

# Functional rare variants influence the clinical response to anti-TNF therapy in Crohn's disease

María Chaparro\*<sup>#</sup> , Adrià Aterido<sup>#</sup>, Iván Guerra, Marisa Iborra\*, Jose Luis Cabriada, Luis Bujanda\*, Carlos Taxonera , Valle García-Sánchez\*, Ignacio Marín-Jiménez, Manuel Barreiro-de Acosta, Isabel Vera, Maria Dolores Martín-Arranz, Borja Hernández-Breijo, Francisco Mesonero, Laura Sempere, Fernando Gomollón\*, Joaquín Hinojosa, Fernando Bermejo, Belén Beltrán\*, Ainhoa Rodríguez-Pescador, Jesús María Banales\*, David Olivares, Patricia Aguilar-Melero, Luis Menchén, Rocío Ferreira-Iglesias, Isabel Blazquez Gómez, Beatriz Benitez García, Luis G Guijarro, Alicia C Marin\*, David Bernardo\*, Sara Marsal, Antonio Julia\*\* and Javier P Gisbert\*<sup>\*\*\*</sup>  on behalf of the PREDICROHN Study Group from GETECCU

## Abstract

**Background:** The effect of low-frequency functional variation on anti-tumor necrosis factor alpha (TNF) response in Crohn's disease (CD) patients remains unexplored. The objective of this study was to investigate the impact of functional rare variants in clinical response to anti-TNF therapy in CD.

**Methods:** CD anti-TNF naïve patients starting anti-TNF treatment due to active disease [Crohn's Disease Activity Index (CDAI > 150)] were included. The whole genome was sequenced using the Illumina HiSeq4000 platform. Clinical response was defined as a CDAI score <150 at week 14 of anti-TNF treatment. Low-frequency variants were annotated and classified according to their damaging potential. The whole genome of CD patients was screened to identify homozygous loss-of-function (LoF) variants. The TNF signaling pathway was tested for overabundance of damaging variants using the SKAT-O method. Functional implication of the associated rare variation was evaluated using cell-type epigenetic enrichment analyses.

**Results:** A total of 41 consecutive CD patients were included; 3250 functional rare variants were identified (2682 damaging and 568 LoF variants). Two homozygous LoF mutations were found in *HLA-B* and *HLA-DRB1* genes associated with lack of response and remission, respectively. Genome-wide LoF variants were enriched in epigenetic marks specific for the gastrointestinal tissue (colon,  $p = 4.11e-4$ ; duodenum,  $p = 0.011$ ). The burden of damaging variation in the TNF signaling pathway was associated with response to anti-TNF therapy ( $p = 0.016$ ); damaging variants were enriched in epigenetic marks from CD8<sup>+</sup> ( $p = 6.01e-4$ ) and CD4<sup>+</sup> ( $p = 0.032$ ) T cells.

**Conclusions:** Functional rare variants are involved in the response to anti-TNF therapy in CD. Cell-type enrichment analysis suggests that the gut mucosa and CD8<sup>+</sup> T cells are the main mediators of this response.

**Keywords:** adalimumab, Crohn's disease, infliximab, genes, tumor necrosis factor alpha, whole-genome analysis

Received: 10 April 2019; revised manuscript accepted: 28 June 2019.

*Ther Adv Gastroenterol*

2019, Vol. 12: 1–14

DOI: 10.1177/  
1756284819867848

© The Author(s), 2019.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

Correspondence to:

**María Chaparro**  
Gastroenterology Unit,  
Hospital Universitario de  
La Princesa, Instituto de  
Investigación Sanitaria  
Princesa (IIS-IP),  
Universidad Autónoma de  
Madrid, 62. 28006 Madrid,  
Spain  
mariachs2005@gmail.com

**Adrià Aterido**  
Rheumatology Research  
Group, Vall d'Hebron  
Research Institute,  
Barcelona, Spain

Department of Experimental  
and Health Sciences,  
Universitat Pompeu Fabra,  
Barcelona, Spain

**Iván Guerra**  
**Fernando Bermejo**  
Gastroenterology Unit,  
Hospital Universitario de  
Fuenlabrada, Instituto de  
Investigación de Hospital  
La Paz (IdiPaz), Madrid,  
Spain

**Marisa Iborra**  
**Belén Beltrán**  
Gastroenterology Unit,  
Hospital Universitario de  
La Fe, Valencia, Spain

**Jose Luis Cabriada**  
**Ainhoa Rodríguez-Pescador**  
Gastroenterology Unit,  
Hospital Universitario de  
Galdakano, Vizcaya, Spain

**Luis Bujanda**  
**Jesús María Banales**  
Gastroenterology Unit,  
Hospital Universitario  
de Donostia, Instituto  
Biodonostia, UPV/EHU,  
Ikerbasque, Guipúzcoa,  
Spain

**Carlos Taxonera**  
**David Olivares**  
Gastroenterology Unit,  
Hospital Universitario  
Clínico San Carlos and  
IdISSC, Madrid, Spain

**Valle García-Sánchez**  
**Patricia Aguilar-Melero**  
Gastroenterology Unit,  
Instituto Maimónides de  
Investigación Biomédica  
de Córdoba (IMIBIC),  
Hospital Universitario  
Reina Sofía/Universidad de  
Córdoba, Spain

**Ignacio Marín-Jiménez**  
**Luis Menchén**  
Gastroenterology Unit,  
Hospital Universitario  
Gregorio Marañón e IISGM,  
Madrid, Spain

**Manuel Barreiro-de Acosta**  
**Rocío Ferreiro-Iglesias**  
Gastroenterology Unit,  
Hospital Universitario  
Clínico de Santiago,  
Santiago de Compostela,  
Spain

**Isabel Vera**  
**Isabel Blazquez Gómez**  
Gastroenterology Unit,  
Hospital Universitario  
Puerta de Hierro  
Majadahonda, Madrid, Spain

**Maria Dolores Martín-Arranz**  
**Beatriz Benitez García**  
Gastroenterology Unit,  
Hospital Universitario La  
Paz, Madrid, Spain

**Borja Hernández-Breijo**  
Universidad de Alcalá,  
Alcalá de Henares, Spain  
Immuno-Rheumatology  
Research Group. IdiPaz.  
Hospital Universitario La  
Paz, Madrid, Spain

**Francisco Mesonero**  
Gastroenterology Unit,  
Hospital Universitario  
Ramón y Cajal, Madrid,  
Spain

**Laura Sempere**  
Gastroenterology Unit,  
Hospital Universitario  
Alicante, Alicante, Spain

**Fernando Gomollón**  
Gastroenterology  
Unit, Hospital Clínico  
Universitario, Lozano Blesa,  
IIS Aragón, Zaragoza, Spain

**Joaquín Hinojosa**  
Gastroenterology Unit,  
Hospital Universitario  
Manises, Valencia, Spain

**Luis G Guijarro**  
Universidad de Alcalá,  
Alcalá de Henares, Spain

**Alicia C Marín**  
**David Bernardo**  
**Javier P Gisbert**  
Gastroenterology Unit,  
Hospital Universitario de  
La Princesa, Instituto de  
Investigación Sanitaria  
Princesa (IIS-IP),  
Universidad Autónoma de  
Madrid, Spain

**Sara Marsat**  
**Antonio Julia**  
Rheumatology Research  
Group, Vall d'Hebron  
Research Institute,  
Barcelona, Spain

\*Centro de Investigación  
Biomédica en Red de  
Enfermedades Hepáticas y  
Digestivas (CIBEREHD).

#Equally contributed as  
First authors.

\*\*Equally contributed as  
Senior authors

## Introduction

Anti-tumor necrosis factor alpha (TNF) drugs have been shown to be effective for induction and maintenance of remission in Crohn's disease (CD); however, up to 30% of patients do not respond to these drugs.<sup>1-3</sup> In 2014, the use of vedolizumab, which targets  $\alpha 4\beta 7$  integrins, was approved for CD. In 2016, ustekinumab, directed against the p40 subunit shared by interleukins 12 and 23, was approved in patients with CD, while the JAK-inhibitor tofacitinib has just been approved by the European Agency of Medicines for ulcerative colitis (UC) patients, hence increasing the therapeutic arsenal against inflammatory bowel diseases (IBD). All these drugs have high cost, so they pose a great economic burden on health systems. At present, the medical community does not have reliable criteria for selecting which patient will benefit from any of the above-mentioned drugs. Thus, the variables (epidemiological, clinical, analytical, etc.) commonly used to predict patient response to biologic therapy have shown low utility.

Therefore, it is a priority to identify mechanisms involved in the response to each drug, to suggest more rational use of resources and for personalized medicine. Thus, achieving that goal will allow us to indicate the most appropriate treatment to each individual, hence avoiding administering drugs to patients who will not respond (an inadequate use of resources and an unjustified risk of adverse effects).

There is compelling evidence that the clinical response to anti-TNF therapy has a genetic component.<sup>4</sup> To date, several candidate-gene studies have investigated the contribution of common variation to the heterogeneity of anti-TNF response in CD.<sup>5-9</sup> However, the few associations identified so far have not been consistently replicated.<sup>10</sup> The study of the genetic basis of the clinical response to anti-TNF therapy at the genome-wide scale represents a new opportunity to identify genetic factors underlying the anti-TNF response in CD.

The advent of next generation sequencing (NGS) technologies like whole genome sequencing (WGS) is enabling the analysis and identification of low-frequency variation associated with complex traits,<sup>11,12</sup> including the response to anti-TNF agents.<sup>13</sup> In particular, rare variants with a high functional impact can display much larger effects

on complex phenotypes than common genetic variants.<sup>14</sup> From a clinical perspective, loss-of-function (LoF) variants are one of the most interesting forms of rare functional genetic variation.<sup>15</sup> Accordingly, in the pharmacogenomics field, the analysis of LoF variants is starting to provide new insights into the genetic basis underlying the clinical response to therapies in complex diseases. The analysis of low-frequency variation could therefore shed new light on the biological mechanisms that are relevant for anti-TNF response in CD.

In the present study, we sequenced the whole genome of CD patients treated with anti-TNF therapy. At the genome-wide level, we identified LoF variants affecting critical genes for the immune response. Specific analysis of the TNF signaling pathway has also shown the significant burden of damaging variation associated with treatment response. These findings support the importance of functional rare variation for the anti-TNF response in CD, and could help to select those patients who will benefit most from anti-TNF treatment.

## Methods

### Patients

A total of 41 patients diagnosed with CD based on European Crohn's and Colitis Organization criteria that started anti-TNF treatment were included.<sup>16</sup> Patients came from the Predicrohn study, which is a multicenter study aiming to optimize treatment with anti-TNF drugs.<sup>17</sup> The trial included 150 CD patients naïve to anti-TNF treatment; only patients with active disease, based on the Crohn's Disease Activity Index (CDAI) at baseline, were eligible for the genetic approach. All patients were naïve to anti-TNF therapy and received 160/80 mg adalimumab at weeks 0 and 2, and 40 mg every other week thereafter, or infliximab 5 mg/kg at weeks 0, 2, 6, and 14, and every 8 weeks thereafter.

### Study design

We conducted a prospective, multicenter cohort study. Treatment (infliximab or adalimumab) was not assigned randomly but based on the clinician's decision. Clinical response was evaluated at week 14 by measuring the change in the CDAI score. Remission was defined as CDAI  $\leq 150$  at week 14, and short-term partial

**Table 1.** Characteristics of the study population.

| Characteristics                    | Responders (n = 31) | Nonresponders (n = 10) | p-value |
|------------------------------------|---------------------|------------------------|---------|
| Men (%)                            | 13 (41.93)          | 8 (80.00)              | 0.066   |
| Location (%)                       |                     |                        |         |
| Ileal                              | 9 (29.03)           | 5 (50.00)              | 0.27    |
| Colonic                            | 5 (16.13)           | 0 (0.00)               | 0.31    |
| Ileocolonic                        | 15 (48.39)          | 3 (30.00)              | 0.47    |
| Behavior (%)                       |                     |                        |         |
| Inflammatory                       | 20 (64.52)          | 2 (20.00)              | 0.026   |
| Strictureing                       | 4 (12.90)           | 2 (20.00)              | 0.99    |
| Fistulizing                        | 5 (16.13)           | 5 (50.00)              | 0.044   |
| Perianal disease (%)               | 7 (22.58)           | 1 (10.00)              | 0.65    |
| Extraintestinal manifestations (%) | 12 (38.71)          | 2 (20.00)              | 0.45    |
| Smoking habit (%)                  | 13 (41.94)          | 3 (30.00)              | 0.71    |
| Steroids                           | 5 (16.13)           | 1 (10.00)              | 0.98    |
| Immunomodulators (%)               |                     |                        |         |
| Thiopurines                        | 21 (67.74)          | 8 (80.00)              | 0.69    |
| Methotrexate                       | 2 (6.45)            | 0 (0.00)               | 0.99    |
| Anti-TNF type (%)                  |                     |                        |         |
| Adalimumab                         | 13 (41.94)          | 3 (30.00)              | 0.71    |
| Infliximab                         | 18 (58.06)          | 7 (70.00)              | 0.71    |
| Previous Surgery                   | 11 (35.48)          | 6 (60.00)              | 0.27    |

response as a decrease in the CDAI score of at least 70 points at week 14, without achieving remission. Patients that achieved remission, and those with partial response, were considered responders. Patients not meeting those criteria were considered nonresponders. Concomitant treatments were also recorded. Blood samples were obtained at baseline. The distribution of the categorical variables between responders and nonresponders to anti-TNF therapy was compared using the Fisher's exact test (Table 1).

The study was approved by the ethical review board of all participating centers. All patients signed an informed written consent to participate.

#### *WGS of the study population*

WGS of the 41 CD patients included in the present study was performed using the Illumina HiSeq4000 platform at Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain). All DNA samples were sequenced to an average coverage  $>20\times$  using 150 bp paired-end reads. For each sample, we obtained  $>360$  million reads. Duplicate reads were subsequently excluded from the analysis. The remaining reads were aligned to the GRCh38/hg38 assembly of the human reference genome using the Burrows Wheeler algorithm. Variant calling was performed according to the best practices outlined in the Genome Analysis ToolKit (GATK).<sup>18</sup>

In the quality control (QC) analysis, genetic variants with a genotype quality  $\geq 50$ , a read depth  $\geq 20$  and a read position rank sum  $\geq -8$  were selected for the present study. From these, only genetic variants that showed a coverage depth  $\geq 2$  for the alternative allele were included ( $n = 3,564,615$  genetic variants). In order to discard genetic variants that showed a different genotype between the forward and reverse strands, genetic variation showing a Fisher's strand bias  $\geq 60$  and a strand odds ratio  $\geq 3$  were excluded from the study. In addition, in order to keep only those genetic variants that were confidently mapped, we selected only genetic variation that showed a mapping quality score  $\geq 40$  and a mapping quality rank sum  $\geq -10$ . After QC, only variants showing a minor allele frequency  $< 0.05$  were included in downstream analyses.

In order to discard any confounding effect from population stratification in the CD patient cohort, we performed principal component analysis (PCA) incorporating genome-wide association study (GWAS) data from a cohort of 1558 healthy controls from the Spanish population. These individuals were collected and genotyped by the Immune-Mediated Inflammatory Disease (IMID) Consortium, as described previously.<sup>19</sup> After selecting the single nucleotide polymorphisms (SNPs) that were in common between both samples, we evaluated the existence of population stratification using the EIGENSOFT (v4.2) software.<sup>20</sup> Using the first 10 principal components (PCs) of variation over 10 iterations, we did not detect any CD patient showing an outlier genetic background (Figure S1).

#### *Annotation of genetic variation*

Low-frequency variants were annotated using the SnpEff software,<sup>21</sup> and their damaging potential was predicted using PolyPhen-2 and Sift.<sup>22,23</sup> In PolyPhen-2, the functional significance of allele replacement is predicted by a Naïve Bayes classifier approach that also estimates the probability of each variant to be classified as damaging when it is nondamaging [false positive rate (FPR)]. In the present study, low-frequency variants were defined as damaging if classified as possibly or probably damaging by both PolyPhen-2 (FPR  $< 10\%$  and FPR  $< 20\%$  for probably and possibly damaging, respectively) and Sift, as previously described.<sup>13</sup> Based on the annotation displayed in

the Exome Aggregation Consortium (ExAC) Database,<sup>24</sup> damaging variants were characterized according to their impact using the Sequence Ontology tool.<sup>25</sup> Briefly, damaging variants were classified into: start lost: codon variants that change at least one base of the start codon; stop gained: sequence variants that change at least one base of a codon leading to a premature stop codon and shortening of the protein; stop lost: sequence variants that change at least one base of the terminator codon resulting in an elongated transcript; synonymous: sequence variants that do not alter the amino acid sequence encoded; missense: sequence variants that change one or more bases resulting in a different amino acid sequence without altering the length of the protein sequence; splice donor: splice variants that change the two base pair region at the 5' end of an intron; splice region: sequence variants that change the region of the splice site either within 1–3 bases of the exon or 3–8 bases of the intron; 5' UTR: variants mapping on the 5' UTR region; 3' UTR: variants mapping on the 3' UTR region; structural interaction: variants that impact the internal interactions of the resulting polypeptide structure; lincRNA: variants mapping on a long intergenic noncoding RNA region; snRNA: variants mapping on a small nuclear RNA region; sense intronic: variants mapping on a noncoding transcript region located within an intronic region with no overlap with exonic sequences; nonstop decay: variants that might impact the translation of mRNA molecules lacking a stop codon; and nonsense-mediated decay: variants that might impact the mRNA degradation.

According to the methodology implemented in SnpEff, genetic variants that correlate with the complete loss of function of the encoded transcript were classified as LoF. These genetic variants include: stop codon-introducing variants, start codon-introducing variants, splice site-disrupting variants, insertion/deletion variants predicted to disrupt the transcript reading frame, and large deletions that remove either the first exon of the gene or more than 50% of the protein-coding sequence of the affected transcript.<sup>21</sup> Given that stop codon-introducing and frameshift variants are enriched toward the 3' end of the gene and tend to show tolerance to truncation close to the 5' end, putative LoF variants identified in the last 5% of the coding region were not included in the present study.

For each variant predicted as damaging or LoF, we also estimated its putative impact at the protein level using the method implemented in SnpEff.<sup>26</sup> According to this method, three categories were defined: high impact, the variant is assumed to have a disruptive impact in the protein, probably causing protein truncation, loss of function or triggering nonsense mediated decay; moderate impact, the variant is nondisruptive and might change the protein effectiveness; and low impact, the variant is assumed to be mostly harmless or unlikely to change the protein behavior.

#### *SKAT-O rare variant burden test*

The contribution of low-frequency damaging variants from the TNF signaling pathway in the clinical response to anti-TNF therapy was investigated using the SKAT-O set-based collapsing method implemented in EPACTS software.<sup>27</sup> This statistical approach was designed to efficiently unify both the burden and nonburden algorithms for rare variant analysis. While the former method collapses genetic variation into a single burden variable, the latter aggregates the individual variant-score test statistics resulting from the association analysis into a kernel matrix. With this algorithm, the SKAT-O test has been found to maximize the statistical power of the analysis. We used the SKAT-O test to compare the global burden of low-frequency variation at the TNF signaling pathway between responders and nonresponders to anti-TNF therapy. This statistical analysis was adjusted for the clinical variables that showed a significantly different distribution between responder and nonresponder patients (Table 1). The annotation for the TNF signaling pathway was obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) reference database (hsa04668,  $n = 108$  genes).<sup>28</sup>

#### *Epigenetic enrichment analysis*

In order to identify the relevant cell types for anti-TNF response, we performed an epigenetic enrichment analysis. First, we analyzed the enrichment of genome-wide LoF variation ( $n = 568$  SNPs) on cell-type-specific H3K4me3 chromatin marks using the EPIGWAS software.<sup>29</sup> Second, we followed the same procedure to analyze the enrichment of damaging variation at the TNF signaling pathway on cell-type-specific H3K4me3 chromatin marks. Cell type-specific

epigenetic data from H3K4me3 was obtained from the Roadmap Epigenomics project.<sup>30</sup>

The statistical methodology implemented in the EPIGWAS software consists of five consecutive steps. First, the LD between the associated variants and those located within a distance window of 500 kb is computed to define the loci to be analyzed. Second, the regulatory activity of each variant in each cell type is scored as the height of the nearest H3K4me3 peak, divided by the distance between the SNP and the summit of the peak. For each cell type and locus, the value of the variant with the highest score is selected as a measure of the overlap between the locus and the cell type-specific regulation. Third, these scores are normalized using Euclidean methodology. Fourth, the scores from all loci are summed within each cell type in order to compute the cell type regulatory score. In the last step, the statistical significance of cell-type-specific overlap was obtained using a permutation-based approach ( $n = 10,000$  permutations). In each permutation, a matched set of SNPs is selected from a set of independent SNPs from the HapMap Project and the cell type specific regulatory score is computed again for each cell type. The statistical significance of the enrichment analysis is finally defined as the proportion of sets of SNPs matched with a cell type specific regulatory score exceeding the observed score for the associated variants of interest.

## Results

#### *Study population*

A total of 41 CD patients were included. The main characteristics of the study cohort are summarized in Table 1. Mean CDAI was 256 (standard deviation = 57). Most of the patients had ileocolonic location and inflammatory behavior, and over 40% had been operated on due to CD. Approximately 40% received adalimumab and 60% infliximab. At week 14, 25 patients (61%) achieved remission and 6 had partial response (14.6%), and therefore considered responders, and 10 (24.4%) were primary nonresponders. When comparing the distribution of the different epidemiological and clinical variables between responders and nonresponders to anti-TNF therapy, we detected significant differences only in the disease behavior ( $p < 0.05$ , Table 1). In



**Table 2.** Characterization of the damaging low-frequency genetic variation identified in the study population.

| Variant annotation            | Variants* |
|-------------------------------|-----------|
| Start lost                    | 8         |
| Stop gained                   | 2         |
| Stop lost                     | 2         |
| Synonymous                    | 30        |
| Missense                      | 2590      |
| Splice donor                  | 2         |
| Splice region                 | 89        |
| Structural interaction        | 50        |
| Long intergenic noncoding RNA | 56        |
| Sense intronic                | 17        |
| Small nuclear RNA             | 12        |
| 3' UTR                        | 479       |
| 5' UTR                        | 86        |
| Nonsense-mediated decay       | 1         |
| Nonstop decay                 | 9         |

\*Number of low-frequency damaging variants showing the indicated annotation. Given that a single genetic variant can impact different genes, the annotated variants outnumber those predicted as damaging.

addition, there was no difference in CDAI score at baseline between responders and nonresponders (259 versus 246,  $p = 0.5$ ).

#### Classification of low-frequency damaging variants

A total of 14,994,115 genetic variants were identified by WGS, of which 1,738,956 were found to be low-frequency variants. Using PolyPhen-2 and Sift, 2682 low-frequency genetic variants were predicted to be damaging. In order to characterize the effect of the damaging variants, we classified them according to their damaging potential (Table 2). Overall, we found that 1.9% and 98.1% of the low-frequency damaging variants have a high and moderate/low impact at the protein level, respectively.

#### Genome-wide LoF variants associated with anti-TNF therapy response

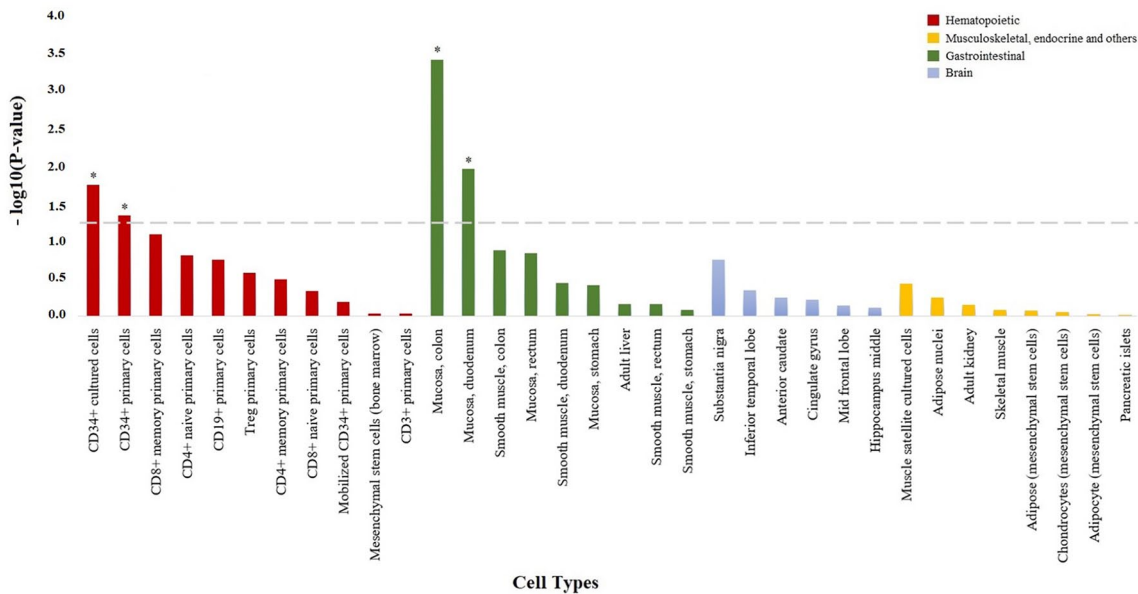
We detected 568 low-frequency LoF variants associated with response to anti-TNF therapy, including 281 and 386 variants in adalimumab- and infliximab-treated patients, respectively ( $n = 99$  overlapping variants). A total of 275 and 384 low-frequency LoF variants were detected in patients with and without prior surgery, respectively ( $n = 92$  overlapping variants). They were all heterozygous LoF except for two variants that had homozygous LoF carriers. These variants mapped to the *HLA-B* (SNP rs41563412) and *HLA-DRB1* (SNP rs1071752) genes. The unique homozygous carrier of the damaging *HLA-B* rs41563412-GCA variant (frameshift mutation, p.Ser33LeufsTer9) did not respond to anti-TNF therapy (i.e. adalimumab) and was not involved in a surgical intervention. Conversely, the three patients that carried one or two copies of the damaging *HLA-DRB1* rs1071752-C variant (synonymous mutation, p.Gln125Gln) were all responders to anti-TNF therapy (i.e. infliximab), and only one of them was not involved in a surgical intervention. Both SNPs were found to be protein-coding variants with a high functional impact at the protein level. The strongest damaging impact was detected in 10 LoF SNPs mapping at *NECAP2*, *HEYL*, *THNSL2*, *NPHP3-ACAD11*, *CLDN18*, *LY6G6C*, *C6orf226*, *RAB2IL1* and *EFCAB5* genes (Table 3). The SNP chr2:88173152 at *THNSL2* gene was not annotated in the ExAC or dbSNP databases,<sup>31</sup> suggesting this could be a new genetic variant.

Using cell-type epigenetic enrichment analysis, we found that the genome-wide LoF variants associated with anti-TNF response were significantly enriched for epigenetic marks characteristic of the colonic ( $p = 4.11e-4$ ) and duodenum mucosa ( $p = 0.011$ , Figure 1). In patients with prior surgery, the genome-wide LoF variants associated with the overall anti-TNF response were significantly enriched in epigenetic marks from the CD34+ hematopoietic precursor cells ( $p = 0.021$ ), whereas a significant enrichment in epigenetic marks from the colonic ( $p = 4.10e-4$ ) and duodenum mucosa ( $p = 7.61e-3$ ) was detected in patients without prior surgery. When performing the cell-type epigenetic enrichment analysis stratified by drug, the infliximab-associated LoF variants showed also a significant enrichment in epigenetic marks from the colonic mucosa ( $p = 6.01e-4$ ) as

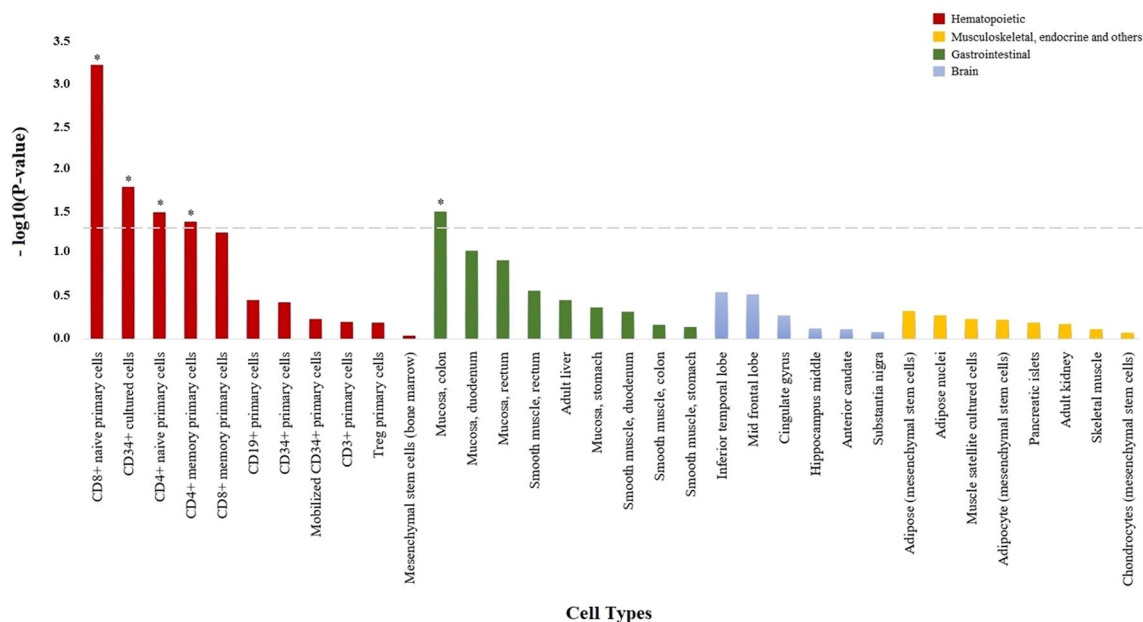
**Table 3.** Characterization of the damaging loss-of-function variants identified in the study population.

| SNP           | Chr | Coord     | A1 | A2 | Gene                | Variant type | Annotation   | CD (%)  | Resp | Nresp |
|---------------|-----|-----------|----|----|---------------------|--------------|--------------|---------|------|-------|
| rs150581659   | 1   | 16447908  | G  | A  | <i>NECAP2</i>       | exonic       | stop gained  | 1 (2.4) | 1    | 0     |
| rs761330653   | 1   | 39639623  | C  | T  | <i>HEYL</i>         | exonic       | start lost   | 1 (2.4) | 1    | 0     |
| rs764641613   | 1   | 39639624  | A  | C  | <i>HEYL</i>         | exonic       | start lost   | 1 (2.4) | 1    | 0     |
| chr2:88173152 | 2   | 88173152  | T  | C  | <i>THNSL2</i>       | exonic       | start lost   | 1 (2.4) | 0    | 1     |
| rs41272317    | 3   | 132618633 | C  | A  | <i>NPHP3-ACAD11</i> | intronic     | splice donor | 1 (2.4) | 0    | 1     |
| rs138856042   | 3   | 137998869 | A  | G  | <i>CLDN18</i>       | exonic       | start lost   | 1 (2.4) | 1    | 0     |
| rs151314696   | 6   | 31719303  | C  | T  | <i>LY6G6C</i>       | exonic       | start lost   | 1 (2.4) | 1    | 0     |
| rs61732354    | 6   | 42890786  | C  | T  | <i>C6orf226</i>     | exonic       | start lost   | 2 (4.8) | 1    | 1     |
| rs192561318   | 11  | 61920191  | A  | G  | <i>RAB3IL1</i>      | intronic     | start lost   | 1 (2.4) | 1    | 0     |
| rs79556405    | 17  | 29941799  | G  | A  | <i>EFCAB5</i>       | intronic     | start lost   | 4 (9.7) | 2    | 2     |

A1, reference allele; A2, alternative allele that causes the LoF of the protein encoded by the indicated gene; CD, number of CD patients carrying the LoF variant; Chr, chromosome; Coord, SNP coordinates in GRCh38/hg38; Nresp, number of patients that carry the LoF variant and did not show a significant clinical response to anti-TNF therapy; Resp, number of patients that carry the LoF variant and responded to anti-TNF therapy; SNP, single nucleotide polymorphism.



**Figure 1.** Enrichment of genome-wide loss-of-function variation in H3K4me3 histone marks from specific cell types. Bar plot showing the statistical significance of the overlap between genome-wide loss-of-function variation and H3K4me3 histone marks from the 34 cell types analyzed. All cell types are shown in the x axis, and those with H3K4me3 histone marks significantly enriched in genome-wide loss-of-function variation are represented with an asterisk ( $p < 0.05$ ).



**Figure 2.** Enrichment of damaging variation at TNF signaling pathway in H3K4me3 histone marks from specific cell types. Bar plot showing the statistical significance of the overlap between damaging variation from the TNF signaling pathway and H3K4me3 histone marks from the 34 cell types analyzed. All cell types are shown in the x axis, and those with H3K4me3 histone marks significantly enriched in damaging variation from the TNF signaling pathway are represented with an asterisk ( $p < 0.05$ ). TNF, tumor necrosis factor alpha.

well as from CD34+ hematopoietic precursor cells ( $p = 0.012$ ). The adalimumab-associated LoF variants showed a significant enrichment in epigenetic marks from the colonic mucosa ( $p = 2.81e-3$ ), colonic smooth muscle ( $p = 0.02$ ), CD4+ T cells ( $p = 0.034$ ) and CD34+ hematopoietic precursor cells ( $p = 0.043$ ).

#### Association of damaging variation at the TNF signaling pathway with anti-TNF response

A total of 19 damaging low-frequency genetic variants were found to map to the TNF signaling pathway (hsa04668). The analysis of this variation using the SKAT-O test showed that the damaging burden is different between anti-TNF responder (partial response or remission) and nonresponder patients ( $p = 0.016$ ). The cell-type enrichment analysis showed an increase in epigenetic marks specific from immune cell types (Figure 2), principally CD8+ T cells ( $p = 6.01e-4$ ) but also CD4+ T cells ( $p = 0.032$ ). Like in the functional characterization of the LoF variants, epigenetic enrichment analysis also indicated an association with gut tissue (colonic mucosa,  $p = 0.031$ ).

The analysis of low-frequency genetic variants mapping to the TNF signaling pathway using the SKAT-O test showed that the damaging burden is different between anti-TNF responder (partial response or remission) and nonresponder patients ( $p = 0.016$ ). When stratifying the analysis by anti-TNF agent, we detected that the burden of damaging variation is significantly different between responders and nonresponders to infliximab therapy ( $p = 9.76e-3$ ). No significant differences were detected among adalimumab-treated patients ( $p > 0.05$ ). The damaging burden was found to be also significantly different between responders and nonresponders to anti-TNF therapy in patients without prior surgery ( $p = 0.011$ ), but not among those patients that had not been previously involved in a surgical intervention ( $p > 0.05$ ).

#### Discussion

In this study, we sequenced the whole genome of CD patients to examine the hypothesis that functional rare variants are associated with anti-TNF response. Using genome-wide and set-based analyses, we found evidence of LoF and damaging



variants influencing the anti-TNF response. Cell type enrichment analysis suggests that these variants affect the cytotoxic response of the adaptive immune system (CD8<sup>+</sup> naïve T-cells) and the gastrointestinal mucosa.

A major finding from this study is that, among the millions of variants of the human genome, the only two homozygous LoF SNPs were found to map to *HLA* genes. These two variants, rs1071752 and rs41563412, are located in an exonic region of the *HLA-DRB1* gene and an intronic region of the *HLA-B* gene, respectively. The proteins encoded by the *HLA-DRB1* and *HLA-B* genes play a central role in the adaptive immune system by presenting antigens to T-cells, including CD4<sup>+</sup> [in a major histocompatibility complex class II (MHC)-II dependent manner as is the case of HLA-DRB1] and CD8<sup>+</sup> T-cells (in an MHC-I dependent manner as the HLA-B), being the adaptive immune response altered in CD patients.<sup>32–34</sup> Although it has been shown that genetic factors underlying clinical phenotypes of CD (i.e. other than anti-TNF response) might not necessarily be shared with CD risk loci, the HLA region itself is indeed one of the strongest risk loci for CD, and multiple genetic association studies have previously identified variation at *HLA-DRB1* and *HLA-B* genes showing a significant association with CD risk.<sup>35–38</sup> To our knowledge, genetic variation at these two genes has not been previously associated with anti-TNF response in CD. Our results support that genetic variation at the *HLA-B* and *HLA-DRB1* disease susceptibility loci could also modulate anti-TNF efficacy, and, therefore, these two genes could be two promising biomarkers for future tailored TNF-blocking therapy in CD patients.

In the genome-wide screening for LoF variants, we have also identified functionally-relevant variants at nine genes, including *NECAP2*, *HEYL*, *THNSL2*, *NPHP3-ACAD11*, *CLDN18*, *LY6G6C*, *C6orf226*, *RAB3IL1*, and *EFCAB5*. *CLDN18* encodes a tight junction protein from the claudin family that regulates the integrity of the intestinal epithelium.<sup>39</sup> The transcriptomic profile of claudins has been found to depend on disease activity, and to vary along the intestinal mucosa.<sup>40</sup> *CLDN18* has been recently associated with the pathogenesis of IBD,<sup>35</sup> where, by modulating the permeability of the epithelial barrier, it may allow the lamina propria immune system to encounter luminal antigens that otherwise would

not be found. Of relevance, the *LY6G6C* gene maps to a cluster of leukocyte antigen genes located in the MHC-III.<sup>41</sup> The MHC-III locus also contains the *TNF* gene, indicating that this genomic region is likely to be important in regulating the inflammatory activity of immune-mediated diseases.<sup>42</sup> Given the physical proximity between the *LY6G6C* and *TNF* genes (i.e. ~140 kb), our results suggest that variation at *LY6G6C* could alter the expression of the *TNF* gene, and, consequently, influence the anti-TNF response.

We also found that the LoF variant at *THNSL2* gene (i.e. chr2:88173152 SNP) is not annotated in the databases of reference for genetic variation.<sup>31</sup> The *THNSL2* gene encodes a threonine synthase-like protein. To date, however, its biological function in humans is still unknown. The scarce information on the biology of the *THNSL2* gene derives from an *in vitro* study showing that *THNSL2* is transcribed into a rare mRNA splice-variant that leads to SOAFT formation, a human T cell secreted cytokine that exacerbates the inflammatory state of autoimmune diseases.<sup>43</sup> In our study cohort, the nonresponse phenotype of the patient carrying the chr2:88173152 SNP suggests that this variant could have a functional effect on the anti-TNF mechanism of action that compromises the treatment response. Future targeted sequencing studies on the *THNSL2* gene are warranted to confirm the involvement of the chr2:88173152 SNP TNF signaling in CD.

Epigenetic mark enrichment analysis of the LoF variants indicates that the gut mucosa is an important mediator of anti-TNF response in CD. Previous genetic studies have shown the importance of the colonic mucosa in the regulation of the inflammatory processes that characterize the pathogenesis of CD.<sup>4</sup> The gastrointestinal mucosa of CD patients is highly enriched in activated TNF-secreting macrophages.<sup>44–46</sup> From a pharmacological perspective, the anti-TNF therapy has been shown to inhibit not only the gut inflammation but also the inflammation-induced angiogenesis in the intestinal mucosa.<sup>47,48</sup> Therefore, our findings are consistent with the critical role of the intestinal mucosa as the location where the immune response is initiated in CD, and, therefore, as the key organ determining the clinical response to anti-TNF therapy.

In this study, we provide the first evidence of association between rare damaging variation in the

TNF signaling pathway and anti-TNF response. A recent sequencing study in rheumatoid arthritis found that rare variants in this pathway do not contribute to treatment response.<sup>13</sup> However, in this latter study, only exomes were sequenced, and, thereby, important functional variation from other genomic regions was missed. Additionally, there is clear evidence that rheumatoid arthritis and CD have very little genetic risk overlap. This also suggests that different biological mechanisms operate through TNF signaling, and, consequently, drug efficacy is mediated by different groups of genes and associated regulatory variants.

The epigenetic enrichment analysis of the TNF signaling pathway indicated that CD8<sup>+</sup> T-cells are important determinants of the efficacy of anti-TNF agents. The inflamed mucosa of CD patients carries large numbers of immune cells, including CD8<sup>+</sup> T cells,<sup>49</sup> which are the main cell type responsible for destroying the intestinal epithelium, leading to an enhanced exposure to the gut microbiota and the subsequent triggering of chronic inflammatory processes.<sup>50-52</sup> As opposed to CD4<sup>+</sup> T-cells, antigen presentation to CD8<sup>+</sup> T-cells is MHC-I restricted. Indeed, the rs41563412 variant located in an intronic region of the HLA-B gene is included within such an MHC-I family. Besides, the unique homozygous carrier of the damaging *HLA-B* rs41563412-*GCA* variant (frameshift mutation, p.Ser33LeufsTer9) did not respond to anti-TNF therapy, suggesting a role of this LoF SNP in modulating the CD8<sup>+</sup> cytotoxic immune response. The relation between CD8<sup>+</sup> T-cells and the response to anti-TNF therapy has been also investigated in previous experimental studies.<sup>46</sup> Using *in vitro* isolated CD8<sup>+</sup> T cells, the administration of anti-TNF therapy was shown to inhibit the proliferation of this T cell subtype both in the gut mucosa and in blood.<sup>53,54</sup> Importantly, the transcriptomic profile of CD8<sup>+</sup> T cells has been associated with the incidence of being refractory to disease treatment.<sup>55</sup> Notably, in our study, the relationship between CD8<sup>+</sup> T-cells and anti-TNF response was restricted to the naïve, but not the memory, subset of CD8<sup>+</sup> T-cells, suggesting a primary defect on this cell fraction. Our findings are in line with these results, and support the predominant role of CD8<sup>+</sup> T cells in the TNF-dependent inflammatory activity in the gut.

Our study has some limitations. First, the response was categorized based on clinical activity indexes. However, we included patients with high activity

index scores to ensure that the symptoms were due mainly to the activity of the disease. In fact, in clinical practice, short-term response is commonly based on patient symptoms. The sample size of the present study is low if compared with genome-wide studies on common variants. However, compared with GWAS, here, we used WGS to identify all possible rare variants. Subsequently, by focusing our analysis only on rare variants with great functional impact, we increase the power of our study. In this respect, this strategy is similar to that used to identify variations associated with rare Mendelian diseases, where a small number of patients have almost exclusive mutations. In this regard, our findings are not trivial at all, as among all genes, the only homozygous LoF homozygosity, a topmost damaging variation, occurred by chance only in HLA genes, which encode key antigen-presenting proteins and are known risk factors in IBD. Finally, given our limited sample size, we restricted rare variation burden analysis specifically to the anti-TNF pathway, thereby preventing a large penalty by testing and analyzing hundreds of other biological pathways.

The analytical strategy implemented in this study is a distinct feature of the present work. In order to identify rare damaging variants underlying the anti-TNF response, we screened the whole genome of anti-TNF-treated CD patients. Compared with array-based genome-wide studies, WGS efficiently captures rare variants that otherwise would have been missed.<sup>56</sup> The analytical approach used for testing association of the TNF signaling pathway is also radically different from single-marker tests. Single-marker tests are underpowered to detect significant associations of low-frequency variants, unless very large sample sizes are used.<sup>57</sup> The set-based collapsing analysis method used here, SKAT-O, groups low-frequency variants into gene sets in order to detect significant overabundance of low-frequency variation associated with complex traits at high statistical power, despite using relatively small cohorts.<sup>27</sup> Given the sample size of our study, we restricted the analysis to the key signaling pathway that is interfered with by anti-TNF drugs. Using this approach, we found a significant burden of rare damaging variants associated with the response to TNF alpha blockers in CD.

In conclusion, we report evidence that functional low-frequency variation is associated with the response to anti-TNF therapy in CD. These findings also provide new insights into the

underlying heterogeneity of CD, revealing the basis of TNF-dependent biological mechanisms. The identification of this diversity at the molecular level will be essential to allow the stratification of patients, and, ultimately, help clinicians choose the most appropriate drug.

### Acknowledgments

María Chaparro and Adria Aterido contributed equally as first authors. Antonio Julia and Javier Gisbert contributed equally as senior authors. The authors thank Dr Simon Heath and Dr Raul Tonda from the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain) for their help in the WGS data processing.

### Author contributions

María Chaparro and Javier P. Gisbert: study design, data collection, data analysis, data interpretation, writing the manuscript.

Antonio Julià and Adrià Aterido: Genetic analysis, data analysis, data interpretation and manuscript writing.

Borja Hernández-Breijo, Luis González Guijarro, Alicia C. Marín and David Bernardo: Sample management.

Remaining authors: Study design, patient inclusion, sample obtaining and data collection.

All authors approved the final version of the manuscript.

María Chaparro and Javier P. Gisbert are the guarantors of the article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by grants from the "Instituto de Salud Carlos III" (FIS.12/02557 and PI13/00041).

### Conflict of interest statement

M Chaparro has served as a speaker, or has received research or education funding from MSD, Abbvie, Hospira, Pfizer, Takeda, Janssen, Ferring, Shire Pharmaceuticals, Dr Falk Pharma, Tillotts Pharma.

I Guerra has served as a speaker for Pfizer and Janssen and consultant and advisory member for Kern Pharma and MSD.

M Iborra has received funding for educational activities, research projects and scientific meetings from Takeda and Abbvie.

JL Cabriada has served as consultant for and received research funding from MSD, Abbvie, Pfizer and Kern Ph.

L Bujanda: none

C Taxonera has served as a speaker, a consultant and advisory member for, or has received research funding from, MSD, Abbvie, Hospira, Pfizer, Takeda, Ferring, Faes Farma, Shire Pharmaceuticals, Dr Falk Pharma, Gebro Pharma.

V García-Sánchez has served as a speaker, a consultant and advisory member for or has received research funding from MSD, Abbvie, Hospira, Kern Pharma, Takeda, Pfizer, Ferring, Faes Farma, Shire Pharmaceuticals, Dr Falk Pharma, Janssen, Gebro Pharma and Otsuka Pharmaceutical.

I Marín-Jiménez has served as a consultant, advisory member, speaker, or has received research funding from Abbvie, Chiesi, Faes Farma, Falk-Pharma, Ferring, Gebro Pharma, Hospira, Janssen, MSD, Otsuka Pharmaceutical, Pfizer, Shire, Takeda, Tillots, and UCB Pharma.

M Barreiro-de Acosta has served as a speaker, a consultant and advisory member for or has received research funding from MSD, Abbvie, Hospira, Pfizer, Kern Pharma, Biogen Takeda, Ferring, Faes Farma, Shire Pharmaceuticals, Dr Falk Pharma, Tillotts Pharma, Chiesi, Gebro Pharma, Otsuka Pharmaceutical and Vifor Pharma.

I Vera has served as a speaker for AbbVie, Takeda and Shire

MD Martín-Arranz has served as a speaker, a consultant and advisory member for MSD; Abbvie, Takeda, Janssen, Ferring, Faes Farma, Shire Pharmaceutical, Tillots Pharma and Chiesi

B.Hernández-Breijo: None

F.Mesonero: None

L Sempere has served as a speaker, a consultant and advisory member for TAKEDA, ABBVIE, KERN, MSD Y TILLOTTS.

F Gomollón has received fees for lectures from Janssen, Abbvie, Takeda, MSD and research grants (Group): MSD, Abbvie. Advisory Committees: None active at present.

J Hinojosa Abbvie, MSD, Takeda, Janssen, Kern, Pfizer-Hospira, Otsuka, Faes, Ferring, Biogen, Shire.

F Bermejo has served as a speaker, a consultant and advisory member for, or has received research funding from, MSD, Abbvie, Pfizer, Hospira, Takeda, Ferring, Faes Farma, Shire, Tillotts, Chiesi and Gebro.

B Beltrán: None

A Rodríguez Pescador

JM Banales: has served as a speaker and advisory member for OWL Metabolomics.

D Olivares: None

P Aguilar-Melero: None

L Menchén has served as a speaker, a consultant and advisory member for or has received research funding from MSD, Abbvie, Hospira, Pfizer, Takeda, Ferring, Faes Farma, Shire Pharmaceuticals, Dr Falk Pharma, Tillotts and General Electric.

R Ferreira-Iglesias has received grants from Abbvie, MSD, Shire, Falk, Tillotts.

I Blazquez Gomez: None

B Benítez García: None

LG Guijarro: None

AC Marin: None


D Bernardo: None

JP Gisbert has served as a speaker, a consultant and advisory member for, or has received research funding from, MSD, Abbvie, Hospira, Pfizer, Kern Pharma, Biogen, Takeda, Janssen, Roche, Ferring, Faes Farma, Shire Pharmaceuticals, Dr Falk Pharma, Tillotts Pharma, Chiesi, Casen Fleet, Gebro Pharma, Otsuka Pharmaceutical, Vifor Pharma.

#### ORCID iDs

M Chaparro  <https://orcid.org/0000-0002-9275-4242>

C Taxonera  <https://orcid.org/0000-0001-9166-7350>

JP Gisbert  <https://orcid.org/0000-0003-2090-3445>

#### Supplemental material

Supplemental material for this article is available online.

#### References

1. Hanauer SB, Feagan BG, Lichtenstein GR, *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet* 2002; 359: 1541–1549.
2. Present DH, Rutgeerts P, Targan S, *et al.* Infliximab for the treatment of fistulas in patients with Crohn's disease. *N Engl J Med* 1999; 340: 1398–1405.
3. 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: 323–333; quiz 591.
4. Goyette P, Boucher G, Mallon D, *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.
5. Baumgart DC and Sandborn WJ. Crohn's disease. *Lancet* 2012; 380: 1590–1605.
6. Cottone M and Criscuoli V. Infliximab to treat Crohn's disease: an update. *Clin Exp Gastroenterol* 2011; 4: 227–238.
7. Reenaers C, Louis E and Belaiche J. Current directions of biologic therapies in inflammatory bowel disease. *Therap Adv Gastroenterol* 2010; 3: 99–106.
8. Lopez-Hernandez R, Valdes M, Campillo JA, *et al.* Genetic polymorphisms of tumour necrosis factor alpha (TNF-alpha) promoter gene and response to TNF-alpha inhibitors in Spanish patients with inflammatory bowel disease. *Int J Immunogenet* 2014; 41: 63–68.
9. Stokkers PC and Hommes DW. New cytokine therapeutics for inflammatory bowel disease. *Cytokine* 2004; 28: 167–173.
10. Rufini S, Ciccacci C, Novelli G, *et al.* Pharmacogenetics of inflammatory bowel disease: a focus on Crohn's disease. *Pharmacogenomics* 2017; 18: 1095–1114.
11. Styrkarsdottir U, Helgason H, Sigurdsson A, *et al.* Whole-genome sequencing identifies rare genotypes in COMP and CHADL associated with high risk of hip osteoarthritis. *Nat Genet* 2017; 49: 801–805.



12. Tachmazidou I, Suveges D, Min JL, *et al.* Whole-Genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am J Hum Genet* 2017; 100: 865–884.
13. Cui J, Diogo D, Stahl EA, *et al.* Brief report: the role of rare protein-coding variants in anti-tumor necrosis factor treatment response in rheumatoid arthritis. *Arthritis Rheumatol* 2017; 69: 735–741.
14. Bomba L, Walter K and Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol* 2017; 18: 77.
15. MacArthur DG and Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet* 2010; 19: R125–R130.
16. Gomollon F, Dignass A, Annesse V, *et al.* 3rd European evidence-based consensus on the diagnosis and management of Crohn's disease 2016: part 1: diagnosis and medical management. *J Crohns Colitis* 2017; 11: 3–25.
17. Hernandez-Breijo B, Chaparro M, Cano-Martinez D, *et al.* Standardization of the homogeneous mobility shift assay protocol for evaluation of anti-infliximab antibodies. Application of the method to Crohn's disease patients treated with infliximab. *Biochem Pharmacol* 2016; 122: 33–41.
18. McKenna A, Hanna M, Banks E, *et al.* The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20: 1297–1303.
19. Julia A, Tortosa R, Hernanz JM, *et al.* Risk variants for psoriasis vulgaris in a large case-control collection and association with clinical subphenotypes. *Hum Mol Genet* 2012; 21: 4549–4557.
20. Price AL, Patterson NJ, Plenge RM, *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904–909.
21. Cingolani P, Platts A, Wang le L, *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; 6: 80–92.
22. Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010; 7: 248–249.
23. Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; 4: 1073–1081.
24. Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536: 285–291.
25. Eilbeck K, Lewis SE, Mungall CJ, *et al.* The sequence ontology: a tool for the unification of genome annotations. *Genome Biol* 2005; 6: R44.
26. MacArthur DG, Balasubramanian S, Frankish A, *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012; 335: 823–828.
27. Lee S, Emond MJ, Bamshad MJ, *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 2012; 91: 224–237.
28. Kanehisa M and Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; 28: 27–30.
29. Trynka G, Sandor C, Han B, *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 2013; 45: 124–130.
30. Bernstein BE, Stamatoyannopoulos JA, Costello JF, *et al.* The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* 2010; 28: 1045–1048.
31. Sherry ST, Ward MH, Kholodov M, *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; 29: 308–311.
32. de Mattos BR, Garcia MP, Nogueira JB, *et al.* Inflammatory bowel disease: an overview of immune mechanisms and biological treatments. *Mediators Inflamm* 2015; 2015: 493012.
33. Roggenbuck D, Reinhold D, Baumgart DC, *et al.* Autoimmunity in Crohn's disease—a putative stratification factor of the clinical phenotype. *Adv Clin Chem* 2016; 77: 77–101.
34. Matricon J, Barnich N and Ardid D. Immunopathogenesis of inflammatory bowel disease. *Self Nonself* 2010; 1: 299–309.
35. Dotti I, Mora-Buch R, Ferrer-Picon E, *et al.* Alterations in the epithelial stem cell compartment could contribute to permanent changes in the mucosa of patients with ulcerative colitis. *Gut* 2017; 66: 2069–2079.
36. Arimura Y, Isshiki H, Onodera K, *et al.* Characteristics of Japanese inflammatory bowel disease susceptibility loci. *J Gastroenterol* 2014; 49: 1217–1230.
37. 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: 306–315.
38. Silverberg MS, Mirea L, Bull SB, *et al.* A population- and family-based study of Canadian families reveals association of HLA DRB1\*0103 with colonic involvement in inflammatory bowel disease. *Inflamm Bowel Dis* 2003; 9: 1–9.
  39. Mitic LL, Van Itallie CM and Anderson JM. Molecular physiology and pathophysiology of tight junctions I. Tight junction structure and function: lessons from mutant animals and proteins. *Am J Physiol Gastrointest Liver Physiol* 2000; 279: G250–G254.
  40. Lameris AL, Huybers S, Kaukinen K, *et al.* Expression profiling of claudins in the human gastrointestinal tract in health and during inflammatory bowel disease. *Scand J Gastroenterol* 2013; 48: 58–69.
  41. Mallya M, Campbell RD and Aguado B. Characterization of the five novel Ly-6 superfamily members encoded in the MHC, and detection of cells expressing their potential ligands. *Protein Sci* 2006; 15: 2244–2256.
  42. Bradley JR. TNF-mediated inflammatory disease. *J Pathol* 2008; 214: 149–160.
  43. Rifas L and Weitzmann MN. A novel T cell cytokine, secreted osteoclastogenic factor of activated T cells, induces osteoclast formation in a RANKL-independent manner. *Arthritis Rheum* 2009; 60: 3324–3335.
  44. Breese EJ, Michie CA, Nicholls SW, *et al.* Tumor necrosis factor alpha-producing cells in the intestinal mucosa of children with inflammatory bowel disease. *Gastroenterology* 1994; 106: 1455–1466.
  45. Lala S, Ogura Y, Osborne C, *et al.* Crohn's disease and the NOD2 gene: a role for paneth cells. *Gastroenterology* 2003; 125: 47–57.
  46. Billmeier U, Dieterich W, Neurath MF, *et al.* Molecular mechanism of action of anti-tumor necrosis factor antibodies in inflammatory bowel diseases. *World J Gastroenterol* 2016; 22: 9300–9313.
  47. Danese S, Sans M, Scaldaferrri F, *et al.* TNF-alpha blockade down-regulates the CD40/CD40L pathway in the mucosal microcirculation: a novel anti-inflammatory mechanism of infliximab in Crohn's disease. *J Immunol* 2006; 176: 2617–2624.
  48. Rutella S, Fiorino G, Vetrano S, *et al.* Infliximab therapy inhibits inflammation-induced angiogenesis in the mucosa of patients with Crohn's disease. *Am J Gastroenterol* 2011; 106: 762–770.
  49. Monteleone G and Caprioli F. T-cell-directed therapies in inflammatory bowel diseases. *Clin Sci (Lond)* 2010; 118: 707–715.
  50. Nancey S, Holvoet S, Graber I, *et al.* CD8<sup>+</sup> cytotoxic T cells induce relapsing colitis in normal mice. *Gastroenterology* 2006; 131: 485–496.
  51. Westendorf AM, Fleissner D, Deppenmeier S, *et al.* Autoimmune-mediated intestinal inflammation—impact and regulation of antigen-specific CD8<sup>+</sup> T cells. *Gastroenterology* 2006; 131: 510–524.
  52. Steinhoff U, Brinkmann V, Klemm U, *et al.* Autoimmune intestinal pathology induced by hsp60-specific CD8 T cells. *Immunity* 1999; 11: 349–358.
  53. Dahlen R, Strid H, Lundgren A, *et al.* Infliximab inhibits activation and effector functions of peripheral blood T cells in vitro from patients with clinically active ulcerative colitis. *Scand J Immunol* 2013; 78: 275–284.
  54. Baert FJ, D'Haens GR, Peeters M, *et al.* Tumor necrosis factor alpha antibody (infliximab) therapy profoundly down-regulates the inflammation in Crohn's ileocolitis. *Gastroenterology* 1999; 116: 22–28.
  55. Lee JC, Lyons PA, McKinney EF, *et al.* Gene expression profiling of CD8<sup>+</sup> T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. *J Clin Invest* 2011; 121: 4170–4179.
  56. Cirulli ET and Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010; 11: 415–425.
  57. Lee S, Abecasis GR, Boehnke M, *et al.* Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014; 95: 5–23.