



SCIENTIFIC REPORTS



OPEN

Mutational profile of rare variants in inflammasome-related genes in Behçet disease: A Next Generation Sequencing approach

Sergio Burillo-Sanz¹ , Marco-Antonio Montes-Cano¹, José-Raúl García-Lozano¹, Lourdes Ortiz-Fernández¹, Norberto Ortego-Centeno², Francisco-José García-Hernández³, Gerard Espinosa⁴, Genaro Graña-Gil⁵, Juan Sánchez-Bursón⁶, María Rosa Juliá⁷, Roser Solans⁸, Ricardo Blanco⁹, Ana-Celia Barnosi-Marín¹⁰, Ricardo Gómez De la Torre¹¹, Patricia Fanlo¹², Mónica Rodríguez-Carballeira¹³, Luis Rodríguez-Rodríguez¹⁴ , Teresa Camps¹⁵, Santos Castañeda¹⁶, Juan-Jose Alegre-Sancho¹⁷, Javier Martín¹⁸ & María Francisca González-Escribano¹

Behçet's disease (BD) is an immune-mediated systemic disorder with a well-established association with HLA class I and other genes. BD has clinical overlap with many autoinflammatory diseases (AIDs). The aim of this study was to investigate the role of rare variants in seven genes involved in AIDs: *CECR1*, *MEFV*, *MVK*, *NLRP3*, *NOD2*, *PSTPIP1* and *TNFRSF1A* using a next generation sequencing (NGS) approach in 355 BD patients. To check global association of each gene, 4 tests: SKAT, CollapseBt, C(α) and weighted KBAC were used. Databases: 1000 Genomes Project Phase 3, Infevers, HGMD and ClinVar and algorithms: PolyPhen2 and SIFT were consulted to collect information of the 62 variants found. All the genes resulted associated using SKAT but only 3 (*MVK*, *NOD2* and *PSTPIP1*) with C(α) and weighted KBAC. When all the genes are considered, 40 variants were associated to AIDs in clinical databases and 25 were predicted as pathogenic at least by one of the algorithms. Including only *MVK*, *NOD2* and *PSTPIP1*, the associated to AIDs variants found in BD were 20 and the predicted as pathogenic, 12. The maxima contribution corresponds to *NOD2*. This study supports influence of rare variants in genes involved in AIDs in the pathogenesis of BD.

¹Department of Immunology, Hospital Universitario Virgen del Rocío (IBiS, CSIC, US), Sevilla, 41013, Spain.

²Department of Internal Medicine, Hospital Clínico San Cecilio, Granada, 18003, Spain. ³Department of Internal Medicine, Hospital Universitario Virgen del Rocío, Sevilla, 41003, Spain. ⁴Department Autoimmune Diseases, Hospital Universitari Clínic, Barcelona, 08036, Spain. ⁵Department of Rheumatology, Complejo Hospitalario Universitario A Coruña, A Coruña, 15006, Spain. ⁶Department of Rheumatology, Hospital Universitario de Valme, Sevilla, 41014, Spain. ⁷Department of Immunology, Hospital Universitari Son Espases, Palma de Mallorca, 07120, Spain. ⁸Department of Internal Medicine, Autoimmune Systemic Diseases Unit, Hospital Vall d'Hebron, Universidad Autonoma de Barcelona, Barcelona, 08035, Spain. ⁹Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, Santander, 39008, Spain. ¹⁰Department of Internal Medicine, Complejo Hospitalario Torrecárdenas, Almería, 04009, Spain. ¹¹Department of Internal Medicine, Hospital Universitario Central de Asturias, Asturias, 33011, Spain. ¹²Department of Internal Medicine, Hospital Virgen del Camino, Pamplona, 31008, Spain. ¹³Department of Internal Medicine, Hospital Universitari Mútua Terrassa, Terrassa, 08221, Spain. ¹⁴Department of Rheumatology, Hospital Clínico San Carlos, Madrid, 28040, Spain. ¹⁵Department of Internal Medicine, Hospital Regional Universitario de Málaga, Málaga, 29010, Spain. ¹⁶Department of Rheumatology, Hospital de la Princesa, IIS-Princesa, Madrid, 28006, Spain. ¹⁷Department of Rheumatology, Hospital Universitario Doctor Peset, Valencia, 46017, Spain. ¹⁸Instituto de Parasitología y Biomedicina "López-Neyra", CSIC, PTS Granada, Granada, 18016, Spain. Correspondence and requests for materials should be addressed to M.F.G.-E. (email: mariaf.gonzalez.sspa@juntadeandalucia.es)

Behçet's disease (BD) [OMIM #109650] is a complex and immune-mediated systemic syndrome first described in 1937 by a Turkish dermatologist, Hulusi Behçet¹. The disease is characterised by inflammatory lesions of blood vessels which lead to a wide range of clinical phenotypes such as recurrent oral and genital ulceration, ocular involvement (mainly uveitis) and skin lesions, amongst others. The aetiology of BD remains obscure, although some evidences suggest that certain infectious agents and environmental factors may trigger the disease in genetically predisposed individuals². BD is a rare disease, more common along the old "Silk Road"; from China to the Mediterranean area³. The geographical distribution, along with the familial aggregation and association with HLA class I molecules, currently support the genetic component of BD. In this sense, the HLA genetic contribution is estimated to be 20%⁴, with HLA-B51 having an OR of 5.8⁵. Recently, new non-HLA susceptibility genes, such as *IL23R* and *IL10*, have been reported to be involved in the susceptibility to the disease. Other evidences support the involvement of innate immunity and inflammation in the pathogenesis of the disease, in this sense, overexpression of pro-inflammatory cytokines, such as IL-1, IL-17, IL-23 and IFN- γ has been found in BD patients with active uveitis⁶.

Genetic studies have provided evidence of the key role that inflammasome-related genes play in the pathogenesis of autoinflammatory diseases (AIDs). Inflammasome is a multimeric protein complex which is assembled by members of the families Nod-like receptors (NLR) and AIM2-like receptors (ALR), such as NLRP3, which is the most extensively studied inflammasome, to initiate defense mechanisms against bacteria. Two signals are required for NLRP3 activation: induction of the *NLRP3* gene expression by sensor proteins (like TLRs and NOD2) and activation of the protein by a pathogen-associated molecular pattern (PAMP) or damage-associated molecular pattern (DAMP). NLRP3 activation triggers NLRP3-inflammasome assembly which results in IL1 β release and pyroptosis⁷. The AIDs affect different organs and systems being some of them monogenic disorders and, therefore, variants in inflammasome-related genes cause the disease. Thus, variants in the *MEFV* gene cause familial mediterranean fever (FME, OMIM #249100); in *NLRP3*, cryopyrin-associated periodic syndrome (CAPS, OMIM #120100); in *TNFRSF1A*, TNF receptor-associated periodic syndrome (TRAPS, OMIM#142680); in *MVK*, Hyper-IgD syndrome (HIDS, OMIM#249100) and in *PSTPIP1*, pyogenic arthritis, pyoderma gangrenosum, and acne syndrome (PAPAS, OMIM#604416). Nevertheless, there are also some syndromes classified as AIDs, which are polygenic and complex, such as Crohn's disease (CD, OMIM #266600), in these cases, the disease is not caused by variants in a specific gene, but it is associated to changes in several genes, for example, CD has been associated to variants of the *NOD2* among other genes. In addition, environmental factors are required in genetically susceptible individuals to trigger these complex diseases.

BD has a considerable clinical overlap with many of the AIDs, therefore it is plausible to hypothesize that variants in genes involved in these diseases, most of them related with the inflammasome, could partly explain the genetic component of BD. The involvement of these genes in BD could have remained hidden in genome wide association studies (GWAS) because, although these studies are effective in the identification of common variants associated to diseases, they are not suitable for detecting associations with rare or infrequent variants. In this sense, the study of rare variants and the study of polymorphisms have different approaches because the case-control studies, usually used to compare distributions in order to investigate the involvement of polymorphisms in susceptibility to complex diseases, are unusefull to analyze rare variants, because of the absence or the extremely low frequency of their minor allele in general population.

The aim of this study was to assess whether rare variants in genes involved in AIDs, most of them related with the inflammasome, contributes to the pathogeny of BD. The study was performed by exploring the mutational spectrum of this set of genes in a cohort of BD patients using a next generation sequencing (NGS) approach.

Methods

Patients. The study population consisted of 355 patients diagnosed with BD (43.7% males) of Spanish European descent. Peripheral blood (PB) samples collected in EDTA tubes were obtained of patients from different Spanish hospitals⁸, all of them fulfilled the 1990 International Study Group classification criteria for this disease⁹. Clinical features of the patient group were: 100% had oral ulcers, 75.3% skin lesions, 59.5% genital ulcers, 54% uveitis, 42% arthritis, 21% vascular, 18.2% neurological, 16.4% positive pathergy test and 15.5% gastrointestinal involvement and 36% were B51 positive. The study was approved by the ethical committees of all centres and hospitals involved (Hospital Universitario Virgen del Rocío, Hospital Clínico San Cecilio, Hospital Universitario Clínico, Complejo Hospitalario Universitario A Coruña, Hospital Universitario de Valme, Hospital Universitario Son Espases, Hospital Vall d'Hebron, Hospital Universitario Marqués de Valdecilla, Complejo Hospitalario Torrecárdenas, Hospital Universitario Central de Asturias, Hospital Virgen del Camino, Hospital Universitari Mútua Terrassa, Hospital Clínico San Carlos, Hospital Regional Universitario de Málaga, Hospital de la Princesa, Hospital Universitario Doctor Peset, Instituto de Parasitología y Biomedicina López-Neyra). All participants signed a written informed consent prior to their enrolment in the study.

Sample preparation. Genomic DNA was extracted from 200 to 400 μ l of PB using 'QIamp DNA mini kit' (Qiagen, Barcelona, Spain) following manufacture's recommendations. DNA preparations were quantified in a Qubit[®] 3.0 fluorometer and diluted to 0.67 ng/ μ l. For convenience, DNA samples obtained from 355 patients were pooled at equimolar concentration in 71 pools (5 samples/pool). Each pool was processed as a single sample in the subsequent sequencing procedures.

Next Generation Sequencing panel-based analyses. A custom-designed NGS gene panel (AmpliSeq[™] software, Ion Torrent, Thermo Fisher Scientific, Waltham, MA) targeting all the coding regions and flanking intronic sequences of 7 AIDs-related genes was used in this study: *NLRP3* (NG_007509.2), *TNFRSF1A* (NG_007506.1), *MVK* (NG_007702.1), *PSTPIP1* (NG_007526.1), *MEFV* (NG_007871.1), *NOD2* (NG_007508.1) and *CECRI* (NG_033943.1). The total target size was 27,375 bases, distributed in 125 amplicons with a length

of 219 ± 41 bases. After preparation of pools, the amplification and barcoding were carried out with the Ion AmpliSeq™ Kit for Chef DL8, according to the manufacturer's recommendations, and it consisted on 19 cycles of PCR in an Ion Chef system. Libraries were quantified by qPCR or with Qubit® 3.0 Fluorometer, diluted to a final concentration of 40 pM and used for template preparation with the Ion PGM™ Hi-Q™ Chef Kit on the Ion Chef system with 200 bases of templated size. Sequencing templates were loaded in Ion 318™ v2BC Chips (8 pools labeled with barcodes were loaded in each chip) and sequenced in an Ion PGM™ Sequencer. The total readings/amplicon were $4,066 \pm 1,483$ bases and 99.3% amplicons had a coverage greater than 100X.

Variant calling and allele quantification. Variant calling was performed with Ion Reporter™ Software v5.0¹⁰, using a somatic mutation detection and quantification pipeline implemented in the software. Readings were aligned to the human genome reference, hg19. Variants reported by Ion Reporter were verified and filtered by visual inspection on Integrative Genomics Viewer IGV v2.3.68, Broad Institute¹¹ in order to verify the missense mutation presence and to perform a confirmation of the reading alignments. Ion Reporter was also used to quantify the number of mutated alleles in each pool. In our study, one variant in a pool corresponds to 1 allele out of the 10 possible (5 patients per pool) and therefore, it would be quantified as 10%.

Validation of the sample pooling results. Ion Reporter was used to quantify the number of mutated alleles/pool (namely, to quantify germline alleles in pooled samples). This software was primary designed to quantify somatic variants in individual samples; therefore, a validation study was carried out for assessing the accuracy of the method. This validation was performed by analyzing 101 variants in samples of 5 patients which had previously been sequenced individually. The quality control results were 100% sensitivity, all variants detected in the analysis of the individual samples were also detected in pooled samples and 100% specificity, none false-positives were present in the analysis of the pools. The global correlation in the allele quantification was $R^2 = 0.91$ (Figure S2).

Classification of variants. Variants found in patients were classified according to their minor allele frequency (MAF) in the entire population of the 1000 Genomes Project Phase 3 (www.ensembl.org) as common, those variants with a $MAF > 0.05$; rare, $0 < MAF < 0.05$; and undetected, undescribed in that database. Clinical significance of the missense variants found was investigated in human mutation databases: Infevers¹², HGMD¹³ and ClinVar¹⁴ in order to know, which variants are considered non-pathogenic polymorphisms, which variants have been described to be associated with AIDs and which variants are not described. Lastly, two bioinformatics approximations, PolyPhen2 (HumVar model) and SIFT, were used to evaluate the functional impact of variants in the proteins¹⁵.

Disease association studies. The association of the rare variants of each gene with the diseases was evaluated for using Variant Tools software^{16,17}. For patient data, the 71 calling files from Ion Reporter were randomly splitted in 5 files with GATK (Broad Institute, MA) to conserve number of samples ($n = 355$), these VCF files were imported to the Variant Tools software. For the controls, data from 107 samples of 1000 genomes project IBS population were retrieved as VCF files using Data Slicer¹⁸, according to bed file containing target regions of our NGS panel. These files were concatenated with the VCF tools software¹⁹ and imported as a single file to Variant Tools. Variant annotation was filtered with NSFP database, including only missense variants with $MAF \leq 0.05$ in controls. These variants were included in the gene association studies performed and analyzed with the Variant Tools implementation of 4 different tests: a burden or aggregation test, SNP-set (Sequence) Kernel Association Test (SKAT)²⁰; a fixed threshold collapsing method, CollapseBt²¹; one test used to contrast the unusual distribution of minor alleles in cases and controls, $C(\alpha)$ ²²; and, lastly, a weighted burden test²³, weighted KBAC²⁴, which is similar to SKAT but scoring the variants taking into account their Polyphen2 score. Variants were grouped by gene name in refGene database. NOD2 p.Arg702Trp and p.Gly908Arg single variant association to BD was evaluated by Chi-square test. A p-value < 0.05 was considered significant.

Association between rare variants (those with $MAF < 0.05$) of each gene and clinical manifestations was investigated after Sanger-sequencing of all the samples of the pool having each rare specific variant in order to know which sample has the variant. Distribution were compared using Chi-square test with Bonferroni's correction, a p-value < 0.05 was considered significant.

Sanger sequencing verification of novel variants. Presence of those variants that were not found in human mutation databases and rare variants (those with $MAF < 0.05$) was verified by Sanger sequencing with BigDye™ Direct Cycle Sequencing kit using primers generated by Thermo Fisher NGS Sanger confirmation software on Ion Reporter, following manufacturer's recommendations. Patients carrying each variant were identified in all the pools using this method. The novel Sanger-verified variants (*CECR1* p.Arg49Trp, p.Ala247Val; p.Met309Ile and p.Val349Ile, *MEFV* p.Gly111Glu, p.Leu367Val; and p.His404Arg, *PSTPIP1* p.Thr68Met, p.Val122Ile p.Asp289His and; *NOD2* p.Leu81Val, p.Leu349Phe, p.Val733Leu and p.Val816Ile, and *TNFRSF1A* p.His155Tyr, p.Ile199Thr, p.Arg312Lys and p.Asn336His) were submitted to Infevers and ClinVar databases.

Data Availability. Complete datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. All the NGS raw sequence reads have been deposited in SRA database at: <https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP105162> New variants have been deposited in ClinVar database at: <https://www.ncbi.nlm.nih.gov/clinvar/submitters/505967/>

Ethical use of human participants statement. All participants signed a written informed consent to publish identified information prior to their enrolment in the study. All methods were performed after obtaining

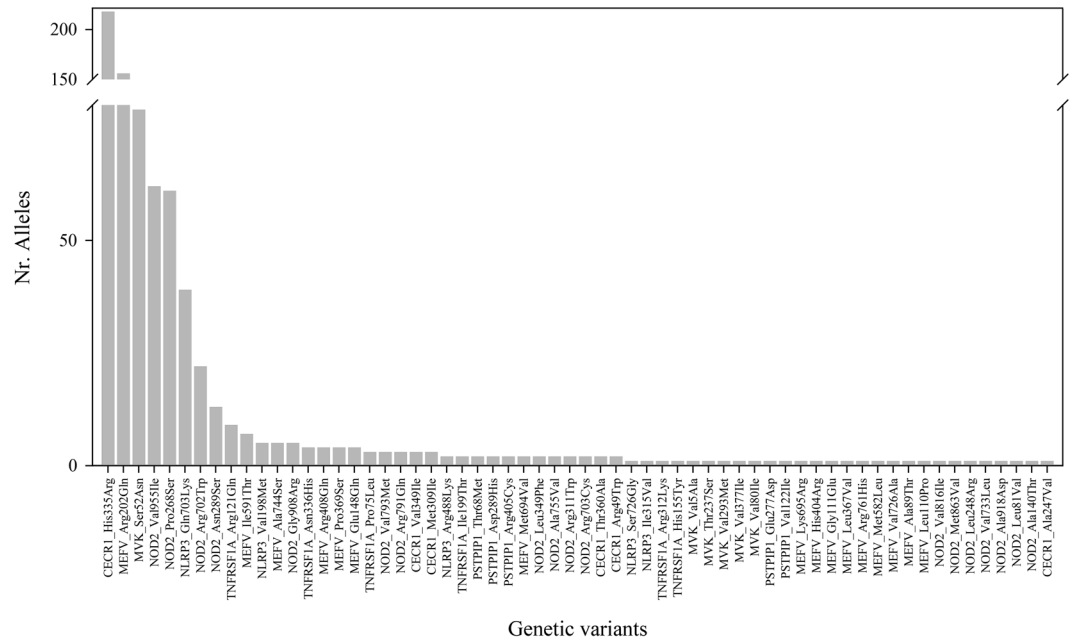


Figure 1. Graphical representation of the number of alleles (y-axis) of each variant detected (listed in x-axis and ordered according to the number of alleles found) in our Behçet disease patient cohort.

the signed informed consent from patients, and in accordance to our institution (Hospital Universitario Virgen del Rocío) ethical committee, and with the Helsinki Declaration of 1975, as revised in 1983.

Results

All the coding regions and flanking intronic sequences of the 7 genes studied were covered with two exceptions: the amplicon of the first half of the *MEFV* exon 5, (aminoacids from 453 to 486) because it did not reach the minimal coverage requirements and a region in *TNFRSF1A* exon 10 (aminoacids 372 and 380) which was not targeted by the primers proposed by the Ion AmpliSeq™ software. In the total of 27,375 bases sequenced in these 7 genes, 62 missense (exonic) single nucleotide variants (SNP) were found (detailed information is displayed in Table S1). Variants were detected in all the genes and the frequency of each of them is displayed in Fig. 1. Neither insertions/deletions with enough quality of sequence, nor intronic variants in the two exon-nearest bases were identified.

Five variants were common SNP ($MAF > 0.05$): 1 in *CECR1*, 2 in *MEFV*, 1 in *MVK* and 1 in *NOD2*; but most of them are rare: 34 had MAF between 0 and 0.05: 2 in *CECR1*, 7 in *MEFV*, 3 in *MVK*, 4 in *NLRP3*, 14 in *NOD2*, 1 in *PSTPIP1* and 3 in *TNFRSF1A*; lastly, 23 variants (including 7 new) had a frequency equal to 0 in the entire population of the 1000 Genomes database: 3 in *CECR1*, 7 in *MEFV*, 2 in *MVK*, 1 in *NLRP3*, 4 in *NOD2*, 3 in *PSTPIP1* and 3 in *TNFRSF1A* (Table 1 and Figure S2A). Rare and not reported variants ($MAF < 0.05$) were considered to analyze the association at gene level taking the IBS population of the 1000 genomes project as control. The results with the 4 tests used are displayed in Table 2 and they are different depending on the test applied: the seven genes could be evaluated and resulted associated with SKAT; when CollapseBt was used, *CECR1* had to be excluded from the analysis and other two genes resulted non-associated: *MEFV* and *TNFRSF1A*; lastly, the results with the $C(\alpha)$ and weighted KBAC tests were the same, *CECR1* could not be evaluated and *MEFV*, *NLRP3* and *TNFRSF1A* did not reach significance whereas *MVK*, *PSTPIP1* and *NOD2* were associated.

In regard to the available clinical information in human mutation databases, 4 out of the 5 missense variants with $MAF > 0.05$ in healthy population are considered non-pathogenic polymorphisms (*CECR1* p.His335Arg, *MEFV* p.Arg202Gln, *MVK* p.Ser52Asn and *NOD2* p.Pro268Ser), whereas one is considered associated to FMF (*MEFV* p.Glu148Gln). Regarding the 34 variants with MAF between 0 and 0.05 in healthy population, 2 have been described as non-pathogenic polymorphisms although functional effects have been reported in some cases^{25–27} (*NLRP3* p.Glu703Lys and *NOD2* p.Val955Ile, MAF in healthy population 0.02 and 0.03, respectively), 28 have been associated with different AIDs and 4 are not found in any clinical database or report. Finally, regarding those 23 variants not detected in healthy population, 9 have been associated with AIDs, and 14 have not been associated with any disease neither in clinical databases nor in journal publications, 7 of which are new (not found in any sequencing database): *MEFV* p.Leu367Val, *NOD2* p.Val733Leu and p.Val816Ile, *PSTPIP1* p.Val122Ile and p.Asp289His and *TNFRSF1A* p.His155Tyr and p.Asn336His (Table 1 and Figure S2B).

Next, two bioinformatic approaches, PolyPhen2 and SIFT, were used to evaluate the functional impact of variants in the proteins. PolyPhen2 predicted some damaging variant in all the 7 genes included, whereas according to SIFT, none of the variants detected in our BD patients in *MVK*, *NLRP3* and *PSTPIP1* were damaging. A total of 25 out of the 62 variants (38.7%) detected were predicted as damaging, 10 only by PolyPhen2, 5 only by SIFT and 10 by both approaches. Both algorithms, predicted *MEFV* p.Glu148Gln as damaging, but the rest of the 4 SNPs with $MAF > 0.05$ in healthy populations and also *NLRP3* p.Glu703Lys and *NOD2* p.Val955Ile as non-damaging.

Gene	Frequencies of variants in 1000 Genomes All Populations Database																	
	MAF > 0.05						0 < MAF < 0.05						MAF = 0					
	Human Mutation Databases						Human Mutation Databases						Human Mutation Databases					
	Non-Pathogenic		Associated to other syndromes		Unknown		Non-Pathogenic		Associated to other syndromes		Unknown		Non-Pathogenic		Associated to other syndromes		Unknown	
CECR1	1	0	0	0	0	0	0	0	0	2	0	0	1	2	0	0	0	
MEFV	1	1	0	0	0	0	7	0	0	0	0	4	3	0	0	0	0	
MVK	1	0	0	0	0	0	3	0	0	0	0	2	0	0	0	0	0	
NLRP3	0	0	0	0	1	3	0	0	0	0	0	1	0	0	0	0	0	
NOD2	1	0	0	0	1	12	0	0	0	0	0	1	4	0	0	0	0	
PSTPIP1	0	0	0	0	0	1	1	0	0	1	1	0	1	2	0	0	0	
TNFRSF1A	0	0	0	0	0	2	1	0	0	1	1	0	0	3	0	0	0	
Total	4	1	0	0	2	28	4	0	0	9	14	0	9	14	0	0	0	
PolyPhen2 prediction																		
Damaging						Damaging						Damaging						
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	
0	4	1	0	0	0	0	2	12	16	1	3	0	0	3	6	3	11	
SIFT prediction																		
Damaging						Damaging						Damaging						
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	
0	4	1	0	0	0	0	2	7	21	0	4	0	0	2	7	5	9	

Table 1. Classification of the variants found in the present study in BD patients. The number of variants in each group is displayed.

Gene	P-values			
	¹ SKAT	¹ CollapseBt	² C(α)	^{2,3} Wighted KBAC
CECR1	3.08E-06	NA	NA	NA
MEFV	0.028	0.35	0.786	0.47
MVK	9.49E-05	3.49E-12	0.0004	0.00039
NLRP3	0.020	0.036	0.078	0.064
NOD2	0.002	0.0008	0.038	0.0014
PSTPIP1	0.0001	0.01	0.008	0.016
TNFRSF1A	0.0001	0.821	0.73	0.65

Table 2. Analysis of the association with BD of rare variants in seven genes related to AIDs. These tests analyze the association of rare variants (MAF < 0.05) grouped by gene. ¹Asymptotic and ²permutated P-values. ³Variants weighted by their Polyphen2 score. The P-values < 0.05 were considered significant and they are shown in bold. NA: Not available, the gene could not be evaluated.

Regarding rare variants associated to AIDs in clinical databases or reports, the prediction of PolyPhen2 and SIFT was damaging for 41% and 24% respectively, whereas for the unknown variants, it was 22% and 27% (Table 1 and Table S1). With respect to the damaging prediction for the 10 variants described in healthy population but not found in clinical databases or reports, 2 were predicted by PolyPhen2: CECR1 p.Ala247Val and PSTPIP1 p.Thr68Met, 2 by SIFT: CECR1 p.Arg49Trp and NOD2 p.Leu81Val and 1 for both: MEFV p.His404Arg. In regards to the prediction for the 7 new variants, 1 was predicted as damaging by PolyPhen2, PSTPIP1 p.Asp289His, whereas SIFT predicted 2: NOD2 p.Leu349Phe and TNFRSF1A p.Asn336His.

Lastly, to evaluate the contribution to the disease of these 7 genes, the percentage of pathogenic alleles in our BD patients was calculated taking into account the results of: 1) the gene association tests, 2) the variants included in clinical databases associated to AIDs, and 3) the prediction of PolyPhen2 and SIFT. When all the 40 variants associated to AIDs in clinical databases in these 7 genes are included in the estimation, the total number of mutated alleles in our patients is 219 (out of 710 possible, 30.84%). The maxima contribution corresponds to NOD2 (16.61% of the variants), followed by NLRP3 (6.76%), MEFV (4.4%) and the rest, TNFRSF1A (1.69%), MVK (0.70%) and PSTPIP1 (0.42%) and CECR1 (0.28%). When only the MVK, NOD2 and PSTPIP1 are included, the number of mutated alleles in the 20 variants reported as associated to AIDs is 126 (17.75%). A total 71 alleles (10.0%) were predicted as pathogenic by PolyPhen2 and/or SIFT but only 44 alleles (6.20%) were predicted by the two algorithms, corresponding the maxima contribution to NOD2 in all the cases (Table 3).

NOD2 p.Asn289Ser, p.Arg702Trp and p.Gly908Arg are loss of function variants which are damaging according with PolyPhen2 and associated to CD, although, they have been reported as possible protective in BD^{28, 29}. These variants have a relatively high frequency, thus, in order to improve the knowledge of the effect of these

Variant	Number of alleles in BD patients (Total of alleles 710)	AID
CECR1_Arg49Trp¹	2	Unknown
CECR1_Ala247Val²	1	Unknown
CECR1_Thr360Ala³	2	Sneddon's syndrome
MEFV_Ala89Thr³	1	FMF
MEFV_Leu110Pro¹	1	FMF
MEFV_Glu148Gln³	5	FMF
MEFV_Pro369Ser³	3	FMF
MEFV_His404Arg³	1	Unknown
MEFV_Arg408Gln	3	FMF
MEFV_Met582Leu	1	Recurrent Arthritis, FMF
MEFV_Ile591Thr	7	FMF
MEFV_Met694Val	2	FMF
MEFV_Lys695Arg²	1	FMF
MEFV_Val726Ala	1	FMF
MEFV_Ala744Ser	5	FMF
MEFV_Arg761His	1	FMF
MVK_Val5Ala	1	HIDS
MVK_Val80Ile	1	HIDS
MVK_Thr237Ser²	1	HIDS
MVK_Val293Met²	1	HIDS
MVK_Val377Ile	1	HIDS
NLRP3_Val198Met	5	FCAS
NLRP3_Ile315Val	1	MAGIC Syndrome
NLRP3_Arg488Lys	2	FCAS
NLRP3_Gln703Lys	39	FCAS
NLRP3_Ser726Gly²	1	CINCA/NOMID
NOD2_Leu81Val¹	1	Unkown
NOD2_Ala140Thr	1	CD
NOD2_Leu248Arg³	1	CD
NOD2_Asn289Ser [*]	13	CD
NOD2_Arg311Trp³	2	CD
NOD2_Leu349Phe¹	2	Unkown
NOD2_Arg702Trp^{3*}	22	CD
NOD2_Arg703Cys³	2	CD
NOD2_Ala755Val²	2	CD
NOD2_Arg791Gln	3	Spondylarthropathy
NOD2_Val793Met	3	CD
NOD2_Met863Val	1	CD
NOD2_Gly908Arg^{3*}	5	CD
NOD2_Ala918Asp²	1	CD
NOD2_Val955Ile	62	CD
PSTPIP1_Thr68Met²	2	Unknown
PSTPIP1_Glu277Asp	1	PAPASH
PSTPIP1_Asp289His²	2	Unknown
PSTPIP1_Arg405Cys	2	Idiopathic juvenile arthritis
TNFRSF1A_Pro75Leu²	3	TRAPS
TNFRSF1A_Arg121Gln	9	TRAPS
TNFRSF1A_Asn336His¹	4	Unkown

Table 3. Number of alleles of the different variants found in our BD patients reported in clinical database and/or pathogenic according to PolyPhen2 and/or SIFT. Variations pathogenic according to the bioinformatic prediction are displayed in bold. ¹SIFT, ²PolyPhen2 and ³both. ^{*}*NOD2* loss of function variants reported as associated to risk in CD but as possible protective in BD.

variants in BD, their frequencies in the patient group were compared with those obtained in ethnically matched healthy controls. No significant differences were observed in any case (Table S2). Even excluding the alleles provided by these 3 loss of function variants, *NOD2* remain as the major contributor (11% of the alleles).

Finally, Table S3 summarizes the clinical features of patients with rare variants (MAF < 0.05) in each gene, no statistically significant differences were found in any case ($p_c > 0.05$).

Discussion

The most relevant result of this study is the description of a considerable number of rare variants, in inflammasome-related genes in patients with a polygenic, complex and immune-mediated disease, such as BD, using a NGS approach. These exonic variants are extremely rare or even absent worldwide in healthy population and most of them are considered causative in monogenic AIDs.

Several studies have investigated the association of BD with specific missense variants in two of these genes, *MEFV* and *NOD2*³⁰. In addition, the entire coding sequence of *NOD2* has also been investigated^{28,29}, although NGS was used only in one study²⁹. The previous NGS study, reported no association of any of the 5 of the Japanese studied genes (*NLRP3*, *TNFRSF1A*, *PSTPIP1*, *MEFV* and *NOD2*) but association of the two genes (*MEFV* and *NOD2*) investigated in Turkish. Study of the association of rare variants in complex diseases has difficulties because of the lack of statistical power. In order to avoid this problem, specific tests grouping variants to check the association of the entire gene have been designed. *NOD2*, resulted associated with all the tests used, nevertheless, association of *MEFV* was detected only using SKAT; differences among tests depend on the statistical power, type I errors likelihood, linkage disequilibrium (LD), etc. Regarding the rest of the genes studied, association of *PSTPIP1* and *MVK* was detected with all of the tests, whereas non-association of *NLRP3* and *TNFRSF1A* was observed with C(α) and weighted KBAC, finally, *CERC*, had to be removed from these two last tests.

Another point that hinders the address of these studies is to determine which variants could influence the pathology and which not. In general, variants common in healthy population are polymorphism but there are exceptions such as *MEFV* p.Glu148Gln (MAF in global healthy population = 0.13) very common in Asian populations, although rare in European (MAF = 0 in the IBS population) which is found in 14% of Spanish FMF patients^{31,32}. In regard to the rare variants, there is a great diversity and population differences. For example, the previous NGS study reported 27 variants with MAF < 0.03, but only 4 were shared with those found in the present suggesting that these variants are specific of each population. In addition, the low frequency of these variants entails a lack of information in clinical databases and ignorance of functional implications based only in the prediction algorithms available. PolyPhen2 is an algorithm with a high specificity; in fact, in our study it predicts as benign all the variants known to be polymorphisms, but its sensitivity is low, many of the variants associated to AIDs are classified as benign by PolyPhen2, even the highly pathogenic *MEFV* p.Met694Val the most common variant in FMF Spanish patients³². Regarding to SIFT, also has a lack of sensitivity, since variants with demonstrated pathogenicity, such as *MEFV* Met694Val and Val726Ala, and *MVK* Thr237Ser and Val377Ile, are predicted as tolerated. Rate of concordance between both algorithms in our study was 75.8% and, although in general, SIFT had a lower sensitivity than Polyphen2, discordances in both senses were found. Combination of both algorithms might be a good option, especially in those many cases in which other functional information is not available. One of the most important limitations of our study is precisely the lack of functional studies and the difficulty to address them easily because of the multiple variants found. Therefore, some of these variants may contribute to the disease onset, but others could be only SNPs without any effect, so that functional and familiar segregation studies of each variant should be carried out in order to assess their phenotypic effects and penetrance. In fact, we can not discard influence of rare variants with clinical features because, although no statistically significant differences were found in any case, sample size is too low to discard association.

In addition, our study used IBS population as control and, although this population is theoretically ethnically matched with our BD cohort, samples were studied in a different way (individually, not pooled) and using a different platform. Therefore, although previous studies suggest that data obtained from both platforms are comparable³³, and analysis of the distribution of common SNPs along with verification by Sanger sequencing of variants with MAF < 0.05 suggests that this possibility is unlikely, stratification problems can not be completely ruled out.

All these reasons make difficult the quantification of the weight of each gene in the disease, although, *NOD2* was the most contributor with all the conditions used in our study. *NOD2* encodes a protein involved in bacterial recognition and loss of function variants, such as p.Arg702Trp, p.Gly908Arg and p.Leu1007fsinsC, has been associated with CD^{34,35}. These loss of function *NOD2* variants have been reported as possible protective in BD, nevertheless, no evidence of this role was found in our cohort. Additionally, the distribution of the p.Asn289Ser, which is not associated to CD but which has a marked reduction of NF- κ B activation³⁶, is not different in BD Spanish patients and controls. From these 3 variants, p.Arg702Trp was not included in the association study of the gene because of its MAF > 0.05. Together with *NOD2*, *PSTPIP1* and *MVK* are associated with the disease with all the tests used. *PSTPIP1* is also part of inflammasome interacting with *MEFV*, variants in this gene cause PAPAS, an autosomic dominant syndrome and PAPAS-related variants enhance the release of IL1 β ³⁷. Regarding *MVK*, it encodes an enzyme involved in cholesterol synthesis. Loss of function variants in this gene gives rise to the accumulation of mevalonic acid and mevalonolactone that co-stimulate inflammasome. The more than 120 variants described to date in this gene have been associated with HIDS and mevalonate kinase deficiency^{38–40}. Association of the rest of the genes has a less support in our study, this is the case of *MEFV* which encodes a pyrin that is a component of *NLRP3* inflammasome complex⁴¹. The result of no-association of *MEFV* when CollapseBt was used could be partially explained because some of the variants of this gene are described as part of the same complex allele^{12,31}, being the only showing evidences of LD in our study and, as a consequence, some genetic test could overestimate the association of *MEFV*. In addition, discrepancies among populations could be based in ethnical differences. In this sense, previous NGS study in Turkish patients concluded that the non-synonymous *MEFV* variants are, in general, associated to the disease although the main weight was attributed to p.Met694Val, nevertheless, only 2 alleles of this variant were found in our cohort. Additionally, despite the relatively high number of variants found in this gene in both studies, only 8 (including founder *MEFV* variants) are shared. Variants in this gene cause FME, those better known as high pathogenic are 4: p.Met680Ile, p.Met694Val, p.Met694Ile and

p.Val726Ala⁴² that have been reported in BD patients^{30,43,44}. These variants were found in our cohort together with other 14 variants, including the well-known non-pathogenic polymorphism p.Arg202Gln, although the *MEFV* most common variant in our study (excepting p.Arg202Gln) was p.Ile591Thr, which is an uncommon variant previously associated to FMF³². Therefore, further studies are needed to clarify the role of this gene in BD. Association of *NLRP3* was detected with SKAT and CollapseBt but did not reach significance with C(α) and weighted KBAC, perhaps because of lack of statistical power. This gene encodes a component of the inflammasome and its variants have been associated with diverse autoinflammatory syndromes. *TNFRSF1A*, which encodes a TNF receptor and cause TRAPS, was found associated only with SKAT. Lastly, *CECRI*, which could only be analyzed with SKAT, encodes an adenosine deaminase (*ADA2*) and has been associated with Cat Eye Syndrome, Sneddon's syndrome and Polyarteritis nodosa. Variants in this gene cause hypogammaglobulinemia and decreased B-cell and monocyte differentiation.

In summary, many rare variants in AIDs related genes are found in BD patients. Some of these variants may play a role in the pathogenesis of this disease, although others could be polymorphisms without any significance in the susceptibility. Specifically, this study supports influence of variants of *NOD2*, *PSTPIP1* and *MVK* in the pathogenesis of BD. Among these genes, *NOD2* would be the major contributor. Altogether these results support a contribution of the innate immune system in the pathogenesis of BD.

References

- Behcet, H. Uber rezidivierende aphthose, durch ein virus verursachte geschwure am mund, am auge und an den genitalien. *Dermatol. Wochenschr.* **105**, 1152–1157 (1937).
- Mendes, D. *et al.* Behcet's disease—a contemporary review. *J Autoimmun* **32**, 178–188 (2009).
- Sakane, T., Takeno, M., Suzuki, N. & Inaba, G. Behcet's disease. *N. Engl. J. Med.* **341**, 1284–91 (1999).
- Yazici, H., Fresko, I. & Yurdakul, S. Behcet's syndrome: disease manifestations, management, and advances in treatment. *Nat. Clin. Pract. Rheumatol.* **3**, 148–55 (2007).
- de Menthon, M., Lavalley, M. P., Maldini, C., Guillevin, L. & Mahr, A. HLA-B51/B5 and the risk of Behcet's disease: a systematic review and meta-analysis of case-control genetic association studies. *Arthritis Rheum.* **61**, 1287–96 (2009).
- Chi, W. *et al.* Upregulated IL-23 and IL-17 in Behcet patients with active uveitis. *Invest. Ophthalmol. Vis. Sci.* **49**, 3058–64 (2008).
- Man, S. M. & Kanneganti, T.-D. Regulation of inflammasome activation. *Immunol. Rev.* **265**, 6–21 (2015).
- Ortiz-Fernández, L. *et al.* Genetic Analysis with the Immunochip Platform in Behcet Disease. Identification of Residues Associated in the HLA Class I Region and New Susceptibility Loci. *PLoS One* **11**, e0161305 (2016).
- Criteria for diagnosis of Behcet's disease. International Study Group for Behcet's Disease. *Lancet (London, England)* **335**, 1078–80 (1990).
- Scientific, T. F. Ion Reporter™ Software. 2015 Available at: <https://ionreporter.thermofisher.com/ir/>. (Accessed: 8th January 2016).
- Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
- Sarrauste de Menthère, C. *et al.* INFEVERS: the Registry for FMF and hereditary inflammatory disorders mutations. *Nucleic Acids Res.* **31**, 282–5 (2003).
- Institute of Medical Genetics in Cardiff. The Human Gene Mutation Database. Available at: <http://www.hgmd.cf.ac.uk/ac/index.php>. (Accessed: 1st August 2016).
- Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–8 (2016).
- Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–9 (2010).
- Wang, G. T. *et al.* Variant association tools for quality control and analysis of large-scale sequence and genotyping array data. *Am. J. Hum. Genet.* **94**, 770–83 (2014).
- San Lucas, F. A., Wang, G., Scheet, P. & Peng, B. Integrated annotation and analysis of genetic variants from next-generation sequencing studies with variant tools. *Bioinformatics* **28**, 421–2 (2012).
- 1000 Genomes Data Slicer. Available at: http://browser.1000genomes.org/Homo_sapiens/UserData/SelectSlice. (Accessed: 1st October 2016).
- Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–8 (2011).
- Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
- Li, B. *et al.* Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–21 (2008).
- Neale, B. M. *et al.* Testing for an Unusual Distribution of Rare Variants. *PLoS Genet.* **7**, e1001322 (2011).
- Madsen, B. E. & Browning, S. R. A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. *PLoS Genet.* **5**, e1000384 (2009).
- Liu, D. J. & Leal, S. M. A Novel Adaptive Method for the Analysis of Next-Generation Sequencing Data to Detect Complex Trait Associations with Rare Variants Due to Gene Main Effects and Interactions. *PLoS Genet.* **6**, e1001156 (2010).
- Hedegaard, C. J., Enevold, C., Sellebjerg, F., Bendtzen, K. & Nielsen, C. H. Variation in *NOD2* augments Th2- and Th17 responses to myelin basic protein in multiple sclerosis. *PLoS One* **6**, e20253 (2011).
- Blomgran, R. *et al.* Common Genetic Variations in the *NALP3* Inflammasome Are Associated with Delayed Apoptosis of Human Neutrophils. *PLoS One* **7**, e31326 (2012).
- Verma, D. *et al.* The Q705K polymorphism in *NLRP3* is a gain-of-function alteration leading to excessive interleukin-1 β and IL-18 production. *PLoS One* **7**, e34977 (2012).
- Ahmad, T. *et al.* *CARD15* polymorphisms in Behcet's disease. *Scand. J. Rheumatol.* **34**, 233–7 (2005).
- Kirino, Y. *et al.* Targeted resequencing implicates the familial Mediterranean fever gene *MEFV* and the toll-like receptor 4 gene *TLR4* in Behcet disease. *Proc. Natl. Acad. Sci.* **110**, 8134–8139 (2013).
- Baruch, Y. *et al.* *MEFV*, *TNFRSF1A* and *CARD15* mutation analysis in Behcet's disease. *Clin Exp Rheumatol* **29**, S24–27 (2011).
- Bernot, A. *et al.* Non-founder mutations in the *MEFV* gene establish this gene as the cause of familial Mediterranean fever (FMF). *Hum. Mol. Genet.* **7**, 1317–25 (1998).
- Toutou, I. The spectrum of Familial Mediterranean Fever (FMF) mutations. *Eur. J. Hum. Genet.* **9**, 473–83 (2001).
- Erguner, B., Ustek, D. & Sagiroglu, M. S. Performance comparison of Next Generation sequencing platforms. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2015**, 6453–6456 (2015).
- Lesage, S. *et al.* *CARD15/NOD2* mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am. J. Hum. Genet.* **70**, 845–57 (2002).
- Hugot, J. P. *et al.* Association of *NOD2* leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
- Chamailard, M. *et al.* Gene-environment interaction modulated by allelic heterogeneity in inflammatory diseases. *Proc. Natl. Acad. Sci. USA* **100**, 3455–60 (2003).

37. Yu, J.-W. *et al.* Pyrin Activates the ASC Pyroptosome in Response to Engagement by Autoinflammatory PSTPIP1 Mutants. *Mol. Cell* **28**, 214–227 (2007).
38. Mandey, S. H. L., Schneiders, M. S., Koster, J. & Waterham, H. R. Mutational spectrum and genotype-phenotype correlations in mevalonate kinase deficiency. *Hum. Mutat.* **27**, 796–802 (2006).
39. Drenth, J. P. *et al.* Mutations in the gene encoding mevalonate kinase cause hyper-IgD and periodic fever syndrome. International Hyper-IgD Study Group. *Nat. Genet.* **22**, 178–81 (1999).
40. Houten, S. M. *et al.* Organization of the mevalonate kinase (MVK) gene and identification of novel mutations causing mevalonic aciduria and hyperimmunoglobulinaemia D and periodic fever syndrome. *Eur. J. Hum. Genet.* **9**, 253–9 (2001).
41. Ozen, S. & Bilginer, Y. A clinical guide to autoinflammatory diseases: familial Mediterranean fever and next-of-kin. *Nat. Rev. Rheumatol.* **10**, 135–47 (2014).
42. French, F. M. F. Consortium. A candidate gene for familial Mediterranean fever. *Nat. Genet.* **17**, 25–31 (1997).
43. Wu, Z. *et al.* Association between MEFV Mutations M694V and M680I and Behçet's Disease: A Meta-Analysis. *PLoS One* **10**, e0132704 (2015).
44. Tasliyurt, T., Yigit, S., Rustemoglu, A., Gul, U. & Ates, O. Common MEFV gene mutations in Turkish patients with Behçet's disease. *Gene* **530** (2013).

Acknowledgements

This work was supported by Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III (ISCIII, 13/01118 and PI16/01373), Fondos FEDER and Plan Andaluz de Investigación (CTS-0197). SBS is the recipient of a fellowship (ISCIII, 13/01118).

Author Contributions

S.B.S.: Experimental procedures, data analysis, manuscript preparation and submission. M.F.G.E.: Experimental design, data analysis, manuscript preparation. M.A.M.C.: Experimental support, data analysis, manuscript preparation. J.R.G.L.: Experimental support, manuscript preparation. L.O.F.: Samples and clinical data provider. N.O.C.: Samples and clinical data provider. F.G.H.: Samples and clinical data provider. G.E.: Samples and clinical data provider. G.G.G.: Samples and clinical data provider. J.S.B.: Samples and clinical data provider. M.J.: Samples and clinical data provider. R.S.: Samples and clinical data provider. R.B.: Samples and clinical data provider. A.B.M.: Samples and clinical data provider. R.G.T.: Samples and clinical data provider. P.F.: Samples and clinical data provider. M.R.C.: Samples and clinical data provider. L.R.R.: Samples and clinical data provider. T.C.: Samples and clinical data provider. S.C.: Samples and clinical data provider. J.A.S.: Samples and clinical data provider. J.M.: Samples and clinical data provider.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-09164-7](https://doi.org/10.1038/s41598-017-09164-7)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017