Published in final edited form as:

J Pediatr Gastroenterol Nutr. 2023 September 01; 77(3): 354–357. doi:10.1097/MPG.0000000000003873.

Exome Sequencing Implicates *DGKZ, ESRRA,* and *GXYLT1* for Modulating Granuloma Formation in Crohn's Disease

R. Alan Harris[1,#], **Allyson H Bush**[2,#], **Todd N Eagar**[3,4], **Justin Qian**[2], **Michael P Greenwood**[3,4], **Antone R Opekun**[2], **Robert Baldassano**[5], **Stephen L Guthery**[6], **Joshua D Noe**[7], **Anthony Otley**[8], **Joel R. Rosh**[9], **Subra Kugathasan**[10], **Richard Kellermayer**[2,11,*]

[1]Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX

[2]Department of Pediatrics, Section of Gastroenterology, Hepatology and Nutrition, Baylor College of Medicine/Texas Children's Hospital, Houston, TX

[3]Department of Pathology and Genomic Medicine, Houston Methodist Hospital, Houston, TX

[4]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY

[5]Division of Gastroenterology, Hepatology and Nutrition, University of Pennsylvania, Children's Hospital of Philadelphia, PA

[6]Department of Pediatrics, University of Utah and Intermountain Primary Children's Hospital, Salt Lake City, UT

[7]Department of Pediatrics, Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Medical College of Wisconsin, Milwaukee, WI

[8]IWK Health/Dalhousie University, Halifax, Nova Scotia, Canada

[9]Goryeb Children's Hospital/Atlantic Children's Health, Morristown, NJ

[10]Departments of Pediatrics and Human Genetics at Emory University School of Medicine, Atlanta, GA

[11]Children's Nutrition and Research Center, Houston, TX

[#]contributed equally

Abstract

Non-caseating granulomas may indicate a more aggressive phenotype of Crohn's disease (CD). Genetic associations of granulomatous CD (GCD) may help elucidate disease pathogenesis.

Whole-exome sequencing (WES) was performed on peripheral blood derived DNA from 17 pediatric patients with GCD and 19 with non-GCD (NGCD), and from an independent validation cohort of 44 GCD and 19 NGCD cases. PLINK analysis was used to identify single nucleotide polymorphisms (SNPs) differentiating between groups, and subgroup allele frequencies were also

[*]Address for Correspondence: Richard Kellermayer, MD, PhD, Section of Pediatric Gastroenterology, Hepatology & Nutrition, Baylor College of Medicine, 6621 Fannin St., CC1010.00, Houston, TX 77030-2399, Voice: 713-798-0319 Fax: 832-825-3633 kellerma@bcm.edu.

compared to a public genomic database (gnomAD). The CADD scoring tool was used to predict deleteriousness of SNPs. HLA haplotype findings were compared to a control group (n=8496).

PLINK based analysis between GCD and NGCD groups did not find consistently significant hits. GnomAD control comparisons, however, showed consistent subgroup associations with *DGKZ, ESRRA,* and *GXYLT1*, genes that have been implicated in mammalian granulomatous inflammation.

Our findings may guide future research and precision medicine.

Keywords

Crohn disease; granuloma; genetics; inflammatory bowel disease; Mycobacterium

INTRODUCTION

Non-caseating granulomas are a hallmark histopathological finding of Crohn's disease (CD), although not necessary for diagnosis. Studies have suggested that the presence of granulomas may indicate a more aggressive CD phenotype associated with a complicated clinical course, including stricturing and/or penetrating disease, need for biologic therapy, and need for surgery (1, 2, 3). Studies in children with CD demonstrated that the presence of granulomas may be associated with more perianal fistula formation (4), C-reactive protein (CRP) elevation and microbiome separation (5), more extensive disease and more hospitalizations, with questions as to whether a granulomatous subtype is more likely to require surgical intervention (6, 7). Even microgranulomas, smaller collections of histiocytes, have been associated with more frequent escalation to anti-tumor necrosis factor drugs in children with CD (8). Identification of genetic associations of granulomatous CD (GCD) may therefore help uncover disease pathogenesis, which in turn may optimize treatments and guide novel therapeutics to combat CD complications.

There is limited genetic information about granulomatous CD. A study by Brinar, *et al.* (9) genotyped for 79 single nucleotide polymorphisms (SNPs) tagging autophagy genes and found an association between several of those (in *ATG4, ATG2A, ATG4D,* and *FNBP1L*) and the presence of granulomas in surgically-treated CD patients. Apart from this work, where post-hoc selection examined those patients with a complicated phenotype (i.e., pathology samples were taken from surgical specimens), we found limited literature examining the genetic associations of GCD. Amongst those is the lack of correlation between *NOD2* variants and granuloma formation in adult and pediatric patients with CD. (10)

The aim of this study was to determine the extent of gene exome variation between pediatric patients primarily distinguished by a pathognomonic submucosal granuloma detected at diagnosis of CD, and those without a granuloma found at diagnosis.

METHODS

Patients

Discovery patients were selected from two IRB approved studies at Baylor College of Medicine: DIGESTIVE DISEASE CENTER – INFLAMMATORY BOWEL DISEASE RESEARCH TISSUE BANK (H-17654; P30 DK56338) and PRO-KIIDS PILOT STUDY TO EXAMINE SPECIFIC GENOTYPES AND PROTEOTYPES THAT INCREASE THE RISK FOR COMPLICATED CROHN'S DISEASE (H-43617). The patients in both protocols gave informed consent for performing genetic studies linked to their disease. The validation cohort granuloma and exome data were collected from participating RISK Consortium sites. The RISK cohort was a multicenter study that enrolled treatment naïve pediatric patients aged 3 to 17 years with CD and non-IBD controls from 28 sites with the United States and Canada from 2008 to 2012 (11). None of the discovery cohort patients were included into the validation cohort.

GCD was defined as a single or more non-caseating granuloma pathognomic for CD, detected in any of the mucosal biopsy specimens obtained during diagnostic endoscopy of a given treatment-naïve patient, based on the official pathology report.

Exome sequencing

Whole-exome next-generation sequencing (WES) was performed from peripheral blood-derived DNA. DNA was isolated using standard DNA isolation methods and analyzed for purity and molecular weight using PicoGreen and gel imaging. After DNA quality control tests, Illumina sequencing libraries with incorporated barcodes were produced following standard procedures (12) using the HGSC VCRome Kit capture reagent. Bar-coded samples were pooled and captured together. The resulting pools, enriched for the human exome by the capture process, were sequenced using the Illumina NovaSeq instrument.

SNP calling and annotation

The Illumina 150bp paired-end reads were aligned to the human GRCh38 reference genome assembly using BWA MEM with an average on-target sequence depth of 63.20X across the samples. Picard MarkDuplicates version 1.105 (http://broadinstitute.github.io/picard/) was used to identify and mark duplicate reads. The GATK v 4.1.2.0 best practices pipeline (13, 14) was used to identify single nucleotide polymorphisms (SNPs). Variant Effect Predictor software (VEP) (15) was used to annotate variants based on merged Ensembl and RefSeq gene models. The potential deleteriousness of SNPs was determined by the Combined Annotation Dependent Depletion (CADD) scoring tool (16).

HLA haplotype analysis

HLA typing was generated from WES data using HLA-HD (17). Alleles sharing the same amino acid sequence for the peptide binding domains were grouped for analysis (P groups) (http://hla.alleles.org/). Allele-frequency comparisons were made using HLA-B, HLA-DRB1, HLA-DQA1/DQB1 or HLA-DPA1/DPB1 typing data from individuals typed at Houston Methodist Hospital between 2018–2020.

Data analysis

The chi-square test is a statistical test used to determine if there is a significant difference between the expected frequencies and the observed frequencies in a categorical data set. PLINK (18) was used to identify SNPs that were overrepresented in comparisons between granulomatous Crohn's disease (GCD) and non-granulomatous Crohn's disease (NGCD) patients based on chi-square allelic tests. Due to the small, but highly selective, sample sizes the level of significance for exome wide overrepresentation was relaxed to $p<0.01$ for the PLINK-based group comparison.

GCD and NGCD allele frequencies were also compared to publicly available control genomic data in the Genome Aggregation Database (gnomAD) v3.1.2 (n= 76,156) (19). We did this additional analysis because we recognized that both our discovery and validation cohorts were relatively small in size providing limited power, and wished to further investigate if there may be different genetically mediated pathophysiologic contributions in GCD and NGCD based on a the separate analysis of SNPs from each group vs a large group of controls (i.e., gnomAD). For gnomAD comparisons the chi-square test was used to identify significant ($p<1\times10^{-8}$) allelic differences between GCD or NGCD patients and the gnomAD allele frequencies.

For comparisons to gnomAD, we performed separate chi square tests on the discovery and validation cohorts. We considered SNPs with the exact same genomic coordinates and consistent results in both cohorts as validated and removed from consideration SNPs unique to a single cohort. This approach is based on the assumption that variants with the exact same coordinates are more likely to be true associations and are less likely to be false positives.

RESULTS

Seventeen patients with GCD and 19 with non-granulomatous CD (NGCD) were studied in discovery. The subgroups were similar in respect to commonly examined clinical, laboratory and endoscopic features of CD with the exception of granuloma detected in the diagnostic biopsy samples (Supplementary Table 1). Notably, the average number of biopsies obtained were similar between the groups, limiting the potential for sampling-error based misclassification. In the independent validation cohort, there were 44 GCD and 19 NGCD cases, which were collected through voluntary participation by RISK consortium members. There were no consistently significant genetic variations between the discovery and validation cohorts comparing the two groups by PLINK.

Since our PLINK based approach did not yield significant results, we decided to pursue an alternative method. We compared SNP allele frequency between GCD and gnomAD (see materials and methods), and between NGCD and gnomAD. The larger sample size of the gnomAD data set provided more statistical power than the relatively small sample size of the discovery and validation cohorts. A rather conservative approach was taken by setting our SNP selection criteria to those with allelic frequency difference compared to gnomAD of $p<1\times10^{-8}$ and with sequencing data at the particular SNP available in at least 90% of patients in both the discovery and validation cohorts. Variant allele frequency in each cohort

was required to be at least two, with singletons being removed. Thereafter, we selected only those SNPs that met our selection criteria in both of the cohorts and consistently associated with only GCD or NGCD, respectively. In order to concentrate on potentially functional SNPs, a CADD PHRED score of 20 or greater was required. A CADD PHRED score of 20 predicts the SNP is among the 1% most deleterious substitutions in the human genome. This conservative approach led to the identification of six GCD specific SNPs, two of which were linked to three non-HLA genes, *DGKZ* and *ESRRA/PRDX5* respectively (Table 1). Additionally, there were three SNPs, which highly significantly and consistently associated with NGCD, two of which were linked to non-HLA genes, *GXLT1* and *TTC28* (Table 1). Of note, only 2 variant alleles of *TTC28* were present in the discovery cohort of NGCD cases, making this SNP a less likely candidate. We focused on non-HLA gene associations because our independent HLA haplotype analysis did not identify consistent and significant HLA-variation in GCD and NGCD compared to organ donor controls.

DISCUSSION

This is the first report on exome-wide variation in diagnostic mucosal biopsy based granulomatous Crohn's disease. The work was performed in pediatric cases where diagnostic delay and comorbidities less commonly influence intestinal pathology compared to adults (20). Consequently, histopathology based separation might better delineate individual pathogenesis in this age group of patients. While our study was limited by the number of cases studied, the discovery/validation design and the conservative selection criteria lend strength to the results. Based on our findings, *DGKZ, ESSRA, PRDX5* and *GXLT1* are implicated as potential genetic contributors to granuloma development in CD. Intriguingly, all of these genes have been indicated to play a role in mammalian granulomatous inflammation.

SNP rs114974750 is a missense variant in *DGKZ* (diacylglycerol kinase zeta), which, promotes TGF-beta signaling, and has been found to be differentially expressed in human macrophages in response to *Mycobacterium avium* strain SE01 (21), which is taxonomically very close to *Mycobacterium paratuberculosis* (MAP). MAP has been described as the pathogen for Johne's disease, a wasting granulomatous intestinal inflammation in cattle, and has been considered in the microbial origins of CD in some patients (22). In contrast, SNP rs181558534 is a missense variant in *GXYLT1,* which was associated with non-granulomatous CD. *GXYLT1* has been indicated to modulate NOTCH1 signaling, and induce resistance to Johne's disease in cattle (23). Therefore, *GXYLT1* inhibition could plausibly increase the risk of granuloma formation (similarly as in Johne's disease in cattle) in the background of CD.

SNP rs201336331 is a missense variant in *ESRRA* (estrogen-related receptor α), which has been found to critically influence the severity of granulomatous inflammation of the lungs in response to *Mycobacterium tuberculosis* infection in mice (24). *Esrra−/−* mice showed more severe pathological responses and granulomatous lesions in the lungs compared to *Esrra+/+* mice. *ESRRA* was also recently shown to modulate intestinal inflammatory responses through activation of autophagic flux in association with gut microbiota (25).

SNP rs201336331 is also 2,324 bp upstream of *PRDX5*, but it does not lie in a promoter or enhancer associated with *PRDX5* based on either Ensembl regulatory features (26) or GeneHancer (27). This protein is one of the six mammalian peroxiredoxins and mainly functions as a cytoprotective antioxidant enzyme. Sarcoidosis has been linked with a *PRDX5* risk locus (rs479777; chr11:64340005) by fine mapping and expression analysis (28). This locus is only 24 Kbp downstream of the GCD rs201336331 risk locus in our study. Importantly, a CD risk allele (rs694739; chr11:64329761) has been described in association with *ESRRA/PRDX5* by Franke, et al in 2010 (29), which is 14 Kbp downstream of our GCD rs201336331 risk locus.

The functional relevance of our candidate risk allele compendium fosters confidence in the impact of our findings. We trust that larger scale, but similar genetic analyses will further our understanding of the genetic background for GCD. Furthermore, basic and translational studies on the candidate genes identified herein could set the nidus for granuloma-based CD subtype treatment and prevention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding and Acknowledgements

This work was supported by the ProKIIDS Network of the Crohn's and Colitis Foundation [#585708 to RK]. We are grateful for philanthropic funds from the Wagner Family led Gutsy Kids Fund and the generous families contributing to that, and for funds from the Klaasmeyer family for PSC research to RK.

Conflict of interest:

AO reports Advisory Board for AbbVie; research funding from AbbVie, Janssen, Takeda, Lilly, Pfizer

REFERENCES

1. Heresbach D, Alexandre JL, Branger B, et al. Frequency and significance of granulomas in a cohort of incident cases of Crohn's disease. Gut 2005;54:215–222. [PubMed: 15647184]
2. Hong SW, Yoon H, Shin CM, et al. Clinical significance of granulomas in Crohn's disease: A systematic review and meta-analysis. J Gastroenterol Hepatol 2020;35:364–373. [PubMed: 31522456]
3. Johnson CM, Hartman DJ, Ramos-Rivers C, et al. Epithelioid Granulomas Associate with Increased Severity and Progression of Crohn's Disease, Based on 6-Year Follow-Up. Clin Gastroenterol Hepatol 2018;16:900–907.e1. [PubMed: 29277619]
4. Lawrence LS, Heider A, Singer AAM, et al. Granulomas in Diagnostic Biopsies Associated with High Risk of Crohn's Complications-But May Be Preventable. Inflamm Bowel Dis 2022;28:523–530. [PubMed: 33999198]
5. Kellermayer R, Mir SA, Nagy-Szakal D, et al. Microbiota separation and C-reactive protein elevation in treatment-naïve pediatric granulomatous Crohn disease. J Pediatr Gastroenterol Nutr 2012;55:243–250. [PubMed: 22699834]
6. Rothschild B, Rinawi F, Herman Y, et al. Prognostic significance of granulomas in children with Crohn's disease. Scand J Gastroenterol 2017;52:716–721. [PubMed: 28346037]
7. Markowitz J, Kahn E, and Daum F. Prognostic significance of epithelioid granulomas found in rectosigmoid biopsies at the initial presentation of pediatric Crohn's disease. J Pediatr Gastroenterol Nutr 1989;9:182–186. [PubMed: 2809938]

8. Moran CJ, Pankaj A, Kaplan JL, et al. Microgranulomas at diagnosis are associated with need for antitumor necrosis factor escalation in pediatric Crohn Disease. J Pediatr Gastrenterol Nutr 2022; 74(5):617–620.
9. Brinar M, Vermeire S, Cleynen I, et al. Genetic variants in autophagy-related genes and granuloma formation in a cohort of surgically treated Crohn's disease patients. J Crohns Colitis 2012;6:43–50. [PubMed: 22261526]
10. Shaoul R, Karban A, Weiss B, et al. NOD2/CARD15 mutations and presence of granulomas in pediatric and adult Crohn's disease. Inflamm Bowel Dis. 2004 Nov;10(6):709–14. [PubMed: 15626887]
11. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 2014;15:382–392. [PubMed: 24629344]
12. Lupski JR, Gonzaga-Jauregui C, Yang Y, et al. Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. Genome Med 2013;5:57. [PubMed: 23806086]
13. 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. [PubMed: 20644199]
14. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–498. [PubMed: 21478889]
15. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17:122. [PubMed: 27268795]
16. Rentzsch P, Witten D, Cooper GM, et al. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019;47: D886–D894. [PubMed: 30371827]
17. Kawaguchi S, Higasa K, Shimizu M, et al. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. Hum Mutat 2017;38:788–797. [PubMed: 28419628]
18. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575. [PubMed: 17701901]
19. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–443. [PubMed: 32461654]
20. Schoepfer A, Santos J, Fournier N, et al. Systematic Analysis of the Impact of Diagnostic Delay on Bowel Damage in Paediatric Versus Adult Onset Crohn's Disease. J Crohns Colitis 2019;13:1334–1342. [PubMed: 31002741]
21. Blumenthal A, Lauber J, Hoffmann R, et al. Common and unique gene expression signatures of human macrophages in response to four strains of Mycobacterium avium that differ in their growth and persistence characteristics. Infect Immun 2005;73:3330–3341. [PubMed: 15908359]
22. Proietti E, Fuhler GM, and Peppelenbosch MP. Mycobacterium Avium Subspecies Paratuberculosis Infection and Biological Treatment of IBD: Cause or Consequence? J Crohns Colitis 2021;15:1247–1249. [PubMed: 33735385]
23. Mallikarjunappa S, Brito LF, Pant SD, et al. Johne's Disease in Dairy Cattle: An Immunogenetic Perspective. Front Vet Sci 2021;8:718987.
24. Kim SY, Yang CS, Lee HM, et al. ESRRA (estrogen-related receptor α) is a key coordinator of transcriptional and post-translational activation of autophagy to promote innate host defense. Autophagy 2018;14:152–168. [PubMed: 28841353]
25. Kim S, Lee JY, Shin SG, et al. ESRRA (estrogen related receptor alpha) is a critical regulator of intestinal homeostasis through activation of autophagic flux via gut microbiota. Autophagy 2021;17:2856–2875. [PubMed: 33172329]
26. Zerbino DR, Wilder SP, Johnson N, et al. The ensembl regulatory build. Genome Biol 2015;16:56. [PubMed: 25887522]
27. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) 2017; bax028.
28. Fischer A, Schmid B, Ellinghaus D, et al. A novel sarcoidosis risk locus for Europeans on chromosome 11q13.1. Am J Respir Crit Care Med 2012;186:877–885. [PubMed: 22837380]

29. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet 2010;42:1118–1125. [PubMed: 21102463]

What is known:

- Granulomas are a pathognomic histologic finding for Crohn's disease in the background of inflammatory bowel disease
- Granulomas may indicate a more progressive disease phenotype

What is new:

- Polymorphisms in genes relevant for granulomatous inflammation may be more common in patients with granulomatous Crohn's disease
- Identification of genes associated with granulomatous Crohn's disease may help guide personalized medical care

Table 1.

Single nucleotide polymorphisms (SNPs) that significantly ($p<10^{-8}$) and consistently were enriched in either granulomatous or non-granulomatous Crohn's disease (CD) compared to the gnomAD database in both the discovery and validation cohorts

SNP_ID	Discovery Cohort			Validation Cohort			gnomAD	Discovery	Validation	Annotations			
chrom:position: refAllele:altAllele	Ref allele count	Alt allele count	Alt allele frequency	Ref allele count	Alt allele count	Alt allele frequency	Alt allele frequency	Chi-square p	Chi-square p	Gene	VEP Consequence	CADD Phred	dbSNP ID
Granulomatous CD (GCD)													
11:46347717:G:A	25	9	0.2647	47	39	0.4535	0.0136	1.67E-32	5.58E-259	*DGKZ*	missense_variant	23	rs114974750
11:64315797:G:A	21	13	0.3824	62	26	0.2955	0.0459	3.38E-19	1.12E-27	*ESRRA PRDX5*	missense_variant upstream_gene_variant	23.1	rs201336331
6:31271830:G:A	29	5	0.1471	78	10	0.1136	0.0058	5.34E-22	1.35E-35	*HLA-C*	missense_variant	23.2	rs41542423
6:31356822:T:G	29	5	0.1471	78	10	0.1136	0.0059	1.23E-21	6.53E-35	*HLA-B*	missense_variant	20.6	rs1050538
6:31271766:C:T	28	6	0.1765	80	8	0.0909	0.0073	3.96E-26	1.02E-17	*HLA-C*	missense_variant	23.8	rs1050428
6:32589726:A:G	29	5	0.1471	80	8	0.0909	0.0080	5.16E-16	6.51E-16	*HLA-DRB1*	missense_variant	22.2	rs201726340
Non-Granulomatous CD (NGCD)													
12:42144532:C:A	33	5	0.1316	23	15	0.3947	0.0063	2.89E-18	5.53E-184	*GXYLT1*	missense_variant	23.2	rs181558534
6:31356864:T:A	28	10	0.2632	25	13	0.3421	0.0315	1.52E-14	1.39E-25	*HLA-B*	missense_variant	22.2	rs1050518
22:28679634:C:T	34	2	0.0556	27	11	0.2895	0.0011	1.39E-13	0	*TTC28*	synonymous_variant	20.2	rs76130400

J Pediatr Gastroenterol Nutr. Author manuscript; available in PMC 2024 September 01.