Network Analysis of Common Genes and Transcriptional Factors between Celiac Disease and Inflammatory Bowel Diseases

Fereshteh Izadi1*, Mohammad Hasan Soheilifar2, Hoda Keshmiri Neghab3, Mahya Soheilifar4, Gholamreza Esmaeeli Djavid3

ABSTRACT

BACKGROUND
Understanding the associations among different disorders remarkably improves their diagnosis and treatments. Celiac disease is the most complicated and prevalent form of immune-mediated diseases. On the other hand, inflammatory bowel diseases lead to inflammation of the intestine with an unknown cause. Although inflammatory bowel diseases have been often thought of as an autoimmune disorder, they can be triggered by whatever that can lead to the inflammation in the whole bowel. Henceforth, both aforementioned diseases are related to autoimmune attacks and cause a sort of inflammatory event, which exploring trade-off among them supposedly will lead to discovering important genes and, in turn, to the possible common therapeutic protocols. In the current study, we aimed to determine the correlation between the common genes in celiac disease and inflammatory bowel diseases.

METHODS
314 and 851 genes correlated with celiac disease and inflammatory bowel diseases respectively extracted from DisGeNET were subjected to an in-silico data analysis framework to mine prognosticates genes and the associated pathways.

RESULTS
149 shared genes between these diseases regulated by highlighted transcription factors NFKB1, IRF1, STAT1, HSF1, GATA3 were characterized as discriminating molecules, which by further screening were enriched in pathways mostly involved in apoptosis, T cell activation, and cytokine, chemokine, and interleukin signaling.

CONCLUSION
We observed that the identified common genes were associated with a wide range of pathogenic mechanisms underlying these diseases.

KEYWORDS: Celiac disease, Inflammatory bowel disease, Disease-associated genes, Network analysis

disease (IBD) includes two major types; Crohn’s disease (CD) and ulcerative colitis (UC), which are immune-mediated diseases that cause damage to the gastrointestinal tract with equal prevalence in both sexes. UC is a worldwide chronic inflammatory disorder of the colon that leads to typical ulcers in the mucosa of the rectum and colon. On the other hand, CD is a chronic inflammatory state that can affect the gastrointestinal system from the mouth to anus. Vitamin D levels, diet, hormone use, stresses caused by industrialization, and cigarette smoking has been considered as risk factors of IBD. The relationships between different autoimmune diseases have previously led to few studies in which the underlying molecular similarities among these disorders were subjected to an investigation. Most patients with CeD have gastrointestinal anomalies; however, kidney, skin, nervous system, and the reproductive systems are also affected by the consequences of CeD.

It has been well documented that histocompatibility antigens class II, HLA-DQ2, and HLA-DQ8 undergo extensive changes in patients with CeD. Overactivation of the immune system in CeD and IBD either by gluten exposure or through the alteration in gut microbiota leads to an increase in the intraepithelial T lymphocytes and inflammatory cytokines, which in turn leads to the intestinal atrophy followed by malabsorption in affected patients.

Although the association between CeD and IBD has been shown in a number of studies, data mining methods can be considered as a powerful approach toward understanding the etiology of these diseases. Considering the rapid growth of omics profiling, data integration and computational approaches are prominent contributors to unraveling disease complexities. In this study, the possibility of the presence of gene expression similarities between CeD and IBD patients was investigated. We mainly aimed to elucidate the underlying relationships between CeD and IBD using in-silico analytical approaches. The co-expressed genes and pathways that, according to our previous knowledge, assumed to play key roles in the pathogenesis of CeD and IBD were subsequently prioritized. Here we tried to provide a systematic view of common genes and molecular pathways implicated in both disorders.

MATERIALS AND METHODS

Firstly the curated gene-disease relations in CeD and IBD were retrieved from DisGeNET v2.0 server (http://www.disgenet.org/web/DisGeNET). These gene-disease associations have been collected from several databases, including UniProt, human CTD, PsyGeNET, Orphanet, and HPO. Afterward a list of 9905 regulatory links between human gene-transcription factors was obtained from TRRUST database (http://www.grnpedia.org/trrust/). These regulatory connections have been collected from PubMed articles and experimentally validated transcriptional regulations consisting of 821 human transcription factors and 2,159 target genes of TFs. Subsequently, using GeneMANIA (https://genemania.org/), a network was constructed using the common co-expressed genes and pathways of CeD and IBD. Moreover, highly connected clusters within GeneMANIA constructed network were identified using MCODE (http://baderlab.org/Software/MCODE/). Only MCODE score > 2 was considered as significant. The resulted sub-networks were visualized in Cytoscape 3.4.0 (http://www.cytoscape.org/). Pathway analysis was conducted on the resulted clusters using PANTHER with its default parameters (http://pantherdb.org/). To identify the transcription factors that regulate the expression of genes within each cluster, iRegulon Cytoscape plugin (http://iregulon.aertslab.org/) was employed. This way detects the putative regulons and co-factors associated with CeD and IBD. Finally, the Gemma server (http://www.chibi.ubc.ca/Gemma/home.html) was sought to determine the differential expression of shared genes implicated in CeD and IBD.

RESULTS

DisGeNET contains a list of diseases associated genes that have been collected based on the presence of genetic overlaps between diseases. In this list, among the 4,753,986 potential disease associations, 13,064 diseases were found to share at least one gene with other diseases. Among these associations, CeD was significantly associated (combined score at \( p < 0.05 \)) with atherosclerosis, multiple sclerosis, and hepatitis C, while in IBD, the significant associations were seen with stomach neoplasms, osteoporosis, post-menopausal syndrome, and sepsis. The combined score is calculated by taking the log of the \( p \) values generated by the Fisher exact test and multiplying that by the z-score
of the deviation from the expected rank. Both CeD and IBD, with a partial similar genetic background, target the gut, which is usually followed up by chronic inflammation. Since the ubiquity of CeD in patients with IBD remains unclear, we, therefore, were interested in mining CeD-IBD correlations at the transcriptome level to better understand the underlying reasons for this overlap. Only CeD-IBD associated genes retrieved from DisGeNET database with at least one evidence from a database of protein families (Pfam) were chosen for further analysis. Out of 325 CeD and 880 IBD associated genes, 314 and 851 genes, respectively, were selected from which 149 genes were common between CeD and IBD. As mentioned earlier, GeneMANIA server was used to establish a common network of co-expressed genes and pathways in CeD-IBD by utilizing the common identified 149 genes. Afterward, the GeneMANIA network was further clustered into the highly interconnected gene clusters by MCODE, which totally resulted in the eight sub-networks (figure 1A). The MCODE scores varied from 2.8-28.8. According to the previous studies, immune response associated transcription factors (NFKB1, IRF2, VDR), interleukins (IL10, IL6, IL2, and IL2RA), epidermal growth factor receptor (EGFR), and interferon-gamma (IFNG) were more interesting. The genes within the identified sub-networks were enriched in the inflammatory response (FAS), apoptotic pathway (CD14, IRF1), and signaling pathways (EGFR, CCL5, CCR5). In pathway analysis, signaling pathways like inflammation mediated by cytokine and chemokines, interleukin, toll receptor, and apoptosis seemed to be more biologically associated with CeD and IBD (figure 2). We additionally investigated the 149 common genes to discover disease-associated motifs and master regulators by iRegulon and TRRUST (figure 3). Among the
eight identified sub-networks except for sub-network 1 in which IRF2 was found as the most probable regulator and NFkB1 placed as the second regulator, NFkB1 was found to be the global regulator of 149 common genes in other sub-networks. We also examined if the gene-transcription factor links from iRegulon were also present as experimentally validated links in TRRUST. As mentioned in the method section, finally, the differential expressed genes from these 149 common genes were determined using Gemma server. Only 39 genes were detected as differentially expressed in CeD-oriented experiments in Gemma in different cell types (figure 4A) while all of the 149 genes were identified as differentially expressed in IBD experiments under different time points, genetic modifications, phenotypes, cell types, and medical procedures. The expression of these genes did not change in different diets and clinical treatments of patients with IBD (figure 4B).

DISCUSSION

CeD, IBD, and irritable bowel syndrome (IBS) are considered as the most crucial immune-mediated intestinal complications. IBD is a chronic, systemic inflammatory disease with gastrointestinal complaints, which is commonly treated by methylated thiopurine metabolites, like 6-methyl mercaptopurine. CeD is a genetic disease, which causes various manifestations such as gut lesions, triggering by environmental (gluten, viral infection, etc) and genetic factors (aberrations of LHA and non-HLA genes), with a worldwide prevalence, especially in western countries with a sex bias toward women (p = 0.001). CeD is further identified by severe autoimmune responses leading to small intestinal mucosa damages. Patients with CeD explicitly subject to a 5-10-fold risk of developing IBD. In spite of the large number of studies that have helped to expand our knowledge about IBD and CeD, questions such as how these disorders are developed and who are susceptible to these autoimmune disorders have not been answered yet. In this article, we used freely available data sets and bioinformatic tools to unravel genes whose interactions are probably implicated in the pathogenesis of both CeD and IBD disorders. Furthermore, we tried to elucidate the common underlying molecular pathways between CeD and IBD. Interestingly, most identified genes were enriched in chemokine and cytokine signaling pathways. The levels of specific cytokines and chemokines have been found to be elevated in patients with IBD. In accordance, the downregulation of chemokines CCL2CCL5 and CXCL8 in patients with IBD has been reported. JAK/STAT signaling pathway for which a few numbers of 149 genes were enriched probably reveals a circuit of STAT1, TNF, and IFN in the development of the inflammation. GeneMANIA constructed a co-expression network that demonstrated an equal distribution of interleukins in almost all of the eight sub-networks (figure 1A). This implies a vital role of interleukins in immune-mediated disorders. Noteworthy, in comparison with healthy individuals, interleukins, especially inflammatory cytokines like IL1B...
**Fig. 4:** Expression profile of differentially expressed genes in celiac disease (A) and inflammatory bowel diseases (B) from the Gemma server.
exhibited more than a two-fold increase in the expression in both CeD and IBD patients (figure 4). Among the CeD and IBD common genes, a number of interferon genes (IFNA1, IFNA13, IFNG) were identified that are under the regulation of NFKB1, IFR1, STAT1, and STAT3. These interferons participate in the inflammatory and apoptotic process (figure 1B, figure 2), among which IFNG showed a mild to high downregulation in CeD and IBD (≥2-fold). Further, by inspecting 1KB from the up-streams of CeD-IBD associated genes to characterize the disease-specific cis-regulatory motifs, NFKB1 and IRF1 were identified as the most significant regulators of the common genes. NFKB1 showed ≥2 fold down-regulation in patients with CeD and IBD, mostly in cell type-dependent experiments. However, in patients with IBD, in some cases <2 fold change up-regulation was observed (figure 4B). IRF1 was also detected as down-regulated in both CeD and IBD. NFKB1 transcription factor with the highest connectivity (figure 1B), has a confirmed contribution in the regulation of interleukin genes. For instance, TFN, IL2, and IL21, which were previously indicated to induce proliferation of innate intraepithelial lymphocytes, were detected as upregulated in CeD and IBD. NFKB as an indispensable regulator of the adaptive immune system, regulates the activation, proliferation, and survival of lymphocytes. NFKB pathway has also been shown to be constitutively upregulated in patients with CeD. In NFKB signaling pathway, there are clusters of differentiation (CDs) genes such as CD14, which is the marker molecule for monocytes and macrophages involved in proximal signaling. TNF and NOS2, from which only TNF indicated an expressional alteration in CeD and IBD are also involved in proximal signaling. Thus, the results indicate the dysregulation of effectors associated with proximal pathways involved in B cell signaling in patients with CeD and IBD. Another aspect of the identified shared genes was the abundance of human leukocyte antigen (HLA-A, HLA-G, HLA-DRB1) that, based on the TRRUST TF-target relationships, were mainly regulated by IRF1(activation) and VDR (unknown). However, in motif analysis, IRF1 obtained a higher score than VDR as a master regulator of HLAs. HLA-mediated inflammatory reactions are thought to be associated with CeD, and IBD in addition to other autoimmune disorders. IRF1 contributes to inducing gluten intolerance, which in turn is repressed by STAT1. Accordingly, FOXP3, which is essential in adaptive immunity (regulating the differentiation of T-cells) and is respectively activated and suppressed by NFKB1 and IRF1, showed up-regulation in IBD. VDR was shown to be expressed in the duodenal mucosa of patients with CeD, exhibiting the mucosal damages. As predisposing genetic factors, HLAs contribute to the genetic pathogenesis of autoimmune disorders, including CeD and IBD at 30-50%. Some HLAs indicated up-regulation in the present study. The major type of chronic IBD, Crohn’s disease has been shown to be related to specific HLA class I and II from which the expression of HLA-G was absent in CD and, as a result, can be used as a diagnostic marker to discriminate between CD and UC in cases of indeterminate colitis.

FAS has been demonstrated to be associated with the severity of atrophy in CeD. As a molecule with high connectivity in the constructed network (figure 1B) was also found to mediate cell apoptosis by contributing with IL10, interferon genes, T-cell receptors (CD94), and TNF-α. This could give a partial view of initiating malignant and non-malignant proliferation in T-cells due to the inflammatory events that occurred in CeD and IBD. Among 149 genes, IL23A and CD28 have exhibited altered expression in both children and adult patients with CeD, while STAT3 was only differentially expressed in adults. Inflammation mediated by chemokines and interferons was demonstrated as the most significant pathways contained most of CeD and IBD shared genes. In concordance, the abundance of interleukin genes is still noteworthy, so that interleukins have confirmed roles in several immune and inflammatory events that are occurred in autoimmune diseases like CeD. In keeping with this, further studies could help in evaluating the prognostic role of interleukins in cross-talk of in CeD-IBD. Altogether, several genes and transcription factors were emphasized in this study, which their interactions with each other result in the development of CeD and IBD. This mostly occurs through signaling pathways and T-B cell activations. Furthermore, grouping the genes in chemokine and cytokine signaling pathways concerning the up-regulation of chemokines and cytokines, especially in patients with IBD, might be biologically relevant to
the necessity in declining the expression of them to inhibit inflammatory responses in these patients as a therapeutic strategy.

However, this analysis had some limitations; CeD related genes were retrieved from adult and children patients with CeD. In addition, genes implicated in the autosomal recessive form of IBD and IBD forms 1-27 have been used in this study. Although CeD can emerge at any age there are some variations in CeD in children and in adults. One, therefore, should be more distinct from driving the exact types of disease-disease links using the strict criteria. Moreover, linking the results to the available GWAS data for autoimmune diseases and exploring the genetic variations or the methylation status of CpG islands in the up-stream of shared genes will seemingly provide a more systemic understanding of CeD and IBD correlations.

CONCLUSIONS

Utilizing CeD and IBD trade-off can be helpful in explaining disease inverse co-incidence. Genes obtained by bioinformatics data analysis like NFKB1, IRF1, HLA, and interleukin genes could be suggested for future studies for proofing their roles by confident experiments.

ACKNOWLEDGMENTS

We would like to thank Dr. Nooshin Omranian, scientific staff in Systems Biology and Mathematical Modelling Group, Max Planck Institute for Molecular Plant Physiology, Potsdam, Germany, for her helpful comments.

ETHICAL APPROVAL

There is nothing to be declared.

CONFLICT OF INTEREST

The authors declare no conflict of interest related to this work.

REFERENCES


Transcriptional interaction of Celiac Disease and IBD


