Inherited determinants of Crohn’s disease and ulcerative colitis phenotypes: a genetic association study

Isabelle Cleynen*, Gabrielle Boucher*, Luke Jostins*, L Philip Schumm, Sebastian Zeissig, Tariq Ahmad, Vibeke Andersen, Jane M Andrews, Vita Annese, Stephan Brand, Steven R Brant, Judy H Cho, Mark J Daly, Marla Dubinsky, Richard H Duerr, Lynnette R Ferguson, Andre Franke, Richard B Geary, Philippe Guyette, Hakon Hakonarson, Jonas Halfvarson, Johannes R Hov, Hailang Huang, Nicholas A Kennedy, Limas Kucpinski̧as, Ian C Lawrence, James C Lee, Jack Satsangi, Stephan Schreiber, Emilie Théâtre, Andrea E van der Meulen-de Jong, Rinse K Weersma, David C Wilson, International Inflammatory Bowel Disease Genetics Consortium†, Miles Parkes, Severine Vermeire, John D Rioux, John Mansfield, Mark S Silverberg, Graham Raford-Smith, Dermot P B McGovern*, Jeffrey C Barrett*, Charlie W Lees*

Summary

Background Crohn’s disease and ulcerative colitis are the two major forms of inflammatory bowel disease; treatment strategies have historically been determined by this binary categorisation. Genetic studies have identified 163 susceptibility loci for inflammatory bowel disease, mostly shared between Crohn’s disease and ulcerative colitis. We undertook the largest genotype association study, to date, in widely used clinical subphenotypes of inflammatory bowel disease with the goal of further understanding the biological relations between diseases.

Methods This study included patients from 49 centres in 16 countries in Europe, North America, and Australasia. We applied the Montreal classification system of inflammatory bowel disease subphenotypes to 34 819 patients (19713 with Crohn’s disease, 14683 with ulcerative colitis) genotyped on the Immunochip array. We tested for genotype–phenotype associations across 156 154 genetic variants. We generated genetic risk scores by combining information from all known inflammatory bowel disease associations to summarise the total load of genetic risk for a particular phenotype. We used these risk scores to test the hypothesis that colonic Crohn’s disease, ileal Crohn’s disease, and ulcerative colitis are all genetically distinct from each other, and to attempt to identify patients with a mismatch between clinical diagnosis and genetic risk profile.

Findings After quality control, the primary analysis included 29 838 patients (16 902 with Crohn’s disease, 12 597 with ulcerative colitis). Three loci (NOD2, MHC, and MST1 3p21) were associated with subphenotypes of inflammatory bowel disease, mainly disease location (essentially fixed over time; median follow-up of 10·5 years). Little or no genetic association with disease behaviour (which changed dramatically over time) remained after conditioning on disease location and age at onset. The genetic risk score representing all known risk alleles for inflammatory bowel disease showed strong association with disease subphenotype (p=1·65 × 10–79), even after exclusion of NOD2, MHC, and 3p21 (p=9·23 × 10–18). Predictive models based on the genetic risk score strongly distinguished colonic from ileal Crohn’s disease. Our genetic risk score could also identify a small number of patients with discrepant genetic risk profiles who were significantly more likely to have a revised diagnosis after follow-up (p=6·8 × 10–4).

Interpretation Our data support a continuum of disorders within inflammatory bowel disease, much better explained by three groups (ileal Crohn’s disease, colonic Crohn’s disease, and ulcerative colitis) than by Crohn’s disease and ulcerative colitis as currently defined. Disease location is an intrinsic aspect of a patient’s disease, in part genetically determined, and the major driver to changes in disease behaviour over time.

Funding International Inflammatory Bowel Disease Genetics Consortium members funding sources (see Acknowledgements for full list).

Copyright © Cleynen et al. Open Access article distributed under the terms of CC BY.

Introduction

Crohn’s disease and ulcerative colitis, the two major forms of inflammatory bowel disease, affect about one in 200 people in developed countries, with a rising incidence and prevalence in developing countries.1 Many patients with inflammatory bowel disease have a lifetime of debilitating physical symptoms (eg, urgent diarrhoea, rectal bleeding, vomiting, anorexia, and lethargy), which frequently lead to poor psychosocial wellbeing with wide ranging consequences for academic attainment, employment, relationships, and sexual health.2 Furthermore, the financial costs of inflammatory bowel disease are substantial and are estimated at more than US$2·2 billion per year in the USA alone.3 Inflammatory bowel disease is characterised by an exaggerated mucosal immune response to luminal gut contents in genetically susceptible individuals.4 In Crohn’s disease, inflammation can occur in any part of the gastrointestinal tract, whereas ulcerative colitis is typically confined to the colon. The universally adopted Montreal
Evidence before this study
We searched PubMed for genotype-phenotype association studies in inflammatory bowel disease, published between Jan 1, 1996, and Oct 17, 2014, with the search terms “inflammatory bowel disease” AND “genotype phenotype” AND “genetic association” AND (“disease course” or “disease extent” or “location behaviour”). We found 31 studies with sample sizes of 66 to 2804 patients, and studying between one and 163 genetic variants. Most of these studies implicated NOD2 and HLA in subphenotypes of inflammatory bowel disease. Many studies, especially the early ones, studied only NOD2 and HLA, and none included genetic variants not previously implicated in risk for inflammatory bowel disease.

Added value of this study
This study is the largest genotype-phenotype study of inflammatory bowel disease by at least a factor of ten, and is among the largest studies of genetic determinants of clinical subphenotypes of any complex disease. We have refined the classification distinguishes clinical subphenotypes in Crohn’s disease by disease location and behaviour, and age of onset, and in ulcerative colitis by disease extent and age of onset.10-12 Molecular studies have suggested that ileal and colonic Crohn’s disease are distinct entities because variants in NOD2 are associated with small bowel disease and HLA alleles with colonic disease.10-12 However, current recommendations do not advocate the use of these established markers in making treatment decisions, nor for choosing patients for clinical trials.20-21 The natural history and clinical course of inflammatory bowel disease is very heterogeneous: up to 20% of patients with ulcerative colitis need colectomy for medically refractory disease, and more than 50% of patients with Crohn’s disease need surgery within 10 years of diagnosis;22 however, up to 50% of patients with ulcerative colitis and 30% with Crohn’s disease will have a fairly indolent disease course without the need for immunosuppression or surgery.23-24

Inflammatory bowel disease has been at the vanguard of progress in understanding the genetic framework of complex diseases, with 163 susceptibility loci identified so far.19 Most inflammatory bowel disease loci confer risk of both ulcerative colitis and Crohn’s disease, but typically show distinct effect sizes in the two disorders. These findings suggest that genetic variation might define molecular subtypes independent of traditional and clinically defined diagnostic entities, allowing new insights into the molecular basis of these subphenotypes.

Our international study of around 30 000 patients with inflammatory bowel disease genotyped by microarray is the largest genotype–subphenotype study in the disease done so far. We have used genetic risk scores to study genetic heterogeneity underpinning the natural history of inflammatory bowel disease. This analysis rejects the current binary classification of Crohn’s disease and ulcerative colitis as distinct and homogeneous clinical entities in favour of a continuum of illness better fit by a three-category model (ie, ileal Crohn’s disease, colonic Crohn’s disease, and ulcerative colitis). We show that these risk scores have clinical potential (although they are currently only weak predictors), and believe they might have widespread applicability in other diseases.

Methods
Study design and patients
We acquired phenotype data for 34 819 patients, including 19 713 with Crohn’s disease and 14 683 with ulcerative colitis. The cohort included all patients in different centres over the years who had inflammatory bowel disease as per Lennard-Jones’ criteria.34 All inclusion criteria are included in appendix A. After quality control (appendix A), the primary analysis included 29 838 patients (16 902 with Crohn’s disease, 12 937 with ulcerative colitis, 255 with indeterminate colitis, and 84 missing an exact diagnosis). This study includes patients from population-based registries, and secondary and tertiary-referral centres at 49 sites in 16 countries in Europe, North America, and Australasia, most of which have been previously described (appendix B).34 Confirmation of diagnosis of inflammatory bowel disease and assignment of clinical subphenotypes were done by clinicians specialising in inflammatory bowel disease or trained phenotypers through case note reviews of clinical, radiological, histopathological, and endoscopic reports, and classified per the Montreal classification criteria (see appendix A for details).35 For behaviour and surgery in Crohn’s disease, and colectomy in ulcerative colitis, Kaplan-Meier Articles

157

www.thelancet.com Vol 387 January 9, 2016

Inflammatory Bowel Disease Service, Department of Gastroenterology and Hepatology, Royal Adelaide Hospital, Adelaide, Australia (Prof J Andrews MD); School of Medicine, University of Adelaide, Adelaide, Australia (Prof J Andrews); Unit of Gastroenterology, Istituto di Ricovero e Cura a Carattere Scientifico-Casa Sollievo della Sofferenza (IRCCS-CS) Hospital, San Giovanni Rotondo, Italy (V Annese MD); Azienda Ospedaliero Universitaria (AOU) Careggi, Unit of Gastroenterology SOD2, Florence, Italy (V Annese); Department of Medicine II, University Hospital Munich-Grosshadern, Ludwig Maximilians-University, Munich, Germany (Prof S Brand MD); Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA (S R Brant MD); Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA (S R Brant); Department of Genetics, Yale School of Medicine, New Haven, CT, USA (Prof H Cho MD); Broad Institute of MIT and Harvard, Cambridge, MA, USA (Prof M J Daly PhD, H Huang PhD); Department of Pediatrics, Cedars Sinai Medical Center, Los Angeles, CA, USA (M Dubinsky MD); Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA (Prof H H Duer MD); Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA (Prof H H Duer); School of Medical Sciences, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand (Prof L Ferguson PhD); Department of Medicine, University of Otago, Christchurch, New Zealand (Prof R G Gasney PhD); Department of Gastroenterology, Christchurch Hospital, Christchurch, New Zealand (Prof R G Gasney); Center for Applied Genomics, Istituto di Ricovero e Cura a Carattere Scientifico-Casa Sollievo della Sofferenza (IRCCS-CS) Hospital, San Giovanni Rotondo, Italy (V Annese MD); Azienda Ospedaliero Universitaria (AOU) Careggi, Unit of Gastroenterology SOD2, Florence, Italy (V Annese); Department of Medicine II, University Hospital Munich-Grosshadern, Ludwig Maximilians-University, Munich, Germany (Prof S Brand MD); Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA (S R Brant MD); Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA (S R Brant); Department of Genetics, Yale School of Medicine, New Haven, CT, USA (Prof H Cho MD); Broad Institute of MIT and Harvard, Cambridge, MA, USA (Prof M J Daly PhD, H Huang PhD); Department of Pediatrics, Cedars Sinai Medical Center, Los Angeles, CA, USA (M Dubinsky MD); Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA (Prof H H Duer MD); Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA (Prof H H Duer); School of Medical Sciences, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand (Prof L Ferguson PhD); Department of Medicine, University of Otago, Christchurch, New Zealand (Prof R G Gasney PhD); Department of Gastroenterology, Christchurch Hospital, Christchurch, New Zealand (Prof R G Gasney); Center for Applied Genomics,
Articles

To learn about the relative phenotypic variance explained by different risk factors in adult inflammatory bowel disease, we fitted a model to predict Crohn’s disease location that included both demographic predictors (smoking status, age at diagnosis, and year) and genetic predictors (SNPs at NOD2, MST1, and the HLA cluster as well as the genetic risk score). Variance explained on the logit scale by each predictor was calculated with the McKelvey–Zavoina pseudo $R^2$. Centres with a high proportion (>60%) of missing data for smoking status were removed. To reduce the effect of changes in clinical practice and smoking rates, only patients born between 1955 and 1985 were included.

In addition to looking at single SNPs, we also combined information from 193 SNPs and 23 HLA types previously associated with inflammatory bowel disease to generate genetic risk scores (appendix A), which provide better predictive accuracy than individual SNPs. To assess classification accuracy, we re-ran the risk score analyses with a cross-validation strategy, in which models were fitted in non-UK origin samples and assessed by how well they classified UK samples.

To assess if the risk score can be used to identify misclassified patients, we selected 97 outlier patients that fell in the extreme tail of the scores for the opposite phenotype (log Crohn’s disease versus ulcerative colitis [CD vs UC] score < –2 for Crohn’s disease outliers and log CD vs UC score ≥ 2 for ulcerative colitis outliers), as well as 95 randomly selected cases with non-outlier scores matched by recruitment centre. Clinicians from each centre were then asked to re-phenotype both outlier and non-outlier patients in a masked fashion. The CD versus UC risk score was chosen for this experiment because it had the strongest association with Crohn’s disease location and behaviour.

Statistical analysis

The median effect size of known inflammatory bowel disease risk variants was about OR 1·1, with a median minor allele frequency of roughly 30%. The sample size of our study gave us high power to detect an effect of equivalent magnitude of Crohn’s disease location (power of 67% for ileal vs non-ileal disease), Crohn’s disease behaviour (94% for complicated vs non-complicated disease) and ulcerative colitis disease extent (84% for extensive vs non-extensive disease) at genome-wide significance. Binary and linear genotype–phenotype analyses were done with PLINK version 1.07, and multinomial and ordinal regression with a custom program, Trinculo version 0.4 (appendix A). Survival analysis and risk prediction were done with R-2.15.1 using the packages “survival” and “Mangrove”, respectively. Data handling and plotting was done with R.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the
data in the study and had final responsibility for the decision to submit for publication.

Results

Our primary analyses were done on matched genotype and phenotype data from 29 838 patients of European ancestry (appendix A) with inflammatory bowel disease (16 902 with Crohn’s disease, 12 597 with ulcerative colitis; table 1) with a total of 217 195 patient-years of follow-up (median per patient: 11 years for Crohn’s disease and 10 years for ulcerative colitis). Demographic features of the study population agreed with previously published results: patients with Crohn’s disease were more likely to be younger at diagnosis, female, smokers, and have affected family members than were patients with ulcerative colitis (table 1). Extensive disease was more common in those diagnosed at a younger age in both Crohn’s disease and ulcerative colitis, whereas disease behaviour was relatively unaffected by age at diagnosis (appendix A). Reaffirming the progressive nature of Crohn’s disease, the proportion of patients with strictureing (B2) or penetrating (B3) disease increased from less than 30% (n/N) at diagnosis to 43% (n/N) at 5 years, 56% (n/N) at 10 years, and 74% (n/N) at 30 years (figure 1, which shows the progression in B1, B2, and B3 disease individually [smoothed estimates over intervals]). By contrast, disease location showed little variation during the same period (figure 1). With the exception of the population-based cohorts from Scandinavia, survival analyses of time to development of complicated disease (B2, B3) or first surgery in Crohn’s disease were highly consistent across the different countries of origin despite different health-care systems and methods of sampling (appendix A). In Crohn’s disease, time from diagnosis to progression (complicated disease or surgical intervention) was significantly shorter in purely ileal (L1) compared with ileocolonic (L3) or colonic (L2) disease (p<10⁻¹⁰⁰; figure 1; appendix A). Overall, 7257 (52%) of 13 862 patients with Crohn’s disease had undergone surgery by the time of last follow-up. In ulcerative colitis, in which the overall rate of colectomy was 22% 10 years after diagnosis, time to surgery was shorter in patients with extensive disease (E3) than in those with left-sided disease (E2) or proctitis (E1; p=8×10⁻⁸⁴; figure 1).

We tested genetic variants for association with age at diagnosis and time to surgery in all patients with inflammatory bowel disease; disease location and behaviour in Crohn’s disease; and disease extent in ulcerative colitis (table 2 and table 3). Across all analyses, three loci achieved genome-wide significance (p<5×10⁻⁸): 3p21 (MST1), NOD2, and the MHC. No additional signals were noted after replication of suggestive loci (p<1×10⁻⁸) in an independent GWAS cohort (appendix B). Although NOD2 was strongly associated with Crohn’s disease location, behaviour, and age at diagnosis, adjustment for the other phenotypes showed that the

| Phenotypes | Crohn’s disease (n=16 902) | Ulcerative colitis (n=12 597) | Inflammatory bowel disease* (n=29 838) |
|------------|----------------------------|---------------------------------|----------------------------------------|
| **Demographics** | | | |
| Sex | | | |
| Male | 7227 (44%) | 6339 (51%) | 13 738 (47%) |
| Female | 9257 (56%) | 6027 (49%) | 15 448 (53%) |
| Missing | 418 (3%) | 231 (2%) | 652 (2%) |
| Age at diagnosis (years) | | | |
| Median (quartiles) | 25 (19–36) | 31 (22–24) | 28 (20–40) |
| <17 (A1) | 25 (18%) | 123 (11%) | 3903 (15%) |
| 17–40 (A2) | 9166 (64%) | 6594 (58%) | 15 854 (61%) |
| >40 (A3) | 2626 (18%) | 3469 (31%) | 6141 (24%) |
| Missing | 2542 (15%) | 1301 (10%) | 3940 (13%) |
| Family history | | | |
| Yes | 3471 (27%) | 2232 (21%) | 5778 (24%) |
| No | 9575 (73%) | 8260 (79%) | 18 005 (76%) |
| Missing | 3856 (23%) | 2105 (17%) | 6055 (20%) |
| Smoking status | | | |
| Smoker | 3339 (28%) | 1162 (12%) | 4512 (21%) |
| Ex-smoker | 1665 (14%) | 2739 (29%) | 4436 (20%) |
| Non-smoker | 6752 (58%) | 5853 (60%) | 12 770 (59%) |
| Missing | 5166 (41%) | 2843 (23%) | 8120 (27%) |

| Disease location | | | |
| ileal (L1) | 3872 (31%) | ·· | ·· |
| colorectal (L2) | 2933 (24%) | ·· | ·· |
| ileocolonic (L3) | 5520 (44%) | ·· | ·· |
| Other | 154 (1%) | ·· | ·· |
| Upper GI (L4) | 1695 (14%) | ·· | ·· |
| Missing | 2777 (18%) | ·· | ·· |

| Disease extent | | | |
| Proctitis (E1) | ·· | 1271 (12%) | ·· |
| Left-sided (E2) | ·· | 4087 (33%) | ·· |
| Extensive (E3) | ·· | 5212 (48%) | ·· |
| Other | ·· | 205 (2%) | ·· |
| Missing | ·· | 1822 (14%) | ·· |

| Disease behaviour | | | |
| Inflammatory (B1) | 6196 (50%) | ·· | ·· |
| Strictureing (B2) | 3250 (26%) | ·· | ·· |
| Penetrating (B3) | 3054 (24%) | ·· | ·· |
| Missing | 2762 (18%) | ·· | ·· |

| Surgery† | | | |
| Yes | 7257 (52%) | 1932 (18%) | ·· |
| No | 6605 (48%) | 8575 (82%) | ·· |
| Missing | 3040 (18%) | 2090 (17%) | ·· |

| GI=gastrointestinal. *Includes 255 patients with indeterminate colitis and 84 patients with missing exact diagnosis. †Excludes data obtained with patient questionnaires (2658 patients). §Surgery in ulcerative colitis refers to colectomy. Denominators for data are: 12 485 for disease location (L1, L2, L3, and other); and 12 597 for disease behaviour in E1, E2, and E3. |

**Table 1: Phenotype distribution of primary cohort**
Figure 1: Evolution of clinical subphenotypes
(A) Proportion of patients with Crohn’s disease who have inflammatory (Montreal classification B1), stricturing (B2), or penetrating (B3) disease over time from diagnosis to most recent follow-up. (B) Proportion of patients with Crohn’s disease who have ileal (L1), colonic (L2), or ileocolonic (L3) disease over time from diagnosis to most recent follow-up. (C) Survival plot of time from diagnosis of Crohn’s disease to resectional surgery stratified by disease location. (D) Survival plot of time from diagnosis of ulcerative colitis to colectomy stratified by disease extent (extensive disease, E3; non-extensive disease, E1 and E2).

Table 2: Associations between genotype and age at diagnosis achieving genome-wide significance

| MAF | Age at diagnosis of IBD p value (β (SE)) | Age at diagnosis of Crohn’s disease p value (β (SE)) | Age at diagnosis of ulcerative colitis p value (β (SE)) |
|-----|------------------------------------------|-----------------------------------------------------|-----------------------------------------------------|
| 3p21 (MST1) |                                           |                                                     |                                                     |
| rs35261698 | 0.306 | 6.14 × 10⁻⁶* | -0.06 (0.01)* |
| rs217252  | 0.288 | 1.35 × 10⁻⁶* | -0.06 (0.01)* |
| rs2197999 | 0.281 | 2.73 × 10⁻⁶* | -0.06 (0.01)* |
| 6p21 (MHC) |                                           |                                                     |                                                     |
| rs3125674 | 0.116 | 3.42 × 10⁻⁷ | -0.03 (0.01) |
| rs41519651 | 0.034 | - | - |
| rs129891 | 0.209 | 1.15 × 10⁻⁶ | -0.05 (0.01) |
| rs268832 | 0.393 | 7.42 × 10⁻⁶* | -0.05 (0.01)* |
| rs482044 | 0.401 | - | - |
| 16q12 (NOD2) |                                           |                                                     |                                                     |
| rs2066844 (p.R702W)  | 0.045 | 3.18 × 10⁻⁶ | -0.08 (0.02) |
| rs2066845 (p.G908R) | 0.016 | 2.10 × 10⁻⁶ | -0.01 (0.03) |
| rs2066847 (p.L1007fsX) | 0.024 | 6.46 × 10⁻⁶* | -0.16 (0.02)* |

Loci are listed by single nucleotide polymorphism. Age at diagnosis assessed by linear regression analysis on normalised data for Crohn’s disease and ulcerative colitis; IBD assessed by meta-analysis of Crohn’s disease and ulcerative colitis data. Effect size is given as standard deviation unit (standard error of effect). MAF = minor allele frequency. IBD = inflammatory bowel disease. *Non-significant associations (p value >0.05). †Genome-wide significant associations. ‡The most significant association pews per subphenotype, if genome-wide significant.
association of NOD2 with behaviour was driven almost entirely by its phenotypic correlation with location and age at diagnosis (figure 2).

We noted complex and correlated HLA signals for susceptibility to inflammatory bowel disease overall and age at onset, as well as Crohn’s disease location and behaviour, extent of ulcerative colitis, and surgery (figure 2; appendix A). Notably, alleles associated with the strongest shared risk allele for Crohn’s disease and ileocolonic association with NOD2 with behaviour was driven almost entirely by its phenotypic correlation with location and age at diagnosis (figure 2).

We noted complex and correlated HLA signals for susceptibility to inflammatory bowel disease overall and age at onset, as well as Crohn’s disease location and behaviour, extent of ulcerative colitis, and surgery (figure 2; appendix A). Notably, alleles associated with the strongest shared risk allele for Crohn’s disease and ileocolonic association with NOD2 with behaviour was driven almost entirely by its phenotypic correlation with location and age at diagnosis (figure 2).

| MAF | Crohn’s disease | Ulcerative colitis |
|-----|----------------|--------------------|
|     | Disease location |                  |                  |
|     | OR (95% CI), ileocolonic vs colonic | OR (95% CI), ileal vs colonic | OR (95% CI), ileal vs colonic |
|     | p value | p value | p value |
| 3p21 (MST1) | n2127252 | 0.288 | 3.10 × 10⁻¹⁴* | 1.07 (1.00-1.13) | 1.08 (1.02-1.15) | 1.10 (1.08-1.13) |
|     |           |         |         |         |         |         |
| 6p21 (MHC) | n3115674 | 0.116 | 3.00 × 10⁻¹⁰* | 0.88 (0.80-0.97) | 0.81 (0.72-0.91) | 4.00 × 10⁻¹⁰* | 0.89 (0.82-0.96) | 5.22 × 10⁻¹⁰* | 1.43 (1.30-1.58)* |
|     | n4151651 | 0.034 | 2.42 × 10⁻⁹* | 0.71 (0.62-0.83)* | 0.58 (0.50-0.68)* | 2.50 × 10⁻¹⁰* | 0.87 (0.77-0.98) | 6.05 × 10⁻¹⁰* | 1.72 (1.47-2.00)* |
|     | n6930777 | 0.112 | 8.13 × 10⁻⁹* | 0.68 (0.62-0.76)* | 0.58 (0.52-0.65)* | 2.00 × 10⁻¹⁰* | 0.89 (0.82-0.96) | 2.49 × 10⁻¹⁰* | 1.36 (1.21-1.52) |
|     | n3128981 | 0.209 | 8.00 × 10⁻¹⁰* | 0.92 (0.87-0.98) | 2.00 × 10⁻¹⁰* | 0.92 (0.87-0.98) | 3.22 × 10⁻¹⁰* | 1.24 (1.17-1.32)* |
|     | n9268332 | 0.393 | 1.40 × 10⁻¹⁰* | 1.03 (0.97-1.09) | 0.94 (0.87-1.02) | 4.00 × 10⁻¹⁰* | 0.93 (0.89-0.97) | 6.59 × 10⁻¹⁰* | 1.12 (1.06-1.19) |
|     | n482044 | 0.401 | 2.38 × 10⁻¹⁰* | 1.15 (1.08-1.22)* | 1.25 (1.16-1.35)* | 8.46 × 10⁻¹⁰* | 1.11 (1.07-1.15) | 1.57 × 10⁻¹⁰* | 0.88 (0.83-0.93) |
|     | n7705575 | 0.439 | 1.00 × 10⁻¹⁰* | 1.23 (1.16-1.30)* | 1.33 (1.23-1.44)* | 2.82 × 10⁻¹⁰* | 1.16 (1.12-1.21)* | 3.24 × 10⁻¹⁰* | 0.92 (0.87-0.98) |
|     | n16q2 (NOD2) | n2066544 | 0.045 | 2.50 × 10⁻¹⁰* | 1.61 (1.43-1.81)* | 1.94 (1.72-2.18)* | 1.76 × 10⁻¹⁰* | 1.21 (1.12-1.31) | 4.67 × 10⁻¹⁰* | 1.10 (1.03-1.18) |
|     |           | n2066545 | 0.016 | 2.77 × 10⁻¹⁰* | 1.59 (1.31-1.93)* | 1.82 (1.50-2.11)* | 7.17 × 10⁻¹⁰* | 1.28 (1.14-1.44) | 2.87 × 10⁻¹⁰* | 1.17 (1.06-1.30) |
|     |           | n2066547 | 0.024 | 1.01 × 10⁻¹⁰* | 1.89 (1.62-2.21)* | 2.50 (2.14-2.92)* | 5.73 × 10⁻¹⁰* | 1.31 (1.21-1.44)* | 2.04 × 10⁻¹⁰* | 1.31 (1.22-1.40)* |

| OR value | p value |
|----------|---------|
| 1.00 | 1.00 |
| 1.00 | 1.00 |
| 1.00 | 1.00 |
| 1.00 | 1.00 |
| 1.00 | 1.00 |

rs3115674 was independently associated with Crohn’s disease behaviour (p=1.56 × 10⁻⁰⁹; figure 2; appendix A). On the basis of sample size, our primary analysis had similar power to detect associations to disease location (ileal vs colonic) as the first International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) meta-analysis on Crohn’s disease.38 However, with the exception of NOD2, MHC, and MST1, we do not report significant associations between subphenotypes and individual SNPs, including those robustly associated with disease susceptibility. We noted, however, that many known risk loci showed nominal evidence for association to a range of subphenotypes, so we posited that genetic risk scores representing the combined effect of many individually weak signals might be a more powerful approach to study the genetic underpinnings of subphenotypes for inflammatory bowel disease. We
calculated different inflammatory bowel disease risk scores constructed from all available data (strength and direction of association) on the lead SNPs from each of the 163 known inflammatory bowel disease susceptibility loci. Although all of the risk scores were associated with Crohn’s disease and ulcerative colitis (CD vs UC score; figure 2). Importantly, this CD versus UC score retained significance even after NOD2, MHC, and MST1 were removed (appendix A), lending support to the notion that the genetic risk score offers more information about the genetic substructure of inflammatory bowel disease than individual SNP associations alone. The strongest correlations in our study were between the CD versus UC risk score and Crohn’s disease location and behaviour (figure 2; p=1.65 x 10^-78, or p=9.23 x 10^-18 after genome-wide significant loci were removed). Risk scores that incorporated imputed HLA types that have been implicated in risk for inflammatory bowel disease significantly improved the genetic risk scores compared with those using SNPs only (appendix A).

Having shown the genetic risk score to be a useful measurement of inflammatory bowel disease subphenotype, we used it to study the genetic relation between ileal Crohn’s disease, colonic Crohn’s disease,
and ulcerative colitis. The CD versus UC risk score placed colonic Crohn’s disease between ileal Crohn’s disease and ulcerative colitis (figure 3). Several other risk scores supported this relation, and although partly driven by the highly location-specific NOD2 variants, a risk score with NOD2 removed showed a similar pattern (appendix A). Additionally, statistical model selection across SNPs and HLA types strongly favoured a model in which colonic Crohn’s disease is intermediate between ileal Crohn’s disease and ulcerative colitis over one that grouped both Crohn’s disease subphenotypes as a single category (appendix A). To test whether this finding extends to other subtypes of inflammatory bowel disease, we applied the genetic risk score to two intermediate forms of the disease: ileocolonic Crohn’s disease (L3), in which the disease affects both small and large bowel, and colonic inflammatory bowel disease unclassified, in which the clinical and histological appearances are indistinguishable between Crohn’s disease and ulcerative colitis. The CD versus UC score placed ileocolonic Crohn’s disease as intermediate between ileal (L1) and colonic (L2) Crohn’s disease, and colonic inflammatory bowel disease unclassified between ulcerative colitis and colonic Crohn’s disease (figure 3).

Despite the statistical significance of the associations between genetic risk score and subphenotype, the small effect sizes translated into fairly low predictive accuracy when tested by cross-validation (appendix A). The risk score that was most significantly associated with location of Crohn’s disease in the primary analysis (CD vs UC) gave an area under the receiver-operating characteristic (ROC) curve of only 0.60 (95% CI 0.57–0.63) for distinguishing between colonic (L2) and ileal (L1) Crohn’s disease in cross-validation, and even a specifically constructed ileal versus colonic score achieved an area
under the curve of only 0.63 (0.59–0.66; appendix A). Comparison of the clinical characteristics of patients with Crohn’s disease that fell into the extreme tails of the genetic risk score (low: log[CD vs UC] ≤–2; high: log[CD vs UC] ≥3) showed significant differences in disease location and behaviour (appendix A), suggesting that although a genetic risk score might not be able to classify all patients, it could be informative at the extremes.

Overall, we can only explain a little of the phenotypic variance in the adult population with either classical or genetic predictors. The combination of smoking and the strongest genetic predictors explains only 6.8% of the variance for disease location in Crohn’s disease (table 4 and table 5), and 1.1% for disease extent in ulcerative colitis (table 5).

To assess the possible clinical usefulness of the genetic risk score, we reassessed the cases of Crohn’s disease with a low CD versus UC risk score (more ulcerative colitis-like), and cases of ulcerative colitis with a high CD versus UC risk score (more Crohn’s disease-like; figure 4). Masked re-phenotyping of these cases raised doubts about the original diagnosis in 27% of the outlier cases compared with 8% of non-outlier cases (corrected for disease location, p=6.8×10–⁴; appendix A). This finding suggests that we can indeed use genetics to identify small numbers of misclassified patients.

### Discussion

The successful identification of genetic variants associated with complex diseases such as inflammatory bowel disease has raised the exciting possibility of a more personalised approach to clinical management. In inflammatory bowel disease, this quest is particularly urgent because of the substantial heterogeneity in disease course, and individual response to therapy.

Past studies in inflammatory bowel disease have established a genetic component of disease subphenotype, but these studies have been limited to a handful of candidate regions in modest numbers of patients. Our study, involving the universal application of standardised phenotyping by trained personnel on nearly 30 000 patients with inflammatory bowel disease from 49 centres worldwide, combined with matching genotypes from more than 150 000 variants, represents the definitive investigation to date into the genetic basis of subphenotypes of inflammatory bowel disease.

The only genome-wide significant associations we noted were between age of onset and disease location with variants at NOD2, MHC, and 3p21. Although the associations between ileal Crohn’s disease and NOD2, and those between colonic Crohn’s disease and the MHC, have been previously described, our study has dissected the phenotype–location associations for the first time. Importantly, in Crohn’s disease, we showed that NOD2 is not associated with stricturing disease after accounting for disease location. These findings, and the rarity of long-term change in disease location compared with behaviour, suggest that location is a fundamental biological aspect of a patient’s disease, whereas behaviour (like surgery or treatment history) is a marker of disease progression.
Our results for ulcerative colitis accord with the previously reported independent associations between the MHC and both extensive disease and colectomy.\(^7\)\(^,\)\(^9\) Notably, the strongest associations with extensive ulcerative colitis are variants on the ancestral 8.1 HLA haplotype. This haplotype is a known recessive risk for primary sclerosing cholangitis,\(^2\)\(^,\)\(^4\) a disease often associated with an extensive but quiescent form of ulcerative colitis.\(^4\) Although variants associated with susceptibility to Crohn’s disease and those associated with susceptibility to ulcerative colitis are both predictive of disease location, the ulcerative colitis-associated variants are the most predictive. The variants associated with disease susceptibility are also slightly predictive of age at diagnosis.

Composite genetic risk scores from all 163 independent susceptibility signals were strongly associated with all our main subphenotypes, and these findings remained significant after excluding NOD2 and MHC. This result accords with a similar finding for genetic risk scores in bipolar disorder,\(^6\) and hints at the possibility that such approaches might be broadly applicable for studying clinical heterogeneity of common diseases. This finding suggests that many or most risk variants for inflammatory bowel disease do contribute weakly to subphenotype. The relative dearth of individual single-nucleotide polymorphism associations with subphenotypes in our study, by contrast with those reported in similarly powered studies of inflammatory bowel disease susceptibility, suggests that the genetic variants studied here have a small effect, and that environmental factors (such as diet, microbiota, and smoking) might be strong contributors to the subphenotypes. However, an intriguing possibility, supported by the notable absence of any functional or pathway enrichment in the components of the genetic risk scores, is that current phenotypic classifications do not correspond strongly to underlying molecular entities. Of particular note is the genetic distinction seen between ileal Crohn’s disease, colonic Crohn’s disease, and ulcerative colitis. These disease types were identified as equally distinct entities on a genetic continuum: on multiple risk scores, colonic Crohn’s disease was genetically intermediate between ileal Crohn’s disease and ulcerative colitis, a finding that remained significant after excluding NOD2. This result substantiates the view that colonic versus ileal disease, rather than disease extent, is the primary clinical unit of Crohn’s disease classification, and is further supported by the finding that both the genetic risk score and clinical complication rate in patients with ileocolonic disease is intermediate between that of patients with ileal disease and colonic disease. It will be of great interest, and potential clinical use, to see if application of these risk scores helps to classify the 10% of patients who are currently designated colonic inflammatory bowel disease unclassified.

The composite scores were also able to identify small numbers of patients with outlier scores who were much more likely to be misdiagnosed than a typical patient. These findings support the possible clinical usefulness of composite scores of multiple genetic variants, each of small effect. For example, genetic outliers could be excluded or specifically targeted for clinical trials to select more homogeneous groups. If these data were readily available, they might affect clinical decisions or inform risk–benefit discussions with patients. For example, the type of surgery offered to patients with refractory colitis crucially depends on whether they have ulcerative colitis or colonic Crohn’s disease; poor outcomes (including pelvic sepsis, incontinence, and sexual dysfunction) of ileal pouch–anal anastomosis reconstruction surgery are much more common in patients with colonic Crohn’s disease. Use of genetic risk scores to identify possible misdiagnoses in this group of patients could help to reduce this problem. Our data also suggest that genetic risk scores could augment biomarkers, such as faecal calprotectin, currently used for patient stratification in inflammatory bowel disease.

A limitation of our study is that the genetic variants tested were restricted to those present (and that passed quality control) on the Immunochip platform, designed for replication and fine mapping of potential immune-mediated disease loci. Therefore, there still might be important loci that determine disease behaviour, location, and age at onset but are independent of those that confer risk for inflammatory bowel disease (or other immune-mediated diseases), and had limited or absent coverage on the Immunochip. A high-powered GWAS, designed to assess possibly overlooked genetic determinants for these outcomes of phenotype expressivity, which uses our large collection of cooperatively phenotyped cases of inflammatory bowel disease with genomic DNA (and the pre-existing Immunochip genotypes), will likely be of great value.\(^8\)

In summary, our research represents the largest genotype–phenotype study in inflammatory bowel disease done so far. Associations achieving genome-wide significance were identified at only three loci, suggesting that new clinical phenotypic classifications might need to be explored for inflammatory bowel disease, and the relation examined between subphenotypes and other omic profiles and environmental factors, including the microbiota.\(^9\) However, our data suggest that on the basis of genetic factors, inflammatory bowel disease is better classified into three distinct groups (ileal Crohn’s disease, colonic Crohn’s disease, and ulcerative colitis), and we would recommend that clinicians adopt this nomenclature in regular practice. We also show that, although genetic risk scores do not yet have widespread clinical use, they are already valuable in some contexts in inflammatory bowel disease, and their study in other complex disease phenotypes is warranted.
Acknowledgments
We thank all the individuals who contributed samples and the physicians and nursing staff who helped with recruitment worldwide. We also thank Prof Gil McVean (professor of Statistical Genetics, Wellcome Trust Centre of Human Genetics, University of Oxford, UK) for thorough review of all statistical methods. UK case collections were supported by the National Association for Colitis and Crohn’s disease; Wellcome Trust grants 098051 (JC) and 098759 (LJ); Medical Research Council UK the Catherine McEwan Foundation; Peninsula College of Medicine and Dentistry, Exeter; the National Institute for Health Research, through the Comprehensive Local Research Network, and through Biomedical Research Centre awards to Guy’s and St Thomas’ National Health Service Trust, King’s College London, Addenbrooke’s Hospital, University of Cambridge School of Clinical Medicine and to the University of Manchester and Central Manchester Foundation Trust. The Wellcome Trust Case Control Consortium projects were supported by Wellcome Trust grants 085148/Z/08/Z, 085475/B/08/Z, and 085475/Z/08/Z. North American collections and data processing were supported by funds to the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) IBD Genetics Consortium, which is funded by the following grants: DK062431 (SRB), DK062422 and DK062423 (SRB and Ferring), HK062421 (RHD), HK062422 (JDR), HK062423 (MSS), HK067984 and HK084554 (MJD). Additional funds were provided by funding to [HJC] HK062429-S1 and Crohn’s & Colitis Foundation of America, Senior Investigator Award [S-2229], and RHD (CA14743).

Declaration of interests
IC, GB, LJ, IJP, DM, JCB, and CWL contributed equally to the manuscript.

Contributors
IC, GB, LJ, DBPM, JCB, and CWL contributed equally to the manuscript. CWL, JCB, DBPM, GR-S, MSS, JM, JDR, SV, and MP conceived, designed, and managed the study and managed the funding. IC, GB, LJ, IJP, LDR, DBPM, JCB, and CWL were involved in manuscript preparation. IC, GB, LJ, PG, MJ, HHu, JDR, and JCB performed or supervised statistical and computational analyses. SZ, TA, Váund, JMA, Vann, SB, SRB, JHC, MD, RHD, LRF, RBG, HHu, JH, HHu, NK, JCL, JCL, JS, SJ, ET, AfdV M-dj, RKW, DCW, MP, SV, JDR, JM, MSS, GR-S, DBPM, and CWL were involved in patient recruitment and assembling phenotypic data. IC, LPS, SZ, RHD, AF, ACL, ET, RKW, JDR, and DBPM established DNA collections, genotyping, and data management. LPS assembled, managed, and validated all submitted phenotypic data. SZ, TA, Váund, JMA, Vann, SB, SRB, JHC, MD, AF, RBG, PG, HHu, JH, HHu, NK, JCL, JCL, JS, ET, DCW, MP, SV, JDR, JM, MSS, GR-S, DBPM, and CWL were involved in patient recruitment and assembling phenotypic data. All authors read and approved the final manuscript before submission.

References
1 Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 2012; 142: 46–54, e42.
2 Neovius M, Arkema EV, Blomqvist P, et al. Patients with ulcerative colitis miss more days of work than the general population, even following colectomy. Gastroenterology 2013; 144: 536–41.
3 Everhart JE, Ruhl CE. Burden of digestive diseases in the United States part 1: overall and upper gastrointestinal diseases. Gastroenterology 2009; 136: 376–86.
Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature 2011; 474: 307–17.

Kirschner BS. Ulcerative colitis in children. Pediatr Clin North Am 1996; 43: 235–54.

Moum B, Ekbom A, Vatn MH, et al. Inflammatory bowel disease: re-evaluation of the diagnosis in a prospective population based study in south eastern Norway. Gut 1997; 40: 328–32.

Pera A, Bellando P, Caldera D, et al. Colonoscopy in inflammatory bowel disease. Diagnostic accuracy and proposal of an endoscopic score. Gastroenterology 1987; 92: 181–85.

Satsangi J, Silverberg MS, Vermeire S, et al. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. Gut 2006; 55: 749–53.

Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated view of autoimmune disease. Nat Rev Immunol 2005; 5: 5–36.

Crohn BB, Ginzburg L, Oppenheimer GD. Regional ileitis; a pathologic and clinical entity. JAMA 1932; 99: 1323–29.

Dalziel TK. Thomas Kennedy Dalziel 1861–1924. Chronic interstitial enteritis. J R Soc Med 1989; 32: 1076–78.

Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated view of autoimmune disease. Nat Rev Immunol 2005; 5: 5–36.

Cleynen I, Gonzalez JR, Figueroa C, et al. Genetic factors and clinical manifestations of Crohn’s disease. Gastroenterology 2002; 122: 867–74.

Ahmad T, ArmuZZi A, Bunce M, et al. The molecular classification of the clinical manifestations of Crohn’s disease. Gastroenterology 2002; 122: 854–66.

Brant SR, Picco MF, Achkar JP, et al. Defining complex contributions of NOD2/CARD15 genes, age at onset, and tobacco use on Crohn’s disease phenotypes. Inflamm Bowel Dis 2003; 9: 281–89.

Cleynen I, Gonzalez JR, Figueroa C, et al. Genetic factors conferring an increased susceptibility to develop Crohn’s disease also influence disease phenotype: results from the IBDchip European Project. Gut 2013; 62: 1566–65.

de la Concha EG, Fernandez-Arques M, Lopez-Nava G, et al. Susceptibility to severe ulcerative colitis is associated with polymorphism in the central MHC gene IKBK. Gastroenterology 2000; 119: 1091–55.

Economou M, Trikalinos TA, Loizou KT, Tsianos E, Ioannidis JPA. Differential effects of NOD2 variants on Crohn’s disease risk and phenotype in diverse populations: a meta-analysis. Gut 2009; 58: 1–9.

Newman B, Silverberg MS, Gu X, et al. CARD15 and HLA DRB1 alleles influence susceptibility and disease localization in Crohn’s disease. Am J Gastroenterol 2004; 99: 306–15.

Dignass A, Lindsay JO, Sturm A, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 2: current management. J Crohns Colitis 2012; 6: 991–1030.

Dignass A, Van Assche G, Lindsay JO, et al. The second European evidence-based consensus on the diagnosis and management of Crohn’s disease: current management. J Crohns Colitis 2010; 4: 28–62.

Mowat C, Cole A, Windsor A, et al. Guidelines for the management of inflammatory bowel disease in adults. Gut 2011; 60: 571–92.

Hanauer SB, Feagan BG, Lichtenstein GR, et al. Maintenance infliximab for Crohn’s disease: the ACCENT I randomised trial. Lancet 2002; 359: 1541–49.

Hanauer SB, Sandborn WJ, Rutgeerts P, et al. Human anti-tumor necrosis factor monoclonal antibody (adalimumab) in Crohn’s disease: the CLASSIC I trial. Gastroenterology 2006; 130: 323–33.

Reinisch W, Sandborn WJ, Hommes DW, et al. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. Gut 2011; 60: 780–87.

Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med 2005; 353: 2462–76.

Sandborn WJ, van Assche G, Reinisch W, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology 2012; 142: 257–65, e1–3.

Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med 2013; 369: 699–710.

Sandborn WJ, Feagan BG, Rutgeerts P, et al. Vedolizumab as induction and maintenance therapy for Crohn’s disease. N Engl J Med 2013; 369: 711–21.

Frolkis AD, Dykeman J, Negron ME, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. Gastroenterology 2013; 145: 996–1006.

Solberg IC, Lygren I, Jahnsen J, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). Scand J Gastroenterol 2009; 44: 431–40.

Solberg IC, Vatn MH, Hoie O, et al. Clinical course in Crohn’s disease: results of a Norwegian population-based ten-year follow-up study. Clin Gastroenterol Hepatol 2007; 5: 1430–18.

Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012; 491: 319–24.

Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl 1989; 170: 2–6; discussion 16–19.

Goyette P, Boucher G, Mallon D, et al. High-density mapping of the MHC identifies a shared role for HLA-DRBI*01:03 in inflammatory bowel diseases and heterogeneous advantage in ulcerative colitis. Nat Genet 2015; 7: 172–79.

Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–75.

Jostins L, Levine AP, Barrett JC. Using genetic prediction from known complex disease loci to guide the design of next-generation sequencing experiments. PLoS One 2013; 8: e76328.

Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association studies and common disease: findings and implications. Nat Rev Genet 2009; 10: 771–82.

Faulcon WA Jr, Loftus EV Jr, Harmsen WS, et al. The natural history of corticosteroid therapy for inflammatory bowel disease: a population-based study. Gastroenterology 2001; 121: 255–60.

Halfvarson J, Bodin I, Tysk C, et al. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. Gastroenterology 2003; 124: 1767–73.

Henckaerts L, Van Steen K, Verstreken I, et al. Genetic risk profiling and prediction of disease course in Crohn’s disease patients. Clin Gastroenterol Hepatol 2009; 7: 972–80, e2.

Donaldson PT, Farrant JM, Wilkinson ML, et al. Dual association of HLA DR2 and DR3 with primary sclerosing cholangitis. Hepatology 1991; 13: 129–33.

Sparkland A, Saarinen S, Boberg KM, et al. HLA class II haplotypes in primary sclerosing cholangitis patients from five European populations. Tissue Antigens 1999; 53: 459–69.

Loftus EV Jr, Harwood GC, Loftus CG, et al. PSC-IBD: a unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. Gut 2005; 54: 91–96.

Harnshere ML, O’Donovan MC, Jones IR, et al. Polygenic dissection of the bipolar phenotype. Br J Psychiatry 2011; 198: 284–88.

Chen GB, Lee SH, Brion MJ, et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Hum Mol Genet 2014; 23: 4710–20.

Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012; 13: R79.

www.thelancet.com