DNA polymorphisms predict time to progression from uncomplicated to complicated Crohn's disease

Drobez, Cvetka Pernat; Repnik, Katja; Gorenjak, Mario; Ferkolj, Ivan; Weersma, Rinse K.; Potocnik, Uros

Original article

DNA polymorphisms predict time to progression from uncomplicated to complicated Crohn's disease

Cvetka Pernat Drobež[a], Katja Repnik[b,c], Mario Gorenjak[b], Ivan Ferkolj[d], Rinse K. Weersma[e] and Uroš Potočnik[b,c]

Objective Most patients with Crohn's disease (CD) are diagnosed with the uncomplicated inflammatory form of the disease (Montreal stage B1). However, the majority of them will progress to complicated stricturing (B2) and penetrating (B3) CD during their lifetimes. The aim of our study was to identify the genetic factors associated with time to progression from uncomplicated to complicated CD.

Patients and methods Patients with an inflammatory phenotype at diagnosis were followed up for 10 years. Genotyping was carried out using Illumina ImmunoChip. After quality control, association analyses, Bonferroni's adjustments, linear and Cox's regression, and Kaplan–Meier analysis were carried out for 111 patients and Manhattan plots were constructed.

Results Ten years after diagnosis, 39.1% of the patients still had the inflammatory form and 60.9% progressed to complicated disease, with an average time to progression of 5.91 years. Ileal and ileocolonic locations were associated with the complicated CD ($P = 1.08E - 03$). We found that patients with the AA genotype at single-nucleotide polymorphism rs16857259 near the gene *CACNA1E* progressed to the complicated form later (8.80 years) compared with patients with the AC (5.11 years) or CC (2.00 years) genotypes ($P = 3.82E - 07$). In addition, nine single-nucleotide polymorphisms (near the genes *RASGRP1*, *SULF2*, *XPO1*, *ZBTB44*, *HLA DOA/BRD2*, *HLA DRB1/HLA DQA1*, *PPARA*, *PUDP*, and *KIAA1614*) showed a suggestive association with disease progression ($P < 10^{-5}$). Multivariate Cox's regression analysis on the basis of clinical and genetic data confirmed the association of the selected model with disease progression ($P = 5.73E - 16$).

Conclusion Our study confirmed the association between the locus on chromosome 1 near the gene *CACNA1E* with time to progression from inflammatory to stricturing or penetrating CD. Predicting the time to progression is useful to the clinician in terms of individualizing patients' management. *Eur J Gastroenterol Hepatol* 00:000–000 Copyright © 2017 Wolters Kluwer Health, Inc. All rights reserved.

Introduction

Crohn's disease (CD), a subtype of inflammatory bowel diseases (IBDs), is a chronic inflammatory condition of the gastrointestinal tract characterized by periods of relapses and remissions [1]. In CD, subclinical inflammation often persists, and there is a progressive evolution to irreversible bowel damage, which is inaccessible to medical therapy and frequently requires surgical intervention [2]. The ultimate goal in the treatment of CD with medications is to prevent the progression of the disease before bowel damage develops [3]. The challenge remains to identify patients who will benefit most from early intensive therapy, while sparing those who will derive minimal benefit from

such treatment [4]. Therefore, it would be useful to predict at the time of diagnosis as to which patients are likely to develop the complicated course of disease. Several tools have been proposed to predict the occurrence of complications, including clinical characteristics, and fecal, serologic, and genetic markers. Clinical factors associated with complicated CD include young age at diagnosis, small bowel location (ileal and/or ileocolonic), upper gastrointestinal extent, stricturing or penetrating behavior, perianal disease, severe endoscopic lesions, and smoking [5–8]. Genetic markers seem to be more useful as predictors of the clinical course of CD as they are stable over time and unaffected by disease behavior [9]. With the advent of genome-wide association studies (GWAS), more than 200 susceptibility loci have been identified for IBD, including 37 specific for CD [10,11]. An association between genetic markers and complicated CD phenotype has been suggested in several studies; however, they mostly focused on *NOD2* [12,13] and other candidate genes such as *DLG5*, *ATG16L1*, *PRDM1*, *IRGM*, *TNFSF15*, *C13ORF31*, *JAK2*, and *IL23R* [14,15]. Very few studies so far have investigated genetic markers as predictors of time to conversion from uncomplicated to complicated CD [9,15,16]. Most of them were looking for correlations between single-nucleotide polymorphisms (SNPs) and time to progression to complicated disease, focusing on a single candidate gene [9,15]. The association of *NOD2* variants with a shorter time to onset of complicated disease has been suggested [15],

although this association was not confirmed in recent GWAS [16].

In the present study, we aimed to identify biomarkers that can predict the group of CD patients expected to have a rapid progression from uncomplicated inflammatory to complicated stricturing or penetrating disease. We studied a precisely defined cohort of CD patients, who only had the inflammatory phenotype at diagnosis, were followed up for 10 years, and for whom the time to progression had been defined precisely.

Patients and methods

Patients and data collection

We reviewed the medical files of 198 Slovenian patients with CD, who had been recruited from two main University Medical Centers in Slovenia – that is, Ljubljana and Maribor. All clinical data for the patients enrolled in the study were collected retrospectively. All patients enrolled in the study were Central European Caucasians of Slavic origin, representing an ethnically and genetically homogenous population, diagnosed with CD between 1978 and 2006. Only patients with the uncomplicated inflammatory form of the disease at diagnosis were included and were followed up yearly for 10 years, starting at the first year from the time of diagnosis. All patients agreed to participate in the study by providing a written consent. Confirmation of the CD diagnosis and assignment of clinical phenotypes were performed by a senior clinician specializing in IBD through case note reviews of clinical, radiological, histopathological, and endoscopic reports. All patients were examined at least annually by a gastroenterologist and for every examination, the medical record was obtained. At the regular annual exam, a global medical assessment was carried out and the full blood count and C-reactive protein as a marker of disease activity were analyzed [17]. If the patient developed borborygmus and cramping pain, which disappeared after defecation, or if the patient developed abdominal pain with tension of the abdominal wall and with signs of systemic inflammation, endoscopic and radiological exams (mostly computed tomography) were performed to confirm progression from uncomplicated to complicated disease. Disease phenotype was defined according to the patient's medical records and from the questionnaire, filled out by each patient. Details included sex, age, family history of IBD, age at diagnosis, disease location, and behavior and data on smoking. The diagnosis of CD was made on the basis of endoscopic, radiological, and histological criteria [18]. The disease location (L1–L4), behavior (B1–B3), and age at diagnosis (A1–A3) were described at diagnosis and during the follow-up period according to the Montreal classification [19]. A family history of the disease was defined as the presence of IBD in first-degree or second-degree relatives. Patients were defined as current smokers if they smoked at least seven cigarettes per week, as never-smokers if they had never smoked, and as ex-smokers if they had stopped smoking at least 6 months before the diagnosis of CD [20]. Information on therapy, including corticosteroids at first flare as well as systemic corticosteroids, immunosuppressors, and biological therapy during follow-up was collected from the medical records.

After reviewing 198 medical records, 37 patients with incomplete medical data (on either the diagnosis period or the time interval during the 10-year study period) or loss of follow-up were excluded. A total of 161 patients were studied. Patient characteristics at the time of diagnosis are presented in Table 1.

Two relevant outcomes were considered. If the development of complicated CD was observed, time to progression (in years) from inflammatory (B1) to stricturing (B2) or penetrating (B3) disease was determined. The B2/B3 status was determined for each patient using endoscopy and radiology as complementary techniques to define the site and extent of strictures and penetrating lesions according to the criteria of international guidelines for routine clinical practice [21,22]. Strictures (B2) were defined as the occurrence of constant luminal narrowing shown by radiological, endoscopic, and/or surgical–pathologic methods with obstructive signs/symptoms or prestenotic dilatation [23]. Nonperianal penetrating disease (B3) was defined as the occurrence of enteroenteric, enterovesical, or enterocutaneous fistulas or abscesses [24]. Perianal disease (Montreal classification p) was classified separately and defined as the occurrence of perianal fistulas and/or abscesses, including rectovaginal fistulae. The development of perianal disease was not considered as a change in behavior, but as a modifier of disease behavior.

Experiments were conducted with the understanding and written consent from each individual. The study was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). The Slovenian National Committee for Medical Ethics approved the study with the consensus '21k/11/11'.

DNA isolation and genotyping

DNA samples were obtained from 12 ml of peripheral blood. Peripheral blood lymphocytes were collected using FicollPaque PLUS (GE Healthcare, Uppsala, Sweden) and DNA was isolated from lymphocytes using the TRI reagent (Sigma-Aldrich, St Louis, Missouri, USA) according to the manufacturer's instructions. All patients were genotyped using Illumina ImmunoChip (iCHIP) (Illumina Inc., San Diego, California, USA) according to the manufacturer's protocol and as described previously [16,25]. Genotyping clusters of the SNPs included in the current analysis were checked manually using the GenomeStudio software by Illumina Inc. Individuals with a call rate less than 95% and/or discordant sex information, and SNPs with a call rate less than 98% were excluded from further analysis.

Statistical analysis

Quality control was performed using the following parameters: missing rate per SNP less than 0.05, missing rate per individual less than 0.02, heterozygosity per individual ± 0.2, missing rate per SNP difference less than 0.02, SNP pruning to remove linkage disequilibrium r^2 more than 0.05, and low-frequency SNPs with minor allele frequency less than 0.05. After quality control, the data set comprised 169.548 SNPs and 111 individuals with corresponding progression data that were extracted for further analysis. The power analysis for quantitative trait was carried out using SPSS 22.0 (IBM Inc., Armonk, New York, USA) and

Table 1. Patients' characteristics at baseline as prognostic factors for disease progression over a follow-up period of 10 years

Characteristics	Baseline [n (%)]	B1 after 10 years [n (%)]	B2 or B3 after 10 years [n (%)]	OR	95% CI	P-value
N (%)	161 (100)	63 (39.1)	98 (60.9)	–	–	–
Sex						
Male	68 (42.2)	22 (32.4)	46 (67.6)	1.649	0.859–3.166	0.144
Female	93 (57.8)	41 (44.1)	52 (55.9)			
Age at diagnosis						
A1	24 (14.9)	10 (41.7)	14 (58.3)	0.883	0.366–2.132	0.823
A2	114 (70.8)	40 (35.1)	74 (64.9)	1.773	0.890–3.532	0.113
A3	23 (14.3)	13 (56.5)	10 (43.5)	0.437	0.179–1.069	0.105
Disease location						
L1	34 (21.1)	12 (35.3)	22 (64.7)	1.230	0.560–2.705	0.694
L2	54 (33.5)	31 (57.4)	23 (42.6)	0.317	0.160–0.625	**1.08E − 03**[a]
L3	73 (45.3)	20 (27.4)	53 (72.6)	2.532	1.305–4.913	**6.08E − 03**
L4	7 (4.3)	1 (14.3)	6 (85.7)	4.043	0.475–34.414	0.248
Perianal disease						
Yes	6 (3.7)	3 (1.9)	3 (1.9)	0.632	0.123–3.232	0.680
No	155 (96.3)	60 (37.2)	95 (59.0)			
Smoking						
Current smoker	64 (39.8)	24 (38.1)	40 (40.8)	1.121	0.586–2.144	0.745
Ex-smoker	4 (2.4)	2 (3.2)	2 (2.1)	0.635	0.087–4.630	0.645
Never-smoker	93 (57.8)	37 (58.7)	56 (57.1)	0.937	0.493–1.780	0.871
Family history						
Yes	25 (15.5)	10 (40.0)	15 (60.0)	0.958	0.401–2.289	1.000
No	136 (84.5)	53 (39.0)	83 (61.0)			

CI, confidence interval; OR, odds ratio.
Bold values indicate statistical significance, $P < 0.05$.
[a]L1 + L3 versus L2.

R 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria). Heredity was represented by variance explained by additive effects under an additive model and was calculated using analysis of variance in SPSS 22.0. Subsequently, a post-hoc power analysis was carried out using R 3.4.1 and using implemented qchisq for the threshold and pchisq for the actual power calculation. Association analyses and Bonferroni's adjustments were carried out using PLINK v1.07 (*http://zzz.bwh.harvard.edu/plink/*) [26], and Manhattan plots were constructed using R and the qqman package [27]. ImmunoChip-wide significance level was set to $P = 3.8E – 07$. For most significant SNPs, linear regression, Cox's regression, and Kaplan–Meier (for the most significant SNP after Bonferroni's correction) were performed using SPSS IBM Statistics 22.0 (IBM Inc.). For linear regression, variables with variance inflation factor more than 10 and condition index more than 30 were excluded from analysis. Cox's regression was also performed using previously excluded variables.

Results

Clinical predictors of progression to complicated disease

During the follow-up period of 10 years, out of 161 patients with stage B1 at diagnosis and enrolled in our follow-up study, 63 (39.1%) patients remained in stage B1, whereas 98 (60.9%) patients progressed to complicated B2 or B3 stage during the follow-up period of 10 years. The average time to progression from B1 to B2 or B3 was 5.91 years, with an SD of 2.7 years. Of 98 patients with progression, 96 (99%) patients underwent surgery during the 10-year follow-up period, whereas in the group of patients showing no progression in the 10-year follow-up period, only one patient was operated on because of bleeding. On comparing patients' characteristics at the baseline (sex, age at diagnosis, disease location, perianal disease, smoking, and family history) according to disease progression, we found an association between location and progression. Patients with ileal and ileocolonic location (L1 + L3) progressed to the B2 or the B3 phenotype within the follow-up period of 10 years more frequently compared with the patients with a colonic L2 location ($P = 1.08E – 3$, odds ratio = 0.317, 95% confidence interval = 0.160–0.625; Table 1). A slight trend toward significance was observed between the age at diagnosis and disease progression ($P = 0.113$, odds ratio = 1.773, 95% confidence interval = 0.890–3.532). 64.9% of patients with A2 progressed to the B2 or the B3 phenotype compared with 51.1% of patients with A1 or A3. The association between the time to progression and therapy (immunosuppressors and biological therapy) was confirmed (Table 2). Biological therapy and immunosuppressors, introduced at least 1 year before the phenotype change, delayed the progression of the disease for 3 years ($P = 5.35E–03$ and $6.00E–06$, respectively; Table 2). The Kaplan–Meier log-rank analysis confirmed delayed progression of the disease in patients receiving biological therapy (log-rank $P = 6.0E–06$; Fig. 1a) and in patients receiving immunosuppressors (log-rank $P = 7.07E–04$; Fig. 1b).

Genetic predictors of progression to complicated disease

After quality control, genotypes for 169.548 SNPs were available for 111 patients. Out of 111 patients with available genotypes, 65 (58.6%) progressed from the uncomplicated (B1) to the complicated form (B2 or B3) of CD. The average time to progression from the B1 to the B2 or the B3 phenotype for patients with genotypes passing QC was 6.12 ± 2.82 years.

Ten independent loci with P less than $9.9E10 – 5$ were found to be associated with disease progression (Table 3). The strongest statistically significant GWAS association between the genotype and progression from the B1 to the

Table 2. Therapy and time to progression in patients who progressed to the B2 or the B3 phenotype during a follow-up period of 10 years

Therapies	Patients treated before progression		Patients with no treatment or treated after progression		
	n (%)	Median time to progression (years)	n (%)	Median time to progression (years)	P-value
Corticosteroids at first flare	54 (55.1)	6	44 (44.9)	6	0.485
Systemic corticosteroids	85 (86.7)	6	13 (13.3)	4	0.261
Immunosuppressors	38 (38.8)	8	60 (61.2)	5	6.00E − 06
Biological therapy	14 (14.3)	9	84 (85.7)	6	5.35E − 03

Bold values indicate statistical significance, $P < 0.05$.
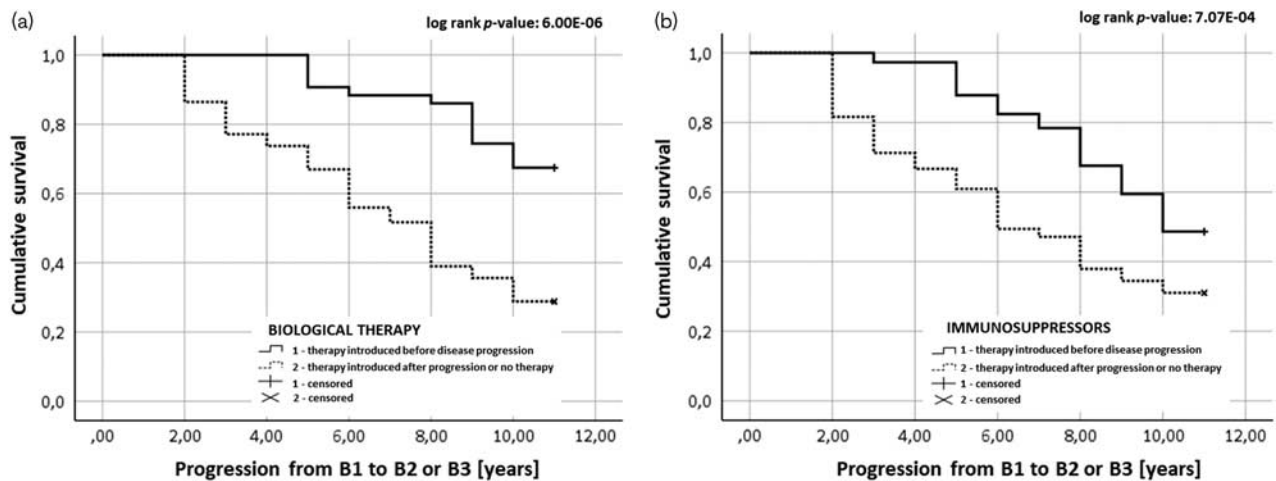
Fig. 1. (a) Association between biological therapy and the time to progression from B1 to B2 or B3. (b) Association between immunosuppressors and the time to progression from B1 to B2 or B3.

**Table 3.** Top 10 loci associated with disease progression from the B1 to the B2 or the B3 phenotype

Chromosomes	SNP ID	Genotype	Genotype frequency	Mean time to progression from B1 to B2 or B3 (years)	SD (years)	P-value[a]	Locus (candidate gene)
1	rs16857259	CC	0.009	2.00	0.00	3.82E − 07	CACNA1E
		CA	0.162	5.11	3.16		
		AA	0.829	8.80	2.85		
15	rs16967072	TT	0.000	NA	NA	6.55E − 06	RASGRP1
		TC	0.081	3.67	1.66		
		CC	0.919	8.54	3.04		
20	rs2664588	TT	0.234	5.85	3.48	1.31E − 05	SULF2
		TC	0.469	8.44	3.09		
		CC	0.297	9.49	2.22		
2	rs7600515	CC	0.162	5.50	3.71	1.54E − 05	XPO1
		CG	0.387	7.95	3.03		
		GG	0.451	9.26	2.63		
11	rs17605375	CC	0.009	2.00	0.00	2.06E − 05	ZBTB44
		CT	0.171	5.74	3.19		
		TT	0.820	8.71	2.96		
6	rs206768	CC	0.171	6.05	3.31	2.22E − 05	HLA DOA/BRD2
		CT	0.432	7.71	3.30		
		TT	0.396	9.52	2.48		
6	rs12660719	GG	0.009	2.00	0.00	3.31E − 05	HLA DRB1/HLA DQA1
		GA	0.045	3.00	1.41		
		AA	0.946	8.45	3.04		
22	rs4253662	AA	0.000	NA	NA	6.80E − 05	PPARA
		AG	0.153	5.35	3.08		
		GG	0.847	8.65	3.01		
X	rs11095248	TT	0.100	5.46	2.62	8.72E − 05	PUDP
		TC	0.146	6.50	3.95		
		CC	0.755	8.78	2.88		
1	rs7515553	GG	0.225	9.52	2.22	9.83E − 05	KIAA1614
		GA	0.532	8.48	3.23		
		AA	0.243	6.15	3.18		

SNP, single-nucleotide polymorphism.
[a]P-values below 9.9E10 − 5 indicate a less stringent suggestive association threshold and serve as an indicator of a suggestive association that requires further validation.
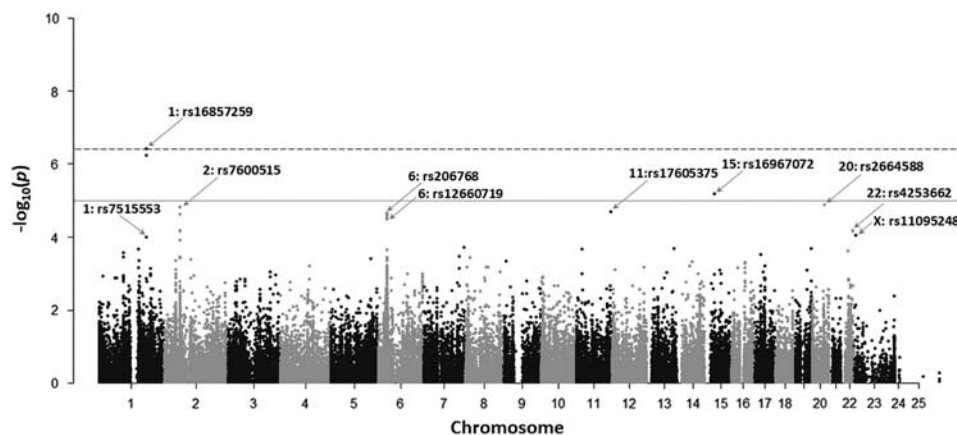
Fig. 2. Genome-wide association studies results (Manhattan plot). Dashed line (- - -) indicates Bonferroni's significance threshold $P = 3.8E - 07$. The dotted line (•••••) indicates suggestive P-value ($P = E - 05$).
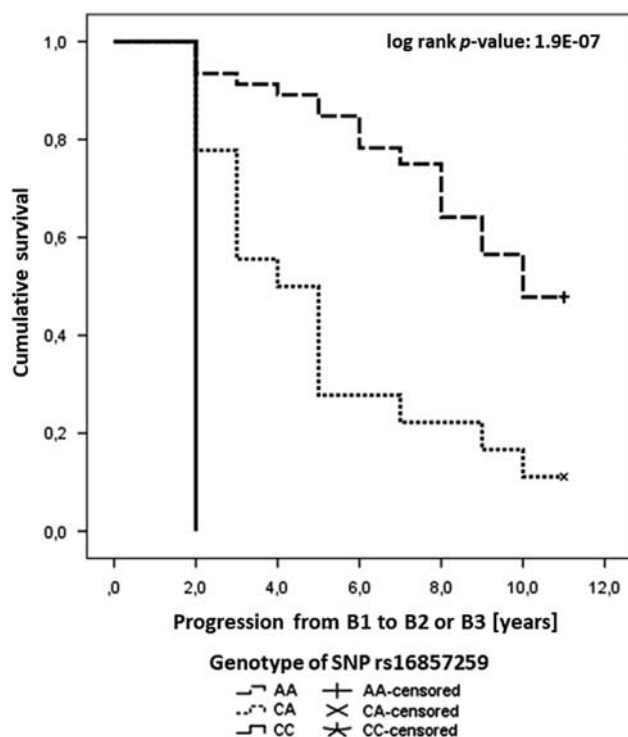
Fig. 3. Kaplan–Meier curve for single-nucleotide polymorphism rs16857259.

B2 or the B3 phenotype was found for SNP rs16857259 on chromosome 1. Patients with the CC genotype progressed to the B2 or the B3 phenotype on average after 2 years, patients with the AC genotype after 5.1 years, and patients with the AA genotype after 8.8 years ($P = 3.82E - 07$). Other loci associated with disease progression were found on chromosomes 15 (rs16967072), 20 (rs2664588), 2 (rs7600515), 11 (rs17605375), 22 (rs4253662), X (rs11095248), and 1 (rs7515553) (Table 3 and Fig. 2). Two independent loci from the HLA region (chromosome 6) were also associated with disease progression, the SNP rs206768 upstream of *BRD2* and downstream of *HLA DOA*, and the SNP rs12660719 upstream of *HLA DRB1* and downstream of *HLA DQA1*.

In both cases, the ancestral allele, T for rs206768 and A for rs12660719, was associated with longer time to progression.

Association of genetic and clinical factors with time to progression

Linear regression, which included top 10 SNPs associated with disease progression and selected clinical parameters – sex, location, age at diagnosis, family history, smoking, use of corticosteroids, biological therapy, and immunosuppressors, explained 69.7% of disease progression variability ($P = 1.5E - 16$). The SNPs rs16857259 ($P = 3.16E - 4$), rs7600515 ($P = 4.0E - 03$), rs2664588 ($P = 6.0E - 03$), rs7515553 ($P = 4.0E - 03$), rs12660719 ($P = 2.4E - 4$), and rs11095248 ($P = 1.7E - 02$) were confirmed as independent prognostic factors of disease progression. None of the clinical parameters were associated independently with disease progression in the predicted model.

Furthermore, Cox's regression was performed. The covariates included were the top five SNPs associated with disease progression, biological therapy, and immunosuppressors. The regression model was significant ($P = 5.73E - 16$) and confirmed the independent hazard for the cumulative survival of the SNPs rs16857259 ($P = 0.005$), rs7600515 ($P = 0.010$), rs17605375 ($P = 0.001$), rs16967072 ($P = 0.001$), and rs2664588 ($P = 0.016$), but not biological therapy ($P = 0.200$) or immunosuppressors ($P = 0.355$). In addition, a univariate Kaplan–Meier log-rank analysis was carried out for all the SNPs associated with disease progression. Figure 3 presents the Kaplan–Meier curve for the most significant SNP rs1685725, for which patients with the CC genotype and AC genotypes progressed much faster from the B1 phenotype to the B2 or B3 phenotypes compared with patients with the AA genotype, with a log-rank significance of $P = 1.9E - 07$ (Fig. 3). From the Kaplan–Meier curve, we found that the highest difference between the genotypes could be observed in the first 5 years, which is similar for all the SNPs associated with disease progression.

The post-hoc power analysis showed an overall study power of 72%.

Discussion

Our study represents the first comprehensive identification of genetic factors associated with time to progression from uncomplicated to complicated CD using the previously developed custom-designed Illumina ImmunoChip [10], designed for comprehensive in-depth replication of GWAS results and fine mapping of susceptibility loci implicated in immune-mediated disorders. We found that time to progression is partially genetically predisposed. In our study, the strongest and most consistent association between the time to progression and the gene variants was found for the SNP rs16857259 near the gene *CACNA1E*. Carriers of the C allele for the SNP rs16857259 progressed faster from the uncomplicated inflammatory to the complicated stricturing and penetrating CD. The SNP rs16857259 is located 5 kb upstream of the *CACNA1E* gene on chromosome 1q25.3. The *CACNA1E* gene encodes $Ca_V2.3$, a subunit of Ca^{2+}-channels, which is expressed in neuronal cells and pancreatic β cells. The SNPs in the *CACNA1E* gene have been associated previously with type 2 diabetes, specifically with interfering with insulin action [28–30], which can be of interest for CD, as patients with CD show increased insulin secretion [31]. Interestingly, *CACNA1E* was the most important of the 99/4 (maximum/minimum) transcripts in peripheral blood, which can be used to differentiate between non-IBD colitis/healthy controls and CD patients by 100% specificity and sensitivity [32]. In addition, calcium channels have also been associated with different cancer types, including colon [33] and prostate [34] cancer and nephroblastoma (Wilms' tumor) [35], indicating the implication of Ca^{2+}-mediated intracellular signaling pathways in tumorigenesis as well as proliferation. In Wilms' tumors, *CACNA1E* was overexpressed in the relapsing form compared with the nonrelapsing form of disease [35]. Torkamani *et al.* [36] carried out a pathway analysis approach to characterize the likely polygenic basis of seven common diseases, including CD, and observed pathway hits in calcium signaling, specifically at RHOA, AKAP6, and calcium channels CACNA1C and CACNA2D3, all of which modulate calcium levels, which, in turn, through calcineurin signaling, regulate the transcriptional activity of another associated gene in the calcium signaling pathway, NF-AT (NFATC1). In addition, allele C of SNP rs16857259 identified in our study as the most promising predictor of disease progression is in high linkage disequilibrium with allele A of SNP rs10797715. Allele A of SNP rs10797715 correlates with a higher expression of gene *CACNA1E*, suggesting that higher expression of gene *CACNA1E* is associated with more rapid disease progression in CD patients. Next, the association with disease progression was confirmed for a locus on chromosome 15 near the gene *RASGRP1*. Interestingly, the SNP rs16967103, only 12.616 bp upstream of the SNP rs16967072, which was associated with disease progression in our study, was associated with CD in a GWAS [10], making this region an interesting target, especially as RASGRP1 has an important immune function [37]. Further, the region on chromosome 20 near the *SULF2* gene (rs2664588) showed a promising association with time to progression from the uncomplicated to the complicated phenotype. Interestingly, the *SULF2* gene was recently found to be implicated in antibacterial autophagy

in PBMCs from healthy individuals homozygous for the ATG16L1*300T or *300A alleles [38], of which *300A is a well-known CD risk variant [39,40] and affects the response efficiency to biological treatment with adalimumab in CD patients [41]. Next to *CACNA1E*, *RASGRP1*, and *SULF2* loci, suggestive associations between seven additional loci and disease progression were observed in our study. Among these, two loci are a part of the HLA region, for which a strong association with IBD has already been confirmed [42], and that has also been replicated in CD [43]. Ina addition, the association between the SNP rs77005575 from the HLA region and disease behavior has been observed in a GWAS [16]. We confirmed the association between disease progression and the SNP rs12660719, which is located only 3.725 bp downstream from rs77005575. Patients with GG or AG genotypes of SNP rs12660719 progressed faster from the uncomplicated to the complicated form of CD.

To date, genetic studies of disease behavior have been rare and the design of these studies has varied widely. Most of them have looked for correlations between the SNPs and the time to progression, focusing on single candidate genes [9,15]. Henckaerts *et al.* [9] examined the influence of 50 CD-associated risk loci on changes in disease behavior and found an association between the SNP rs1363670 near *IL12B* and the stricturing disease behavior and shorter time to strictures, especially in patients with ileal involvement, and concluded that CD-associated SNPs play a role in disease progression. Further, Cleynen *et al.* [15] carried out a similar study, where they investigated the influence of more than 40 SNPs on the clinical course of CD and highlighted *NOD2* and early immunomodulatory use as the clinically most meaningful predictors for the clinical course. However, a recently published genome-wide genotype–phenotype study showed that *NOD2* is not associated with the stricturing disease after accounting for disease location [16]. Interestingly, a recently published study also suggests that disease-associated genes are not the same genes as those associated with disease progression [44]. Our study makes a similar conclusion as the strongest correlation was observed for a gene not associated previously with CD. These findings confirm the need for replication studies as well as homogenous and unified cohorts of patients to find reliable genetic markers of disease progression.

It is well known that the behavior of CD changes over the course of the disease [45] and that different phenotypes require specific diagnostic solutions [46]. In our study, 60.9% of CD patients showed a change in the phenotype during the follow-up period of 10 years. This is consistent with the observation based on two studies from referral centers, whereas, after 10 years, 70% of patients had the stricturing or the penetrating form of the disease [24,45], whereas the corresponding figure in a population-based 10-year follow-up study was only 53% [47]. However, the results are difficult to compare because in the studies from referral centers, as in our study, patients with a more severe course of disease are likely to be over-represented. One of the most important clinical risk factors for CD progression is disease location, which was also confirmed in our study. Patients with ileal and ileocolonic locations progressed to the complicated phenotype more frequently compared with patients with the colonic location. In numerous studies, a strong association between the ileal,

ileocolonic, and upper gastrointestinal involvement and CD progression has been observed [48–51]. Disease location is considered a fundamental biological aspect of the CD patient's disease and a major driver to changes in disease behavior over time [16].

On comparing the patients' characteristics at baseline according to disease progression, we found no association between sex, perianal disease, smoking, family history, the need for corticosteroid use, and CD progression. Several studies have confirmed that patients who present at a younger age (particularly <40 years) have a higher risk of developing complicated disease [16,47,52]. In our study, an association between the age at diagnosis and disease progression has been observed. So far, it has been shown that women may have a higher risk of intestinal resection and requirement for surgery compared with men [47,52]. Further, multiple studies have associated smoking with a higher risk of developing complicated disease in CD patients [53,54]. However, our study did not show any association between sex and disease progression, or with the smoking status and disease progression, similar to another study specifically designed to explore the influence of smoking on the clinical phenotype [55]. On comparing the family history of IBD and disease behavior, no major differences were observed in our study, or in other studies [56]. The use of systemic steroids for treating the first flare at diagnosis was suggested recently as an important risk factor for a disabling course of CD [47,52]. Our study found no association between the use of systemic steroids at first flare and the progression of the disease. In our opinion, early use of systemic steroids should only be considered as a surrogate marker of active disease, rather than a predictor of complicated disease. The associations between systemic corticosteroid use and CD progression have been reported previously [57–59]. However, our study did not confirm an influence of systemic corticosteroid use on time to progression. Early intervention with immunosuppressors and biological therapy can prevent CD progression and avoid complications of the disease [60–63]. Our study confirmed the protective effect of immunosuppressors and biological therapy on time to progression. Immunosuppressors and biological therapy delayed the progression of the disease by 3 years. This is consistent with the observation that immunosuppressors and biological agents started before the behavior change delay phenotype progression of CD [57].

A major advantage of our study was the determination of exact time (to a 1-year resolution) of progression from the uncomplicated to the complicated form of disease during a follow-up period of 10 years. The selected monitored period in our study is reasonable as it has been shown that disease progression from the inflammatory to the stricturing or the penetrating form occurs more frequently during the first 5 years of the disease [47,57]. To the best of our knowledge, in other genetic studies, the exact time to progression was not determined; rather, an approximation or other adjustments were made. For instance, in the largest genotype–phenotype study of IBD, time to progression was estimated as time to first event according to the year of diagnosis and the year of last review (censored data) [16]. We are, however, aware of the limitation of our study that the cohort size is sufficient only to detect SNPs and associations with the highest contribution to disease progression

and that our study most likely missed SNPs with smaller contribution as well as produced a list of suggestive associations that need further confirmation in independent cohorts. Genetic studies in cohorts with available data on the exact time to progression, similar to the data used in our study, are currently lacking and are highly warranted to enable independent replication studies and meta-analyses for the identification of most reliable genetic predictors of disease progression in the near future.

In conclusion, the results of our study confirm genetic biomarkers as useful predictors of those CD patients who are expected to experience rapid progression from uncomplicated inflammatory to complicated stricturing and penetrating CD. Several regions showed a suggestive association of disease progression with the SNP rs16857259 near the gene *CACNA1E*, which is the most promising candidate.

Conflicts of interest

There are no conflicts of interest.

References

1　Colombel JF, Narula N, Peyrin-Biroulet L. Management strategies to improve outcomes of patients with inflammatory bowel diseases. *Gastroenterology* 2017; 152:351–361.e5.

2　Pariente B, Mary JY, Danese S, Chowers Y, de Cruz P, D'Haens G, *et al*. Development of the Lémann index to assess digestive tract damage in patients with Crohn's disease. *Gastroenterology* 2015; 148:52–63.e3.

3　Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007; 369:1641–1657.

4　Devlin SM, Cheifetz AS, Siegel CA, group B. Patient-specific approach to combination versus monotherapy with the use of antitumor necrosis factor α agents for inflammatory bowel disease. *Gastroenterol Clin North Am* 2012; 41:411–428.

5　Zallot C, Peyrin-Biroulet L. Clinical risk factors for complicated disease: how reliable are they? *Dig Dis* 2012; 30 (Suppl 3):67–72.

6　Romberg-Camps MJ, Dagnelie PC, Kester AD, Hesselink-van de Kruijs MA, Cilissen M, Engels LG, *et al*. Influence of phenotype at diagnosis and of other potential prognostic factors on the course of inflammatory bowel disease. *Am J Gastroenterol* 2009; 104:371–383.

7　Lakatos PL. Prevalence, predictors, and clinical consequences of medical adherence in IBD: how to improve it? *World J Gastroenterol* 2009; 15:4234–4239.

8　Louis E, Mary JY, Vernier-Massouille G, Grimaud JC, Bouhnik Y, Laharie D, *et al*. Maintenance of remission among patients with Crohn's disease on antimetabolite therapy after infliximab therapy is stopped. *Gastroenterology* 2012; 142:63–70.e5. quiz e31.

9　Henckaerts L, van Steen K, Verstreken I, Cleynen I, Franke A, Schreiber S, *et al*. Genetic risk profiling and prediction of disease course in Crohn's disease patients. *Clin Gastroenterol Hepatol* 2009; 7:972–980.e2.

10　Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, *et al*. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012; 491:119–124.

11　Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, *et al*. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015; 47:979–986.

12　Adler J, Rangwalla SC, Dwamena BA, Higgins PD. The prognostic power of the NOD2 genotype for complicated Crohn's disease: a meta-analysis. *Am J Gastroenterol* 2011; 106:699–712.

13　Lichtenstein GR, Targan SR, Dubinsky MC, Rotter JI, Barken DM, Princen F, *et al*. Combination of genetic and quantitative serological

immune markers are associated with complicated Crohn's disease behavior. *Inflamm Bowel Dis* 2011; 17:2488–2496.

14   Weersma RK, Stokkers PC, van Bodegraven AA, van Hogezand RA, Verspaget HW, de Jong DJ, et al. Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort. *Gut* 2009; 58:388–395.

15   Cleynen I, Gonzalez JR, Figueroa C, Franke A, McGovern D, Bortlik M, 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; 62:1556–1565.

16   Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* 2016; 387:156–167.

17   Boirivant M, Leoni M, Tariciotti D, Fais S, Squarcia O, Pallone F. The clinical significance of serum C reactive protein levels in Crohn's disease. Results of a prospective longitudinal study. *J Clin Gastroenterol* 1988; 10:401–405.

18   Dignass A, Van Assche G, Lindsay JO, Lémann M, Söderholm J, Colombel JF, 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.

19   Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 2006; 55:749–753.

20   Cosnes J, Carbonnel F, Beaugerie L, Le QY, Gendre JP. Effects of cigarette smoking on the long-term course of Crohn's disease. *Gastroenterology* 1996; 110:424–431.

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

22   Van Assche G, Dignass A, Panes J, Beaugerie L, Karagiannis J, Allez M, et al. The second European evidence-based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *J Crohns Colitis* 2010; 4:7–27.

23   Gasche C, Scholmerich J, Brynskov J, D'Haens G, Hanauer SB, Irvine EJ, et al. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000; 6:8–15.

24   Cosnes J, Cattan S, Blain A, Beaugerie L, Carbonnel F, Parc R, et al. Long-term evolution of disease behavior of Crohn's disease. *Inflamm Bowel Dis* 2002; 8:244–250.

25   Fransen K, Mitrovic M, van Diemen CC, Thelma BK, Sood A, Franke A, et al. Limited evidence for parent-of-origin effects in inflammatory bowel disease associated loci. *PLoS One* 2012; 7:e45287.

26   Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81:559–575.

27   Turner D. S qqman: an R package for visualizing GWAS results using Q–Q and manhattan plots. *biorXiv* 2014; doi: 10.1101/005165.

28   Muller YL, Hanson RL, Zimmerman C, Harper I, Sutherland J, Kobes S, et al. Variants in the Ca V 2.3 (alpha 1E) subunit of voltage-activated Ca^{2+} channels are associated with insulin resistance and type 2 diabetes in Pima Indians. *Diabetes* 2007; 56:3089–3094.

29   Trombetta M, Bonetti S, Boselli M, Turrini F, Malerba G, Trabetti E, et al. CACNA1E variants affect beta cell function in patients with newly diagnosed type 2 diabetes. the Verona newly diagnosed type 2 diabetes study (VNDS) 3. *PLoS One* 2012; 7:e32755.

30   Holmkvist J, Tojjar D, Almgren P, Lyssenko V, Lindgren CM, Isomaa B, et al. Polymorphisms in the gene encoding the voltage-dependent Ca(2+) channel Ca (V)2.3 (CACNA1E) are associated with type 2 diabetes and impaired insulin secretion. *Diabetologia* 2007; 50: 2467–2475.

31   Bregenzer N, Hartmann A, Strauch U, Schölmerich J, Andus T, Bollheimer LC. Increased insulin resistance and beta cell activity in patients with Crohn's disease. *Inflamm Bowel Dis* 2006; 12:53–56.

32   Sipos F, Galamb O, Wichmann B, Krenács T, Tóth K, Leiszter K, et al. Peripheral blood based discrimination of ulcerative colitis and Crohn's disease from non-IBD colitis by genome-wide gene expression profiling. *Dis Markers* 2011; 30:1–17.

33   Wang XT, Nagaba Y, Cross HS, Wrba F, Zhang L, Guggino SE. The mRNA of L-type calcium channel elevated in colon cancer: protein distribution in normal and cancerous colon. *Am J Pathol* 2000; 157:1549–1562.

34   Wissenbach U, Niemeyer B, Himmerkus N, Fixemer T, Bonkhoff H, Flockerzi V. TRPV6 and prostate cancer: cancer growth beyond the prostate correlates with increased TRPV6 Ca^{2+} channel expression. *Biochem Biophys Res Commun* 2004; 322:1359–1363.

35   Natrajan R, Little SE, Reis-Filho JS, Hing L, Messahel B, Grundy PE, et al. Amplification and overexpression of CACNA1E correlates with relapse in favorable histology Wilms' tumors. *Clin Cancer Res* 2006; 12:7284–7293.

36   Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 2008; 92:265–272.

37   Salzer E, Cagdas D, Hons M, Mace EM, Garncarz W, Petronczki OY, et al. RASGRP1 deficiency causes immunodeficiency with impaired cytoskeletal dynamics. *Nat Immunol* 2016; 17:1352–1360.

38   Begun J, Lassen KG, Jijon HB, Baxt LA, Goel G, Heath RJ, et al. Integrated genomics of Crohn's disease risk variant identifies a role for CLEC12A in antibacterial autophagy. *Cell Rep* 2015; 11:1905–1918.

39   Glas J, Konrad A, Schmechel S, Dambacher J, Seiderer J, Schroff F, et al. The ATG16L1 gene variants rs2241879 and rs2241880 (T300A) are strongly associated with susceptibility to Crohn's disease in the German population. *Am J Gastroenterol* 2008; 103:682–691.

40   Lapaquette P, Brest P, Hofman P, Darfeuille-Michaud A. Etiology of Crohn's disease: many roads lead to autophagy. *J Mol Med (Berl)* 2012; 90:987–996.

41   Koder S, Repnik K, Ferkolj I, Pernat C, Skok P, Weersma RK, et al. Genetic polymorphism in ATG16L1 gene influences the response to adalimumab in Crohn's disease patients. *Pharmacogenomics* 2015; 16:191–204.

42   Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet* 2015; 47:172–179.

43   Brant SR, Okou DT, Simpson CL, Cutler DJ, Haritunians T, Bradfield JP, et al. Genome-wide association study identifies African-specific susceptibility loci in African Americans with inflammatory bowel disease. *Gastroenterology* 2016; 152:206–217.

44   Lee JC, Biasci D, Roberts R, Gearry RB, Mansfield JC, Ahmad T, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet* 2017; 49:262–268.

45   Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr El Yafi FA, Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut* 2001; 49:777–782.

46   Vucelic B. Inflammatory bowel diseases: controversies in the use of diagnostic procedures. *Dig Dis* 2009; 27:269–277.

47   Solberg IC, Vatn MH, Høie O, Stray N, Sauar J, Jahnsen J, et al. Clinical course in Crohn's disease: results of a Norwegian population-based ten-year follow-up study. *Clin Gastroenterol Hepatol* 2007; 5:1430–1438.

48   Gathungu G, Kim MO, Ferguson JP, Sharma Y, Zhang W, Ng SM, et al. Granulocyte-macrophage colony-stimulating factor autoantibodies: a marker of aggressive Crohn's disease. *Inflamm Bowel Dis* 2013; 19:1671–1680.

49   Gearry RB, Kamm MA, Hart AL, Bassett P, Gabe SM, Nightingale JM. Predictors for developing intestinal failure in patients with Crohn's disease. *J Gastroenterol Hepatol* 2013; 28:801–807.

50   Lazarev M, Huang C, Bitton A, Cho JH, Duerr RH, McGovern DP, et al. Relationship between proximal Crohn's disease location and disease behavior and surgery: a cross-sectional study of the IBD Genetics Consortium. *Am J Gastroenterol* 2013; 108:106–112.

51   Thia KT, Sandborn WJ, Harmsen WS, Zinsmeister AR, Loftus EV. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. *Gastroenterology* 2010; 139:1147–1155.

52   Beaugerie L, Seksik P, Nion-Larmurier I, Gendre JP, Cosnes J. Predictors of Crohn's disease. *Gastroenterology* 2006; 130:650–656.

53   Rubin DT, Hanauer SB. Smoking and inflammatory bowel disease. *Eur J Gastroenterol Hepatol* 2000; 12:855–862.

54   Lawrance IC, Murray K, Batman B, Gearry RB, Grafton R, Krishnaprasad K, et al. Crohn's disease and smoking: is it ever too late to quit? *J Crohns Colitis* 2013; 7:e665–e671.

55   Aldhous MC, Drummond HE, Anderson N, Baneshi MR, Smith LA, Arnott ID, et al. Smoking habit and load influence age at diagnosis and disease extent in ulcerative colitis. *Am J Gastroenterol* 2007; 102:589–597.

56   Henriksen M, Jahnsen J, Lygren I, Vatn MH, Moum B, Group IS. Are there any differences in phenotype or disease course between familial and sporadic cases of inflammatory bowel disease? Results of a population-based follow-up study. *Am J Gastroenterol* 2007; 102:1955–1963.

57   Magro F, Rodrigues-Pinto E, Coelho R, Andrade P, Santos-Antunes J, Lopes S, et al. Is it possible to change phenotype progression in

Crohn's disease in the era of immunomodulators? Predictive factors of phenotype progression. *Am J Gastroenterol* 2014; 109:1026–1036.

58   Peyrin-Biroulet L, Harmsen WS, Tremaine WJ, Zinsmeister AR, Sandborn WJ, Loftus EV. Surgery in a population-based cohort of Crohn's disease from Olmsted County, Minnesota (1970–2004). *Am J Gastroenterol* 2012; 107:1693–1701.

59   Lakatos PL, Golovics PA, David G, Pandur T, Erdelyi Z, Horvath A, *et al*. Has there been a change in the natural history of Crohn's disease? Surgical rates and medical management in a population-based inception cohort from Western Hungary between 1977–2009. *Am J Gastroenterol* 2012; 107:579–588.

60   D'Haens G, Baert F, van AG, Caenepeel P, Vergauwe P, Tuynman H, *et al*. Early combined immunosuppression or conventional management

in patients with newly diagnosed Crohn's disease: an open randomised trial. *Lancet* 2008; 371:660–667.

61   Schreiber S, Reinisch W, Colombel JF, Sandborn WJ, Hommes DW, Robinson AM, *et al*. Subgroup analysis of the placebo-controlled CHARM trial: increased remission rates through 3 years for adalimumab-treated patients with early Crohn's disease. *J Crohns Colitis* 2013; 7:213–221.

62   Colombel JF, Sandborn WJ, Reinisch W, Mantzaris GJ, Kornbluth A, Rachmilewitz D, *et al*. Infliximab, azathioprine, or combination therapy for Crohn's disease. *N Engl J Med* 2010; 362:1383–1395.

63   Khanna R, Bressler B, Levesque BG, Zou G, Stitt LW, Greenberg GR, *et al*. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* 2015; 386:1825–1834.