Introducing GATA3 as a prominent player in Crohn’s disease

Sina Rezaei-Tavirani1, Nastaran Asri2, MohammadAli Emamhadi3, Somayeh Jahani-Sherafat1, Ali Seyed Salehi1, Zahra Gholamrezaei1, Elena Lak2
1Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3Forensic Medicine Department, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4Laser Application in Medical Sciences Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
5Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT
Aim: This study was aimed at gene assessment of Crohn's disease (CD) through protein-protein interaction (PPI) network analysis to find crucial genes.

Background: CD is a major subtype of inflammatory bowel diseases (IBD), which affects gastrointestinal tract. PPI network analysis is a suitable tool to clarify a critical gene as a drug target or diagnostic biomarker for these types of diseases.

Methods: Gene expression profile GSE126124 of 20 CD patients and 20 healthy controls was obtained from the Gene Expression Omnibus (GEO) database. RNA profile of peripheral blood mononuclear cells (PBMCs) and colon biopsy samples of the studied groups was investigated. Crucial genes were selected and analyzed via the PPI network by Cytoscape software. Gene ontology enrichment for the hubs, bottlenecks, and hub-bottlenecks was performed via CluGO plugin of Cytoscape software.

Results: Eighty-one differentially expressed genes (DEGs) among 250 initial DEGs were highlighted as significant by FC>2 and p-value ≤ 0.05, and 69 significant DEGs were used for PPI network construction. The network was characterized by poor connections, so 20 top neighbors were added to form a scale-free network. The main connected component included 39 query DEGs and 20 added first neighbors. Three clusters of biological processes associated with crucial genes were identified and discussed.

Conclusion: The results of this study indicated that GATA3 has a key role in CD pathogenesis and could be a possible drug target or diagnostic biomarker for Crohn’s disease.

Keywords: Crohn’s disease, Genes, Gene ontology, GATA3 transcription factor.

Introduction
Inflammatory bowel diseases (IBD) are chronic relapsing disorders with the ability to affect the entire intestine accompanied by long-term morbidity (1-3). The two main subgroups of IBD are Crohn's disease (CD) and ulcerative colitis (UC) (4, 5). CD affects gastrointestinal (GI) tract and causes abdominal pain, fever, and clinical signs of bowel obstruction (5-8). A combination of genetic and environmental risk factors associated with CD susceptibility and numerous studies have pointed to the role of different gene groups in the pathophysiology of this disease (9-12). Different genetic studies using molecular biology, high throughput methods, and bioinformatics approaches tried to discover the molecular mechanism and biochemical pathways involved in CD development (13, 14). Protein-protein interaction (PPI) network analysis is an approach to evaluate dysregulated genes...
GATA3 and Crohn’s Disease

(queried genes) in patients compared to healthy individuals through a scale-free network. The topological properties of the created scale-free network contained useful information to discriminate queried dysregulated genes (15, 16). Centrality measures such as “Degree centrality (K)” and “Betweenness centrality (BC)” are used in analyzing the role of network elements. Degree refers to the number of the first neighbor of a node (equal to the number of links between the studied nodes with its first neighbors) and BC is a function of the shortest path where the node is involved. Degree and BC are widely used to analyze the studied network (17). Based on the degree value and BC amount, three types of critical nodes can be determined: the hub, bottleneck, and hub-bottleneck nodes. The top nodes based on degree value and BC are known as hub and bottleneck, respectively. Hub-bottleneck nodes are regarded as the common hub and bottleneck elements of the network.

This study aimed to find a possible new biomarker for Crohn’s disease through analyzing and screening related genes. The crucial genes were enriched via the gene ontology method.

Methods

Gene expression profile GSE126124 of a total of 20 CD patients (10 males and 10 females) and 20 healthy controls (10 males and 10 females) were extracted from Gene Expression Omnibus (GEO) database. RNA profile of peripheral blood mononuclear cells (PBMCs) and colon biopsy samples from patients with CD vs. controls was investigated. The age of patients and controls were between 8-18 years old. Gene expression distributions were evaluated through boxplot analysis using GEO2R. The top 250 differentially expressed genes (DEGs) based on p-value amounts were selected for more investigation. Eighty-one DEGs that were characterized by Fold change (FC) > 2 and p-value ≤ 0.05 were considered as significant DEGs. The significant DEGs plus the 20 first neighbors were included in a PPI network by Cytoscape software v 3.7.2 and STRING. The constructed network was analyzed by Network analyzer (an application of Cytoscape) to identify crucial genes. Degree and Betweenness centrality (BC) were considered to find critical DEGs. The top 10% of the main connected component network based on degree value and BC were considered as hub and bottleneck nodes, respectively. The nodes that were both hub and bottleneck were determined as hub-bottlenecks. Gene ontology enrichment for the hubs, bottlenecks, and hub-bottlenecks was performed via the CluGO plugin of Cytoscape software. The significant biological terms based on p-value ≤ 0.05 were identified. The terms were created in three groups, which were connected to hubs, bottlenecks, and hub-bottleneck.

Results

Gene expression profiles of 20 CD patients and 20 controls were matched by boxplot analysis (Figure 1).

Figure 1. Distribution of gene expression changes in CD patients and controls. Blue-colored bars: CD patients | Pink-colored bars: Controls

Since the middle line of samples was matched, the studied profiles were comparable. The top 250 DEGs, based on p-value, were identified by the GEO2R analyzer. Among them, 81 DEGs were characterized as the significant DEGs by FC>2 and p-value ≤ 0.05. Sixty-nine (85%) of the 81 significant DEGs were recognized by the STRING database and were candidate to construct the PPI network. The created network was not considered as scale-free because of the
weak interactions between the included DEGs. As most of the queried DEGs were isolated, 20 first neighbors were added to the queried DEGs to form a scale-free network (Figure 2).

![Figure 2](image-url)

**Figure 2.** The main constructed PPI network. The 59 genes (39 queried DEGs and 20 first neighbors) were constructed as a PPI network. The network is a layout by the degree value of the nodes. The larger size of the node corresponds to a higher value of the degree. The color change from purple to green indicates an increment of degree value.

Thirty isolated query DEGs were excluded from further investigation, and therefore the constructed scale-free network included 59 linked DEGs, 31 of which were identified as connected components. Hence, the main connected component included 39 query DEGs and 20 added first neighbors, and the central nodes were selected among the 39 query DEGs for further evaluations. To investigate which nodes play a more important role in the main connected component constructed network, two central parameters (degree and BC) were considered. The top 10% (4 individual) of nodes based on degree value were introduced as the hub nodes. The bottleneck nodes were identified as the top 10% of nodes based on betweenness centrality. The hubs (GATA3, IL-7, CD28, CD5) and bottlenecks (TXN [Thioredoxin], ETS-1, GATA3, CCR7 [CC-chemokine receptor 7]) are presented in Table 1 and 2, respectively.

### Table 1. Top 10% of queried nodes based on the highest degree value presented as hub nodes.

| R | Name | K  | BC    |
|---|------|----|-------|
| 1 | GATA3* | 28 | 0.038 |
| 2 | IL-7  | 26 | 0.007 |
| 3 | CD28  | 26 | 0.006 |
| 4 | CD5   | 25 | 0.005 |

* hub-bottleneck

### Table 2. Top 10% of queried nodes based on the highest betweenness centrality presented as bottleneck nodes.

| R | Name   | K  | BC    |
|---|--------|----|-------|
| 1 | TXN    | 9  | 0.069 |
| 2 | ETS-1  | 19 | 0.042 |
| 3 | GATA3* | 28 | 0.038 |
| 4 | CCR7   | 23 | 0.036 |

* hub-bottleneck

These seven genes have been considered crucial genes, and since GATA3 is common in hubs and bottlenecks, it is termed hub-bottleneck. Three clusters of biological processes related to the seven crucial genes were determined (Figure 3). For better understanding, the biological processes were screened to find the role of crucial genes (Table 3).

### Discussion

The proteomic studies have provided valuable information about the molecular mechanism of
disorders leading to the development of therapeutic and
diagnostic methods (18). Results of proteomic or
genomic studies contain a large amount of differentially
expressed genes/ proteins (DEGs/ DEPs) between
patients and healthy controls. As a bioinformatics
approach, PPI network analysis is closely tied to
proteomics and genomics investigations (19, 20). In
this study, PPI network analysis is applied to screen the
known proteins or genes related to Crohn’s disease.
Among the 81 queried DEGs related to Crohn’s
disease, 7 genes (IL-7, GATA3, CD28, CD5, TXN,
ETS-1, and CCR7) were introduced as dysregulated
genes.

Dysregulation of cytokines and growth factors play a
major role in the malfunction of the immune system in
inflammatory bowel disease (21). IL-7 signals are
involved in the development of chronic intestinal
inflammation and human intestinal diseases (22). As
depicted in figure 3 and table 3, IL-7 is related to the T
cell differentiation in the thymus, which is the
dominant class of biological terms and contains 16
biological terms (about 89% of all known terms).
According to the results of studies, the lymphocyte
population is reduced in the peripheral blood of CD
patients, which is due to a decrease in IL-7 levels (23).
Kader et al. (24) reported significantly higher levels of
IL-7 in CD patients in remission compared to patients
with active form.

GATA3 is the only hub-bottleneck node that is
characterized by degree value=28 and BC=0.038. Hub-
bottleneck nodes are known as the key element of the
PPI network. There are several studies on the role of
the hub-bottleneck nodes in the incidence and
development of diseases (16). Like IL-7, GATA3 is
involved in T cell differentiation in the thymus.
Moreover, GATA3 plus ETS-1 are the two
dysregulated genes associated with positive regulation
of pri-miRNA transcription by RNA polymerase II.
GATA3 is a transcription factor with the ability to
control Foxp3 (Forkhead box P3) expression and
regulatory T (Treg) cells function. GATA3 is an
essential factor for immune tolerance, and its removal
is associated with autoimmune diseases (25). Li et al.
(26) reported decreased Gata3+ Foxp3+ cell population
in severe CD patients compared to controls.

Research has shown that TXN is involved in various
biological pathways and processes (27). TXN activities
include the protection of proteins against aggregation
and oxidative stress, helping cells to encounter
environmental stresses. They also have a direct role in
apoptosis, inflammatory processes and cell growth (28-
31). Nitric oxide pathway is a well-recognized
biological term in Crohn’s disease (32). In our
investigation, the role of TXN in response to nitric
oxide is highlighted. Oxidative stress has an important
role in IBD pathogenesis and reduced and oxidized
thioredoxin have a key role in intestinal redox biology

---

**Table 3.** The biological terms and pathways concerning hubs, bottlenecks, and hub-bottleneck genes including IL-7, GATA3, TXN, ETS-1, CCR7, CD28, and CD5.

| R  | Goterm                                                                 | gene name       |
|----|------------------------------------------------------------------------|-----------------|
| 1  | response to nitric oxide                                               | CCR7, TXN       |
| 2  | positive regulation of pri-miRNA transcription by RNA polymerase II   | ETS-1, GATA3    |
| 3  | T cell differentiation in thymus                                        | CCR7, CD28, GATA3, IL-7 |
| 4  | inflammatory response to antigenic stimulus                            | GATA3           |
| 5  | interleukin-4 production                                               | CD28, GATA3     |
| 6  | acute inflammatory response to antigenic stimulus                      | CCR7, GATA3     |
| 7  | lymphocyte costimulation                                               | CCR7, CD28      |
| 8  | regulation of inflammatory response to antigenic stimulus             | CCR7, CD28, CD5 |
| 9  | regulation of interleukin-4 production                                 | CD28, GATA3     |
| 10 | hypersensitivity                                                       | CCR7, CD28      |
| 11 | positive regulation of inflammatory response to antigenic stimulus    | CCR7, CD28      |
| 12 | positive regulation of interleukin-4 production                        | CD28, GATA3     |
| 13 | T cell costimulation                                                   | CCR7, CD28, CD5 |
| 14 | T cell selection                                                       | CCR7, CD28, GATA3 |
| 15 | negative T cell selection                                              | CCR7, CD28      |
| 16 | thymic T cell selection                                                | CCR7, CD28, GATA3 |
| 17 | selective expression of chemokine receptors during T-cell polarization | CCR7, CD28      |
| 18 | negative thymic T cell selection                                       | CCR7, CD28      |
was highlighted as the most important CD-related gene with a dramatic difference in expression appearing as a single hub-bottleneck of the network.

Through network analysis, we can introduce IL-7, GATA3, TXN, ETS-1, CCR7, CD28, and CD5 as Crohn’s disease-related critical genes and possible biomarker panel. GATA3 was highlighted as the key element among them, and it was concluded that GATA3 might be a possible drug target or diagnostic biomarker for the disease. Notably, these differentially expressed genes can be investigated in greater detail in further studies with larger number of CD patients.

Acknowledgment

This project is financially supported by Shahid Beheshti University of Medical Sciences.

Conflict of interests

The authors declare that they have no conflict of interest.

References

1. Mortazavi S, Zali M, Raouf M, Nadji M, Kowsarian P, Nowroozi A. The Prevalence of Human Papillomavirus in Cervical Cancer in Iran. Asian Pac J Cancer Prev 2002;3:69-72.
2. Farroo A, Naderi N, Moghaddam SJ, Noorinayer B, Firouzi F, Aghazadeh R, et al. The frequency of C3435T MDR1 gene polymorphism in Iranian patients with ulcerative colitis. Int J Colorectal Dis 2007;22:99-1003.
3. Hendrickson BA, Gokhale R, Cho JH. Clinical aspects and pathophysiology of inflammatory bowel disease. Clin Microbiol Rev 2002;15:79-94.
4. Taghipour N, Aghdasi HA, Haghighi A, Mossafa N, Tabaei SJ, Rostami-Nejad M. Potential treatment of inflammatory bowel disease: a review of helminths therapy. Gastroenterol Hepatol Bed Bench 2014;7:9-16.
5. Moura FA, Goulart MOF. Inflammatory Bowel Diseases: The Crosslink between risk factors and antioxidant therapy. In: Gracia Pimentel AM, Rocha R, Santana GO. Crohn's disease of esophagus, stomach and duodenum. World J Gastrointest Pharmacol Ther 2019;10:35-49.
6. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet 2012;380:1590-605.
7. Pimentel AM, Rocha R, Santana GO. Crohn's disease of esophagus, stomach and duodenum. World J Gastrointest Pharmacol Ther 2019;10:35-49.
8. Rostami-Nejad M, Yazdi MH, Nikfar S, Rezaie A, Abdollahi M. Potential Vaccines for Treating Crohn's Disease. Iran Biomed J 2020;24:1-14.
Gastroenterol Hepatol Bed Bench 2020;13(Suppl.1):S53-S59

9. Goudarzi M, Goudarzi H, Alebouyeh M, Azimi Rad M, Shayegan Mehr FS, Zali MR, et al. Antimicrobial susceptibility of clostridium difficile clinical isolates in iran. Iran Red Crescent Med J. 2013;15:704-11.

10. Loddo I, Romano C. Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. Front Immunol 2015:6.

11. Bianco AM, Girardelli M, Tommasini A. Genetics of inflammatory bowel disease from multifactorial to monogenic forms. World J Gastroenterol 2015;21:12296-310.

12. Oshima S, Watanabe M. Genetic and environmental factors drive personalized medicine for Crohn's disease. J Clin Investig 2018;128:4758-60.

13. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. Nat Clin Pract Gastro 2006;3:390-407.

14. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliae B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterol Hepatol Bed Bench 2014;7:17.

15. Rezaei Tavirani M, Mansouri V, Rezaei Tavirani S, Hesami Tackallou S, Rostami-Nejad M. Gliosarcoma protein-protein interaction network analysis and gene ontology. Int J Cancer Manag 2018;11.

16. Zamanian Azodi M, Peyvandi H, Rostami-Nejad M, Safaie A, Rostami K, Vafaee R, et al. Protein-protein interaction network of celiac disease. Gastroenterol Hepatol Bed Bench. 2016;9:268-277.

17. Brandes U. A faster algorithm for betweenness centrality. J Math Sociol 2001;25:163-70.

18. Amiri-Dashatan N, Koushki M, Abbassadeh H-A, Rostami-Nejad M, Rezaei-Tavirani M. Proteomics Applications in Health: Biomarker and Drug Discovery and Food Industry. Iran J Pharm Res 2018;17:1523-36.

19. Vella D, Marini S, Vitali F, Di Silvestre D, Mauri G, Bellazzi R. MTGO: PPI Network Analysis Via Topological and Functional Module Identification. Sci Rep 2018;8:5499.

20. Rao VS, Srinivas K, Sujini GN, Kumar GNS. Protein-protein interaction detection: methods and analysis. Int J Proteomics 2014;2014:147648.

21. Xu XR, Liu CQ, Feng BS, Liu ZJ. Dysregulation of mucosal immune response in pathogenesis of inflammatory bowel disease. World J Gastroenterol 2014;20:3255-64.

22. Watanabe M, Yamazaki M, Kanai T. Mucosal T cells as a target for treatment of IBD. J Gastroenterol 2003;38.

23. Andreu-Ballester JC, Perez-Griera J, Garcia-Ballesteros C, Amigo V, Catalan-Serra I, Monforte-Albalat A, et al. Deficit of Interleukin-7 in Serum of Patients with Crohn's Disease. Inflamm Bowel Dis 2012;19:E30-31.

24. Kader HA, Tchernev VT, Satyaraj E, Lejnine S, Kotler G, Kingsmore SF, et al. Protein microarray analysis of disease activity in pediatric inflammatory bowel disease demonstrates elevated serum PLGF, IL-7, TGF-beta1, and IL-12p40 levels in Crohn's disease and ulcerative colitis patients in remission versus active disease. Am J Gastroenterol 2005;100:414-23.

25. Wang Y, Su MA, Wan YY. An essential role of the transcription factor GATA-3 for the function of regulatory T cells. Immunity 2011;35:337-48.

26. Li J, Ueno A, Iacucci M, Fort Gasia M, Jijon HB, Panaccione R, et al. Crossover Subsets of CD4+ T Lymphocytes in the Intestinal Lamina Propria of Patients with Crohn’s Disease and Ulcerative Colitis. Dig Dis Sci 2017;62:2357-68.

27. Lee S, Kim SM, Lee RT. Thioredoxin and thioredoxin target proteins: from molecular mechanisms to functional significance. Antioxid Redox Signal 2013;18:1165-207.

28. Rand JD, Grant CM. The thioredoxin system protects ribosomes against stress-induced aggregation. Mol Biol Cell 2006;17:387-401.

29. Masutani H, Ueda S, Yodoi J. The thioredoxin system in retroviral infection and apoptosis. Cell Death Differ 2005;12:991-8.

30. Masoodi M, Zali MR, Ehsani-Ardakani MJ, Mohammad-Alizadeh AH, Aiassofi K, Aghazadeh R, et al. Abdominal pain due to lead-contaminated opium: a new source of inorganic lead poisoning in Iran. Arch Iran Med 2006;9:72-5.

31. Powis G, Mustacich D, Coon A. The role of the redox protein thioredoxin in cell growth and cancer. Free Radic Biol Med 2000;29:312-22.

32. Soulii I, Toumi R, Rafi H, Chafia T. Cytokines and Nitric Oxide in Immunopathogenesis of IBD and Potential Therapeutic Approaches. In: Huber S, editor. New Insights into Inflammatory Bowel Disease. 1st edition. London: InTech; 2016. p:125-141.

33. Biasi F, Leonarduzzi G, Oteiza PI, Poli G. Inflammatory bowel disease: mechanisms, redox considerations, and therapeutic targets. Antioxid Redox Signal 2013;19:1711-47.

34. Tian T, Wang Z, Zhang J. Pathomechanisms of Oxidative Stress in Inflammatory Bowel Disease and Potential Antioxidant Therapies. Oxid Med Cell Longev 2017;2017:4535194.

35. Tamaki H, Nakamura H, Nishio A, Nakase H, Ueno S, Uza N, et al. Human thioredoxin-1 ameliorates experimental murine colitis in association with suppressed macrophage inhibitory factor production. Gastroenterology 2006;131:1110-21.

36. Ge Y, Sun M, Wu W, Ma C, Zhang C, He C, et al. MicroRNA-125a suppresses intestinal mucosal inflammation through targeting ETS-1 in patients with inflammatory bowel diseases. J Autoimmun 2019;101:109-20.

37. Li L, Miao X, Ni R, Miao X, Wang L, Gu X, et al. Epithelial-specific ETS-1 (ESE1/ELF3) regulates apoptosis of intestinal epithelial cells in ulcerative colitis via accelerating NF-κB activation. Immunol Res 2015;62:198-212.

38. Genua M, Sgambaro A, Danese S. Editorial: CCR7 is required for leukocyte egress in an experimental model of Crohn's disease-like ileitis. J Leukoc Biol 2015;97:1000-2.
39. Danese S, Gasbarrini A. Chemokines in inflammatory bowel disease. J Clin Pathol 2005;58:1025-7.

40. Worbs T, Förster R. A key role for CCR7 in establishing central and peripheral tolerance. Trends Immunol 2007;28:274-80.

41. Kawashima D, Oshitani N, Jinno Y, Watanabe K, Nakamura S, Higuchi K, et al. Augmented expression of secondary lymphoid tissue chemokine and EBI1 ligand chemokine in Crohn's disease. J Clin Pathol 2005;58:1057-63.

42. Ledbetter JA, Linsley PS. CD28. In: Delves PJ, Editor. Encyclopedia of Immunology. 2nd edition. New York: Academic Press; 1998. p:482-83.

43. van Nieuwenhuijze A, Liston A. The Molecular Control of Regulatory T Cell Induction. In: Liston A, Editor. Progress in Molecular Biology and Translational Science 1st edition. New York: Academic Press; 2015. p:69-97.

44. De Tena JG, Manzano L, Leal JC, San Antonio E, Sualdea V, Álvarez-Mon M. Active Crohn's disease patients show a distinctive expansion of circulating memory CD4+CD45RO+ CD28 null T cells. J Clin Immunol 2004;24:185-96.

45. Soldevila G, Raman C, Lozano F. The immunomodulatory properties of the CD5 lymphocyte receptor in health and disease. Curr Opin Immunol 2011;23:310-8.

46. Voisinne G, Gonzalez de Peredo A, Roncagalli R. CD5, an Undercover Regulator of TCR Signaling. Front Immunol 2018;9.

47. Dalloul A. CD5: a safeguard against autoimmunity and a shield for cancer cells. Autoimmun Rev 2009;8:349-53.