Background:
Crohn’s disease is the most challenging, common type of autoimmune disorders due to which, intestine is inflamed but its causes have not been defined. Although Crohn’s disease has been often thought of as an autoimmune disorder, it can be triggered by whatever that lead to the inflammation in the whole bowel. Henceforth, exploring trade-off among this disease and genomic variants supposedly will enhance the identification of important genes and in turn to the possible therapeutic protocols. The aim of the present study was to identify any association between Crohn’s disease and single nucleotide polymorphisms.

Materials and Methods:
We retrieved single nucleotide polymorphism of genome-wide association studies (GWAS) on Crohn’s disease among which we extracted neighboring genes and expression variants in addition to pathways in which the affected genes are enriched.

Results:
Using GWAS data can help to explain the disease incidence. We noticed that genes harboring single nucleotide polymorphisms in Crohn’s disease were mainly enriched in interleukin pathways in addition to fatty acid and choline metabolism.

Conclusion:
The affected genes obtained by data analysis like fatty acid desaturase (FADS) and solute carrier family 22 (SLC22) could be proposed for future studies in order to clarifying their interactions by confident experiments and cross-sectional studies.

Keywords: Crohn’s disease, eQTL, Single nucleotide polymorphism
related protein 16-1 (ATG16L1) and Immunity-related GTPase family M protein (IRGM) play role for autophagy in patients with CD (5,6). An exaggerated immune response in patients with CD by the alteration in gut microbial community has been shown to be associated with Toll-like receptor 4 (TLR4), caspase-associated recruitment domain (CARD), interleukin 23R (IL-23R), Signal transducer and activator of transcription 3 (STAT3) and human leukocyte antigen (HLA), tumor necrotic superfamily member 15 (TNFSF15), interferon regulatory factor (IRF) and protein tyrosine phosphatase (PTPN22) genes involved in both innate and adaptive immune reactions, respectively. Most of the many disease-related single nucleotide polymorphisms (SNPs) discovered through the genome-wide association studies (GWAS) are supposed to be influential not only in regulating the expression of neighboring genes but also via another genomic contexts such as outside of protein-coding regions (7,8). GWAS characteristically aim in finding a statistically significant difference in genotypic frequency between a large numbers of individuals with a particular genotype for instance individuals with diseases compared with control population by means of SNP loci distributed in genomes. The SNPs that indicate significant links with disease condition in contrast to normal condition characterize the regions of genomes where likely to be harbored with biomarkers underlying the assessed diseases. Thereof, exploring the impacts of allelic variation on transcriptome using expression quantitative trait locus (eQTL) helps to understand how SNPs lie outside of protein coding regions account for the discovered pathogenic effects. Despite the comprehensive variant mapping by GWAS (9,10) in identifying more than 100 gene loci related to IBD, the pathogenesis of these conditions remains largely unknown. Several genetic and environmental factors have already been reported in correlation with CD, while a number of genome-wide linkage studies suggested different genomic loci as possible candidate genes for susceptibility to IBD (11-15).

Although the association between CD and SNPs has been presented in a number of studies (16-23), data mining methods can be still considered as a powerful approach toward the understanding of the etiology of CD.

In this study, the impact of the SNPs on patients with CD was investigated. We mainly aimed to elucidate the underlying relationships between CD and SNPs further eQTLs using in silico analytical approaches. The neighbor genes of genomics risk factors and underlying pathways that according to our previous knowledge thought to play important roles in the pathogenesis of CD were subsequently prioritized. Here we tried to provide a systematic view of genome-wide associations implicated in CD.

MATERIALS AND METHODS

Firstly, CD risk genomic variants were retrieved from the National Human Genome Research Institute (NHGRI), European Bioinformatics Institute (EBI) catalog of published GWAS (https://www.ebi.ac.uk/gwas/home). The potential regulatory changes triggering by these risk factors including transcription factor binding sites and histone modifications were then investigated in JASPAR (http://jaspar.genereg.net/), and TRANSFAC (http://genexplain.com/transfac/) databases and Roadmap Epigenomics HM ChIP-seq (www.roadmapepigenomics.org), respectively. Next, the distribution of these risk variants was explored by dbGaP database (https://www.ncbi.nlm.nih.gov/gap). The genes neighbor of these risk factors were further functionally enriched in GO molecular labels and biological pathways by DAVID (https://david.ncifcrf.gov/) and KEGG (http://www.genome.jp/kegg/) databases to find potential impacts of the risk alleles on the affected genes. Additionally, the CD GWAS risk variants were screened through the ExSNP database (http://www.exsnp.org/DZeQTL) to determine high-confidence eQTLs at linkage disequilibrium r2 > 0.8 and at 0.05 centi-Morgan distance to the risk SNPs. We finally took an intersection between the affected genes by these eQTL and curated CD related genes 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 the HPO. The shared genes affected by eQTLs were consequently fed into pathwAX web server (http://pathwax.sbc.su.se/) to find a network crosstalk of significant pathways. PathwAX contains KEGG pathway information in addition to networks of gene-gene links in model organisms.
Genomic Polymorphisms in Crohn’s Disease

RESULT

596 SNP-disease associations from 29 CD studies were retrieved from NHGRI at P value 5.00E-08 (supplementary table 1). We only selected SNPs with risk allele frequency > 0.05 of these, 292 mapped neighbor genes were extracted for further investigations. To determine if these genes are related to share diseases, we performed genotype-phenotype interaction analysis by dbGaP database. The genes were immediately grouped as to be related to CD, colitis, ulcerative, celiac disease, diabetes mellitus type 1, and IBD (table 1). The position matrix modelling of JASPAR and TRANSFAC databases showed TEFAP2C, LEF1, and SPI transcription factor binding sites affected by risk alleles at the upstream of 292 unique genes (figure 1a). Moreover, potential histone modifications at H4K20, H3K27, and H3K9 as transcriptional regulatory features were observed (figure 1b). The gene ontology molecular function, categorized the 292 mapped neighbor genes mainly under oncostatine-M, interleukin, chemokine, and CD40 receptors terminologies (figure 2a). Additionally, cytokines, interleukins, interferons, and immune systems were identified as enrichments of pathway themes (figure 2b). We then intersected 292 unique genes affected by risk variants with 899 CD associated genes through the DisGeNET database with at least one evidence from pfam (Supplementary table 2). The common genes were depleted in pathways such as colorectal cancer (figure 3). DisGeNET contains a list of diseases associated genes that has been collected based on the presence of genetic overlaps between diseases. In this list, among the 4,753,986 potential diseases associations, 13,064 diseases were found to share at least one gene with other diseases. Among these associations, CD (mesh: D003424) was significantly associated with mental retardation, skin diseases, tuberculosis, leukemia (b-cell, chronic), osteoporosis (postmenopausal), sepsis, hepatitis b (chronic), and pulmonary disease (chronic obstructive) at P value < 0.05. However, we only focused on colorectal cancer as a depleted pathway (figure 3) to find which gene links have not been significant enough to be related to colorectal cancer whereby these genes are not affected by colorectal cancer. Whereby, CREB5, EIF3C, DUSP5, RXRA, SMND1, TAGLN2, AKAP11, ENTPD7, DCLRE1B, CUTC, IFNAR1, ALDH2, USP1, PTK2B, CNTNAP2, ACSL6, HMHA1, TEC, PRDX5, PPM1, RBMX, NHP2L1, UBE2D1, RPL7, and TERF1 were characterized in colorectal depleted pathway. This procedure assures the approximation of CD associated genes involved in molecular networks, which are totally independent from colorectal

Table 1: The distribution of genes neighboring Crohn’s disease risk variants by dbGaP database. The potential impact of risk variants on the gene products as genotype and potential phenotypes as diseases has been investigated.

<table>
<thead>
<tr>
<th>Term</th>
<th>Adjusted P value</th>
<th>Z-score</th>
<th>Combined Score</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crohn’s disease</td>
<td>6.88E-27</td>
<td>-2.30498</td>
<td>150.5680109</td>
<td>PUS10, TNFSF15, IL23R, STAT3, MST1, ILTN1, PTPN22, CDKAL1, NOD2, FUT2, NELL1, LACC1, ATG16L1, FGFR10P, BSN, NKKX2-3, ZNF365</td>
</tr>
<tr>
<td>Colitis, Ulcerative</td>
<td>1.05E-12</td>
<td>-2.00191</td>
<td>64.0045137</td>
<td>IL10, C1orf106, PUS10, RTEL1, TNFRSF6B, TNFSF15, IL23R, CARD9, MST1, BTNL2, BSN, NKKX2-3</td>
</tr>
<tr>
<td>Celiac disease</td>
<td>1.96E-11</td>
<td>-2.01583</td>
<td>57.73037652</td>
<td>PUS10, ATXN2, ZMIZ1, TAGAP1L, PYPY3C, PTPN2, BACH2, HLA-DQA1, ICOSLG</td>
</tr>
<tr>
<td>Diabetes mellitus, type 1</td>
<td>1.49E-10</td>
<td>-1.95469</td>
<td>51.45321396</td>
<td>NOTCH4, PTPN22, BTLN2, TYK2, RASGRP1, BACH2, INS, IL2RA, HLA-DRA, SH2B3, PTPN2, HLA-DQA1, HLA-DQB1, MICB</td>
</tr>
<tr>
<td>Inflammatory bowel diseases</td>
<td>1.62E-09</td>
<td>-2.18517</td>
<td>51.82159741</td>
<td>ATXN2L, CYLD, TEC, TNFSF15, ATG16L1, ZMIZ1, IL23R, NCF4, SAG, NOD2</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>9.86E-09</td>
<td>-1.79875</td>
<td>39.07605826</td>
<td>NOS2, NOTCH4, IL23R, LST1, HLA-C, PBX2, IL12B, TYK2, TNF, PSORS1C1</td>
</tr>
<tr>
<td>Leprosy</td>
<td>3.76E-08</td>
<td>-1.57738</td>
<td>31.91207441</td>
<td>LACCC1, TNFSF15, RIPK2, HLA-C, HLA-DRA, HLA-DRB1, PSORS1C1, HLA-DQB1</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>5.02E-08</td>
<td>-1.66668</td>
<td>33.01439042</td>
<td>ZMIZ1, IL2RA, NOTCH4, STAT3, CPAMD8, TAGAP1, HLA-DRA, BTLN2, HLA-DQA1, PSORS1C1, HLA-DQA1, TNFSF1A</td>
</tr>
<tr>
<td>Vitiligo</td>
<td>7.38E-07</td>
<td>-0.95705</td>
<td>16.27380929</td>
<td>ZMIZ1, RNASET2, IL2RA, PTPN22, BTLN2, LLP</td>
</tr>
</tbody>
</table>

*The combined score is computed 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.
Fig. 1: Gene ontology enrichment and pathway analysis of genes neighboring Crohn’s disease risk alleles. (a) Functional classification of biological processes, (b) biological pathways by Enrichr and KEGG databases, respectively with default setting. The bars have been arranged top to down illustrating the number of genes and significance level assigned to each GO molecular terminology and biological pathway.

Fig. 2: Functional annotation of SNPs. (Right) the position matrix modelling (PMM) of the influence of SNPs on transcription factor binding sites from JASPAR vertebrate and TRANSFAC. (Left) the influence of SNPs on epigenetic modifications by Roadmap Epigenomics HM ChIP-seq. In, genes are in rows and transcription factors and histones are in columns.

Fig. 3: PathwayX results for the affected genes by Crohn’s disease related eQTLs. The table and pie chart summarize the pathway distribution. The table shows enriched depleted pathways at q-value 0.05 as defined cut-off threshold. Darker shades in colored boxes within the table indicate higher connectivity (links) that a query gene has.
We then seek the NHGR CD associations across 12,825 eQTL from ExSNP, to find eQTLs. Thereof, 82 SNPs were found to be mutual in NHGR GWAS and ExSNP databases. These SNPs targeted 165 and 51 unique genes from GWAS and ExSNP as eQTLs of these 17 genes where found to be affected as CD associated common eQTLs (table 2). Pathway analysis of the eQTLs showed fatty acid and choline metabolism by implicating FADS2, FADS1, SLC22A4, and SLC22A5 at adjusted \( P \) value < 0.05 (table 3). The genomic distribution of eQTLs and neighbor genes has been illustrated in figure 4.

**DISCUSSION**

In this article, 29 GWAS data sources have been employed to investigate the impacts of SNPs in the pathogenesis of CD in addition to elucidating the molecular pathways underlying this disorder.
Interestingly, 292 neighbor genes of 596 SNP-CD associations were enriched in chemokines (GO:0008009, GO:0031727, GO:0031731) and interleukins (GO:0004920). Oncostatin-M receptor (GO:0004924) was demonstrated to be the most significant GO biological label (figure 2a). Oncostatin-M has been recently reported with both the pro-inflammatory and anti-inflammatory effects (24,25). In pathways analysis, interferons, interleukins, and immune systems were more noticeable (figure 2b). Accordingly, the levels of specific cytokines and chemokines have been found to be elevated in patients with IBD (26-28). In concordance, the significant pathways relating to interleukins imply on an indispensable role of interleukins in immune-mediated disorders when interleukins have confirmed roles in several immune and inflammatory events that is occurred in autoimmune diseases. Thus, further studies could help to evaluate the prognostic roles of interleukins in crosstalk of CD and genomic variants. As depicted in figure 2, 596 risk factors trigger potential regulatory changes mostly in binding sites of some transcription factors including TFAP2C at the top in upstream of 292 affected genes (rows of left clustergrammer). Transcription factor AP2C (TFAP2C) has been revealed to play a central role in controlling multiple pathways of estrogen signaling hereby growing evidence proposes that estrogen modulates gut inflammation through the alteration of estrogen receptors in peripheral blood T lymphocytes (29,30).

In the next step, the potential impact of risk variants on the gene products as genotype and potential resulting phenotypes, here CD, has been investigated by dbGaP database. As shown in table 1, 596 SNP of 29 CD studies extracted from NHGRI at $P$ value 5.00E-08 have been linked to CD as the
most significant and later with the other types of autoimmune-diseases including colitis ulcerative, celiac disease, diabetes mellitus type1, IBD, psoriasis, leprosy, multiple sclerosis, and vitiligo. Regarding diabetes mellitus type1, TNF-α has been shown in developing insulin resistance during obesity (31) and in autoimmune destruction of pancreatic β-cells (32) that could explain the effects of TNF-α antagonism on glucose metabolism. Systemic changes of microbiota products may have promoted insulin resistance (33). Regarding psoriasis, the most important susceptibility locus is known to be located on the gene encoding TNF-α providing new venues to the associations of autoimmune-diseases (34). HLA-DR-DQ, RIPK2, CCDC122-LACC1, and NOD2 have been detected as leprosy susceptibility, related to the NOD2 signaling pathway and are CD susceptibility loci. NOD2 is the first recognized susceptibility locus in CD, which has been reported to trigger an immune response to bacterial cell wall. However, in this study HLAs seem to be the most frequent susceptibility loci based on the dbGaP database (table 1). Accordingly, predisposing genetic factors, and HLAs contribute to the genetic pathogeny of autoimmune disorders such as IBD at 30%-50% (35,36). As the main type of chronic IBD, CD has been shown to be associated with specific HLA class I and II (37). Keeping with these, probably genomic risk variants are targeting a circuit of interferons coding genes responsible for inflammation development (38). In the light of previous studies, SNPs found in a cohort of 29 GWAS likely discover susceptibility variants for pathways contributing in intestinal inflammation whereby anti-TNF therapy could be suggested as a clinical target. Further, 82 CD-associated mutual SNPs identified in the NHGR GWAS and ExSNP databases targeting 165 and 51 unique genes, respectively were introduced as eQTLs of those 17 genes were considered as CD associated eQTLs (table 2). The pathway analysis of these 17 genes affected by eQTLs exhibited fatty acid and choline metabolism pathways with the implication of FADS2, FADS1, SLC22A4, and SLC22A5 at adjusted P value < 0.05 (table 3). Serum levels of choline and its derivatives are lower in patients with IBD compared with healthy individuals as a result of pro-inflammatory cytokine levels (39). Regarding the fatty acid metabolism, accumulating evidence has highlighted the co-existence of non-alcoholic fatty liver disease and IBD, which supposedly is contributed by agents such as a severe alteration in intestinal permeability, gut dysbiosis, and chronic inflammatory response (40). As illustrated in figure 4a, most of the aforementioned eQTLs have been located on 5q mainly distributed through the intronic regions (figure 4b). These results suggest that eQTLs especially ones located in intronic regions could practically contribute in inflammatory events. In the present study, fatty acid desaturase 1 (FADS1) with rs102275, rs4246215, and rs4077515 has been identified as a gene neighboring 3 eQTLs associated to CD and in conjugation with FADS2 contribute to the biosynthesis of unsaturated fatty acids (hsa01040) and fatty acid metabolism (hsa01212) pathways. A number of FADs play a pivotal role in the production of omega-6 and omega-3 polyunsaturated fatty acid (PUFAs), for instance eicosapentaenoic, docosahexaenoic acids, and arachidonic acid obtained by the dietary contains fatty acid (41,42). The SNPs within the FADs may impact on the FADS2 expression and consequently the levels of PUFAs within cells and the inflammatory signaling through eicosanoids, and these pathways have been suggested as resolving mediators in chronic inflammatory disorders like atherosclerosis and other immune-mediated disorders (43). The diet-eQTL links could be biologically relevant to a sort of gene-environment interactions underlying CD. SLC22A5 and SLC22A4 risk factors including rs12521868 and rs2188962 as intronic variants were related to choline metabolism in cancer pathway (hsa05231). These genes are members 4 and 5 of the solute carrier family 22 (SLC22A4 [OMIM 604190] and SLC22A5 [OMIM 605956]) encode the organic cation transporters (OCTN) 1 and OCTN2 located through the IBD susceptibility locus IBD5 on chromosome 5q31 have been proposed to be related to CD by several genome screenings. Seemingly, decreasing the OCTN activity or expression may cause the alteration in gut bacterial community, which in turn generates susceptibility to CD (44-46).

Taken together, a number of neighboring genes of genomic risk variants were highlighted in this study whose interactions with each other may contribute to the development of CD. This is mainly occurred through the genes involved in chemokine and cytokine...
pathways, which could be a relevance of the necessity in declining the expression of these chemicals to inhibit inflammatory responses in susceptible individuals as a therapeutic strategy.

However, this analysis is challenged by the disadvantage of inevitable overestimation in computational approaches, thus, employing more stringent parameters in predicting the in silico relationships would be apparently helpful in acquiring more reliable results and overcoming any inaccuracy coming from the nature of computational methodologies at the first come.

CONCLUSION
The main goal of this analysis was exploiting interplay between gene and SNPs, which are thought to be influential in the pathogenesis of CD. Utilizing GWAS data with pooling information of regulatory interactions can help to discover underlying mechanisms and enlighten more molecular underpinnings of inflammatory conditions. We observed that the identified affected genes by genomic risk factors, mainly guided to interleukin signaling pathway and fatty acid and choline metabolism. To summarize, selected genes viz. FADs and SLC22s could be taken into account for future detection and therapeutic targets by experimental investments.

ACKNOWLEDGEMENT
We thank Dr. Nooshin Omranian, scientific staff in Systems Biology and Mathematical Modelling Group, Max Planck Institute for Molecular Plant Physiology, Potsdam, Germany for her precious guidance.

CONFLICT OF INTEREST
The authors declare no conflict of interests related to this work.

REFERENCES
17. Iezzi LE, Medeiros BA, Feitosa MR, Ribeiro de Almeida


