Genetic susceptibility for inflammatory bowel disease across ethnicities and diseases
van Sommeren, Suzanne

DOI:
10.33612/diss.100597247

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 6

Extra-intestinal manifestations and complications in inflammatory bowel disease – From shared genetics to shared biological pathways.

Suzanne van Sommeren*, Marcel Janse*, Juha Karjalainen*, Rudolf Fehrmann, Lude Franke*, Jingyuan Fu* and Rinse K. Weersma*

Inflammatory Bowel Diseases 2014;20(6):987–94
*These authors contributed equally
ABSTRACT

Background: The clinical presentation of the inflammatory bowel diseases (IBD) is extremely heterogeneous and is characterized by various extra-intestinal manifestations and complications (EIM). Increasing genetic insight for IBD and EIM shows multiple shared susceptibility loci. We hypothesize that, next to these overlapping genetic risk loci, distinct disease pathways are shared between IBD and EIM.

Methods: The overlapping genetic risk loci for IBD and its EIM were searched for in literature. We assessed shared disease pathways by performing an extensive pathway analysis by protein-protein interaction (PPI) and co-transcriptional analysis, using both publically available and newly developed databases.

Results: Reliable genetic data was available for primary sclerosing cholangitis (PSC), ankylosing spondylitis (AS), decreased bone mineral density (BMD), colorectal carcinoma (CRC), gallstones, kidney stones and deep venous thrombosis (DVT). We found an extensive overlap in genetic risk loci, especially for IBD and PSC and AS. We identified 370 PPIs, of which 108 are statistically specific. We identified 446 statistically specific co-transcribed gene pairs. The interactions are shown to cluster in specific biological pathways.

Conclusion: We show that the pathogenetic overlap between IBD and its EIM extends beyond shared risk genes to distinctive shared biological pathways. We define genetic background as a risk factor for IBD-EIM alongside known mechanisms as malabsorption and medication. Clustering patients based on distinctive pathways may enable stratification of patients to predict development of EIM.
INTRODUCTION

The inflammatory bowel diseases (IBD) are complex diseases encompassing multiple specific sub-phenotypes. Included in IBD’s disease spectrum are various extra-intestinal manifestations and complications (together named EIM). In addition to having a high incidence in IBD patients, the EIM can actually cause more morbidity than IBD itself. Frequently occurring EIM are immune mediated and involve multiple organ systems like the skin (erythema nodosum, pyoderma gangrenosum), the eyes (uveïtis, episcleritis), the liver (primary sclerosing cholangitis, PSC) or spine (ankylosing spondylitis, AS).1 The skin and eye diseases occur during active intestinal disease, while PSC and AS can also occur while IBD itself is in a quiescent state. Decreased bone mineral density (BMD), gallstones and kidneystones are considered to be mainly driven by the (metabolic) consequences of malabsorption and chronic use of steroids. Longstanding active disease predisposes to the development of colorectal carcinoma (CRC), necessitating frequent surveillance colonoscopies.2 Despite this knowledge, it remains unclear why the IBD phenotype includes so many EIM, in what way disease pathogenesis is overlapping between IBD and EIM and why certain patients develop a particular EIM and others do not. One possible answer to these questions is an overlap in genetic architecture; this would explain why these EIM co-occur so often with IBD and moreover variation in the genetic make-up in individuals could partly explain the variability in disease phenotype. It has been demonstrated before that many immune mediated diseases share a genetic background, and because IBD and its EIM co-occur so frequently we hypothesize that they have a common genetic background.3

It has become clear that IBD is not caused by an abnormality in a single gene leading to a single uniform disease, but is a consequence of the perturbations of complex pathways leading to multiple specific sub-phenotypes.2 Recently there has been tremendous progress in unraveling the genetic background of ulcerative colitis and Crohn’s disease, to date 163 independent genetic susceptibility loci have been identified. The identified single genetic risk factors have been shown to cooperate in disease relevant pathways.4 Similarly, genetic variants predisposing to many EIM have been established, of which some are shared with IBD. We hypothesize that, next to these overlapping genetic risk loci, disease pathways are shared between IBD and EIM.

To understand how these genetic variants predispose to IBD-EIM, we performed a survey of overlapping genetic loci. For this we included the following EIM or complications of IBD: ankylosing spondylitis, primary sclerosing cholangitis, decreased bone mineral density, colorectal carcinoma, erythema nodosum, pyoderma gangrenosum, uveïtis, episcleritis, kidney stones, gall stones and venous thrombosis. Next we analyzed co-regulation between all pairs of genes within IBD loci and EIM loci (with available and reliable genetic data), for protein-protein interactions (PPI) and co-transcription using large databases (including > 80,000 human Affymetrix mRNA expression datasets).

We show that the pathogenetic overlap between IBD and its EIM is partly driven by a shared genetic predisposition and extends beyond purely shared risk genes to distinct shared biological pathways.

MATERIAL AND METHODS

Literature search and selection of associated loci

Genetic association data for Inflammatory Bowel Disease was extracted from the most recent analyses.4 An extensive literature search was performed in Pubmed and the GWAS catalogue (www.genome.gov/gwastudies) till October 2011 to assess available data on genetic susceptibility of the EIM. For EIM all performed GWAS and GWAS meta-analyses were included. We also included large candidate studies if limited GWAS data was available. From GWAS we included SNPs with reported genome wide significance (in most studies defined as p < 5 × 10⁻⁸). For candidate studies we included SNPs with
For PSC, we also included the GP-BAR1 locus because this locus showed functional evidence. When there were only small candidate gene studies available or with conflicting results we did not include the identified susceptibility loci for the pathway analyses. All well-established genetic loci in IBD and the EIM were included and genetic overlap was assessed as presented in Supplementary table 1.

**Mapping SNPs to Genes**

For the co-transcriptional and protein-protein interaction analyses all candidate genes in the loci were included. Therefore, we assessed which genes are located at associated loci in IBD and EIM. The disease-associated SNPs were linked to proximate candidate genes in linkage disequilibrium (LD) with them, using a previously described approach. We downloaded the recombination hotspot and LD information from www.hapmap.org for CEU population (release 28) and genome build hg18. The information of gene positions was based on genome build hg18 and was downloaded from UCSC Genome Browser. For each of the associated SNPs, we first defined the disease locus as the region containing the SNPs with LD $r^2 > 0.5$ to the associated SNP and then extended it to the nearest recombination hotspot. This region was further extended 100 kb on each side to include the potential regulatory regions for genes. If any transcript isoform of a gene overlaps with the defined disease locus, this gene was included as a candidate gene. Thus an $i^{th}$ IBD locus with $n$ number of candidate genes was defined as $IBD_i \{g_1, g_2, ..., g_n\}$; and a $j^{th}$ EIM locus with $z$ number of candidate genes was defined as $EIM_j \{g_w, g_u, ..., g_j\}$.

**PPI between IBD and EIM loci**

We used the PPI database that was described by Rossin et al. and Lage et al. and extracted all the direct interactions between the candidate genes in IBD loci and the candidate genes in EIM loci, for example between the given $i^{th}$ IBD loci $IBD_i \{g_1, g_2, ..., g_n\}$ and the given $j^{th}$ EIM loci $EIM_j \{g_w, g_u, ..., g_j\}$. To assess the specificity of the interactions between loci, we further calculated p-values empirically by testing interactions between the $EIM_j$ loci and the mimicked random $IBD_i$ locus with similar number of genes: if $n \leq 10$, the mimicked locus must contain the same $n$ number of genes; if $n > 10$, the mimicked random locus could contain the number of genes within 10% variation (i.e., gene numbers within $0.9 \times n$ and $1.1 \times n$). The random $IBD_i$ locus was mimicked 1,000 times. We then scored, out of 1,000 random loci, how many times we could observe at least one interaction between the $EIM_j$ locus and the random $IBD_i$ locus and calculated the empirical p-value. The significance threshold was controlled at 0.05.
identify such relationships. First, we mapped each probe set on each of the four platforms to a gene and averaged eigenvector coefficients over probe sets mapping to the same gene. Probes with no or unambiguous mapping (due to e.g. cross-hybridization) were excluded from further analysis. Second, we calculated for each pair of human genes a Pearson product-moment correlation coefficient over the 2,200 eigenvector coefficients of each platform that contained probe sets for the genes or, for the mouse and rat platforms, their most similar orthologs. Both probe set and gene ortholog mapping information were downloaded from Ensembl Biomart (Ensembl release 65). The correlation coefficients were further converted to Z-scores to account for different numbers of available eigenvectors for pairs of genes due to missing orthologs. The Z-score for a pair of genes can be positive or negative and describes the similarity of their regulation. The genes (19,997 in total) form the nodes and the correlation coefficients the edges of the network.

We used this co-transcriptional network to find significantly co-transcribed genomic regions implicated by SNPs associated with IBD or EIM. First, we determined regions of interest for each phenotype. For each SNP associated with a phenotype, we defined implicated genes by identifying furthest SNPs in both the 3' and 5' directions in LD with it ($r^2 > 0.5$ based on CEU HapMap 2), and extending the region first to the nearest recombination hotspot and then an additional 100 kb in both directions. Each gene in the co-regulation network overlapping with this region was considered implicated by the SNP and part of the locus. Overlapping loci were merged together.

We then examined the potential co-transcription between the regions associated with IBD and those associated with each of the EIM by finding pairs of genes that show the strongest co-transcription in either direction in each pair of loci between the phenotypes. To ascertain whether genes in a locus were often co-transcribed with genes in other regions, we ran a permutation test with 1,000 permutations randomly picking a region from the genome with a similar centimorgan range and a similar number of genes as the IBD locus and repeating the procedure with the random locus replacing the IBD locus. We then applied a 5% false discovery rate to eliminate potential false positive gene pairs.

**Biological interpretation of interactions**

To identify pathways in which the interacting genes of IBD and the different EIM play roles, we used DAVID Bioinformatics database. DAVID uses several resources like GO-terms, KEGG – the Kyoto encyclopedia of genes and genomes and BioCarta (www.biocarta.com) to cluster genes together in pathways.

**RESULTS**

**Genetic overlap**

We first assessed genetic overlap between IBD and all EIM: AS, PSC, BMD, CRC, erythema nodosum, pyoderma gangrenosum, uveïtis, episcleritis, kidney stones, gall stones and venous thrombosis. Most robust genetic data was available for AS, CRC, BMD and PSC. For gallstones, kidney stones and deep venous thrombosis limited genetic data was available, only candidate gene studies were available, while for the skin EIM (erythema nodosum, pyoderma gangrenosum) and eye EIM (uveïtis, episcleritis) no genetic data was available (table 1). Supplementary table 1 shows all associated loci per EIM. The classical immune-driven diseases AS and PSC have the largest number of risk loci that are shared with IBD: In AS 13 out of 18 known risk loci are shared with IBD, for PSC this is 10 out of 14. In CRC 3 out of 15 risk loci are shared with IBD. In AS 13 out of 18 known risk loci are shared with IBD, for PSC this is 10 out of 14. In CRC 3 out of 15 risk loci are shared with IBD. For gallstones, kidney stones and deep venous thrombosis limited genetic data was available, only candidate gene studies were available, while for the skin EIM (erythema nodosum, pyoderma gangrenosum) and eye EIM (uveïtis, episcleritis) no genetic data was available (table 1). Supplementary table 1 shows all associated loci per EIM. The classical immune-driven diseases AS and PSC have the largest number of risk loci that are shared with IBD: In AS 13 out of 18 known risk loci are shared with IBD, for PSC this is 10 out of 14. In CRC 3 out of 15 risk loci are shared with IBD, for PSC this is 10 out of 14. In CRC 3 out of 15 risk loci are shared with IBD. For gallstones, kidney stones and deep venous thrombosis limited genetic data was available, only candidate gene studies were available, while for the skin EIM (erythema nodosum, pyoderma gangrenosum) and eye EIM (uveïtis, episcleritis) no genetic data was available (table 1). Supplementary table 1 shows all associated loci per EIM. The classical immune-driven diseases AS and PSC have the largest number of risk loci that are shared with IBD: In AS 13 out of 18 known risk loci are shared with IBD, for PSC this is 10 out of 14. In CRC 3 out of 15 risk loci are shared with IBD, for PSC this is 10 out of 14. In CRC 3 out of 15 risk loci are shared with IBD. For gallstones, kidney stones and deep venous thrombosis limited genetic data was available, only candidate gene studies were available, while for the skin EIM (erythema nodosum, pyoderma gangrenosum) and eye EIM (uveïtis, episcleritis) no genetic data was available (table 1).
Chapter 6

PPI and co-transcriptional regulation

We then searched for statistical significant PPIs and co-transcriptional interactions between IBD genes and EIM genes. From the PPI database we identified 370 PPIs between IBD and EIM loci, excluding PPIs between shared IBD-EIM loci. The (extended) HLA locus is involved in 57 out of 370 interacting loci. After 1,000 permutations of testing PPIs between the tested EIM locus and a mimicked IBD locus 86/370 PPIs turned out to be significantly specific (figure 2). Interacting IBD-EIM loci can hold multiple PPIs between different genes in the loci. These 370 interacting loci hold a total of 915 PPIs between IBD and EIM genes (441 of these PPIs involve a gene originating in the HLA locus).

From the co-transcriptional network we created we identified 10,890 significant co-transcribed genes between IBD and its EIM. After 1,000 permutations 446 turned out to be significantly specific (Supplementary figure 1).

Pathway annotation

We then annotated the biological pathways in which the interacting genes of IBD and the different EIM play roles, by using the DAVID Bioinformatics database. To ensure the reliability of the observed pathways, an extensive literature search was performed. Due to the small amount of associated loci in gallstones, venous thrombosis and kidney stones, few interactions were found and no defined pathways were established. For AS, PSC, CRC and BMD we find several interesting pathways where genes with interactions via PPI or co-transcription play roles. These are highlighted in figure 3 and we discuss them in detail further on.

DISCUSSION

We aimed to give an overview of the current genetic data on IBD and its EIM and provide additional downstream pathway analyses using publicly available and new locally developed databases. The data as shown here are first of all important to understand the pathogenetic mechanisms that lead to the co-occurrence of immune mediated extra-intestinal manifestations or the development of long-term complications. Second, in the future clustering patients based on distinctive involved pathways may enable stratification of patients to predict development of EIM and investigate specific screening protocols.

Colorectal Carcinoma

The most prominent shared pathway between IBD and CRC is regulation of the intrinsic epithelial barrier integrity. Central in this pathway is CDH1; encoding E-cadherin, which anchors cells together in adherens junctions and is
Extra-intestinal manifestations and complications in inflammatory bowel disease – From shared genetics to shared biological pathways.

Figure 1. Chromosomes with risk loci for IBD and its EIM.
Each dot represents the position of an associated locus on the chromosome. For shared risk loci the candidate gene is depicted. Complete loci information can be found in supplementary table 1.
associated to both IBD and CRC. Other interacting genes influence epithelial barrier function by regulating the cellular actin cytoskeleton (RHPN2, FMN1, RH0A, ARPC2, CAPN10) or establishing attachment of the cytoskeleton to the basement membrane (DAG1) or in the basement membrane (LAMB1, LAMA5). The TGF-β signaling pathway harbors many genes associated to CRC or IBD, illustrating the crucial regulatory role TGF-β plays in different relevant pathways including epithelial barrier function and the immune system. The regulatory functions of TGF-β are shared with Wnt signaling, which is an established pathway in CRC and has recently been associated with UC-related CRC.
Bone Mineral Density

Wnt signaling genes are abundantly present in BMD associated loci (WNT16, CTNNB1, GPR177, LRP5) and have interactions with IBD genes like CDH1 and SMAD3. Wnt signaling has been shown to influence the RANK/RANKL/OPG (TNFSF11, TNFRSF11A, TNFRSF11B) pathway.\(^4\) TNFSF11 is associated with BMD and was previously identified as a CD locus,\(^4\) however, in the most recent IBD analyses\(^4\) this locus could not be replicated. Therefore, the association of this locus with IBD remains questionable. TNFSF11 has interactions with its receptor TNFRSF11A and the inhibiting decoy receptor TNFRSF11B, which are both associated to BMD. This pathway is involved in both bone

---

**Figure 3. Summary of shared biological pathways in IBD and its EIM.**

Each square represents an associated gene. An IBD gene with multiple interactions across the EIM is depicted multiple times. A blue line represents the most significant co-transcriptional interaction between two genes. PPIs significant after permutations are represented by black lines, non-significant PPIs by grey lines. Pathways are encircled, sub-pathways within a bigger pathway are depicted in gray circles.

---

Extra-intestinal manifestations and complications in inflammatory bowel disease – From shared genetics to shared biological pathways.
homeostasis as regulating T cell–dendritic cell communications, dendritic cell survival and lymph node organogenesis, crucial in IBD pathogenesis. We see many co-regulation between genes in this pathway and genes that are involved in the adaptive immune system (e.g. PTPN22, CCR6). These findings highlight possibilities for e.g. drug repositioning for Denosumab (Amgen/GlaxoSmithKline) which targets TNFSF11 and is a marketed drug for the treatment of postmenopausal women at high risk of fracture with osteoporosis. The gene ITGB7, encoding integrin beta-7, which associates with alpha-4 to form integrin alpha-4/beta-7, resides in a BMD associated locus and interacts with the IBD gene CDH1. Natalizumab (Biogen), registered for use in Crohn’s disease, blocks homing of lymphocytes to vascular endothelial cells of the gastro-intestinal tract via MADCAM1 on the endothelium and integrin alpha-4/beta-7 on lymphocytes. These findings highlight the importance of investigating pathways associated with the broader IBD-EIM phenotype.

Primary Sclerosing Cholangitis and Ankylosing Spondylitis

Most striking in the analysis for PSC and AS is the large number of overlapping loci. Even if we focus on PSC and AS genes that are not shared with IBD we find many interactions in pathways that are shared with IBD. PSC genes and their connected IBD genes play several roles in T cell signaling, like T cell apoptosis (UBASH3A, BCL2L1, FOXO1 and IRF8) and the JAK-STAT signaling pathway (SOCS1, JAK2, STAT3 and TYK2). For AS the largest shared pathway is the T cell apoptosis pathway, with many interactions with IBD genes. The AS associated genes TAPBPL and NPEPPS function in the same process as the AS-IBD shared gene ERAP1, namely the process of protein antigen binding to the MHC-I molecules which is likely to be crucial to both AS and IBD pathogenesis.

Given the fast moving field of IBD genetics we do not provide a definitive analysis. Very recently the number of shared genes between IBD and other diseases has increased with the publication of several papers using a customized GWAS chip (Immunochip, Illumina, San Diego, CA) focussed on immune mediated diseases. At the time of the current analyses these data were not available but the number of shared pathways is expected to increase substantially. Large-scale analyses of completely phenotyped IBD cohorts (including reliable data on EIM) are necessary to further investigate whether patients with specifics EIM are enriched for risk genes or risk pathways as presented here.

In conclusion, we show that the pathogenic overlap between IBD and its EIM or complications extends substantially beyond purely shared risk genes to shared biological pathways. We identified numerous statistically significant interactions clustering in several distinct biological pathways between IBD and the different EIMs. Hereby we further highlight the genetic background as risk factor for IBD-EIM next to known mechanisms as malabsorption, chronic inflammation or medication.

SUPPLEMENTARY DATA

Supplementary data are available online:
Supplementary table 1 http://links.lww.com/IBD/A456
Supplementary figure 1 http://links.lww.com/IBD/A457

REFERENCES

91

Extra-intestinal manifestations and complications in inflammatory bowel disease – From shared genetics to shared biological pathways.


38. Trégouët DA, Heath S, Saut N, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. Blood 2009;113:5298–5303


