.2 and 7.1 d in HC and UC subjects, respectively, compared with 44.3 d in CD. (d and e) Effect of increasing bacterial dose from $10^5$ to $10^8$ on blood flow (d) and bacterial clearance (e). The numbers of subjects studied in the dose response experiment are depicted in e. All results are expressed as mean ± SEM (**, P < 0.01; ***, P < 0.001).
Oral manifestation of CD, particularly aphthous ulcers, are estimated to occur in 20–50% of patients\(^\text{34}\). A prospective endoscopic study identified upper gastrointestinal (GI) manifestations of CD in 55% of 108 untreated, newly diagnosed adult patients with CD, irrespective of symptoms. All selected were free of \textit{H. pylori}, infection with which, if anything, appears to protect against CD\(^\text{220}\). About a quarter of the patients had lesions in both the stomach and duodenum and in about 20% they were in one or other of these organs. In roughly 2% of patients the gastric outlet is obstructed by a granulomatous inflammation requiring surgical intervention\(^\text{221}\). Aphthous ulcers in the oesophagus were present in 7% of these subjects. Most of these lesions exhibited a granulomatous inflammation on histology.

In view of the systemic nature of the impairment of the innate immune system in CD, it is of great interest, although not altogether surprising that patients with CGD\(^\text{217}\) exhibit very similar upper GI pathology. Aphthous ulceration and other oral lesions are common. Oesophageal, gastric and duodenal inflammation were detected in 21%, 74% and 37% of 78 patients\(^\text{217}\). Large bowel lesions were present in the majority\(^\text{217}\) and are indistinguishable from those of CD\(^\text{210,214}\). Between 4%\(^\text{217}\) and 15%\(^\text{214}\) of these patients also develop gastric outflow obstruction.

CGD is a condition in which there is a failure of microbial killing and digestion by neutrophils as a result of an absence of the respiratory burst produced by a NADPH oxidase, NOX2. Consequence, the pH of the phagocytic vacuole is too low for the efficient activity of the neutral protease digestive enzymes released into the vacuole from the cytoplasmic granules, and they fail to kill and digest the microbes\(^\text{245,246}\). The undigested material retained within the tissue is taken up by macrophages, producing the granulomata that give this condition its name.

Neutrophils play an important role in the debridement of wounds\(^\text{247,248}\), an important prelude and necessity for healing. It is possible that the upper GI inflammation that occurs in CD and CGD results from an impaired repair response to trauma and peptic digestion rather than infection in these locations.

**Identifying the molecular cause/s of the CD phenotype**

The abnormality of innate immunity in most cases of CD lies in the macrophages

Defective secretion of pro-inflammatory cytokines in CD may be the explanation for the observed impairment in neutrophil recruitment\(^\text{218,249-252}\). In CD, the neutrophils themselves are normal\(^\text{213}\) and exhibit normal migration \textit{in vitro}\(^\text{214,249,253}\) and will migrate out of skin windows if chemotactic substances are placed over them\(^\text{213}\). In the absence of a primary abnormality of neutrophil function, CD macrophages showed defective secretion of pro-inflammatory cytokines, but normal release of chemokines, in response to stimulation with \textit{E. coli}\(^\text{216}\) (Figure 8). The genes for these pro-inflammatory cytokines were transcribed and translated, but the proteins were misdirected to lysosomal degradation rather than secretion, suggestive of disordered vesicle trafficking.

The question then arises as to how anti-TNF drugs can be effective against a condition in which the secretion of TNF and other pro-inflammatory cytokines is impaired? The answer is in the timing of the different components of the immune system. The call to arms of the innate immune system is a very early and explosive secretion of pro-inflammatory mediators, including TNF. If the clearance of faecal material from the tissues is incomplete, it becomes walled off by macrophages, endotoxin diffuses into the circulation, and cells of the adaptive immune system are recruited\(^\text{250}\). They secrete a wide array of mediators over the next weeks including TNF which acts as an amplifier of the response\(^\text{247}\). It is of interest that in a recent study of high-resolution gene expression profiling using RNA sequencing of inflamed biopsies from patients with CD, UC and controls, levels of pro-inflammatory cytokines like TNF, IL-1β, IL-6 and IL-23 were all elevated to a lesser extent in CD than in UC\(^\text{251}\). The very early secretion of TNF and other mediators is required to prevent the development of the Crohn’s lesions whereas at a later stage it is the TNF and associated mediators that produce the symptoms, which in some cases respond to anti-TNF treatment. This explains why anti-TNF therapeutics can both cause\(^\text{250}\) and alleviate symptoms of the disease\(^\text{260}\).

**What is the molecular cause of the impairment of acute inflammation?**

There is a strong genetic component to the aetiology of CD. The sibling recurrence risk (risk of developing the disease in the context of an affected sibling) is approximately 13–36\(^\text{261}\) and approximately 12%\(^\text{261}\) of CD patients have at least one affected first degree relative. Furthermore, the study of over 300 twin pairs has demonstrated a higher concordance of disease phenotype in monozygotic (30%) compared with dizygotic twins (4%)\(^\text{262}\). While the twin studies support the role of genetic susceptibility, they also indicate the requirement for additional environmental or other factors for the development of overt disease. By far the most likely such factor is an enteric infection of sufficient severity to overwhelm the ability of the innate immune system to adequately clear the faecal debris from the bowel wall. However, this risk may be further modulated by additional environmental factors, such as smoking. This phenomenon, whereby a genetic predisposition to disease manifests in the presence of environmental precipitants is exemplified by alpha-1-antitrypsin deficiency, in which the predisposition to emphysema is exposed by smoking\(^\text{264}\).

Technological advances have provided the means of interrogating the genetic basis of CD.

1. **Linkage and Genome Wide Association studies (GWAS)**

   **Linkage.** Linkage analysis (positional cloning) is a family based technique for identifying the possible location within the genome of causal mutations underlying genetic diseases\(^\text{265}\). This is done by utilising markers of known location across the genome, such as microsatellites or single nucleotide polymorphisms (SNPs). The transmission of the markers through a family (or collection of families) is examined seeking those whose segregation closely follows the inheritance of the disease, thereby focussing attention on a small region (locus) in which the causal mutation might be found\(^\text{262}\). Linkage analysis of affected sibling pairs with CD permitted the identification of a susceptibility locus on chromosome 16 (termed \textit{IBD1}) in which mutations in the gene \textit{NOD2} were subsequently identified\(^\text{266,267}\). \textit{NOD2} mutations remain the most strongly associated common genetic variants associated with CD. Linkage is only
a powerful tool when almost all cases of the disease in the families under study are caused by mutations in the same gene (i.e. there is limited genetic heterogeneity) that are not seen in unaffected family members (i.e. it is of high penetrance). Numerous factors can limit the effectiveness of linkage analysis such as: the presence of unaffected individuals that harbour the mutation (incomplete penetrance); individuals who develop the disease as a result of mutations in another gene or due to environmental factors (pene-nocopies); the requirement for the combined effect of two or more mutations (epistasis); or the requirement of the involvement of some environmental factor such as an infectious trigger (which will effectively reduce penetrance by not facilitating the manifestation of the underlying genetic predisposition in unexposed individuals). All of these factors are likely to have contributed to the limited success of linkage analysis in CD.

GWAS. Genes reside on chromosomes which undergo recombination at meioses. Population level haplotypes arise due to the non-random positioning of crossing-over events. Haplotypes are characterised by a particular set of SNP genotypes. Depending on the ancestral origin and frequency with which a mutation has arisen in the population, it may occur on a particular haplotype and thus the SNP genotypes defining that haplotype will be enriched in patients harbouring the disease-causing mutation. Therefore, when comparing a large population of diseased individuals with healthy controls, SNPs tagging the underlying mutation should be enriched in the affected compared with unaffected individuals. In GWAS, a set of SNPs are genotyped in an attempt to cover the whole-genome and the above comparison made\textsuperscript{99}. One of the major problems with analysing many hundreds of thousands (or millions) of markers across the genome is that the large number of comparisons undertaken risks producing false positives. This necessitates the utilisation of a stringent p-value threshold for significance of p<5×10\textsuperscript{-8}. As a result very large sample sizes are required\textsuperscript{170,271}.

One of the technical strengths of GWAS as an investigative approach is that DNA is easily obtained and once purified it is stable, enabling it to be conveniently stored and transported. Technological advancements have permitted high throughput SNP genotyping of large numbers of samples on an industrial scale. Furthermore, the GWAS approach has the advantage of being comprehensive (compared with candidate gene studies) and objective (at least up until the stage of data interpretation). There are however a number of limitations, for example incomplete genomic coverage. In addition, a major drawback is that the SNPs employed as markers must be relatively common in the general population, in order to give the study adequate statistical power, so this approach is typically unable to identify low frequency mutations, however penetrant or important.

When a SNP is found to be statistically significantly associated with a disease by GWAS, it can be because the polymorphism is itself pathogenic or, more commonly that it is tagging a closely located genetic variant whose genotype correlates with that of the tagging polymorphism (the two variants are in linkage disequilibrium). The precise location of the causal variant(s) underlying the association signal within the identified locus may be interrogated further by fine mapping (in which higher resolution association studies are conducted) or by resequencing the locus looking for plausible pathogenic variants such as coding variants or those that affect gene-expression (eQTLs).

Increasingly large GWAS have been performed on CD and the results meta-analysed\textsuperscript{272,273}. No single, or small number, of penetrant mutations have been found that independently cause the disease. The latest study of over 20,500 CD cases and 41,600 controls of European ancestry identified 145 loci associated with CD at p<5×10\textsuperscript{-8}. The mean OR of the top SNPs representing these 145 loci was 1.16 and the mean control allele frequency was 0.48. Four SNPs had an OR exceeding 1.5 of which three were

**Figure 8. Proinflammatory cytokine secretion by macrophages from CD patients is deficient in response to *E. coli*.** Cytokine and chemokine release expressed as a percentage of that secreted by HC cells (blue bar) from ileal and colonic CD patients (reproduced from 218 with permission from the publisher).
within NOD2 and the fourth was in IL23R. The mean difference in allele frequency between cases and controls was only 0.02274.

The very significant p-values obtained for the associated loci, led to the general perception that the molecular causes of CD have been identified. Individually the CD GWAS loci have very modest effect sizes (i.e. a small difference in frequency in the control and CD populations), consistent with a polygenic model in which it is thought that it is the combination of these minor influences that is causally important. However, in the latest meta-analysis, all 170 significantly associated loci combined account for only 10.9% of the disease “heritability”272.

An important consideration is that about half of the healthy population also carry these variants, although (by definition) each at a slightly lower frequency than in the CD patients (Figure 9). The healthy controls carrying these variants greatly outnumber the patients with CD in the population. With a prevalence of CD of about three patients in 1000275, and taking the NOD2 frameshift mutation as an example because it has the greatest effect size at 3.32, for every 100,000 individuals in the population there will be 99,680 unaffected individuals of whom ~2390 will carry this mutation. In this population there will be ~320 CD cases of which ~48 will have the mutation276. This means that the penetrance of this mutation, with by far the greatest association with the disease, is only 2%. These effect sizes pale into insignificance when compared with the effect size of HLA-B27 in ankylosing spondylitis of approximately 94277,278, and HLA in type 1 diabetes and coeliac disease with effect sizes of approximately 25 and 50, respectively279.

GWAS conducted on Europeans and East Asian populations have yielded noticeably different findings (Figure 10)280. In East Asian populations, the loci with largest effect sizes were those harbouring the genes TNFSF15/TNFSF8 (genes encoding cytokines that belong to the tumour necrosis factor (TNF) ligand family281) and the major histocompatibility complex. Variants in NOD2 and ATG16L1 demonstrated no association with CD in these populations, and the effect size of the IL23R locus was minimal. The significant heterogeneity in common variant CD genetic architecture between different populations provides a further indication that the genes identified by GWAS are unlikely to play a primary causal role in the development of the disease, the manifestations of which are similar in patients regardless of their ethnicity.

GWAS have been performed for many different diseases and IBD associated loci have been shown to be shared with several other immunologically mediated diseases including Rheumatoid arthritis, Systemic lupus erythematosus (SLE), Ankylosing spondylitis, Coeliac disease and Sarcoidosis282. These associations are not surprising as comorbidities of some of these conditions are well recognised in the context of IBD283–285 and because almost all are associated with an increased incidence of similar pathologies such as arthritis284, uveitis286,287 and bowel inflammation288–291.

Many of the associated genes common to these conditions have been implicated in pathways leading to activation or regulation of the immune response292. It is possible that these genes were highlighted as being common to the chronic inflammatory conditions because they lead to more florid manifestations, causing signs and symptoms of the disease in those individuals with an underlying predisposition, thereby bringing them to the attention of the medical profession.

Despite their unimpressive effect sizes, the main GWAS CD associated molecules have generated considerable attention and are hence worth of a brief summary.

**NOD2**. NOD (Nucleotide-binding Oligomerisation Domain) 2 is a member of an extended family of inflammatory and immune proteins in plants (the resistance (R) genes293), Drosophila (Toll-like receptors294) and animals (NOD families). These proteins combine a central nucleotide-binding domain (NOD) with a C-terminal leucine-rich repeat (LRR) motif and an N-terminal caspase recruitment domain (CARD) or equivalent.

In general these proteins recognise a signal from an invading organism in their leucine-rich domain (LRR) domain that induces a polymerisation that triggers a signalling cascade which terminates in the production and release of pro-inflammatory molecules. NOD2 is activated by muramyl dipeptide (MDP) a component of the cell wall of both Gram negative and Gram positive bacteria. It seems to be taken into the cells within endocytic vacuoles; presumably the organisms are then digested within this compartment and the solubilised MDP is moved into the cytoplasm by peptide transporters like SLC15A3295. Very recently it has been demonstrated that NOD1 and NOD2, do not only respond to bacterial stimuli, but are also important mediators of ER-stress-induced inflammation (described in more detail below)296.

---

**Figure 9.** The allele frequency in Crohn’s disease patients and healthy controls for top 20 of 144 CD associated GWAS SNPs sorted by odds ratio. The data were taken from the European cohort in 275. Loci harbouring genes of interest have been indicated.
In the resting state NOD2 is doubled back on itself in an auto-inhibited conformation in the cytoplasm until activated by the attachment of MDP upon which opens it allowing self-oligomerisation and the binding of ATP\textsuperscript{296,297}. A series of phosphorylation steps then end in the translocation of NF-κB to the nucleus and the production of pro-inflammatory cytokines and antimicrobial peptides\textsuperscript{298}. This theoretical model of NOD2 function as a pattern recognition receptor capable of inducing pro-inflammatory cytokine secretion has been validated by \textit{in vivo} studies in humans in which the application of MDP to skin windows induced the production and release of pro-inflammatory cytokines in healthy and CD patients without NOD2 mutations, but not in those carrying the CD-associated mutations\textsuperscript{215}. The impaired secretion of these inflammatory mediators into skin windows in the absence of MDP, in CD patients without mutations in NOD2 in this study, demonstrates that other pro-inflammatory signals and pathways must be abnormal in these subjects, indicating that there are at least two parallel routes initiating the inflammatory response.

Expression of NOD2 is largely restricted to peripheral blood monocytes\textsuperscript{299} and to Paneth cells at the base of intestinal crypts\textsuperscript{299}. Monocytes constitute approximately 5% of the circulating leukocytes and are generally regarded as functioning predominantly as circulating precursors of macrophages, without much in the way of a distinct set of functions of their own. It seems intuitively unlikely that a highly specialised cell with an active NADPH oxidase and granules containing myeloperoxidase, all of which are lost with the transformation to macrophages, would be produced to act predominantly as a stem cell.
It is generally assumed that pro-inflammatory cytokines are secreted by tissue macrophages, but these are quite widely dispersed amongst tissues and could not accumulate as rapidly as needed at inflammatory sites as monocytes that are carried there in capillaries perfusing the region\(^{103,104}\). Monocytes are rapidly recruited to sites of acute inflammation where they extravasate into the tissues\(^{89,90}\) and make a large contribution to the production of pro-inflammatory cytokines\(^{85-88}\) before being transformed into inflammatory macrophages.

**Autophagy and ATG16L1, CALCOCO2/NDP52, LRRK2 and Optineurin.** First described in the new-born mouse kidney by Clark in 1957\(^{86}\), and reviewed by De Duve and Wattiaux\(^{107}\), autophagy was initially described as a process in the cytoplasm of cells directed to the remodelling of tissues and the removal or damaged or effete organelles and proteins, and to partial self-digestion under starvation conditions.

To undertake this process, cells must first identify the region of cytoplasm, effete organelles, or invading microbes, as objects for engulfment. This is achieved by labelling the surface of the target with chains of a small protein, ubiquitin\(^{80,89}\). The ubiquitinated material is then enclosed by a double membranous structure, produced from elongated vesicles at the Golgi apparatus or (ER)\(^{110}\), the ends of which then fuse to form the characteristic vacuole with a double membrane. This autophagocytic vacuole then fuses with lysosomes containing enzymes that digest the inner membrane and its contents\(^{111}\).

Membrane vesicle extrusion, ubiquitination and fusion of granules with vesicles are general biological processes that are involved in many diverse cellular functions in addition to autophagy. Most of the cellular machinery required for these processes in autophagy was identified in mutant *Saccharomyces cerevisiae*\(^{132}\) and homologues were then found in *Drosophila* and man. The molecular basis of human autophagy has largely been investigated in pro-myelocytic HL60 and the Human Embryonic Kidney (HEK293) cell lines\(^{113,114}\). Thus the autophagy molecules have been identified in assays that measure autophagy in primitive cells. This does not mean that these proteins are necessarily exerting their effects exclusively through autophagy in more mature cells and tissues.

What is probably of particular relevance to CD is the specialised type of autophagy, known as xenophagy, that is designed to cope with bacteria that escape into the host cytosol (such as *Listeria* or *Shigella* or *Listeria*) or those that reside in a modified intracellular vacuole (such as *Salmonella* and *Mycobacteria*), which is important for the entrapment and lysosome-mediated degradation of bacterial pathogens\(^{115-117}\). CD resulting from abnormal xenophagy is seen in Niemann-Pick disease type C1 and XIAP deficiency with NOD2 variants in Crohn’s disease\(^{118}\). Other proteins implicated in xenophagy are the autophagy receptor CALCOCO2/NDP52\(^{115}\), Leucine-rich repeat kinase 2 (LRRK2)\(^{120}\) and Optineurin\(^{21,32}\).

Despite the current interest in autophagy and in exploring the removal of intracytoplasmic organisms from intestinal cells and macrophages by xenophagy, it must not be forgotten that the vast majority of bacteria entering the body are phagocytosed, killed and digested by neutrophils, and that xenophagy only has to deal with a tiny minority escaping this process.

**ATG (autophagy-related) 16L1.** Attention was focussed upon ATG16L1 when an association was demonstrated with CD in a GWAS study\(^{123}\). The SNP was a common, non-synonymous variant resulting in an amino acid change from threonine to alanine at position 300 of the full-length protein (T300A) with an odds ratio of only 1.26\(^{127}\). The ATG16L1 protein was shown to be mainly expressed in the thymus, prostate, liver, kidney and colon with little in the small bowel, and equal levels were found in the bowel from healthy and CD subjects with and without the polymorphism.

ATG16L1 is recruited to the ER at the initiation of the formation of the autophagosome at this site\(^{111}\). It is involved in binding to ubiquitinated cytosolic *Salmonella* whilst these are enveloped in the phagophore to form the autophagosome\(^{124}\) and could be important for the clearance of bacteria within cells\(^{114,125-127}\). This protein could be related to CD through the impaired clearance of intracellular organisms. It has been reported that the cellular architecture of Paneth cells is grossly distorted in knock-out mice severely depleted in this protein\(^{128}\) and in CD patients homozygous for the T300A polymorphism. In the mice, there were many fewer granules in these cells, with their normal contents lying predominantly in the cytoplasm, together with multiple large vesicular structures that might represent the membranes that would normally surround granules. The rest of the bowel looked normal\(^{123}\). It was also reported that CD patients with the T300A polymorphism demonstrated similar morphological changes with disorganized, or diminished, granules or exhibiting diffuse cytoplasmic lysozyme staining. This was claimed to be the first indication that Atg16L1 has a specific role in humans and mice in regulating the specialized properties of Paneth cells, and provides a novel and relevant mouse model that emulates one of the many diverse pathological hallmarks of human CD. Unfortunately the morphology of Paneth cells was not examined in healthy humans with the same polymorphism, which would not have been difficult given that 27% of this population is homozygous for the “risk” allele and 50% heterozygous, and which makes a pathological role for this SNP highly improbable\(^{127}\).

**IRGM.** Immunity related p47 guanosine triphosphatases (GTPases) or IRGs are a group of 47–48 kDa proteins that are implicated in the eradication of organisms like mycobacteria that are taken into cells within phagocytic vacuoles. IRGM has been characterised as an autophagy related molecule\(^{129}\) on the basis that it was associated with vacuoles thought to contain mitochondria, a characteristic in common with autophagocytic vacuole. The evidence for the mitochondrial contents was the accumulation in the vacuoles of Mitotracker Red, a dye that partitions in mitochondria by virtue of the charge across their membranes. Mitochondria are marked for degradation by ubiquitination when they have lost their membrane potential\(^{129}\) and this situation would not be improved by degradation within an autophagosome, so it is unlikely that such structures would attract Mitotracker Red. On the other hand, potentials are developed across the membrane of the phagocytic vacuole by the NADPH oxidase\(^{131}\) and by electrogenic proton pumping of the vacuolar V-ATPases\(^{132}\). The classification of IRGM as primarily involved...
in the autophagocytic processes on the evidence provided for mitophagy should be reconsidered.

The IRGs were discovered in mice as interferon-gamma inducible elements that were important for resistance to microbes engulfed into phagocytic vacuoles. There are 23 of these molecules in mice but only one or very few in humans\(^{33}\). In human cells IRGM is largely expressed in immune cells whereas in mice it is almost exclusively in macrophages in which its expression is massively increased by LPS (\(^{15}\)GeneAtlas MOE430 gcrma, Probeset: 1418825_at). Phagosomal maturation\(^{33,36}\) occurs through sequential fusion of the phagosome with early endosomes and lysosomes, resulting in acidification of the phagosome, release of hydrolytic proteases and death and digestion of the phagosome-bound pathogen. IRGs are located in the ER or Golgi compartments and after the host cell is infected, they are transported to pathogen-containing phagosomal/ vacuoles, where they modulate the formation or processing of the phagosome, underpinning pathogen survival either directly or by facilitating the action of intracellular effector molecules.

**XBP1 (X-box binding protein 1) and ER stress, or unfolded protein response (UPR).** The ER is the principal organelle involved in the synthesis, maturation, and post- or co-translational modification of secreted and membrane proteins, as well as in various metabolic processes including dynamic ion storage and biogenesis of membrane structures\(^{36}\). Properly folded proteins are then directed to the Golgi apparatus, to other intracellular organelles, and to the extracellular surface by the secretory pathway\(^{36}\). The ER stress or unfolded protein response develops when the protein synthetic process becomes disordered; leading to the accumulation of dysfunctional unfolded protein as a result of genetic or adverse environmental factors. To deal with this, the unfolded protein triggers the unfolded protein response (UPR)\(^{37,38,34}\), which includes reducing protein synthesis by downregulating mRNA translation, synthesising more molecular chaperones to assist protein folding, and degrading misfolded proteins which are ubiquitinated and directed to the proteasomes. XBP1 is a transcription factor that plays a central role in activating these ER stress responses\(^{42}\), and unexpectedly, NOD1 and NOD2 have been shown to be important mediators of this process\(^{39}\).

Major demands are made on the protein synthetic machinery in rapidly turning over cells like the intestinal mucosa, and this is particularly true of cells like Paneth and goblet cells that secrete proteins in addition to attending to their own homeostatic requirements. When additional requirements for protein synthesis are called upon, induced, for example, by infection or chemical toxins, defects in the ER stress response could weaken the mucosa leading to ulceration\(^{34}\).

**LRRK2.** A considerable amount of work has been done on this protein because mutations in LRRK2 cause familial and sporadic Parkinson’s disease (PD). It is a large, 280kDa, protein with GTPase and kinase domains, the latter being constitutively active in PD. It is found in immune cells, in lamina propria macrophages, B-lymphocytes and dendritic cells, and levels are markedly increased in the bowel in CD\(^{35}\) and in microglia in the nervous system\(^{36}\). It interacts with the small GTPases Rab32 and Rab38 with which it co-locates to transport vesicles and recycling endosomes\(^{37}\) and it is important for the elimination or intracellular *Salmonellae*\(^{38}\) and *Legionella*\(^{39}\). Rab32 and Rab38 play an important role in the biogenesis and traffic of melanosomes and lysosomes and this system is disordered in Hermansky-Pudlak syndrome\(^{40}\), accounting for the characteristic partial albinism. If LRRK2 and its associated proteins are important for immunological resistance to the development of CD then it might be expected that CD would be more common in conditions in which the LRRK2 system is disordered, which is in fact the case. A clear association exists between CD and PD\(^{40,30}\) and CD and Hermansky-Pudlak\(^{51}\). At least two recent studies have discovered a link between mutations in LRRK2 and CD in Ashkenazi Jews (submitted for publication).

**Summary of outcome of GWAS studies in CD**

The GWAS studies have provided a series of clear answers. No single gene, or a small number of genes, has been identified that is causal for CD. More than 170 GWAS hits combined contribute to about 10% of the “heritability” of CD. With the average individual contribution of only 0.1% it is unlikely that these variants will individually have major effects on cellular function, either in the CD patients or in experimental systems.

Collectively, genes associated with CD by GWAS are enriched for Gene Ontology (GO) annotations (statements describing the functions of specific genes) related to host-microbe interactions, the regulation of cytokine production, lymphocyte activation and the response to molecules of bacterial origin\(^{52}\). This is in keeping with the large body of data that is accumulating as to the role of the innate immune system and its interaction with intestinal microbes in the causal mechanisms in CD.

**Other investigations to identify causal molecules**

**Macrophage expression profiling**

In the knowledge that the release of pro-inflammatory cytokines by macrophages from CD subjects is depressed as a result of impaired vesicle trafficking\(^{38}\), an attempt was made to identify genes contributing to this deficiency by looking for outlier levels of gene expression in these cells. The most commonly under-expressed gene identified was Optineurin, low expression levels of which were found in 10% of CD patients studied\(^{52}\). ADAMDEC1 was under-expressed in about 7% of patients.

Optineurin (Optn)\(^{51,52}\) is a 67 kDa protein, ubiquitously expressed, and its expression can be induced by TNFα and interferons, possibly as a result of NfκB activation, and it is localised in the cytosol and Golgi apparatus. In essence it is a linker, or adaptor, molecule that has several binding partners including Rab8, Huntingtin, the gene for which is mutated in Huntington disease, and Myosin VI, a multifunctional motor protein. Rab8 is a small GTPase involved in vesicular trafficking between the trans-Golgi network (TGN) and the plasma membrane. The function of Huntingtin itself is unknown but it is associated with several factors involved in vesicle trafficking. Myosin VI is attached by OPTN to the Golgi apparatus and it participates in the transport of vesicles and their protein cargos from the from the Trans-Golgi network to be released at the cell surface. OPTN also contains a ubiquitin binding domain with the ability to bind polyubiquitinated cargoes and transport them to autophagosomes via its microtubule-associated protein 1 light chain 3-interacting domain\(^{52}\).
Macrophages from patients with low expression of OPTN secreted abnormally low levels of pro-inflammatory cytokines as do macrophages from OPTN knockout mice\(^\text{354}\). mRNA expression levels of these cytokines were normal, consistent with deranged secretion rather than synthesis. These mice were more susceptible to infection with Citrobacter, E. coli and Salmonella\(^\text{355}\), and showed reduced levels of TNFα in their serum, diminished neutrophil recruitment to sites of acute inflammation and greater mortality, than wild-type mice. OPTN-knockdown zebrafish infected with Salmonella also had a higher mortality\(^\text{356}\).

ADAMDEC1 (ADAM-like Decysin-1) is a member of the ADAM (A Disintegrin And Metalloproteinase) family, the expression of which is restricted to the macrophage/dendritic cell populations of the gastrointestinal tract. Its biological function is unknown but it has been hypothesised to play a role in immunity. Reduced ADAM-DEC1 expression in macrophages from a subgroup of CD patients has provided evidence of a potential role in bowel inflammation\(^\text{357}\). Adamdec1\(^\text{−/−}\) mice were more susceptible to the induction of bacterial and chemical induced colitis and they cleared Citrobacter rodentium less efficiently than wild-type mice after infection\(^\text{358}\,359\).

**DNA sequencing**

The development, availability and falling costs of high throughput DNA sequencing, has provided the means of directly identifying causal mutations in human disease\(^\text{360}\,361\). Several studies employing such technology have been undertaken in CD\(^\text{362}\,363\) and many are likely to appear over the ensuing years. High-throughput DNA sequencing has also been absolutely crucial for the diagnosis of the rare primary immunodeficiencies\(^\text{364}\) that produce bowel inflammation, rather than CD, as described earlier.

Initially the DNA of GWAS associated loci was sequenced, and a small number of rare variants were identified, most notably in the genes CARD9 and IL23R\(^\text{361}\,\text{362}\).

The major problem in identifying causal genes by sequencing is the considerable individual variation in DNA sequence. Asymptomatic individuals carry, on average, approximately 100 genuine loss of function variants with ~20 genes completely inactivated\(^\text{365}\). This makes it very difficult to identify the disease causing mutation/s in any one individual.

Of note, an ongoing study in which whole-genome sequencing has been undertaken in 2,697 CD cases and 3,652 healthy controls failed to identify a single variant at genome-wide significance that had not already been identified by GWAS\(^\text{364}\).

Alternative approaches have been taken to overcome the difficulty of the interpretation of individual variation. Several studies have focussed on the analysis of Ashkenazi Jews (AJ) because they have a roughly fourfold increased incidence of CD and demonstrate genetic homogeneity, having arisen from approximately 350 individuals about 30 generations ago\(^\text{366}\). Chuanget al. sequenced the exomes of 50 AJ CD patients and prioritised low frequency coding variants which were then genotyped in approximately 3,000 AJ CD cases and 3,000 controls. They identified a frameshift mutation in CSF2RB as a strong causal candidate which was associated with CD at \(p<3.5\times10^{-6}\) and an OR of 1.5\(^\text{367}\). This variant is rare in the non-AJ population.

Levine et al. utilised an alternative family based approach\(^\text{368}\). They characterised two very large AJ families with >800 and >200 members with 54 and 26 affected cases respectively, sequenced the exomes of all cases and imputed the genotypes of the unaffected family members. Low frequency coding variants that were predicted to be damaging and were enriched in affected compared with unaffected individuals were prioritised. In the large family they independently identified the identical frameshift mutation in CSF2RB as concurrently reported by Chuanget al. as a likely causal variant. Other strong candidate genes included NLRP2, a NOD-like receptor and a component of the inflamma-some; ZC3H18 which is involved in IKK and NFκB activation\(^\text{369}\), a pathway of established importance in CD; and MEGF10, a phagocytic receptor involved in apoptosis\(^\text{370}\).

**CSF2RB**

Having been identified independently by two groups CSF2RB must be considered as a causal gene for CD in AJs\(^\text{371}\). CSF2RB is the common or shared β subunit of the receptors for granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin (IL)-3, and IL-5\(^\text{372}\). The distinct α chains of these receptors provide cytokine specificity whilst the β chain is responsible for high-affinity binding and is the major downstream signalling component of the receptor complexes.

GM-CSF is produced by myeloid cells, dendritic cells (DCs), T cells, B cells, and several non-immunological cells including epithelial cells\(^\text{373}\) following exposure to inflammatory stimuli to promote the production and function of myeloid haemopoietic cells including haemopoietic progenitor cells and differentiated cells such as basophils, neutrophils, eosinophils, macrophages and certain dendritic cells\(^\text{374}\) to deal with the cause of the inflammation.

IL-3 is predominantly produced by activated T cells, natural killer (NK) cells and mast cells. It acts on the early stages of haemopoesis in synergy with other cytokines to induce progenitors of various lineages but it is a very important stimulus for the generation of mast cells and the regulation of mast cell function as well as basophil production and activation.

IL-5 stimulates mainly the production and function of eosinophils. The major source of IL-5 is T-cells with relatively lower amounts produced by mast cells and eosinophils\(^\text{375}\).

**DUOX2**

Levine et al. also identified a damaging missense mutation in DUOX2 that impaired the function of the protein and showed a possible epistatic interaction with NOD2\(^\text{376}\). DUOX2 is a member of the large NADPH oxidase (NOX) family of enzymes\(^\text{375}\). Its expression in the bowel epithelium is induced by the microbiota\(^\text{376}\). It generates \(\text{H}_2\text{O}_2\) at the mucosal surface and this acts as substrate for lactoperoxidase catalysed oxidation of thiocyanate to microbicidal hypothiocyanite\(^\text{376}\). It might also attract neutrophils to inflammatory sites\(^\text{377}\). Knockdown of the DUOX2 homologue in invertebrates and mice resulted in an impaired tolerance to enteric bacteria\(^\text{378}\). Of
relevance, given the possible NOD2 epistatic interaction observed, a physical and functional interaction between these proteins has been demonstrated in epithelial and HEK293 cells.\textsuperscript{579}

**Future treatment options**

Treatment of CD poses a conundrum. The logical approach to correcting the underlying problem would be to develop means of enhancing innate immunity, although at present no such range of drugs is currently available. It would be dangerous to attempt to do this in the presence of ongoing bowel inflammation but it could be useful to maintain patients in remission after they had been cleared of disease by surgical resection, or through the use non-immunosuppressant therapies such as elemental diets.\textsuperscript{380} This approach was attempted with levamisole as the immunostimulant, with varying results.\textsuperscript{382,383} One problem with this form of treatment is that it is important to ensure that the patients are in remission before it is commenced; otherwise it is likely to exacerbate the inflammation. In two studies levamisole induced a severe reversible polyarthropathy,\textsuperscript{384,385} indicating that the drug was in fact altering the immunological inflammatory axis, and providing clues as to IBD associated arthritis, and to the immunopathology of the idiopathic arthritides in general.

Current drug and biological treatments are largely immunosuppressant. This, to varying degrees of efficiency, damps down the secondary inflammation induced by the retained foreign material within the tissues. Anti-TNF treatments can be very helpful but do not provide a comprehensive answer. Only one third of patients will be in remission after one year on these treatments.\textsuperscript{386} Immunosuppressant treatment further compromises the underlying innate immune deficit to mucosal damage, thereby increasing the likelihood of further infection and the influx of bowel contents into the tissues, possibly converting CD from a sporadic to a chronic condition.

The primary pathology in most cases of CD appears to affect macrophages recruited from the blood as monocytes.\textsuperscript{387} Advances in gene editing with the CRISPR-Cas technology make the corrective treatment of CD a real possibility in the relatively near future. Once a primary causal mutation has been identified, and validated in animal models, bone marrow could be extracted, edited and reinfused into a conditioned patient in much the same way as is being applied to gene therapy for primary immunodeficiencies.\textsuperscript{389}

**Conclusion**

After almost a century of concerted effort in clinical investigations combined with recent technological developments in genomic medicine, a consensus view is developing as to the causes of CD.

A genetic predisposition to the condition can exist, as evidenced by twin and family studies. Studies in patients have clearly demonstrated that there is a defect in the acute inflammatory response resulting in an impaired recruitment of neutrophils to inflammatory sites and a consequent delay in the clearance of bacteria from them. This results from an initial blunted response by monocytes and macrophages leading to deficient secretion of pro-inflammatory cytokines.

CD is a syndrome and the molecular basis of this deficiency will vary considerably. It might reflect the multifactorial extreme end of a normal distribution, or the critical loss of one or more important molecules. Molecules contributing to the former will be very difficult to identify whereas some of those playing a more singular role have been highlighted by linkage, GWAS and DNA sequencing studies. These extend across the spectrum of the interface between microbes and cells: from the derangement of cellular homeostasis, probably in intestinal and immune cells, through impairment of the ER stress response (XBP1 malfunction); defective signalling (aberrant NOD2); depressed killing and digestion of organisms within phagocytic vacuoles (IRGM) or xenophagic recovery of those that escape into the cytoplasm (ATG16L1); failure of vesicle trafficking resulting in disorganised lysosomal biology of reduced cytokine secretion (LRRK2, Optineurin) or ineffective signalling of cytokines on target cells because of aberrant receptors (CSF2RB). In addition, primary abnormalities of neutrophils, the final effector cells, constitute the ultimate predisposition.

However severe the covert underlying predisposition, it requires a triggering factor to be expressed, and all the evidence points to this being an enteric infection. The identity of the organism is of secondary importance to the fact that the mucosa is breached, allowing ingress of the faecal contents into the bowel wall. Increased frequency or potency of these infections, as a result of factors implicated in the “Hygiene Hypothesis” or through increased spread by travel, or changes in sexual practices, could account for the increasing incidence of CD in the Western world.

It is the abnormal response to the penetrating faeces that is the common denominator of the CD syndrome, and it is the failure to eliminate the foreign material in the tissues that leads to the classical chronic granulomatous inflammation, and subsequently to an adaptive immune response. The counter intuitive, but logical, conclusion is that a disease characterised by grossly exuberant inflammation can result from an initial failure of innate inflammation. A similar mechanism might be responsible for other chronic inflammatory conditions of unknown aetiology, like for example sarcoidosis, ankylosing spondylitis, psoriasis, rheumatoid arthritis and SLE.

---

**Competing interests**

No competing interests were disclosed.

**Grant information**

I received financial support from the Medical Research Council, Wellcome Trust, Charles Wolfson Charitable Trust and the Irwin Joffe Memorial Trust.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Acknowledgements**

I thank Holm Ulhig and Amnon Sonnenberg for permission to publish figures from their papers. I am indebted to Adam Levine for assistance with the preparation of this manuscript.


References


118. Fujiya M, Ueno N, Kohyp Y: Probiotic treatments for induction and maintenance of remission in inflammatory bowel diseases: a meta-analysis of randomized...


Open Peer Review

Current Peer Review Status:  ✔️  ❎

Version 1

Reviewer Report 27 October 2016

https://doi.org/10.5256/f1000research.10455.r17018

© 2016 Lee J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

James C. Lee

Department of Medicine, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

This review presents a thorough overview of the pathogenesis of Crohn's disease, drawing on a broad range of evidence, from epidemiology to human immunology to genetics. It is impressive to have collated such diverse topics into a single review, and therefore understandable that some areas are discussed in more detail than others. Overall this is a valuable contribution to the literature.

Of note, some parts of this review read more like a perspective piece and draw conclusions that I suspect are not widely held, although they are supported by the evidence presented. There is also a considerable focus on the author's own work in several sections, although this is understandable and it is valuable to have these experiments discussed in the context of the broader literature regarding Crohn's disease pathogenesis.

The author is fairly dismissive of the role of genetics, and more specifically of the significance of disease-associated genetic variants, based on their small odds ratios and frequency within the healthy population. While I understand this opinion, I think it is important to clarify a couple of these points for a general audience. First, it is important to note that one of the founding premises of GWAS was that it was designed to identify genes that are involved in pathways relevant to disease pathogenesis. It was not expected - based on the common disease-common variant hypothesis - that any of the SNPs or genes would by themselves be causal. To take the example of impaired autophagy - something that has been commonly reported in Crohn's patients - there are several ways in which this might occur. For instance some patients may carry NOD2 or IRGM or ATG16L1 mutations, while in others it may be due to environmental factors such as smoking - see Monick et al. (2010). Based on the multi-hit model the author proposes, the key point is that "hits" are simultaneously required in several pathways in conjunction with an environmental trigger in order for a disease to occur. Because the exact nature of each hit is likely to vary between patients, it is not surprising that univariate genetic analysis identifies small odds ratios at individual SNPs (many of which may have similar effects on a given pathway). Moreover, the size of these effects may be different when gene-environment and epistatic interactions are taken into account. The value of this data, therefore, is that it points towards the specific pathways that are likely to be involved in disease pathogenesis, rather than identifying causal genes.
This is perhaps best illustrated in cardiovascular medicine, where genetic variants in the gene that
encodes HMG-CoA reductase have been significantly associated with LDL cholesterol levels, but the
odds ratios for these associations are very small (presumably because there are many other causes of
elevated cholesterol, Kathiresan et al. (2008)). Despite this, pharmacological targeting of this enzyme with
statins represents the most effective treatment for hypercholesterolaemia. As such, a small odds ratio at
individual SNP should not be confused with the implicated gene or pathway not being biologically or
therapeutically important.

Other points:
- Discussion of the ATG16L1 variant would benefit from including the recent evidence linking the
  variant to susceptibility to caspase-mediated degradation (Murthy et al. (2014)).
- It would be worth highlighting the cross-talk between ER stress and autophagy, particularly in the
  context of Crohn's (Adolph et al. (2013)).
- p.2: IL-23 is thought to be more involved in the expansion and survival of Th17 than in transforming
  naive T cells into Th17 cells
- p.7 typo: Danziel should be Dalzeil

References
1. Adolph TE, Tomczak MF, Niederreiter L, Ko HJ, Böck J, Martinez-Naves E, Glickman JN,
   Tschurtscenthaler M, Hartwig J, Hosomi S, Flak MB, Cusick JL, Kohno K, Iwawaki T, Billmann-Born S,
   (7475): 272-6 PubMed Abstract | Publisher Full Text
2. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, Hirschhorn JN, Berglund G,
   5425-35 PubMed Abstract | Publisher Full Text
   Graham RR, van Lookeren Campagne M: A Crohn's disease variant in Atg16l1 enhances its degradation

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that
it is of an acceptable scientific standard, however I have significant reservations, as outlined
above.
I am most grateful to Dr Lee for taking the time and trouble to thoroughly review my manuscript and for his comments.

I would like to address the issue of the relationship between GWAS and the identification of causal genetic variants. It is simple to establish causality for dominant, highly penetrant genes like polycystic renal disease, polyposis coli and Huntington's disease. However it is more difficult when penetrance is incomplete, when one is reliant upon the odds ratio of the frequency of the variants in affected as compared with unaffected individuals. In general very high odds ratios would strongly suggest causality. Genes with very high odds rations have been identified by GWAS. For example:

Birdshot chorioretinopathy

However, the most important consideration is not whether or not the design of GWAS was to identify causal genes, but what the common perception is, and in many cases it is the perception that GWAS do in fact generally identify causal genes:

The genetic background of Inflammatory Bowel Disease: From correlation to causality.
Uniken Venema WT et al.

Insights into kidney diseases from genome-wide association studies.
Wuttke M and Köttgen A

Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture.
Goddard ME et al.

A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants.
Fritsche LG et al.

Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep.
Suárez-Vega A et al.

It was therefore very important to state in the review that GWAS have not identified genes that cause Crohn's disease.

Under the Summary of outcome of GWAS studies in CD I have now stated:

The GWAS studies have provided a series of clear answers. No single gene, or a small number of genes, has been identified that is causal for CD. More than 170 GWAS hits combined contribute to about 10% of the "heritability" of CD. With the average individual contribution of only 0.1% it is unlikely that these variants will individually have major effects on cellular function, either in the CD patients or in experimental systems.
One of the founding premises of GWAS was that it was designed to identify genes that are involved in pathways relevant to disease pathogenesis. The value of this data, therefore, is that it points towards the specific pathways that are likely to be involved in disease pathogenesis, rather than identifying causal genes.

**Competing Interests:** None
For pre-submission enquiries, contact research@f1000.com