Eguchi *et al. BMC Bioinformatics* (2018) 19:264
https://doi.org/10.1186/s12859-018-2251-x

BMC Bioinformatics

RESEARCH ARTICLE Open Access

An integrative network-based approach to identify novel disease genes and pathways: a case study in the context of inflammatory bowel disease

Ryohei Eguchi[1†], Mohammand Bozlul Karim[1†], Pingzhao Hu[2,3,4†], Tetsuo Sato[5,1], Naoaki Ono[1], Shigehiko Kanaya[1] and Md. Altaf-Ul-Amin[1*†]

Abstract

Background: There are different and complicated associations between genes and diseases. Finding the causal associations between genes and specific diseases is still challenging. In this work we present a method to predict novel associations of genes and pathways with inflammatory bowel disease (IBD) by integrating information of differential gene expression, protein-protein interaction and known disease genes related to IBD.

Results: We downloaded IBD gene expression data from NCBI's Gene Expression Omnibus, performed statistical analysis to determine differentially expressed genes, collected known IBD genes from DisGeNet database, which were used to construct a IBD related PPI network with HIPPIE database. We adapted our graph-based clustering algorithm DPClusO to cluster the disease PPI network. We evaluated the statistical significance of the identified clusters in the context of determining the richness of IBD genes using Fisher's exact test and predicted novel genes related to IBD. We showed 93.8% of our predictions are correct in the context of other databases and published literatures related to IBD.

Conclusions: Finding disease-causing genes is necessary for developing drugs with synergistic effect targeting many genes simultaneously. Here we present an approach to identify novel disease genes and pathways and discuss our approach in the context of IBD. The approach can be generalized to find disease-associated genes for other diseases.

Keywords: Disease gene, Inflammatory bowel disease, Gene expression, Protein-protein interaction

Background

Inflammatory bowel disease (IBD) causes chronic inflammation of some or all part of the digestive tract. There are two major subtypes of IBD: ulcerative colitis (UC) and Crohn's disease (CD). Both types usually involve severe diarrhea, pain, fatigue and weight loss. IBD can bring severe situations and can lead to life-threatening complications. IBD is still not curable since there are no suitable drugs and targets for curing the disease.

IBD is an idiopathic, chronic and often disabling inflammatory disorders of the gastrointestinal tract characterized by dysregulated mucosal immune response. IBD can result in life threatening bleeding, sepsis and bowel obstruction. The pathogenesis of IBD is still elusive and therefore needs to be understood for developing cure for IBD. Genome-wide association studies (GWAS), have significantly advanced our understanding on the importance of genetic susceptibility in IBD. The GWAS performed to date together with a meta-analysis of several GWAS have identified a total of 163 IBD loci [1]. These studies mainly focused on the common genetic variants (single nucleotide polymorphisms (SNPs)). These risk loci are asscciated to a handful of candidate genes which have small contributory effects in IBD.

*Correspondence: amin-m@is.naist.jp
†Md. Altaf-Ul-Amin, Pingzhao Hu, Ryohei Eguchi and Mohammand Bozlul Karim contributed equally to this work.
[1]Graduate School of Science and Technology & NAIST Data Science Center, Nara Institute of Science and Technology, Nara, Japan
Full list of author information is available at the end of the article

Significant interest has been developed for inventing new methods based on integrating omics data for identifying disease causal genes. For example, network-based classification approaches have been developed to integrate gene expression and protein interaction data to predict breast cancer metastasis [2, 3], multiple sclerosis relapse and remissions [4] and autoimmune disease [5]. Other studies also identified subnetwork modules from integrating protein interaction data with GWAS signals for complex diseases [6].

During the past decade, a huge pile of biological data has been generated from various large-scale omics studies, prompting the scientific community to gain deeper insight into underlying biological mechanisms of different diseases. One of the interesting topics is to find disease-gene associations. Broadly speaking, a disease-gene association can be a connection reported in the literature, such as a genetic association (i.e., mutations in a given gene may lead to a specific disease), or inferred from other sources [7]. Similarities between disease symptomes and gene functions could be used to predict disease-causing genes by text mining [8]. The human diseasome was constructed by connecting diseases to shared disease-causing genes [9]. Understanding of disease relationships has been explored using different types of omics data such as biological pathways [10], transcriptome data [11, 12], biomedical ontologies [13, 14], and genome-wide association study (GWAS) data [14–17]. Recently, large-scale biological data have been analyzed based on networks, and network topology has been utilized to provide insights into diseases and their associations with genes [9, 18–20]. Because the interactions between bio-molecules play crucial roles in the cell, the topology of biological networks is likely to have various biological and clinical applications [21, 22].

Cellular functions rely on the coordinated actions of multiple genes, proteins, and metabolites. Therefore, organizing biological information in the context of networks is important for deep understanding of biological systems. Discovery of modules in biological networks helps isolate systems with disease related properties and reduces interactome complexity [23]. Proteins rarely act alone as their functions tend to be regulated. Many molecular processes within a cell are carried out by molecular machines that are built from a large number of protein components organized by their protein-protein interactions (PPIs). The disease proteins (the product of disease genes) are not scattered randomly in the interactome but tend to interact with each other. Because of incompleteness of disease genes and PPI data, the known disease genes usually fail to form observable modules in PPI networks. Out of 299 diseases only 20% of the respective known disease gene from some type of modules [24]. To compensate for such gaps to a certain extent, In the present work we focus on finding novel IBD associated genes and pathways by integrating IBD gene expression, PPIs, and known IBD genes by adapting the DPClusO network clustering algorithm we published previously.

Results and Discussion

The method adopted in the present work has been illustrated in Fig. 1. Based on the IBD gene expression data downloaded from NCBI's Gene Expression Omnibus (GSE57945) [25], we got 1197 and 4315 differentially expressed genes (DEGs) (with false discovery rate (FDR) < 0.05) between control and Crohn's disease (CD) as well as control and ulcerative colitis (UC) samples, respectively. The venn diagram of the overlapping genes between these two sets is shown in Fig. 2. CD and UC are closely related diseases, hence, the differentially expressed genes are largely overlapped (1035 overlapped genes). As our focus is to find novel IBD genes and pathways by system level analysis, we took the union set of the differentially expressed genes from these two comparisons, and combined these genes to a single set consisting of 4477 genes. The differentially expressed genes are the potential candidates to be relevant to IBD.

Construction of a disease relevant PPI network

We initially downloaded 866 genes reported in DisGeNet database [26] as IBD genes. We found that 318 of the 866 IBD genes are out of the 4477 differentially expressed genes (DEGs) we identified from gene expression analysis. Let us name these 318 genes as IBD related differentially expressed genes (IDEGs) and the rest 4159 as only differentially expressed genes (ODEGs). In this work we consider these 318 genes as known IBD genes.

We constructed a disease related PPI network based on Human Integrated Protein-Protein Interaction rEference (HIPPE) database [27]. In HIPPE database each interaction is reported with a confidence score. We first extracted the interactions involving ODEGs with a score greater than 0.7, which included 4135 ODEGs. We then retrieved the interactions involving all 318 IDEGs with a score greater than 0.1. Thus we retrieved a total of 38,500 interactions involving IDEGs, ODEGs and other genes (OGs). From these interactions, we empirically selected interactions to construct the final PPI network according to following criterion: IDEG-IDEG:0.1, IDEG-ODEG:0.1, IDEG-OG:0.72, ODEG-IDEG:0.1, ODEG-ODEG:0.1, ODEG-OG:0.85. In summary, we gave the highest priority to interactions involving IDEGs (genes that are both known IBD genes and differentially expressed genes according to the expression data we used). Also, most priority was given to interactions for which both genes are ODEGs (only differentially expressed genes). These genes are likely to

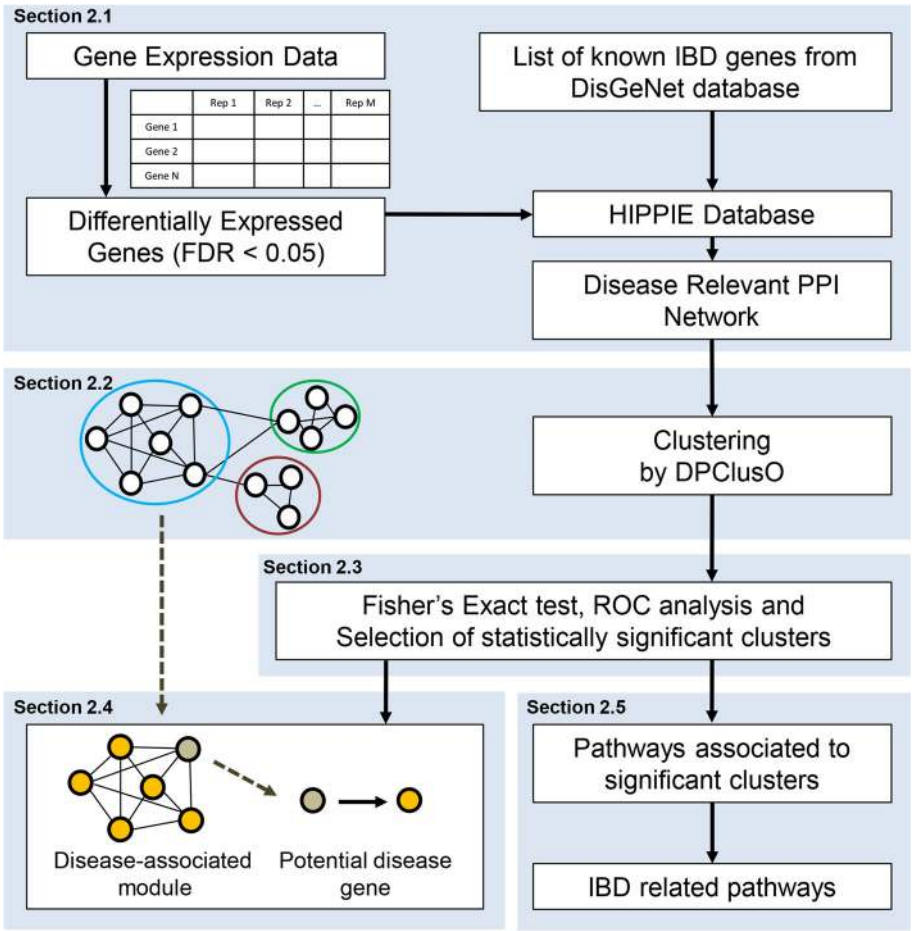

Fig. 1 Flowchart demonstrating the major steps of the proposed approach

contain system level information of molecular mechanism of IBD. The HIPPIE database recommend 0.72 as a good score which we used for IDEG-OG interactions and finally adjusted 0.85 for ODEG-OG interactions to roughly keep similar number of DEGs (IDEGs + ODEGs) and OGs (Other Genes) in the PPI network for the sake of balance and thus extracted unbiased information. Finally we selected 16,429 interactions involving 5056 genes with 291 IDEGs, 2072 ODEGs and 2693 OGs. The degree distribution of the network is shown in Fig. 3. As many other typical PPI networks, the degree distribution of our constructed network followed power law. Some other global network properties of the network include average path length 4.18, clustering co-efficient 0.1 and diameter 11. For such a big network the clustering coefficient of 0.1 is substantialy enough indicating presence of densely connected clusters in the network.

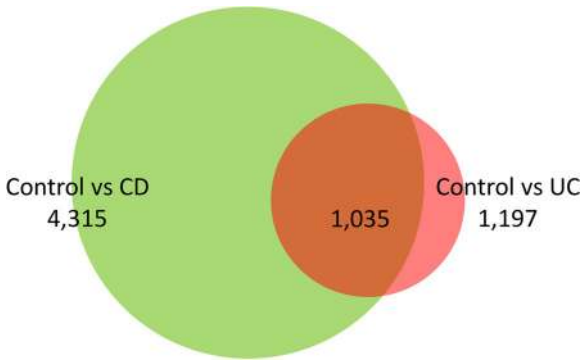

Fig. 2 The venn diagram showing overlapping between differentially expressed genes in case of CD and UC

Clustering of the PPI network

After creating the disease related PPI network we determined clusters in the network by DPClusO algorithm. DPClusO generates overlapping clusters and ensures coverage. For example, each node goes to at least one cluster. We hypothesize that clustering of a disease relevant PPI network helps isolate systems with disease related properties and therefore statistically significant clusters enriched

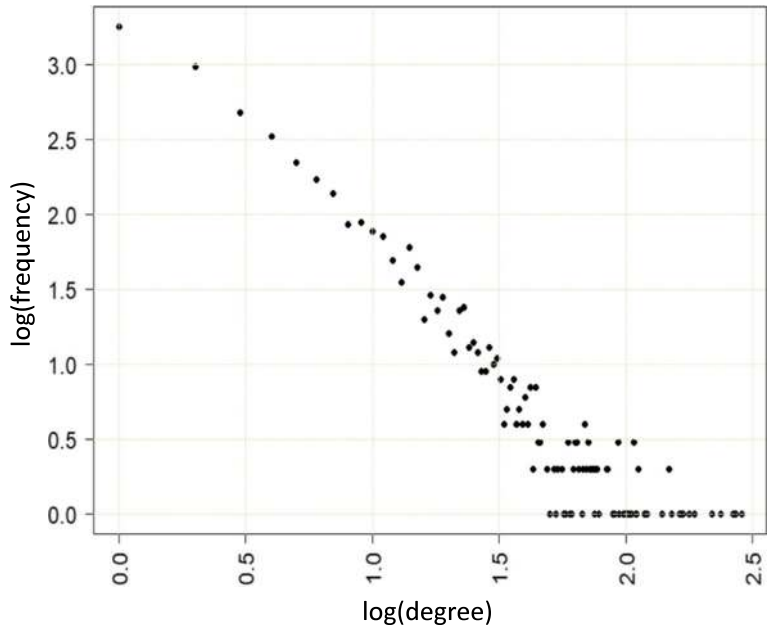

Fig. 3 Degree distribution of the IBD related PPI network follows the power law

with known IBD genes can be used to predict novel IBD genes and pathways based on the associations determined by combined information of IBD gene expression and protein-protein interactions.

We generated 9 sets of clusters from the PPI network by DPClusO algorithm using density values of 0.1, 0.2, 03, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. Table 1 shows characteristics, i.e. the number of clusters, size of the biggest cluster and average cluster size, related to the clusters generated by the 9 different density values. As expected, smaller density value resulted in larger and fewer number of clusters generated. To assess the enrichment of IDEGs in each of the identified clusters we determined Fisher's exact test *p*-values. In this work we proposed to consider statistically significant clusters for predicting novel IBD related genes and pathways. Therefore we assigned a score called SScore (Significance Score) to each gene as a measure of confidence of prediction based on the *p*-values of the clusters they belong to. The definition of SScore is provided in the Methods section. Based on these scores we performed ROC analysis to determine which set of clusters should be used for predicting novel IBD genes.

Table 1 Characteristices of the clusters generated with different input densties using the DPClusO algorithm based on the IBD related PPI network

density	#ofcluster	maxsize	avgsize
0.1	827	70	18.52
0.2	1229	42	10.69
0.3	1779	31	7.18
0.4	2219	21	5.74
0.5	2790	16	4.61
0.6	3597	13	3.53
0.7	4425	11	2.59
0.8	4534	9	2.50
0.9	4775	7	2.31

ROC analysis

In our disease relevant PPI network there are total 5056 genes out of which 291 genes are IDEGs which are among the 318 genes considered as known IBD genes in the present work. We predicted the degree of relevance of the rest 4765 genes with IBD based on SScore. We collected well curated and well studied IBD genes from 3 databases as follows, The Comparative Toxicogenomics Database (CTD) [28], DisGeNet [26], HuGENet [29] and published literatures on results of GWAS [30–33]. The venn diagram of the reported IBD genes in these 4 databases is shown in Fig. 4. It is noticeable that IBD genes listed by these 4 sources are substantially different, indicating the need for finding comprehensive set of potential IBD genes. Although these four sources are not the complete list of IBD genes, they can be used to assess the effectiveness of SScore. The ROC curves corresponding to the 9 sets of clusters are very similar, which imply

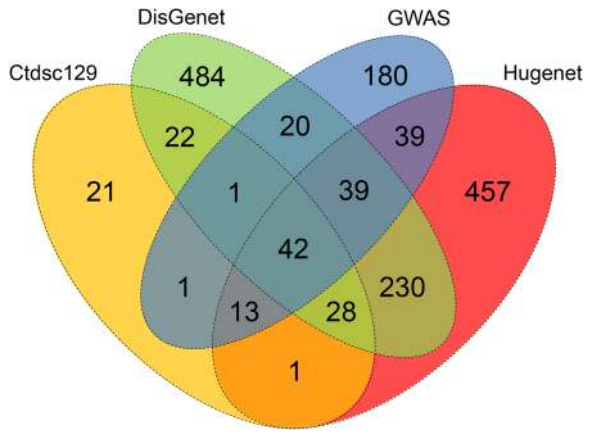

Fig. 4 Venn diagram showing overlapping between IBD genes collected from four different sources

the underlying signal in the carefully constructed PPI network is very strong and DPClusO algorithm has been successful to catch the signal for across a wide range of the input parameter. Figure 5 shows the Area Under the Curve (AUC) for the 9 ROC curves. The AUCs are not very high, which may be due to incomplete information of known good quality IBD genes. We observed that the highest AUC was obtained in the case of the cluster set generated using density $= 0.5$. So we selected the genes included in the statistically significant clusters of this set, adjusted the corresponding p-values for multiple testing [34] and selected the genes having adjusted p-values less than 0.05 as predicted IBD genes.

Prediction and validation

We predicted 909 genes (with adjusted $p - value < 0.05$) included in the clusters selected from the set corresponding to the highest AUC as our predicted IBD genes. These 909 genes are other than the genes considered as known IBD genes (IDEGs) in this work. The list of the 909 predicted IBD genes and corresponding adjusted p-values are shown in Additional file 1. To validate our results we initially searched how many of the predicted genes are exactly matched with well curated known IBD genes. We found 83, 8, 54, 22 of the predicted genes matched with reported IBD genes in (1) HuGeNet, (2) CTD, (3) DisGeNet databases and (4) GWAS results respectively. After considering overlapping between databases, 14.5% of our predicted genes matched with good quality known IBD genes. Given the fact that we made predictions based only on a specific gene expression data and a limited set of known IBD genes, the 14.5% matching with good quality data is significant ($p - value < 3.45 \times 10^{-12}$, p-value determined based on hypergeometric ditsribution assuming total number of human genes as 20000). However, our approach is a computational approach. So, it is rational to compare our result also with computationally predicted IBD genes. In CTD database other than the good quality curated set there is a big set of genes inferred as IBD genes by various methods. When we compare our result with this big set, we find that 93.8% of the genes we predicted matched with reported IBD genes ($p - value < 9.8 \times 10^{-14}$). As we have predicted the genes by wisely integrating the information of gene expression and protein-protein interaction, it is very likely that they are truely related to IBD. One of the predicted genes IL12B is supported by all four above-mentioned sources as an IBD related gene. IL12B and IL23R have been identified as susceptibility genes for IBD by recent genome-wide association studies [35]. Each of the three genes CCR5, IL1R2 and LTA is mentioned as IBD related gene in three of the above mentioned sources. High expression of CCR5 has been reported in active IBD [36]. Epithelial IL1R2 takes part in homeostatic regulation during remission of ulcerative colitis [37]. It has been reported that LTA elicits a strong inflammatory reaction controlled by intestinal dendritic cells [38]. Thus we have found IBD relevance of many other predicted genes by literature review. The proposed method, however is a computational one and the role of the newly predicted genes in IBD pathogenesis should be clarified by further studies.

The degree of relevance of the 909 genes (shown in Additional file 1) predicted by the proposed approach can be evaluated by the corresponding p-values. The top 20 predicted novel IBD genes (not reported in any of the four sources of Fig. 4) based on p-values are IKBKG, BIRC3, BCL10, RNF31, RBCK1, CCRL1, LAMC3, CARD11, KISS1, THBS2, TRAF2, TRAF1, PYCARD, MIS12, ALB, AR, RIPK1, SHARPIN, SNAPIN and ITGA2B. Many of these 20 top IBD risk genes we identified from this study have been found to be associated with IBD. In human, the IKBKG gene encodes NF-κB essential modulator (NEMO) which is an inhibitor of nuclear factor κB kinase subunit gamma (IKK-γ) [39]. NEMO (IKK-γ) is the regulatory subunit of the inhibitor of the I-κB kinase (IKK) complex, that activates NF-κB causing activation of genes involved in inflammation, immunity, cell survival, and other pathways. IBD-like immunopathology can be developed by IKBKG [40]. BIRC2 and BIRC3 are important genes in regulating the expression of proinflammatory cytokines, such as TNF-α, through NF-κB and MAPK pathways [41]. BCL10 is an adaptor protein which is assumed to play role in the PAF-induced inflammatory pathway in human intestinal epithelial cells [42]. RNF31 and HOIL-1L complex functions in linear ubiquitination of proteins in the NF-κB pathway in response to proinflammatory cytokines [43]. CCRL1 acts as a functional receptor for the monocyte chemoattractant protein family of chemokines; elevated chemokine expression is associated with many inflammatory diseases such as IBD,

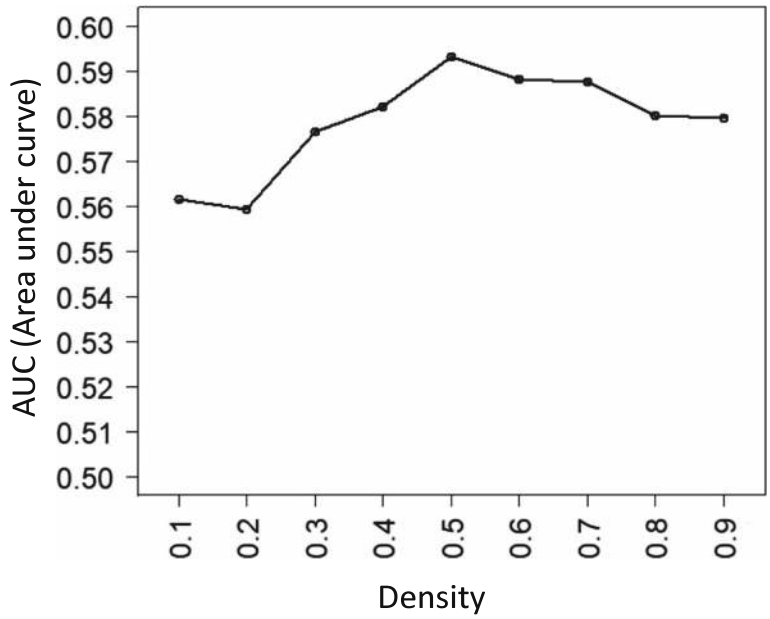

Fig. 5 AUCs corresponding to 9 sets of clusters

rheumatoid arthritis and asthma [44, 45]. As a component of the LUBAC complex, RBCK1 conjugates linear (Met1-linked) polyubiquitin chains to substrates and thus plays imoportant role in NF-κB activation and inflammation regulation [46]. RBCK1-deficiency is associated with autoinflammatory syndrome and immunodeficiency [46]. LAMC3 is expressed saliently at significantly different proportions in low and high coherence expression profiles of IBD patients [47]. The elevated stromal protein thrombospondin-2 (THBS2) has been reported to be a part of a fibroblast-specific inflammation signature [48]. It has been shown that TRAFs are important mediators of innate immune receptor signaling [49]. IBD and IBD recurrence is associated with the overexpression of TRAF2 [50–52]. TRFA1 is reported to be highly expressed in IBD patients [53]. To form the basic Inflammasome subunit, the adaptor protein ASC (encoded by the PYCARD gene) links the NLR sensor to caspase-1 [54]. TNF-α-induced necroptosis is associated with two members of the receptor-interacting protein (RIP) family of kinases – RIPK1 and RIPK3 [55]. Tumor necrosis factor-α (TNF-α) can bind to one of two receptors, TNFR1 or TNFR2; TNFR activation results in the activation of NF-κB leading to the induction of proinflammatory cytokines [55].

Comparison with ToppGene

It has been demonstrated that ToppGene [56] performs better than several other methods such as SUSPECTS [57], PROSPECTOR [58], ENDEAVOUR [59] in candidate gene prioritization. From the ToppGene suite [60] we used ToppGenet which is a web based tool that can take input a set of seed genes and can return a list of genes with closely related roles with a prioritization score. In our work, based on gene expression data and DisGeNet database we considered 318 genes as known IBD genes and based on those we predicted 909 other genes as IBD related genes. We assigned the same 318 genes to ToppGenet and from the output we selected the highest ranking 909 genes which we compared with the 909 genes determined by our approach. For both sets, we determined the number of genes matched with the union of reported IBD genes in 4 sources of Fig. 4. Also we determined the AUCs using prioritization score and SScore in case of ToppGenet and our approach respectively. In case of ToppGenet, we selected network based approach as our approach is also network based. Furthermore, we used 3 available options for ToppGenet as follows: (i) K-Step Markov, (ii) Page rank with priors and (iii) Hits with priors. The comparison results are shown in Table 2. The results show that performance of our approach is comparable in terms of the number of identified genes and better in terms of AUC.

Gene ontology and pathway analysis

As a group the top 20 predicted genes (names mentioned in the previous section) are enriched in some important BP(Biological Process) related GO terms, such as I-κB kinase/NF-κB signaling, positive regulation of immune response, regulation of tumor necrosis factor-mediated signaling pathway and MF(Molecular Function) terms, such as ubiquitin protein ligase binding, identical protein

Table 2 Results of comparison with ToppGene

Parameter of comparison	ToppGenet			Our Method
	K-step Markov	Page rank with priors	Hits with priors	
Number of match	163 >	169 >	102 <	132
AUC	0.4969 <	0.4339 <	0.4831 <	0.5826

binding. We also performed enrichment analysis for all of the 909 genes. Some significant BP related GO terms enriched in these genes are nitrogen compound metabolic process, response to stimulus, immune system process, cell surface receptor signaling pathway, response to stress, response to lipid, positive regulation of leukocyte cell-cell adhesion and MF terms are enzyme regulator activity, kinase activity, protein complex binding, histone deacetylase binding, transcription factor activity, protein binding, protein C-terminus binding. NF-κB pathway mediate events including the activation of genes encoding inflammatory molecules and is found to be chronically active in IBD [61]. All the above mentioned GO terms associated to a group of genes were searched by using the enrichment analysis tool [62] provided in the web page of Gene Ontology Consortium.

As examples we arbitrarily select and show 6 of the statistically significant clusters in Fig. 6(a)-(f). In these clusters 4, 5, 4, 4, 3, 5 genes are IDEGs respectively and 3, 2, 3, 2, 2, 2 genes are reported to be IBD genes by 4 reliable sources as mentioned in Fig. 4. Many of the genes included in these clusters are related to IBD. It has been reported that SOCS deficient mice develop severe colitis (similar to human ulcerative colitis) depending on some factors [63]. Expression of IGF1R in submucosal fibroblast-like cells, subserosal adipocytes and hypertrophic plexus has been confirmed to be CD specific, indicating relations between IGF1R and chronic inflammation [64]. It has been reported that the deficit of PTPN11 is related to the severity of colitis [65]. IRF8 promotes the production of IL12 and IL23 in the development of experimental autoimmune encephalomyelitis and inhibits the

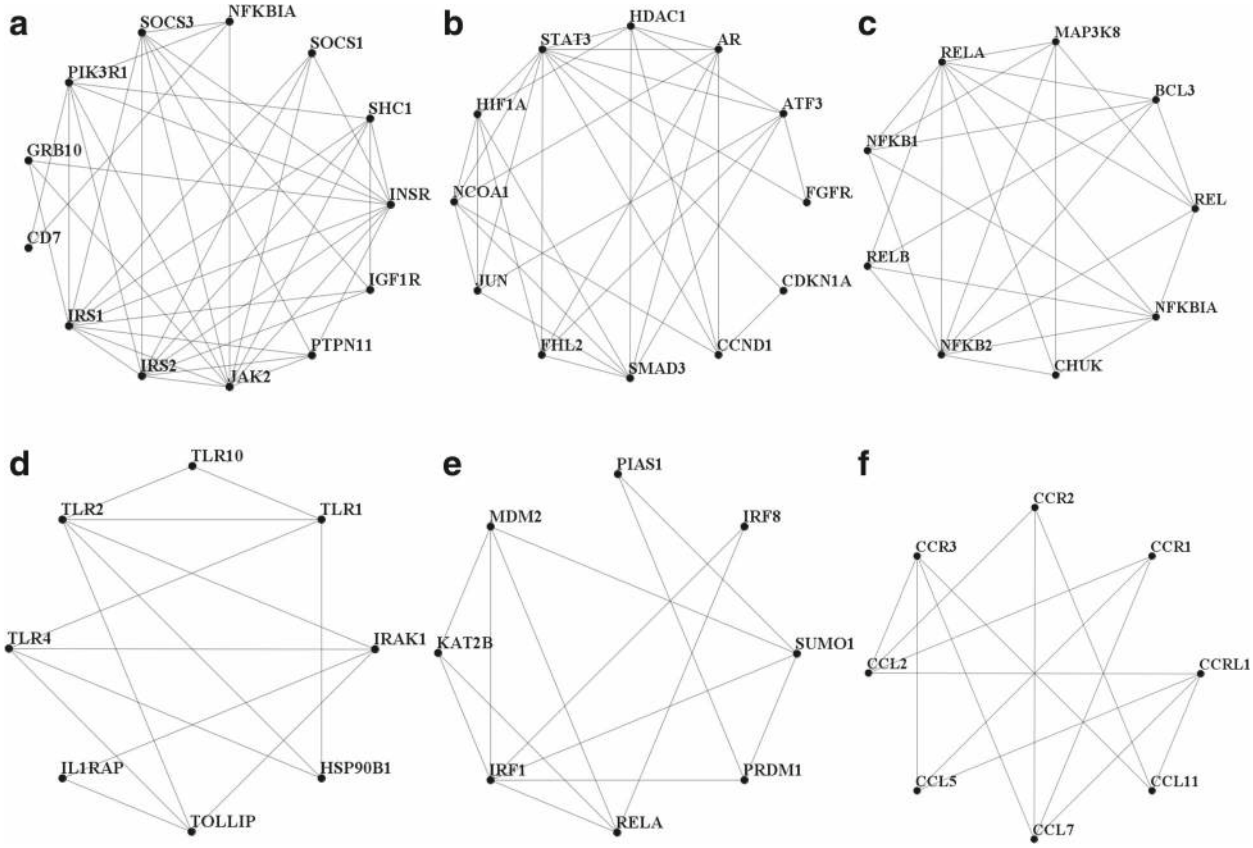

Fig. 6 (**a**)-(**f**) Examples of statistically significant clusters

production of IL27, and thus forms a cytokine environment suitable for differentiation and maintenance of Th1 cells and Th17 cells and also, IRF8 exacerbates inflammation by activating microglia [66]. C-C motif chemokine receptors, CCR1 and CCR3 are membrane proteins that particulaly bind and respond to cytokines of the CC chemokine family [67, 68].

Based on significant p-values, we empirically selected some enriched BP and MF terms for these clusters. Some important BP related GO terms enriched in these clusters (a)-(f) are as follows: (a) cell surface receptor signaling pathway, regulation of cellular response to insulin stimulus, cellular response to hormone stimulus, (b) negative regulation of programmed cell death, response to endogenous stimulus, cell differentiation, (c) regulation of cytokine production, intracellular signal transduction, regulation of type I interferon production, (d) toll-like receptor signaling pathway, activation of innate immune response, inflammatory response, (e) regulation of transcription from RNA polymerase II promoter, negative regulation of transcription, DNA-templated, negative regulation of nitrogen compound metabolic process, (f) chemotaxis, inflammatory response, positive regulation of MAPK cascade and MF related GO terms are as follows: (a) phosphatidylinositol 3-kinase binding, insulin receptor binding, receptor binding (b) transcription factor binding, regulatory region DNA binding, chromatin binding, (c) transcription factor activity, sequence-specific DNA binding, chromatin binding, (d) signal transducer activity, Toll-like receptor binding, (e) SUMO transferase activity, ubiquitin-like protein ligase binding, (f) G-protein coupled receptor binding, cytokine receptor activity.

We hypothesize that clustering disease related PPI network helps isolate systems with disease related properties. Therefore, we selected 442 statistically significant clusters ($p - value < 0.05$). We use these statistically significant clusters to determine IBD related pathways. We separately mappled the genes included in each of the statistically significant clusters to KEGG pathway [69]. For each cluster we determined the top three pathways based on the association of majority number of genes. Additional file 2 shows the selected pathways and enriched GO terms for these clusters. Frequencies of these selected pathways are shown in histograms of Fig. 7. The top 10 pathways with the highest frequency are : (1) MAPK signaling pathway, (2) Chemokine signaling pathway, (3) Cytokine-cytokine receptor interaction, (4) Pathways in cancer, (5) Toll-like receptor signaling pathway, (6) Cell cycle, (7) NOD-like receptor signaling pathway, (8) Apoptosis, (9) Endocytosis, (10) Focal adhesion. Particularly interested pathways associated with IBD in these results are MAPK, Chemokines, Cytokines, Toll-like receptors, and NOD-like receptor pathway. Previous studies have shown that these predicted pathways are highly relevant to IBD.

MAPK signaling pathway are evolutionarily conserved kinase modules whose fanctions are to transmit extracellular signals to various machinery inside the cell that manage fundamental cellular processes such as growth, differentiation, migration, proliferation and apoptosis. Activation of ERK1/2 by growth factors depends on

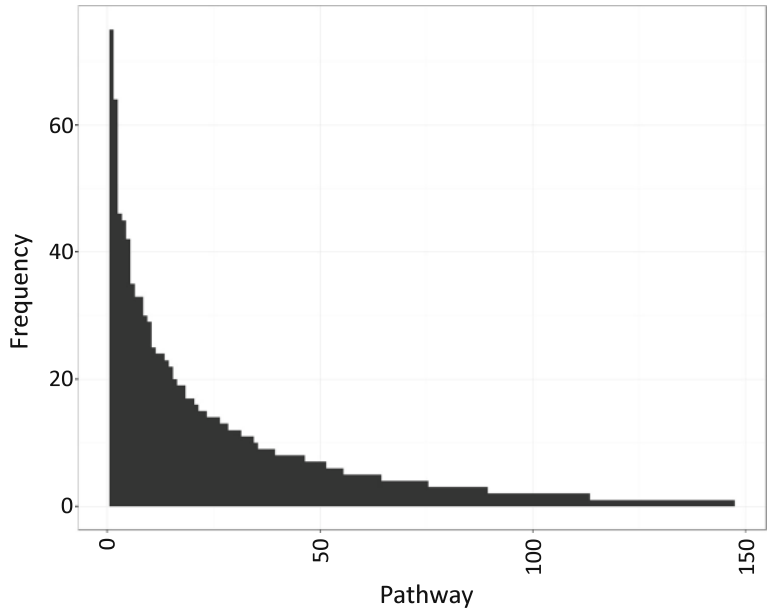

Fig. 7 Frequencies of pathways related to statistically significant clusters

the MAPKKK c-Raf, but other MAPKKKs may activate ERK1/2 in response to pro-inflammatory stimuli [70]. Small chemoattractant peptides called Chemokines provide directional cues for the cell trafficking and therefore are important for protective host response. They are soluble factors which play key roles in regulating immune cell recruitment during inflammatory responses and defense againsst foreign pathogens. Soluble extracellular proteins or glycoproteins known as Cytokines are crucial intercellular regulators and mobilizers of cells involved in inherent as well as adaptive inflammatory host defenses, cell death, cell growth, angiogenesis, differentiation and development and repair processes targeting the restoration of homeostasis. It has been reported that cytokines/chemokines are engaged in not only the initiation but also the persistence of pathologic pain by activating nociceptive sensory neurons. There are inflammatory cytokines engaged in nerve-injury/inflammation-induced central sensitization, and are associated to the development of contralateral hyperalgesia/allodynia [71, 72]. Toll-like receptors (TLRs) are a family of pattern recognition receptors that are best-known for their role in host defence from infection. It has been reported that TLRs play important role in maintaining tissue homeostasis by regulating the inflammatory responses to injury [73]. The intracellular NOD-like receptor (NLR) family contains more than 20 members in mammals and plays a pivotal role in the recognition of intracellular ligands. The activated state of caspase-1 regulates maturation of the pro-inflammatory cytokines IL-1B, IL-18 and drives pyroptosis [74].

Conclusions

We presented a method for predicting IBD related genes and pathways by integrating the information of IBD gene expression and protein-protein interactions and a set of known IBD genes from DisGeNet database. We determined differentially expressed genes (DEGs) based on IBD gene expression data and constructed a IBD relevant PPI network using DEGs and known IBD genes. We extracted high density modules from the PPI network using our graph clustering algorithm DPClusO. We determined modules enrichment with known IBD genes by Fisher's exact test and used those statistically significant modules to predict novel IBD genes and pathways. We compared our results with several other databases and published literatures. We found 93.8% of our predictions are found in these published results. Specially we found our results substantially matched with IBD genes collected in curated databases and high-profile publications.

Furthermore, based on our ranking score, we selected top 20 predicted novel IBD genes and by literature survey we observe that most of these genes are really substantially related to IBD. As a group these 20 genes are enriched in the GO term I-κB kinase/NF-κB signaling. NF-κB pathway mediates events including the activation of genes encoding inflammatory molecules and is found to be chronically active in IBD. Also, based on statistically significant clusters we identified top 10 IBD related pathways which include MAPK signaling pathway, Chemokine signaling pathway, Cytokine-cytokine receptor interaction etc. These pathways play roles in inflammation related diseases including IBD.

Finding disease-causal genes is the part of the process to understand disease mechanism and develop drugs that can provide synergistic effects targeting many genes/proteins simultaneously. This study discussed a computational approach to reach these goals in the context of IBD. The proposed method can also be applied to find disease-causal genes related to other diseases.

Methods

Data collection and preprocessing

We downloaded the IBD gene expression data from NCBI's Gene Expression Omnibus (GSE57945) [25]. The gene expression data was generated using TopHat [75]. The samples were collected for three biological groups: healthy control, Crohn disease and ulcerative colitis [24]. We removed genes with expression values equaling to zero across all samples. The final expression data set included 14664 genes and 322 samples, which included 42 control samples, 218 CD samples, and 62 UC samples. We also downloaded reported IBD genes from several other databases, such as The Comparative Toxicogenomics Database (CTD) [28], DisGeNet [26], HuGENet [29]. The protein-protein interaction data was downloaded from HIPPE database [27].

Identifying differentially expressed genes

We performed differential expression analysis using the R package edgeR, which is based on negative binomial models [76]. We implemented the exact test for a difference in mean between two groups of negative binomial random variables by using edgeR after applying Trimmed Mean of M-value(TMM) normalization [77, 78] to data. False discovery rate (FDR) was estimated from unadjusted p-values using Benjamini Hochberg multiple testing method [34, 79].

Network clustering by DPClusO

DPClusO is a graph clustering algorithm [80], which is the updated version of DPClus algorithm [81]. DPClusO can extract densely connected nodes in a network as a cluster or module. Particularly, it produces overlapping clusters or modules since genes can be disease-causal genes in multiple diseases or have multiple biological functions and are involved in multiple pathways. This algorithm can be applied to an undirected graph $G = (N, E)$ that consists

of a finite set of nodes N and a finite set of edges E. Two important parameters used in this algorithm are density d_k and cluster property cp_{nk}. Density d_k of cluster k is the ratio of the number of edges present in the cluster ($|E|$) and the maximum possible number of edges in the cluster ($|E|_{max}$). The cluster property cp_{nk} of node n with respect to cluster k is expressed by the follow equation:

$$cp_{nk} = \frac{E_{nk}}{d_k \times N_k}$$

N_k is the number of nodes in cluster k. E_{nk} is the total number of edges between the node n and each of the nodes of cluster k.

Fisher's exact test

We evaluated the enrichment of the known IBD genes (referred to as IDEGs in the present work) in the clusters from our PPI analysis using Fisher's exact test. The test is an alternative statistical significance test used in the analysis of 2×2 contingency tables [82, 83].

To do this, for each cluster we determined the values of a, b, c, and d as demonstrated in the following table:

	IBD Genes	Non-IBD Genes	
In Cluster	a	b	$a+b$
Not in Cluster	c	d	$c+d$
	$a+c$	$b+d$	n

Here n is the total number of genes in the network.

SScore

We assigned a score called SScore (Significance Score) to each gene as a measure of confidence of prediction based on the p-values of the clusters they belong to. By definition $SScore = -log(p - value)$. As DPClusO generates overlapping clusters, a gene may belong to more than one clusters and thus may correspond to more than one p-values. We used the lowest p-value corresponding to a gene to calculate its SScore.

ROC Analysis

We evaluated the power of SScore to predict the known IBD genes by performing receiver operating characteristic (ROC) analysis [84, 85]. The ROC curve was created by selecting a series of threshold SScore values to generate True Positive Rate (TPR) and False Positive Rate (FPR). TPR is the proportion of true positive predictions out of all the positive data and FPR is the proportion of false positidve predictions out of all the negative data and can be expressed by the following equations:

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}$$

Corresponding to a certain threshold SScore th, false positive (FP), true positive (TP), false negative (FN) and true negative (TN) are defined as follows: TP is the number of reported IBD genes having $SScore \geq th$, FP is the number of non-IBD genes having $SScore \geq th$, TN is the number of non-IBD genes having $SScore < th$, and FN is the number of reported IBD genes having $SScore < th$.

We observed the performance of SScore to identify known IBD genes by using the Area Under the ROC Curve (AUC) analysis [86]. In term of AUC analysis, we used R package named ROCR [87]. We considered a prediction as 'True' prediction if a gene is reported as IBD gene in any of the following four sources: (1) Human Genome Epidemiology Network (HuGENet), (2) Comparative Toxicogenomics Database (CTD), 3) DisGeNet database and (4) GWAS results [30–33]. Here, FP, TP, FN, TN were calculated based on known information i.e. without having knowledge of all IBD related and unrelated genes. Therefore, the calculated TPR and FPR values were affected by the unknown nature of the TN and FN genes.

Additional files

Additional file 1: List of predicted IBD genes. (XLSX 26 kb)

Additional file 2: Significant clusters with selected pathways and enriched GO terms associated to them. (XLSX 31 kb)

Abbreviations

AUC: Area under the curve; BP: Biological process; CD: Crohn's disease; CTD: The comparative toxicogenomics database; DEG: Differentially expressed gene; GO: Gene ontology; GWAS: Genome-wide association studies; HIPPIE: Human integrated protein-protein interaction rEference; IBD: Inflammatory bowel disease; IDEG: IBD related differentially expressed gene; MF: Molecular function ODEG: Only differentially expressed gene; OG: Other gene; PPI: Protein-protein interaction; SNP: Single nucleotide polymorphisms; UC: Ulcerative colitis

Funding

This work was supported by NAIST Global Collaborative Program 2017 and partially supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (16K07223 and 17K00406), NAIST Big Data Project and by Research Manitoba, Health Sciences Centre Foundation and Mitacs of Canada.

Authors' contributions

Md. A-U-A, PH, RE and MBK designed the research and conducted the experiments. TS, NO and SK guided the research with valuable comments. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Graduate School of Science and Technology & NAIST Data Science Center, Nara Institute of Science and Technology, Nara, Japan. [2]Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Canada. [3]George and Fay Yee Centre for Healthcare Innovation, University of Manitoba, Winnipeg, Canada. [4]Department of Electrical and Computer Engineering,

University of Manitoba, Winnipeg, Canada. [5]Department of Radiological Technology, Gunma Prefectural College of Health Sciences, Gunma, Japan.

Received: 4 October 2017 Accepted: 18 June 2018
Published online: 13 July 2018

References

1. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119.
2. Chuang H-Y, Lee E, Liu Y-T, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007;3(1):140.
3. Auffray C, Fogg D, Garfa M, Elain G, Join-Lambert O, Kayal S, Sarnacki S, Cumano A, Lauvau G, Geissmann F. Monitoring of blood vessels and tissues by a population of monocytes with patrolling behavior. Science. 2007;317(5838):666–70.
4. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A. Global map of physical interactions among differentially expressed genes in multiple sclerosis relapses and remissions. Hum Mol Genet. 2011;20(18):3606–19.
5. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A. Common and specific signatures of gene expression and protein–protein interactions in autoimmune diseases. Genes Immun. 2013;14(2):67.
6. Jia P, Wang L, Meltzer HY, Zhao Z. Pathway-based analysis of gwas datasets: effective but caution required. Int J Neuropsychopharmacol. 2011;14(4):567–72.
7. Sun K, Gonçalves JP, Larminie C, Przulj N. Predicting disease associations via biological network analysis,. BMC Bioinformatics. 2014;15(1):304. https://doi.org/10.1186/1471-2105-15-304.
8. Lage K, Karlberg EO, Størling ZM, Ólason PÍ, Pedersen AG, Rigina O, Hinsby AM, Tümer Z, Pociot F, Tommerup N, Moreau Y, Brunak S. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol. 2007;25(3):309–16. https://doi.org/10.1038/nbt1295.
9. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabasi A-L. The human disease network. Proc Natl Acad Sci. 2007;104(21):8685–90. https://doi.org/10.1073/pnas.0701361104.
10. Li Y, Agarwal P, Ozier O, Baliga N, Wang J. A Pathway-Based View of Human Diseases and Disease Relationships. PLoS ONE. 2009;4(2):4346. https://doi.org/10.1371/journal.pone.0004346.
11. Hu G, Agarwal P, Xu H, Markatou M, Friedman C. Human Disease-Drug Network Based on Genomic Expression Profiles. PLoS ONE. 2009;4(8):6536. https://doi.org/10.1371/journal.pone.0006536.
12. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ. Network-Based Elucidation of Human Disease Similarities Reveals Common Functional Modules Enriched for Pluripotent Drug Targets. PLoS Comput Biol. 2010;6(2):1000662. https://doi.org/10.1371/journal.pcbi.1000662.
13. Finding disease similarity based on implicit semantic similarity. J Biomed Inform. 2012;45(2):363–71. https://doi.org/10.1016/j.jbi.2011.11.017.
14. Žitnik M, Janjić V, Larminie C, Zupan B, Pržulj N. Discovering disease-disease associations by fusing systems-level molecular data,. Sci Rep. 2013;3:3202. https://doi.org/10.1038/srep03202.
15. Huang W, Wang P, Liu Z, Zhang L. Identifying disease associations via genome-wide association studies. BMC Bioinformatics. 2009;10(Suppl 1):68. https://doi.org/10.1186/1471-2105-10-S1-S68.
16. Kim S, Sohn K-A, Xing EP. A multivariate regression approach to association analysis of a quantitative trait network,. Bioinformatics (Oxford, England). 2009;25(12):204–12. https://doi.org/10.1093/bioinformatics/btp218.
17. Lewis SN, Nsoesie E, Weeks C, Qiao D, Zhang L. Prediction of Disease and Phenotype Associations from Genome-Wide Association Studies. PLoS ONE. 2011;6(11):27175. https://doi.org/10.1371/journal.pone.0027175.
18. Lee D-S, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabási A-L. The implications of human metabolic network topology for disease comorbidity,. Proc Natl Acad Sci U S A. 2008;105(29):9880–5. https://doi.org/10.1073/pnas.0802208105.
19. Milenković T, Memišević V, Bonato A, Pržulj N, Butler H. Dominating biological networks. PLoS ONE. 2011;6(8):23016. https://doi.org/10.1371/journal.pone.0023016.
20. Janjić V, Pržulj N, Benson DA, Bryant SH, Canese, et al. The Core Diseasome. Mol BioSyst. 2012;8(10):2614. https://doi.org/10.1039/c2mb25230a.
21. Ideker T, Sharan R. Protein networks in disease. Genome Res. 2008;18(4):644–52. https://doi.org/10.1101/gr.071852.107.
22. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56–68. https://doi.org/10.1038/nrg2918.
23. Leung A, Bader GD, Reimand J. HyperModules: identifying clinically and phenotypically significant network modules with disease mutations for biomarker discovery,. Bioinformatics (Oxford, England). 2014;30(15):2230–2. https://doi.org/10.1093/bioinformatics/btu172.
24. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L. Uncovering disease-disease relationships through the incomplete interactome. Science. 2015;347(6224):1257601.
25. Haberman Y, Tickle TL, Dexheimer PJ, Kim M-O, Tang D, Karns R, Baldassano RN, Noe JD, Rosh J, Markowitz J, et al. Pediatric crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J Clin Investig. 2014;124(8):3617.
26. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes,. Database: J Biol Databases Curation. 2015;2015:028. https://doi.org/10.1093/database/bav028.
27. Schaefer MH, Fontaine J-F, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: Integrating Protein Interaction Networks with Experiment Based Quality Scores. PLoS ONE. 2012;7(2):31826. https://doi.org/10.1371/journal.pone.0031826.
28. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, Mcmorran R, Wiegers J, Wiegers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2017. Nucleic Acids Res. 2017;45:. https://doi.org/10.1093/nar/gkw838.
29. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet. 2008;40(2):124–5. https://doi.org/10.1038/ng0208-124.
30. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015;47(9):979–86. https://doi.org/10.1038/ng.3359.
31. Anderson CA, Boucher G, Lees CW, Franke, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet. 2011;43(3):246–52. https://doi.org/10.1038/ng.764.
32. Franke A, McGovern DPB, Barrett JC, Wang, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010;42(12):1118–25. https://doi.org/10.1038/ng.717.
33. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008;40(8):955–62. https://doi.org/10.1038/ng.175.
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.
35. Glas J, Seiderer J, Wagner J, Olszak T, Fries C, Tillack C, Friedrich M, Beigel F, Stallhofer J, Steib C, et al. Analysis of il12b gene variants in inflammatory bowel disease. PloS one. 2012;7(3):e34349.
36. Ye X, Liu S, Hu M, Song Y, Huang H, Zhong Y. Ccr5 expression in inflammatory bowel disease and its correlation with inflammatory cells and β-arrestin2 expression. Scand J Gastroenterol. 2017;52(5):551–7.
37. Mora-Buch R, Dotti I, Planell N, Calderón-Gómez E, Jung P, Masamunt M, Llach J, Ricart E, Batlle E, Panés J, et al. Epithelial il-1r2 acts as a homeostatic regulator during remission of ulcerative colitis. Mucosal Immunol. 2016;9(4):950.
38. Mohamadzadeh M, Pfeiler EA, Brown JB, Zadeh M, Gramarossa M, Managlia E, Bere P, Sarraj B, Khan MW, Pakanati KC, et al. Regulation of induced colonic inflammation by lactobacillus acidophilus deficient in lipoteichoic acid. Proc Natl Acad Sci. 2011;108(Supplement 1):4623–30.
39. Jin D-Y, Jeang K-T. Isolation of full-length cdna and chromosomal localization of human nf-κb modulator NEMO to Xq28. J Biomed Sci. 1999;6(2):115–20.
40. Uhlig HH, Schwerd T, Koletzko S, Shah N, Kammermeier J, Elkadri A, Ouahed J, Wilson DC, Travis SP, Turner D, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology. 2014;147(5):990–1007.

41. Andreoletti G, Shakhnovich V, Christenson K, Coelho T, Haggarty R, Afzal NA, Batra A, Petersen B-S, Mort M, Beattie RM, et al. Exome analysis of rare and common variants within the NOD signaling pathway. Sci Rep. 2017;7:46454.
42. Borthakur A, Bhattacharyya S, Alrefai WA, Tobacman JK, Ramaswamy K, Dudeja PK. Platelet-activating factor-induced nf-κb activation and il-8 production in intestinal epithelial cells are bcl10-dependent. Inflamm Bowel Dis. 2009;16(4):593–603.
43. Yu Q, Zhang S, Chao K, Feng R, Wang H, Li M, Chen B, He Y, Zeng Z, Chen M. E3 ubiquitin ligase RNF183 is a novel regulator in inflammatory bowel disease. J Crohn's Colitis. 2016;10(6):713–25.
44. Schweickart VL, Epp A, Raport CJ, Gray PW. CCR11 is a functional receptor for the monocyte chemoattractant protein family of chemokines. J Biol Chem. 2000;275(13):9550–6.
45. Wells T, Proudfoot A. Chemokine receptors and their antagonists in allergic lung disease. Inflamm Res. 1999;48(7):353–62.
46. Nilsson J, Schoser B, Laforet P, Kalev O, Lindberg C, Romero NB, Dávila López M, Akman HO, Wahbi K, Iglseder S, et al. Polyglucosan body myopathy caused by defective ubiquitin ligase RBCK1. Ann Neurol. 2013;74(6):914–9.
47. Knecht C, Fretter C, Rosenstiel P, Krawczak M, Hütt M-T. Distinct metabolic network states manifest in the gene expression profiles of pediatric inflammatory bowel disease patients and controls. Sci Rep. 2016;6:32584.
48. Drev D, Bileck A, Erdem ZN, Mohr T, Timelthaler G, Beer A, Gerner C, Marian B. Proteomic profiling identifies markers for inflammation-related tumor–fibroblast interaction. Clin Proteomics. 2017;14(1):33.
49. Häcker H, Tseng P-H, Karin M. Expanding traf function: TRAF3 as a tri-faced immune regulator. Nat Rev Immunol. 2011;11(7):457.
50. Qiao YQ, Shen J, Gu Y, Tong JL, Xu XT, Huang ML, Ran ZH. Gene expression of tumor necrosis factor receptor associated-factor (traf)-1 and traf-2 in inflammatory bowel disease. J Dig Dis. 2013;14(5):244–50.
51. Shen J, Qiao Y, Ran Z, Wang T, Xu J, Feng J. Intestinal protein expression profile identifies inflammatory bowel disease and predicts relapse. Int J Clin Exp Pathol. 2013;6(5):917.
52. Shen J, Qiao Y-q, Ran Z-h, Wang T-r. Up-regulation and pre-activation of traf3 and traf5 in inflammatory bowel disease. Int J Med Sci. 2013;10(2):156.
53. Arch RH, Gedrich RW, Thompson CB. Tumor necrosis factor receptor-associated factors (TRAFs)—a family of adapter proteins that regulates life and death. Genes Dev. 1998;12(18):2821–30. https://doi.org/10.1101/gad.12.18.2821.
54. Ringel-Scaia VM, McDaniel DK, Allen IC. The goldilocks conundrum: Nlr inflammasome modulation of gastrointestinal inflammation during inflammatory bowel disease. Crit Rev™ Immunol. 2016;36(4).
55. Zhao H, Jaffer T, Eguchi S, Wang Z, Linkermann A, Ma D. Role of necroptosis in the pathogenesis of solid organ injury. Cell Death Dis. 2015;6(11):e1975.
56. Chen J, Xu H, Aronow BJ, Jegga AG. Improved human disease candidate gene prioritization using mouse phenotype. BMC Bioinformatics. 2007;8(1):392.
57. Adie EA, Adams RR, Evans KL, Porteous DJ, Pickard BS. Suspects: enabling fast and effective prioritization of positional candidates. Bioinformatics. 2006;22(6):773–4.
58. Tiffin N, Kelso JF, Powell AR, Pan H, Bajic VB, Hide WA. Integration of text-and data-mining using ontologies successfully selects disease gene candidates. Nucleic Acids Res. 2005;33(5):1544–52.
59. Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, Tranchevent L-C, De Moor B, Marynen P, Hassan B, et al. Gene prioritization through genomic data fusion. Nat Biotechnol. 2006;24(5):537.
60. Chen J, Bardes EE, Aronow BJ, Jegga AG. Toppgene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009;37(suppl_2):305–11.
61. Monaco C, Paleolog E. Nuclear factor κb: a potential therapeutic target in atherosclerosis and thrombosis. Cardiovasc Res. 2004;61(4):671–82.
62. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. Panther: a library of protein families and subfamilies indexed by function. Genome Res. 2003;13(9):2129–41.
63. Hanada T, Kobayashi T, Chinen T, Saeki K, Takaki H, Koga K, Minoda Y, Sanada T, Yoshioka T, Mimata H, et al. Ifnγ-dependent, spontaneous development of colorectal carcinomas in socs1-deficient mice. J Exp Med. 2006;203(6):1391–7.
64. El Yafi F, Winkler R, Delvenne P, Boussif N, Belaiche J, Louis E. Altered expression of type i insulin-like growth factor receptor in crohn's disease. Clin Exp Immunol. 2005;139(3):526–33.
65. Spalinger MR, McCole DF, Rogler G, Scharl M. Role of protein tyrosine phosphatases in regulating the immune system: implications for chronic intestinal inflammation. Inflamm Bowel Dis. 2015;21(3):645–55.
66. Yoshida Y, Yoshimi R, Yoshii H, Kim D, Dey A, Xiong H, Munasinghe ea. The transcription factor irf8 activates integrin-mediated tgf-beta signaling and promotes neuroinflammation. Immunity. 2014;40(2):187–98.
67. Domachowske JB, Bonville CA, Gao J-L, Murphy PM, Easton AJ, Rosenberg HF. The chemokine macrophage-inflammatory protein-1α and its receptor ccr1 control pulmonary inflammation and antiviral host defense in paramyxovirus infection. J Immunol. 2000;165(5):2677–82.
68. Manousou P, Kolios G, Valatas V, Drygiannakis I, Bourikas L, Pyrovolaki K, Koutroubakis I, Papadaki H, Kouroumalis E. Increased expression of chemokine receptor ccr3 and its ligands in ulcerative colitis: the role of colonic epithelial cells in in vitro studies. Clin Exp Immunol. 2010;162(2):337–47.
69. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):353–61. https://doi.org/10.1093/nar/gkw1092.
70. Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signalling pathways in cancer. Oncogene. 2007;26(22):3279–90. https://doi.org/10.1038/sj.onc.1210421.
71. Chow MT, Luster AD. Chemokines in Cancer. Cancer Immunol Res. 2014;2(12):1125–1131. http://cancerimmunolres.aacrjournals.org/content/2/12/1125.
72. Zhang J-M, An J. Cytokines, inflammation, and pain,. Int Anesthesiol Clin. 2007;45(2):27–37. https://doi.org/10.1097/AIA.0b013e318034194e.
73. Rakoff-Nahoum S, Medzhitov R. Toll-like receptors and cancer. Nat Rev Cancer. 2009;9(1):57–63. https://doi.org/10.1038/nrc2541.
74. Tervaniemi MH, Katayama S, Skoog T, Siitonen HA, Vuola J, Nuutila K, Sormunen R, Johnsson A, Linnarsson S, Suomela S, Kankuri E, Kere J, Elomaa O. NOD-like receptor signaling and inflammasome-related pathways are highlighted in psoriatic epidermis. Nat Publ Group. 2016. https://doi.org/10.1038/srep22745.
75. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
76. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616.
77. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. Rna-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18(9):1509–17.
78. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of rna-seq data. Genome Biol. 2010;11(3):25.
79. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001;29(4):1165–88.
80. Altaf-Ul-Amin M, Wada M, Kanaya S. Partitioning a ppi network into overlapping modules constrained by high-density and periphery tracking. ISRN Biomath. 2012;2012:11.
81. Altaf-Ul-Amin M, Tsuji H, Kurokawa K, Asahi H, Shinbo Y, Kanaya S. Dpclus: a density-periphery based graph clustering software mainly focused on detection of protein complexes in interaction networks. J Comput Aided Chem. 2006;7:150–6.
82. Fisher RA. On the interpretation of χ 2 from contingency tables, and the calculation of p. J R Stat Soc. 1922;85(1):87–94.
83. Fisher RA. Statistical Methods for Research Workers. New York: Springer; 1992, pp. 66–70. Breakthroughs in Statistics.
84. Metz CE. Basic principles of roc analysis. In: Seminars in Nuclear Medicine. New York: Elsevier; 1978. p. 283–98.
85. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on Machine learning. New York: ACM; 2006. p. 233–40.
86. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology. 1982;143(1):29–36.
87. Sing T, Sander O, Beerenwinkel N, Lengauer T. Rocr: visualizing classifier performance in r. Bioinformatics. 2005;21(20):3940–1.