

Synergistic effects of Heterozygous variants of NOD2, IL10RA, PLA2G6 and COL7A1 causative an extreme early-onset and severe Crohn's disease

Qiang Zhang¹, Xizi Wang¹, Juan Zheng¹, Qiang Lü¹, Rongrong Li¹, XIAODONG JIA¹, and Mingliang Gu¹

¹Liaocheng People's Hospital

May 6, 2022

Abstract

To identify candidate pathogenic genes of early-stage Crohn's disease (CD) and predict potential roles of genetic factors in CD, we performed whole exome sequencing on a child with early-stage Crohn's disease (CD) and her parents (core family), found that the patient carried heterozygous variants of 4 genes: *NOD2* c. 2257 C>T, *IL10RA* c. 301 C>T, *PLA2G6* c. 2029 C>T, *COL7A1* c. 3190 G>A. With joint action of *NOD2*, *IL10RA*, *PLA2G6* and *COL7A1*, excessive intestinal inflammatory response is triggered, resulting in normal intestinal wall tissue damage. Meanwhile, intestinal wall tissue repair is impaired, aggravating inflammation and injury, and leading to severe CD phenotype.

Introduction

Crohn's disease (CD) is an inflammatory bowel disease (IBD) involving any part of the gastrointestinal tract [1]. CD is characterized by intermittent, progressive and destructive transmural inflammation. Patients often present abdominal pain, diarrhea, and repeated anal lesions with delayed healing [2-3]. Since the 21st century, the incidence of IBD has been increased worldwide, with the highest mainly in North America, Northern Europe and Western Europe [4]. Although the prevalence of IBD in Asia is relatively lower than that in western countries, the incidence of CD has been increased rapidly [1]. In China, IBD has become a common disease, with its incidence exhibiting a gradient distribution from the south to north, and from the east to west [5-6].

Genetic susceptibility and host reactions (immune regulation and intestinal flora) contribute to CD [7-8]. Genome-wide association analysis revealed more than 200 loci associated with CD risk, which are mainly located in regulatory regions [9]. For example, *NOD2*, *ATG16L1*, *IRGM*, *LRRK2* and *XBP1* genes are associated with susceptibility to CD. These variants are proposed to cause abnormal secretion activity of Paneth cells, endoplasmic reticulum stress signals that may trigger pathological unfolded protein response (UPR), intestinal barrier dysfunction, and intestinal inflammatory response [10-15]. Abnormalities of STAT3 and NF- κ B pathways lead to intestinal barrier dysfunction, mediating entry of luminal contents into lamina propria, activation of innate and adaptive immunity, and production of pro-inflammatory factors that perpetuate intestinal inflammation [16]. *IL-10* and *IL-10R* genetic variants are associated with early-onset IBD and Mendelian inheritance of high intestinal wall permeability [17]. *IL23R* gene is associated with predisposition to CD [18]. In addition, in genetically susceptible hosts, decreased diversity of host intestinal flora correlates with the development of CD [1]. In individuals, genetic factors interact with intestinal flora or pathogens to trigger a sustained over-active immune response in the intestinal tract, resulting in intestinal tissue damage and aggravating intestinal barrier dysfunction [18]. Although extensive studies have been conducted on CD,

its etiology and pathogenesis are not fully understood. CD remains an incurable and complex disease due to limited prevention strategies and treatment approaches.

We performed whole exome sequencing on a neonatal patient with severe CD and her parents. Candidate pathogenic genes of the neonatal patient were screened and identified using databases of CD pathogenic genes and colorectal tissue-specific genes. Through gene function annotation, pathway analysis and protein structure prediction, biological functions affected by genetic variants were explored to reveal the genetic mechanisms underlying severe CD phenotypes.

Subject and Methods

Research objects

The patient, female, was born at full term in 2010. She was admitted to Beijing Children's Hospital due to diarrhea and anal fissure with perianal abscess more than ten days after birth. Her body temperature was over 38 degrees Celsius and c-reactive protein level was high. After hospitalization, antibiotics were administered. No improvement was observed in perianal abscess ulceration accompanied by inflammation, or anal fissure wound. At age 1, "ileostomy" and "repair of intestinal perforation" were performed. At age 2, high fever, dehydration, anemia and hypoproteinemia repeatedly occur. Colonoscopy identified scattered white scars and patchy hyperemia of the small intestine mucosa, and pebbly nodules of the sigmoid colon mucosa with hyperemia. Pathological findings revealed (1) mild to moderate chronic inflammation of superficial mucosal tissues of the small intestine, accompanied by mild acute inflammation and lymphoid hyperplasia; (2) severe chronic inflammation of sigmoid colon, accompanied by moderate, acute inflammation and lymphoid hyperplasia; (3) congestion and edema of ileocecal intestinal mucosa, partial mucosal degeneration and exudation, scattered inflammatory cell infiltration of intestinal wall, accompanied by mesenchymal hyperemia. At age 3, "total colectomy", "rectum scaphoid fossa fistula" and "jejunal anostomy" were performed. The overall postoperative condition was improved, with normal growth and development. However, fever, infection, dehydration, diarrhea and infections frequently occurred. After symptomatic treatment, her symptoms were gradually relieved. At age 5, pathology revealed (1) mild edema of rectal mucosa with infiltration of chronic inflammatory cells (lymphocytes, plasma cells, and eosinophils), (2) no reduction of goblet cells, (3) extensive Paneth cell metaplasia, (4) mild conjunctival tissue hyperplasia of lamina propria, and (5) no granulomatous lesions.

Her parents reported a family history of immune diseases of varying degrees. Her father suffered from mild asthma as a child, and recovered after puberty. Her mother suffered from mild dermatitis and her grandmother suffered from mild asthma. None of the parents had a family history of IBD. In 2017, blood samples of the patient and her parents were collected and examined at the Joint Laboratory of Translational Medicine Research, Liaocheng People's Hospital. This study was approved by the Ethics Committee of Liaocheng People's Hospital. All subjects signed informed consent, and the child's written informed consent was provided by her legal guardians.

Whole Exome Sequencing

Firstly, genomic DNA was extracted from whole blood by QIAamp DNA blood MIDI Kit (Qiagen, Germany), and quality control was performed by nanodrop (Thermo Fisher Scientific, USA), OD 260/280 ratio of 1.8-2.0.

Secondly, DNA was interrupted into ~200 bp fragments with Covaris S220. Breaking parameters were setup as follows: Duty factor 10%; Peak Incident Power 175; Cycles per Burst 200; Treatment time 360s; and Bath Temperature 4°C-8 °C. Agilent 2100 was used for quality control.

Thirdly, Agilent Sureselect DNA Targeting Sequence Capture Kit was applied for library preparation as follows: (1) end repair was performed on fragmented DNA. The A was added to 3' end, and gap was connected with adapters. After each step, AMPure XP beads was used for purification. (2) Polymerase chain reaction (PCR) was performed with an amplification volume of 50 ul. The program was set up as follows: 98°C pre-denaturation for 2 min; 98°C denaturation for 30 s, 65°C annealing for 30 s, 72°C extension for 1 min, total 10 cycles; 72°C extension 10 min; 4°C, hold. The product was purified with AMPure XP beads. (3)

Amplified DNA was hybridized and placed at 65°C for 16-24 h. (4) After hybridization, stranded penicillin magnetic beads were applied for probe capture and PCR amplification. The amplification volume was 50 ul. The program was set up as follows: pre-denaturation at 98°C for 2 min; denaturation at 98°C for 30 s, annealing at 57°C for 30 s, and extension at 72°C for 1 min; 12 cycles; 72°C extension for 10 min; 4°C, hold. (5) AMPure XP beads were used for purification, Aglinet 2100 was used for quality control. The fragment size was about 250bp-350bp, and library preparation was completed.

Finally, Nextseq 500 (Illumina) was applied for PE75 sequencing.

Screening and functional analysis of candidate pathogenic genes

Trimmatomatic was used to remove the original sequencing connector and low-quality sequences. Filtered sequences were aligned to Human genome HG19 using BWA. Sequences were deduplicated with Picard. Using GATK software, single nucleotide variation (SNV) and indel mutation (In/Del) were analyzed and filtered. The called variants were annotated with ANNOVAR. Loci with a minor allele frequency (MAF) <1% in 1000 Genomes Project (1000G), Exome Aggregation Consortium (ExAC), and ESP6500 databases were screened. Furtherly, sites with non-synonymous variants were selected. SIFT, Polyphen2_HDIV, LRT, MutationTaster and PROVEAN were used to predict potential functions of selected variants. When an individual variant site was indicated to be harmful according to the three databases, its corresponding gene was classified as a “high-risk pathogenic variant”.

GeneCards database was searched for CD-related genes and those with association of >20% were selected as “known CD pathogenic genes”. Colon tissue specific genes were searched in the TiGER database. Those genes overlapped between “high-risk pathogenic variants” and “known CD pathogenic genes” were selected. The protein-protein interaction network was constructed with Cytoscape3.6. The genes with protein-protein interactions were listed as candidate pathogenic genes.

KOBAS was used for signal enrichment analysis of candidate pathogenic genes. Swiss-model was used to predict the 3D structure of candidate pathogenic gene encoding protein. The SwissPDB viewer was used to identify structural changes before and after a mutation was introduced and to estimate alterations in force fields. InterProScan was used to predict protein domains.

Sanger sequencing

Based on the hG19 sequence of reference human genome, primers on both sides of candidate pathogenic gene loci were designed (Table 1). Primers were synthesized by Sangon Biotech (Shanghai). Loci were amplified by PCR and sent to Sangon Biotech for Sanger sequencing.

Results

Identification and functional analysis of candidate pathogenic genes

The average whole exome sequencing depth of the target capture region from the child and her parents was 134×, 89× and 75×, respectively. The sequencing coverage was more than 99.9%. A total of 131169 variant loci were detected in the patient, including 12200 non-synonymous variant loci, as well as insertion/deletion loci in exons and variable splicing regions. Using 1000G database, ExAC database and Esp6500 database, the loci with a MAF>1% were removed. The remaining 676 variant loci were selected. After predicting harmful variant sites, a total of 128 pathogenicity sites were identified, located on 124 genes (Supplementary Table 1).

GeneCards obtained “known CD-related genes”. Totally, 375 genes with disease correlation >20% were identified (Supplementary Table 2). Five genes of the 124 high-risk pathogenic genes carried by the patient were associated with CD: *NOD2*, *IL10RA*, *PRODH*, *PLA2G6*, and *COL7A1*. A total of 199 genes specifically expressed in normal colon were queried by TiGER (Supplementary Table 3). Through protein-protein interaction analysis, *NOD2*, *IL10RA*, *PLA2G6* and *COL7A1* genes were predicted to interact with colon-specific genes (Figure 1, Supplementary Table1).

Sanger sequencing identified that heterozygous variants of *NOD2* c. 2257 C>T, *IL10RA* c. 301 C>T, *PLA2G6* c. 2029 C>T and *COL7A1* c. 3190 G>A were carried by the patient. Her father carried heterozygous variants of *NOD2* c.2257 C>T and *IL10RA* c.301 C>T, while her mother carried heterozygous variants of *PLA2G6* c.2029 C>T and *COL7A1* c.3190 G>A (Figure 2).

Based on analysis of gene functions, *NOD2* might activate MAPK pathway and NF- κ B pathway, and promote the transcriptional expression of immune-related genes, which in turn participate in the gastrointestinal immune response. *IL10RA* is a cell surface receptor for the cytokine IL10. *IL10RA* activates the JAK-STAT pathway, which promotes the expression of anti-inflammatory genes to restrict excessive tissue destruction caused by inflammation. *PLA2G6* is a non-calcium-dependent phospholipase, involved in phospholipid remodeling, cell membrane homeostasis and signal transduction. *COL7A1* encodes $\alpha 1$ chain of collagen type VII that interacts with extracellular matrix to form anchored fibrils. These fibrils contribute to the formation and adhesion of epithelial basement membranes. Based on pathway analysis, these four genes were involved Cytokine-Cytokine receptor interaction 3, TNF signaling pathway, and JAK-STAT signaling pathway. We hypothesized that *NOD2* and *IL10RA* variants might cause excessive intestinal inflammatory response and subsequent damage to the intestinal wall. Variants of *PLA2G6* and *COL7A1* might dysregulate proliferation and differentiation of intestinal epithelial cells, undermine the repair of intestinal wall, and thus aggravate phenotypes of CD.

Protein structure prediction and functional analysis

These variants in coding regions of genes might change proteins' structures, functions and signal transduction, leading to biological dysfunction. Accordingly, *NOD2* R753W, *IL10RA* R101W, *PLA2G6* R677C and *COL7A1* D1064N might affect protein stability (Figure 3).

The 101th amino acid of *IL10RA* is located on the FN3 domain, while the 1064th amino acid of *COL7A1* is located on VWF-A domain (Figure 4). The 753th amino acid of *NOD2* is located on HD2 domain, the 677th amino acid of *PLA2G6* is not located on PNPLA domain (Supplementary Figure 1). FN3 domain is involved in coagulation, inflammation, as well as cell adhesion and migration [21-22]. The VWF-A domain is involved in cell adhesion, migration, pattern formation and signal transduction [23-24]. The ARG to TRP mutation on 101th amino acid of *IL10RA* might result in intestinal disorder and overactivated inflammatory response. The ASP to ASN mutation on 1064th amino acid of *COL7A1* might impair the intestinal wall barrier function and disturbed the intestinal microenvironment, leading to severe CD phenotypes.

Discussion

In this study, whole exome sequencing was performed on the core family of a patient with CD at birth. *NOD2* c. 2257 C>T, *IL10RA* c. 301 C>T, *PLA2G6* C. 2029 c>T, *COL7A1* c. 3190 G>A were identified by deep mining of CD-related genes. It is speculated that a rare, very early-onset and severe CD phenotype attributed to superposition effects of *NOD2*, *L10RA*, *PLA2G6* and *COL7A1* mutated genes.

NOD2 was discovered as a CD-related gene. *NOD2* participates in gastrointestinal immune response, activates MAPK and NF- κ B signaling pathways, and upregulates immune-related genes, such as IL-1, IL-10 and IL-12 [25]. *IL10RA* inhibits the synthesis of pro-inflammatory cytokines [26]. *IL-10RA* gene variant was found in 7 of the 14 children with CD-onset under age 1 [27]. *NOD2* c.2257 C>T and *IL10RA* c.301 C>T variants carried by the patient were inherited from her father. *NOD2* variant correlates with dysregulated NF- κ B signaling pathway, while continuous expression of inflammatory factors leads to excessive gastrointestinal inflammatory response. Meanwhile, *IL10RA* variant relates to reduced or loss of anti-inflammatory function. When the excessive inflammatory response cannot be alleviated, intestinal wall tissue would be injured. Notably, her father suffered from mild asthma as a child and recovered after puberty, without clinical manifestations of intestinal inflammation and CD. Thus, simultaneous *NOD2* and *IL10RA* variants are not sufficient to cause a severe clinical phenotype of CD.

PLA2G6 encodes phospholipase A2 protein, which is involved in phospholipid remodeling, cell membrane homeostasis and signal transduction. *COL7A1* acts as an anchor for the protofilament between the epithelium

and the stroma, which contributes to the formation and adhesion of basement membrane [28]. The frequency of *COL7A1* rs3197999 allele was significantly higher in IBD patients than in healthy subjects [29]. *PLA2G6* c.2029 C>T and *COL7A1* c.3190 G>A variants were inherited from her mother. *PLA2G6* variant may change cell membrane structure, result in disorder of cell surface receptor signal transduction, and affect cell proliferation and differentiation. *COL7A1* variant may cause weakened or lost interaction with extracellular matrix. Subsequently, epithelial basement membrane cannot be formed, while intestinal wall barrier function and repair capability would be reduced or impaired, which aggravates damage to intestinal wall tissue.

Ethics declaration

This study was approved by the Ethics Committee of Liaocheng People's Hospital. All subjects had signed informed consent. Written informed consent of children was provided by their legal guardians.

Data availability

The variation data reported in this paper have been deposited in the Genome Variation Map (GVM) in Big Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Science (under accession number GVM000319, accessible at <http://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000319>)

Author contributions

M.G. designed the study and revised the manuscript. Q.Z. and X.W. interpreted the results and drafted the manuscript. J.Z. collected the samples. J.Z. and R.L. processed the samples. X.W. and Q.L. performed data analyses, and X.J. assisted with the analysis. All authors discussed the results and commented on the article.

References

- Roda, G., Ng, S.C., Kotze, P.G., Argollo, M., Panaccione, R., Spinelli, A., Kaser, A., Peyrin-Biroulet, L., & Danese, S. (2020). Crohn's disease. *Nature Reviews Disease Primers*, 6(1), 22. <https://doi.org/10.1038/s41572-020-0156-2>
- Torres, J., Mehandru, S., Colombel, J.F., & Peyrin-Biroulet, L. (2017). Crohn's disease. *Lancet*, 389(10080), 1741-1755. [https://doi.org/10.1016/S0140-6736\(16\)31711-1](https://doi.org/10.1016/S0140-6736(16)31711-1)
- Li Yim, A.Y.F., Duijvis, N.W., Ghiboub, M., Sharp, C., Ferrero, E., Mannens, M.M.A.M., D'Haens, G.R., de Jonge, W.J., Velde, A.A.T., & Henneman, P. (2020). Whole-Genome DNA Methylation Profiling of CD14+ Monocytes Reveals Disease Status and Activity Differences in Crohn's Disease Patients. *Journal of Clinical Medicine*, 9(4), 1055. <https://doi.org/10.3390/jcm9041055>
- Ng, S.C., Shi, H.Y., Hamidi, N., Underwood, F.E., Tang, W., Benchimol, E.I. Panaccione, R., Ghosh, S., Wu, J.C.Y., Chan, F.K.L., Sung, J.J.Y., & Kaplan, G.G. (2018). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*, 390(10114), 2769-2778. [https://doi.org/10.1016/S0140-6736\(17\)32448-0](https://doi.org/10.1016/S0140-6736(17)32448-0)
- Zhao, J., Ng, S.C., Lei, Y., Yi, F., Li, J., Yu, L., Zou, K., Dan, Z., Dai, M., Ding, Y., Song, M., Mei, Q., Fang, X., Liu, H., Shi, Z., Zhou, R., Xia, M., Wu, Q., Xiong, Z., Zhu, W., Deng, L., Kamm, M.A., & Xia, B. (2013). First prospective, population-based inflammatory bowel disease incidence study in mainland of China: the emergence of 'western' disease. *Inflammatory Bowel Diseases*, 19(9), 1839-1845. <https://doi.org/10.1097/MIB.0b013e31828a6551>
- Kaplan, G.G., & Ng, S.C. (2016). Globalisation of inflammatory bowel disease: perspectives from the evolution of inflammatory bowel disease in the UK and China. *The Lancet Gastroenterology & Hepatology*, 1(4), 307-316. [https://doi.org/10.1016/S2468-1253\(16\)30077-2](https://doi.org/10.1016/S2468-1253(16)30077-2)
- Gordon, H., Moller, F.T., Andersen, V., & Harbord, M. (2015). Heritability in Inflammatory Bowel Disease: From the First Twin Study to Genome-Wide Association Studies. *Inflammatory Bowel Diseases*, 21(6), 1428-1434. <https://doi.org/10.1097/MIB.0000000000000393>

- Liu, J.Z., & Anderson, C.A. (2014). Genetic studies of Crohn's disease: Past, present and future. *Best Practice & Research. Clinical Gastroenterology*, 28(3), 373-386. <https://doi.org/10.1016/j.bpg.2014.04.009>
- Huang, H., Fang, M., Jostins, L., Mirkov, M.U., Boucher, G., Anderson, C.A., Andersen, V., Cleynen, I., Cortes, A., Crins, F., D'Amato M., Deffontaine V., Dmitrieva, J., Docampo, E., Elansary M., Farh K.K.H., Franke, A., Gori A.S., Goyette P., Halfvarson, J., Haritunians, T., Knight, J., Lawrance, I.C., Lees, C.W., Louis, E., Mariman, R., Meuwissen, T., Mni, M., Momozawa, Y., Parkes, M., Spain, S.L., Théâtre, E., Trynka, G., Satsangi, J., van Sommeren, S., Vermeire, S., J Xavier, R.J., Weersma, R.K., Duerr, R.H., Mathew, C.G., Rioux, J.D., McGovern, D.P.B., Cho, J.H., Georges, M., Mark J Daly, M.J., & Barrett, J.C. (2017). Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*, 547(7662), 173-178. <https://doi.org/10.1038/nature22969>
- Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schäffeler, E., Schlee, M., Herrlinger, K.R., Stallmach, A., Noack, F., Fritz, P., Schröder, J.M., Bevins, C.L., Fellermann, K., & Stange, E.F. (2004). NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut*, 53(11), 1658-1664. <https://doi.org/10.1136/gut.2003.032805>
- Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., Kishi, C., Kc, W., Carrero, J.A., Hunt, S., Stone, C.D., Brunt, E.M., Xavier, R.J., Sleckman, B.P., Li, E., Mizushima, N., Stappenbeck, T.S., & Virgin 4th, H.W. (2008). A key role for autophagy and the autophagy gene ATG16L1 in mouse and human intestinal Paneth cells. *Nature*, 456(7219), 259-263. <https://doi.org/10.1038/nature07416>
- Thachil, E., Hugot, J.P., Arbeille, B., Paris, R., Grodet, A., Peuchmaur, M., Codogno, P., Barreau, F., Ogier-Denis, E., Berrebi, D., & Viala, J. (2012). Abnormal activation of autophagy-induced crinophagy in Paneth cells from patients with Crohn's disease. *Gastroenterology*, 142(5), 1097-1099.e4. <https://doi.org/10.1053/j.gastro.2012.01.031>
- Zhang, Q., Pan, Y., Yan, R., Zeng, B., Wang, H., Zhang, X., Li, W., Wei, H., & Liu, Z. (2015). Commensal bacteria direct selective cargo sorting to promote symbiosis. *Nature Immunology*, 16(9), 918-926. <https://doi.org/10.1038/ni.3233>
- Kaser, A., Lee, A.H., Franke, A., Glickman, J.N., Zeissig, S., Tilg, H., Nieuwenhuis, E.E.S., Higgins, D.E., Schreiber, S., Glimcher, L.H., & Blumberg, R.S. (2008). XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell*, 134(5), 743-756. <https://doi.org/10.1016/j.cell.2008.07.021>
- Adolph, T.E., Tomczak, M.F., Niederreiter, L., Ko, H.J., Janne Böck, J., Eduardo Martinez-Naves, E., Jonathan N Glickman, J.N., Tschurtschenthaler, M., Hartwig, J., Hosomi, S., Flak, M.B., Cusick, J.L., Kohno, K., Iwawaki, T., Billmann-Born, S., Raine, T., Bharti, R., Lucius, R., Kweon, M.N., Marciniak, S.J. Choi, A., Hagen, S.J., Schreiber, S., Rosenstiel, P., Kaser, A., & Blumberg, R.S. (2013). Paneth cells as a site of origin for intestinal inflammation. *Nature*, 503(7475), 272-276. <https://doi.org/10.1038/nature12599>
- Willson, T.A., Jurickova, I., Collins, M., & Denson, L.A. (2013). Deletion of intestinal epithelial cell STAT3 promotes T-lymphocyte STAT3 activation and chronic colitis following acute dextran sodium sulfate injury in mice. *Inflammatory Bowel Diseases*, 19(3), 512-525. <https://doi.org/10.1097/MIB.0b013e31828028ad>
- Zhu, L., Shi, T., Zhong, C., Wang, Y., Chang, M., & Liu, X. (2017). IL-10 and IL-10 receptor mutations in very early onset inflammatory bowel disease. *Gastroenterology Research*, 10(2), 65-69. <https://doi.org/10.14740/gr740w>
- Ellinghaus D, Luke Jostins L, Spain SL, Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hübenthal, M., Folseraas, T., Wang, Y., Esko, T., Metspalu, A., Westra, H.J., Franke, L., Pers, T.H., Weersma, R.K., Collip, V., D'Amato, M., Halfvarson, J., Jensen, A.B., Lieb, W., Degenhardt, F., Forstner, A.J., & Hofmann, A. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nature Genetics*, 48(5), 510-518. <https://doi.org/10.1038/ng.3528>

- Xavier, R.J., & Podolsky, D.K. (2007). Unravelling the Pathogenesis of Inflammatory Bowel Disease. *Nature*, 448(7152), 427-434. <https://doi.org/10.1038/nature06005>
- Guha, P., Das, A., Dutta, S., & Chaudhuri, T.K. (2018). A rapid and efficient DNA extraction protocol from fresh and frozen human blood samples. *Journal of Clinical Laboratory Analysis*, 32(1), e22181. <https://doi.org/10.1002/jcla.22181>
- Petersen, T.E., Thogersen, H.C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L., & Magnusson, S. (1983). Partial primary structure of bovine plasma fibronectin: three types of internal homology. *Proceedings of the National Academy of Sciences of the United States of America*, 80(1), 137-141. <https://doi.org/10.1073/pnas.80.1.137>
- Leahy, D.J., Aukhil, I., & Erickson, H.P. (1996). 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell*, 84(1), 155-164. [https://doi.org/10.1016/s0092-8674\(00\)81002-8](https://doi.org/10.1016/s0092-8674(00)81002-8)
- Ruggeri, Z.M., & Ware, J. (1993). von Willebrand factor. *FASEB Journal*, 7(2), 308-316. <https://doi.org/10.1096/fasebj.7.2.8440408>
- Colombatti, A., Bonaldo, P., & Doliana, R. (1993). Type A modules: interacting domains found in several non-fibrillar collagens and in other extracellular matrix proteins. *Matrix*, 13(4), 297-306. [https://doi.org/10.1016/s0934-8832\(11\)80025-9](https://doi.org/10.1016/s0934-8832(11)80025-9)
- Ogura, Y., Bonen, D.K., Inohara, N., Nicolae D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H., Achkar, J.P., Brant, S.R., Bayless, T.M., Kirschner, B.S., Hanauer, S.B., Nuñez, G., & Cho, J.H. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, 411(6837), 603-606. <https://doi.org/10.1038/35079114>
- Moore, K.W., R de Waal Malefyt, R., Coffman, R.L., & O'Garra, A. (2001). Interleukin-10 and the interleukin-10 receptor. *Annual Review of Immunology*, 19, 683-765. <https://doi.org/10.1146/annurev.immunol.19.1.683>
- Shim, J.O., & Seo, J.K. (2014). Very early-onset inflammatory bowel disease (IBD) in infancy is a different disease entity from adult-onset IBD; one form of interleukin-10 receptor mutations. *Journal of Human Genetics*, 59(6), 337-341. <https://doi.org/10.1038/jhg.2014.32>
- Christiano, A.M., Greenspan, D.S., & Lee S. (1994). Cloning of human type VII collagen: complete primary sequence of the α1(VII) chain and identification of intragenic polymorphisms. *The Journal of Biological Chemistry*, 269(32), 20256-20262.
- Narzo, A.F.D., Telesco, S.E., Brodmerkel, C., Argmann, C., Peters, L.A., Li, K., Kidd, B., Dudley, J., Cho, J., Schadt, E.E., Kasarskis, A., Dobrin, R., & Hao, K. (2017). High-Throughput Characterization of Blood Serum Proteomics of IBD Patients with Respect to Aging and Genetic Factors. *PLoS Genetics*, 13(1), e1006565. <https://doi.org/10.1371/journal.pgen.1006565>

Tables

Table 1. Primers used in PCR for amplification of candidate pathogenic genes

Gene	Primer sequence (forward/reverse)	Product size (bp)
<i>NOD2</i>	PF 5'- CTTCAATTGTG- GCAGGCCAGG -3' PR 5'- CAATGTCACCCACAGAGTTG- TAGTC -3'	530

Gene	Primer sequence (forward/reverse)	Product size (bp)
<i>IL10RA</i>	PF 5'- AATGAGCTGCCGTG- GACTAGA -3' PR 5'- AAAGTCCTGCCCTGTTTGGG -3'	480
<i>COL7A1</i>	PF 5'- GGCAGAAAGGTGT- GTCTGGG -3' PR 5'- CATTCAGTGGGAACAGTGGGGAG -3'	530
<i>PLA2G6</i>	PF 5'- CAACACGCACACCCTGAGAT -3' PR 5'- TCCCCACTCATG- CACACTTAGA -3'	480

Figure Legends

Figure 1. Diagram of interaction between candidate pathogenic genes and colon-specific gene proteins.

Figure 2. Genetic patterns and sequencing peaks of candidate pathogenic genes.

Figure 3. Schematic diagrams of 3D structures of candidate proteins. (A-D): The changes in protein structures attributed to NOD2 R753W, IL10RA R101W, PLA2G6 R677C and COL7A1 D1064N. The green solid line represents the amino acids at the variant site; the green dotted line represents hydrogen bonds.

Figure 4. Schematic diagrams of predicted domains of IL10RA and COL7A1 proteins.

Figure S1 Schematic diagrams of predicted domains of NOD2 and PLA2G6 proteins.



