Executive summary

Inflammatory bowel disease (IBD) refers to the separate but related conditions Crohn's disease (CD) and Ulcerative colitis (UC). IBD can appear at any stage in life, but is becoming increasingly common in children and young adults, and requires invasive and costly diagnostic tests. IBD causes unpleasant gastrointestinal symptoms such as abdominal pain and diarrhoea, and chronic inflammation can lead to cancer or damage to the structures of the gut so that for example the bowel may form an opening into the bladder or onto the abdominal wall. IBD places a significant burden on patients and their families due to chronic symptoms, recurrent hospitalisations and surgeries, and this is reflected by significant costs to healthcare providers.

The causes of IBD are not completely understood, but it is thought that a combination of environmental factors and genetic susceptibility, possibly combined with alterations to the bacteria present in the gut, leads to an abnormally prolonged and severe reaction to those bacteria. Despite recent advances in medications for IBD, there is a significant unmet need for further treatment options, for avenues toward personalised care, and for deeper understanding of the disease processes.

IBD-Character is an international collaboration combining leading academic researchers and clinical experts with SME partners in related fields. The consortium collected multiple types of samples from patients (blood, serum, stool, and intestinal biopsies) undergoing investigation for possible IBD. These samples were then analysed by a range of techniques (genomic, epigenomic, proteomic, transcriptomic, and microbial) which is commonly referred to as 'multi-omic'. The strength of this study design is that these data streams can be analysed not only individually, but in numerous combinations, allowing a deeper understanding of the biological processes involved in IBD. The analyses performed have looked at differences between IBD and controls to understand the emergence and persistence of IBD, and have looked at differences between IBD and controls, or between various subgroups of IBD to find biomarkers to identify these groups.

The result with the greatest immediate translational potential is the discovery of biomarkers in blood which identify a subgroup of new IBD patients who are initially stabilised on standard treatment, but who rapidly deteriorate and require hospitalisation, surgery, or escalation of medical treatment. These biomarkers can identify members of this subgroup with an accuracy of up to 95%, raising the possibility of making informed therapeutic decisions on initial treatment in a truly personalised fashion. Members of the consortium are actively pursuing the necessary work to develop these tests for clinical application.

Other significant results from the project include

- replication of our previous findings of DNA methylation changes in blood associated with IBD, and overlaps with some of the DNA methylation changes present in the gut
- correlation of genetic variations with disease-associated changes in RNA and protein expression, and DNA methylation
- the application of a test to categorise disruption to the normal gut bacteria
- the largest study of IBD in twins, which is of great value due to their shared genetic and environmental risk factors

Project context and main objectives

Inflammatory bowel disease (IBD) is an important cause of suffering and distress to many young people in Europe. IBD affects approximately 2.5-3 million people across Europe[1] with direct healthcare costs estimated at approximately €4.6-5.6 billion per year[1]. A recent systematic review showed rising trends in the incidence and prevalence of Crohn's disease (CD) worldwide[2]. Despite advances in therapies, more than 50% of CD patients require surgery and IBD patients also have an increased risk of mortality compared with the general population[3].

Improved patient outcomes are undoubtedly achieved by early diagnosis and treatment. IBD is often difficult to diagnose, and current tests can be invasive and cause radiation exposure. The development of novel, sensitive and specific biomarkers to both diagnose and stratify patients according to risk of severity of disease and treatment response will be a major step forward in the development of a more personalised approach to clinical management. Additionally, in the development of biomarkers, critical insights into the causes of the disease may be gained.

IBD-Character is a multidisciplinary consortium of leading academic and industrial SME researchers in inflammatory bowel disease studying genomics, epigenomics, proteomics and metagenomics.

The main objectives of the study are increased understanding of the processes involved in the development and persistence of IBD; identification of targets for therapeutic intervention; understanding the relationships between the core modalities being assayed; and development of robust biomarkers for clinical application in IBD for early diagnosis or risk stratification by disease course or response to therapy.

This has been achieved by recruitment of 678 patients at diagnosis or exclusion of IBD, in whom blood, stool, and intestinal biopsies have been biobanked, and have been analysed by multiple modalities. A major strength of the IBD-Character study design is the breadth of integrated analysis possible between the data sets, in combination with detailed biographic information, disease characterisation, and clinical follow-up data for participants.

The control subjects for this study are made up of 50 healthy volunteers, and the patients who presented with symptoms requiring investigation, but which were not caused by IBD. Most of these patients had a non-inflammatory cause of their symptoms (e.g. Irritable Bowel Syndrome – IBS), with a small number having either infections or diverticular disease. Recruiting a range of symptomatic non-IBD patients allows for findings to be considered in light of other causes of inflammation and non-inflammatory alterations to bowel habit and diet, as well as comparison with healthy controls.

---

1 Burisch J, Jess T, Martinato M, Lakatos PL. The burden of inflammatory bowel disease in Europe. J Crohns Colitis 2013;7:322–37

2 Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 2012;142:46–54.e42

3 Jess T, Riis L, Vind I, Winther KV, Borg S, Binder V, et al. Changes in clinical characteristics, course, and prognosis of inflammatory bowel disease during the last 5 decades: a population-based study from Copenhagen, Denmark. Inflamm Bowel Dis 2007;13:481–9

Main results

Genotyping

No genotyping results reached genome-wide significance after correction for multiple testing due to the relatively small sample size for a stand-alone GWAS; however, a number of the strongest results do overlap with known IBD risk loci from GWAS meta-analyses, and genetic regions of interest to IBD such as the HLA region. The strongest use for the genetic data generated in IBD-Character has been in conjunction with the other data streams – examining genetic influences on on RNA expression (eQTLs), protein expression (pQTLs), and DNA methylation (meQTLs).

Paired genetic and methylation data showed 2327 FDR significant cis meQTLs indicating a genetic influence on several key loci – RPS6KA2 ($p=8.6\times10^{-34}$), ITGB2 ($p=3.3\times10^{-38}$), and an important replication of the finding published in collaboration with IBD-BIOM (Ventham et al., 2016) of two SNPs previously described as correlated to VMP1/MIR21 methylation (rs8078424, $p=4.4\times10^{-25}$, rs10853015, $p=7.4\times10^{-21}$) which are in LD with an IBD GWAS SNP rs1292053 (Jostins et al., 2012).

We identified 79 cis pQTLs (14 unique proteins and 69 independent SNPs; minor allele frequency>0.05) that survived Holm correction ($p<1.6\times10^{-4}$). VEGF-A showed the most significant association with genotype (rs7767396; effect (β) -0.47; Holm $p=2.4\times10^{-22}$) with a total of 7 significant SNP associations and 3 SNPs in linkage disequilibrium (rs7767396 and rs9472159: dist=7355, $r^2=0.82$, $D'=0.96$; rs9472159 and rs9472158: dist=798, $r^2=0.64$, $D'=1.0$). Other markers included MMP-1, SLAM7 and CCL-23.

Significant correlations between DNA methylation and RNA expression include VMP1 (Holm $p=8.5x10^{-16}$), one of the strongest disease-associated methylation changes in these results, as well as results previously published by consortium members. The methylation changes in VMP1 are around the penultimate exon of the gene and appear to be correlated with the microRNA mir21, and we have previously demonstrated altered mir21 expression in IBD. The finding of a significant correlation with expression of the host gene is intriguing as VMP1 plays a critical role in autophagy - a process well established to be involved in the pathogenesis of IBD (particularly CD).

Methylation

In blood there were 195 probes significantly associated with IBD after Bonferroni correction, including VMP1/MIR21 ($p=3.7\times10^{-20}$), RPS6KA2 ($p=1.1\times10^{-19}$), SBNO2 ($p=2.7\times10^{-19}$), and TNFSF10 ($p=1.1\times10^{-15}$); replicating methylation differences we have previously reported in paediatric CD and adult IBD. Novel findings include PHOSPHO1 (1.3×10^{-15}), MUC4 (5.5×10^{-15}), and CDH24 (1.7×10^{-14}). CD and UC were highly similar with a significant difference at only one probe (NAV2, $p= 6.82\times10^{-8}$). 1709 differentially methylated regions of consecutive FDR significant probes were defined including VMP1/MIR21, ITGB2, TNF and throughout the HLA region.

Linear modelling with rigorous cross-validation has produced accurate biomarkers for diagnosis, and most excitingly for response to treatment. Clinical follow-up data for Edinburgh patients (n=114) was used to define a subgroup in whom treatment escalation with anti-TNF, biologics, or surgery occurred within 1 year. 14 probes were significantly associated with treatment escalation after Bonferroni correctionn and combinations of up to 11 of these probes were capable of

predicting requirement for treatment escalation in a cross-validation subset with an accuracy of 95%.

In the twin cohort, 46 probes exhibited IBD-associated methylation differences surviving Holm correction. These include replications of results from the main IBD-Character 450k blood analysis (RPS6KA2, SLC10A6), loci of established relevance to IBD (JAK3) and novel findings.

In the biopsy data principal component analysis revealed no significant clustering due to recruiting centre, however there was a strong effect of local inflammation, and a strong effect of biopsy location. To simplify analysis and preserve statistical power the main analyses grouped sample locations into ileal and colorectal, CD into L2 and L2/3, and UC into E1/2 and E3. Despite this there are a large number of comparisons which can be made; for example methylation differences associated with location within non-inflamed CD biopsies, or methylation differences between non-inflamed biopsies from CD and non-inflamed biopsies from UC. There are some intriguing significant findings including known genes of interest (e.g. NOD2), overlap with findings from blood (e.g. VMP1/MIR21 and SBNO2), and novel findings. Comparing inflamed ileum and inflamed colorectum from CD patients found 380 DMRs (c.f. 2646 individual probes) including MUC4 (7 probes, min $p=1.9\times10^{-23}$), NOD2 (3 probes, min $p=2.3\times10^{-16}$), and MIR21/VMP1 (12 probes, min $p=3.8\times10^{-15}$).

RNA

In addition to the eQTL analysis above, analysis of the blood RNA expression data showed 2001 significantly (Bonferroni corrected) altered transcripts, correcting for inflammation with hsCRP and albumin. The most significantly dysregulated RNA in blood was CD177 (2.6×10^{-47}). Also of note were the 3rd and 6th most significant findings – calgranulin B ($p=3.8\times10^{-38}$) and C ($p=3.4\times10^{-36}$), and calgranulin A ($p=1.3\times10^{-20}$), which forms the heterodimer calprotectin with calgranulin B was also highly significant.

Protein

The most significant protein findings in biopsies included CXCL9, which could discriminate non-inflamed colonic CD samples from non-IBD samples (FC=2.1) and MMP.10 and MMP.7 which were differentially expressed between inflamed CD and inflamed UC samples (FC=0.31 and 0.32, respectively). In the ileal samples IL-8 could discriminate between non-inflamed CD and non-IBD samples (FC=5.8) and HGF could discriminate between inflamed and non-inflamed CD samples (FC=2.1).

Analysis of serum data revealed that biomarker panels of 16 proteins are capable of distinguishing IBD from symptomatic controls with an accuracy of up to 94.4%. This is significantly more powerful than the 92 protein biomarkers which were envisioned based on an interim analysis. There is substantial redundancy available in choice of protein markers for a final diagnostic assay, which will allow numerous practical concerns to be considered in their selection. An additional analysis of the Edinburgh cohort revealed 15 proteins with significant associations with treatment escalation. The aggregate accuracy in a cross-validation cohort of random models consisting of 10 proteins was 84% (sensitivity 66%, specificity 89%).

Microbiome
In most of the datasets, there is little difference between symptomatic controls (predominantly IBS) and healthy controls (medical student volunteers). However there did appear to be differences in the microbiota profiles between healthy individuals and both IBD and non-IBD patients. Proteobacteria were increased in IBD and non-IBD as compared to the healthy controls ($p<0.02$), while the abundance of Bifidobacterium and Faecalibacterium prausnitzii was decreased ($p<0.02$ and <0.07, respectively).

CD and UC exhibited equal levels of dysbiosis, however the bacteria profiles differed – there was a reduced abundance of Firmicutes, Streptococcus and Clostridia in UC patients compared to CD ($p<0.05$ for all). In UC microbiota profiles were highly dependent upon the disease extent. E1 patients clustered with healthy controls in PLS-DA, and E2/E3 clustered together. Bifidobacterium and Eubacterium were significantly reduced ($p<0.01$), and Escherichia/Proteobacteria were significantly increased ($p<0.01$) in the E2/E3 group as compared to E1.

Multimodal analysis
We have developed a novel pathway modelling approach where we are able to integrate gene expression, methylation, and genotypes that utilises external knowledge databases. The algorithm has been presented at the 19th meeting of the Norwegian Statistical Society (Lindstrøm, 2017). We anticipate further refinements of this novel method which will be of interest to other researchers dealing with complex multi-faceted datasets and anticipate being able to publish results from the IBD-Character dataset giving insight into the pathways involved in IBD susceptibility and treatment response.

Potential impact

There are no diagnostic assays or single biomarkers in clinical use to identify patients with possible IBD, beyond non-specific markers of inflammation, such as C-reactive protein (CRP) and faecal Calprotectin, and there are no available biomarkers which can stratify patients by disease course or treatment response.

The biomarkers we have discovered in serum protein profiles and blood DNA methylation profiles have the potential to make a large impact in the management of IBD. Modifying treatment based on such a biomarker (e.g. pre-emptive treatment with medications considered 'second-line') has the potential to not only reduce the immediate symptomatic burden on patients, but also perhaps delay issues caused by chronic inflammation and repeated abdominal surgery. The effects of this would include improved quality of life for patients, and accompanying societal benefits, but also potentially significant financial savings for health services and governments. The consortium are actively pursuing further development of these biomarkers, and patent applications are currently being planned.

We have also demonstrated biomarkers which can accurately distinguish IBD from other causes of gastrointestinal symptoms. Although we have demonstrated some significant differences between CD and UC, most individual biomarkers perform poorly at distinguishing CD from UC relative to endoscopic investigation, which would limit clinical utility. The novel analytic methods being developed to combine all the individual project components, may offer improved performance.

The IBD-Character cohort is an immensely valuable resource, and as the manuscripts currently in preparation are published and restrictions imposed by the patent application process pass, the data will be made available to other researchers, supporting their ongoing or future research – broadening our understanding of the processes involved in IBD, and discovering targets for future drugs.

The cohort is not merely a snapshot of the participants at diagnosis. We have already been collecting clinical follow-up data, which has resulted in the escalation biomarker discovery. As part of work to replicate this finding due to begin in September 2017, further blood samples and ongoing clinical follow-up data are being sought from some patients recruited into IBD-Character. This will allow us to assess the stability of these findings over time, and will further extend the impact of the IBD-Character project.