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.
Correlation of genetic risk and mRNA expression in a Th17/IL23 pathway analysis in Inflammatory Bowel Disease

Karin Fransen*, Suzanne van Sommeren*, Harm-Jan Westra, Monique Veenstra, Titia Lamberts, Rutger Modderman, Gerard Dijkstra, Jingyuan Fu, Cisca Wijmenga, Lude Franke, Rinse K. Weersma, Cleo C. van Diemen.

*These authors contributed equally
ABSTRACT

Background: The Th17/IL23 pathway has both genetically and biologically been implicated in the pathogenesis of the Inflammatory Bowel Diseases (IBD), Crohn's disease (CD) and ulcerative colitis (UC). So far, it is unknown whether and how associated risk variants affect expression of the genes encoding for Th17/IL23 pathway proteins.

Methods: 10 IBD associated SNPs residing near Th17/IL23 genes were used to construct a genetic risk model in 753 Dutch IBD cases and 1,045 controls. In an independent cohort of 40 CD, 40 UC and 40 controls the genetic risk load and presence of IBD were correlated to qPCR generated mRNA expression of nine representative Th17/IL23 genes in both unstimulated and PMA/CaLo stimulated peripheral blood mononuclear cells (PBMCs). In 1,240 individuals with various immunological diseases with whole genome genotype and mRNA-expression data we also assessed correlation between genetic risk load and differential mRNA-expression and sought for SNPs affecting expression of all currently known Th17/IL23 pathway genes (cis-eQTLs).

Results: The presence of IBD, but not the genetic risk load, was correlated to differential mRNA expression for IL6 in unstimulated PBMCs and to IL23A and RORC in response to stimulation. The cis-eQTL analysis showed little evidence for correlation between genetic risk load and mRNA expression of Th17/IL23 genes, since we identified for only two out of 22 Th17/IL23 genes a cis-eQTL SNP that is also associated to IBD (STAT3 and CCR6).

Conclusion: Our results suggest that only the presence of IBD and not the genetic risk load alters mRNA expression levels of IBD associated Th17/IL23 genes.
Correlation of genetic risk and mRNA expression in a Th17/IL23 pathway analysis in Inflammatory Bowel Disease

INTRODUCTION

Inflammatory bowel diseases (IBD) can be divided into two main forms – Crohn’s disease (CD) and ulcerative colitis (UC). They are disabling diseases characterized by a chronic relapsing inflammatory response to commensal microflora of the gut. The underlying pathogenesis is largely unknown but there is a clear genetic susceptibility. 1,2 Genome wide association (GWA) studies have been extremely successful in identifying risk loci for IBD; single nucleotide polymorphisms (SNPs) in 163 independent risk loci have been associated to IBD. 3 Functional consequences of the associated loci remain largely unknown while unraveling the functional implication of genetic associations is essential for the clarification of the pathogenesis. Since many loci contain multiple genes, one of the first steps in this process is prioritization of candidate genes. Different gene prioritization methods have been applied to GWAS results, such as identifying protein altering coding SNPs, checking expression quantitative trait locus (eQTL) data, protein-protein interactions (such as DAPPLE), text-mining (such as GRAIL) and co-expression of genes with known implicated genes. 4,5 Because these analyses use interconnectivity between genes they are also able to highlight disease associated pathways. These combined methods resulted in the genetic implication of, amongst others, the Th17/IL23 pathway in the pathogenesis of IBD. 6

The Th17/IL23 pathway acts in Th17 cells, which are suggested to play a role in the chronic inflammatory processes. Next to the genetic associations, functional studies highlighted the role of the Th17/IL23 pathway in IBD. First, multiple studies showed elevated messenger RNA (mRNA) expression of genes involved in the Th17/IL23 pathway in colonic biopsies of cases compared to controls and inflamed compared to non-inflamed tissue. 7–11 Second, elevated numbers of mucosal Th17 cells have been measured in active versus quiescent Crohn’s disease. 12

The Th17/IL23 pathway contains at least 22 proteins, 10 of the encoding genes reside in loci that are associated to IBD (‘IBD associated Th17 genes’: IL23R, TYK2, RORC, IL21, IL12B, CCR6, JAK2, IFNγ, SMAD3 and STAT3, figure 1). Despite these breakthroughs, it still remains unclear how the genetic risk variants in the Th17/IL23 pathway exactly contribute to this aberrant function of the Th17/IL23 pathway. The current view is that genetic risk variants alter mRNA expression levels of nearby genes and in this way disturb the function of a pathway. 13–16

In this study we investigated the correlation between the (combined) genetic risk (load) of IBD associated Th17 genes and the mRNA expression profile of the Th17/IL23 pathway.

MATERIALS AND METHODS

More extensive descriptions are available in the supplementary material and methods section.

Samples

All patient samples were collected at the outpatient clinic of the University Medical Centre in Groningen, the Netherlands. IBD patients were diagnosed according to the standard clinical criteria by endoscopy, radiology and histopathology 17 and gave written informed consent. The ‘genetic risk model cohort’, consisting of 1798 individuals, was used to construct and test a genetic risk model for the IBD associated Th17 genes. A separate cohort of 118 individuals (39 CD, 40 UC, 39 controls), the ‘IBD-PBMC cohort’, was used to determine mRNA-expression of nine Th17 representative genes in peripheral blood mononuclear cells (PBMCs), which was then correlated to both disease status and genetic risk load. To prevent influence of immunomodulators and disease activity on the mRNA-expression profile of the patients in the IBD-PBMC cohort, IBD patients did not use any systemic anti-inflammatory drug; only topical mesalasine treatment was allowed, and patients were in clinical remission. 18 Characteristics of these two cohorts are listed in table 1.

A third in-house cohort of 1,240 patients with various immunological diseases with whole genome genotype and mRNA-expression data from whole blood was used to replicate
Figure 1. Schematic representation of the T-helper 17/IL23 pathway.

Antigen presenting cells are activated by extracellular microbes, e.g., fungi, through the toll-like receptor (TLR) and in response, produce several cytokines. First, IL6 and IL1 promote the differentiation of naive T-helper cells into Th17 cells through the 9P130 and the IL6 receptor; second, interleukin-23 alpha (IL23a) and interleukin-12 subunit p40 (IL12b) are produced and form the cytokine complex IL23, which promotes not only differentiation but also proliferation of Th17 cells through the IL23 receptor (IL23R) on Th17 cells. The IL23R consists of 2 subunits, interleukin-12 receptor-beta-1 (IL12RB1) and IL23R and activates together with IL6R and 9p130 the Janus Kinase 2 (JAK2) gene. Simultaneously, Tyrosine Kinase 2 (TYK2) is stimulated through the IL23R. TYK2 and JAK2 stimulate transcription factors signal transducer and activator of transcription 3 (STAT3) and RAR-related orphan receptor C (RORC, RORyt) to produce the proinflammatory cytokines interleukin-21 (IL21), interleukin-22 (IL22), and interleukin-17 (IL17). These cytokines stimulate Th17 cells to maintain the inflammatory response. Transforming growth factor-beta (TGF-b) stimulates RORC directly to produce cytokines and to express chemokine (C-C motif) receptor 6 (CCR6). CCR6 responds to the chemo-attractant chemokine (C-C motif) ligand 20 (CCL20), which is produced to home other Th17 cells. Interferon gamma (IFN-g) downregulates the Th17/IL23 pathway. 

Table 1. Phenotypes of the genetic risk model cohort and the IBD-PBMC cohort.

<table>
<thead>
<tr>
<th></th>
<th>Genetic risk model cohort</th>
<th>IBD-PBMC cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>General characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>423</td>
<td>39</td>
</tr>
<tr>
<td>Average age at inclusion</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>Number of males (percentage)</td>
<td>161 (38%)</td>
<td>17 (43%)</td>
</tr>
<tr>
<td></td>
<td>330 (54%)</td>
<td>18 (45%)</td>
</tr>
<tr>
<td></td>
<td>1,045</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19 (47%)</td>
</tr>
</tbody>
</table>

correlations between genetic risk load and mRNA-expression. Most patients suffered from amyotrophic lateral sclerosis (n = 733) and from chronic obstructive pulmonary disease (n = 452) and only few of UC (n = 48). Prior to analysis of this dataset, a stringent normalization step was performed by correcting gene expression for the first fifty principal components to account for any effects of batches, patient, laboratory specificity, disease specificity and so forth. These components explain most of the variation of the data. This cohort will be referred to as the
Correlation of genetic risk and mRNA expression in a Th17/IL23 pathway analysis in Inflammatory Bowel Disease

Genetic risk model
We hypothesize that the combined effect of multiple SNPs in genes of a pathway has a stronger effect on gene expression of the members of that pathways than single SNPs. Therefore we constructed a genetic risk model for the Th17/IL23 pathway. From the 163 IBD associated loci, the top SNPs of the loci containing genes involved in the Th17/IL23 pathway were selected for the genetic risk model. This resulted in inclusion of 10 independent IBD associated SNPs, residing in 10 loci, all containing one Th17 gene (IL23R, RORC, IL21, IL12B, CCR6, JAK2, IFNy, SMAD3, STAT3, TYK2, figure 1, supplementary table 2).

Genotypes of the 10 IBD associated SNPs were weighted for the magnitude of their association with IBD and included in a risk load per individual. A one-way ANOVA test was used to test the difference in genetic risk load between controls and IBD patients of the genetic risk model cohort. Subsequently a t-test was used to determine differences between CD, UC and controls separately.

Effect of disease status and genetic risk load on Th17 mRNA-expression
Nine genes from different stages in the Th17/IL23 pathway were selected for mRNA-measurements (‘qPCR-Th17 genes’: CCR6, STAT3, IL17F, IL23A, IL12RB1, RORA, IL6, RORC, IL17A (figure 1)). They serve as representatives for the different stages of the pathway. We speculated that disease status and differences in genetic risk load not only might give differential gene expression in ‘resting’ Th17 cells, but can also influence the degree of response of Th17 cells to ex vivo stimulation. Therefore, mRNA expression was determined in both unstimulated and T cell specific stimulated (PMA/CaLo) PBMCs from the IBD-PBMC cohort using qPCR measurements.

To test correlations between disease status (IBD and both subtypes separately) or genetic risk score and differential unstimulated mRNA expression, we performed a linear regression analysis with ACT values (normalized with GAPDH) as outcome variable and either genetic risk load or disease status as a covariate. To test whether disease status or genetic risk load influenced response to stimulation, a linear regression analysis was performed with the ACT values of the qPCR-Th17 genes in stimulated PBMCs as outcome variable, and baseline ACT values and respectively genetic risk load or disease status as covariates. This is a hypothesis driven analysis and p-values are therefore not corrected for multiple testing.

For replication purpose, we used the eQTL cohort to correlate genetic risk load to mRNA expression of the nine Th17 genes analyzed by qPCR.

eQTL analysis Th17/IL23 pathway
In the eQTL cohort all 22 genes involved in the Th17/IL23 pathway were assessed for SNPs close by (i.e. within 250 kb, cis-eQTL) or far away (i.e. > 5 Mb away, trans-eQTL) affecting expression these genes. If an eQTL-SNP for a Th17 gene was identified, the genetic association with IBD was also assessed.

RESULTS

Genetic risk model
The genetic risk model cohort showed a significant difference in genetic risk load of UC cases and CD cases and controls (mean: controls = 3.0; CD = 3.2; UC = 3.2, p-value one-way ANOVA = 4.8 × 10⁻⁸). Subsequently, a t-test revealed a significant difference of IBD cases compared to controls for genetic risk load (p-value = 1.4 × 10⁻⁵), but not between CD and UC cases (p-value = 0.9) suggesting that the Th17/IL23 is genetically important for both subtypes of IBD.

Effect of disease status and genetic risk load on Th17 mRNA-expression
In unstimulated PBMCs we only observed a correlation between disease status and mRNA expression levels of IL6, IL23A, STAT3 and RORC, but no effect of genetic risk load on expression was observed. All correlations between disease status or genetic risk load and mRNA expression of the nine qPCR-Th17 genes are listed in...
Chapter 5

Table 2. Effect of disease presence and genetic risk load on mRNA-expression in unstimulated and stimulated PBMCs.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Unstimulated</th>
<th>Stimulated*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disease status</td>
<td>Genetic risk load</td>
</tr>
<tr>
<td>CCR6</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL23A</td>
<td>lower (p=0.065)</td>
<td>NS</td>
</tr>
<tr>
<td>IL12RB1</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>STAT3</td>
<td>lower (p=0.086)</td>
<td>NS</td>
</tr>
<tr>
<td>IL17F</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL17A</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL6</td>
<td>lower (p=0.003)</td>
<td>NS</td>
</tr>
<tr>
<td>RORC</td>
<td>higher (p=0.085)</td>
<td>NS</td>
</tr>
<tr>
<td>RORA</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

Table 2. IL6 was significantly lower expressed in IBD cases compared to the healthy controls. A trend was observed for higher expression of RORC and lower expression of IL23A and STAT3 in cases compared to controls. When CD and UC were analyzed separately these effects were predominantly seen in CD (supplementary table 4). We also only observed effects of disease presence, and not the genetic risk load on the effect of stimulation. For IL23A the difference in expression after stimulation is significantly larger in IBD cases compared to controls, and for RORC the stimulation effect was significantly smaller in IBD cases compared to controls. A trend toward a less strong response for cases was observed for IL17F (table 2).

In the eQTL cohort, the genetic risk load of each individual was correlated to mRNA expression of all genes present on the expression array in the eQTL cohort. No significant association was observed for any of our nine qPCR-Th17 genes, confirming the lack of association in the IBD-PBMC cohort. Moreover, only one probe, coding for the RNASET2 gene, of all approximately 23,000 probes on the array was significantly correlated to genetic risk load (p-value = 1.78 × 10⁻¹⁵).

Concluding, these analyses suggest that only the presence of IBD, and not genetic risk, influences mRNA-expression of Th17 genes.

eQTL analysis Th17/IL23 pathway

In the eQTL cohort we assessed whether cis- or trans-eQTL SNPs exist for all 22 genes in the Th17/IL23 pathway (figure 2, supplementary table 5). Out of 22 genes involved in the Th17/IL23 pathway, eight genes had one or more cis-eQTL SNP(s). Of these eight genes with cis-eQTL SNPs, four genes have previously been associated with IBD. This means that seven from the eleven IBD associated Th17 genes do not have a cis-eQTL SNP. These results confirm the limited support for genetic effects in Th17 gene expression in IBD as shown by our qPCR analyses.

Gene: Th17 gene. For the unstimulated PBMCs the correlation between disease status and differential mRNA expression is reported, reported P-values from linear regression. ‘Lower’ and ‘higher’ indicate respectively lower and higher mRNA expression in cases compared to controls. For stimulated PBMCs the correlation between disease presence and differential mRNA expression is reported. The linear regression model was adjusted for baseline mRNA expression. ‘larger’ implies that cases had a larger difference in expression in response to stimulation than healthy controls. ‘smaller’ implies that cases had a smaller difference in expression in response to stimulation than healthy controls (supplementary figure 1). P-values are not corrected for multiple testing and only p-values below 0.10 are reported. No significant correlation between genetic risk load and baseline mRNA expression or response to stimulation was found. NS = not significant.
Both STAT3 and TYK2 have cis-eQTL SNPs that are associated with IBD, indicating that the genetic risk effect of these SNPs are, at least partly, driven by eQTL effects. A cis-eQTL SNP for CCR6 is also associated to IBD. However, this SNP also has a much more significant correlation with RNASET2 expression (p-value = $1.78 \times 10^{-15}$), making it questionable if CCR6 is actually the candidate gene in this locus. A cis-eQTL SNP of RORC is in high LD with a reported IBD SNP, indicating that this eQTL SNP drives the association of RORC to IBD. However, the cis-eQTL SNP is only borderline significantly associated to IBD, which contradicts this assumption. Interestingly, we found a SNP that has a strong cis-eQTL effect on IL12RB1, and is moderately associated with CD (p-value = $7.5 \times 10^{-5}$, supplementary table 5). Considering the biological function of this gene in the Th17/IL23 pathway, this might be an interesting new implication of IL12RB1 in the genetics of IBD. No trans-eQTL effects were detected.

**DISCUSSION**

Currently it is hypothesized that, in complex diseases, a substantial amount of associated genetic variants influence gene expression of nearby genes, and in this way contribute to disease pathogenesis. This study is the first to convert all currently known Th17/IL23 IBD risk variants into a risk load and correlate this to mRNA expression of genes in this pathway. For
this study we analyzed mRNA expression levels of nine representative genes in PBMCs of 80 IBD cases and 40 controls, but we did not observe any significant correlations with genetic risk load. In a much larger cohort of 1,240 individuals with various immune related diseases we performed the same analyses. There we identified a strong correlation with RNASET2 mRNA expression with IBD risk load, but not with any of the Th17/IL23 genes. Moreover we sought for individual eQTL SNPs for the 10 IBD associated Th17/IL23 genes in a larger cohort of non-IBD individuals. We observed that only two out of the 10 IBD associated Th17/IL23 genes had a convincing cis-eQTL SNP that also increases IBD risk.

We were surprised not to find any correlation between genetic risk load and mRNA expression in our initial cohort. The genetic risk model, constructed from IBD associated Th17/IL23 genetic variants, was highly significant associated to IBD, underscoring the additive effect of the variants on disease risk. Furthermore, since we used PBMCs from patients that did not use any systemic anti-inflammatory treatments, expression patterns were not influenced by this. Additionally, for these genes no eQTL has been described in other publicly available eQTL datasets. One could speculate that our findings result because of lack of power in the initial IBD-PBMC cohort. Although we had sufficient power to find that mRNA expression levels were influenced by IBD status, the genetic effects may be smaller and thus more difficult to pick up. More convincingly, we also did not observe significant correlations between genetic risk load or single SNPs and mRNA expression in the much larger eQTL cohort of 1,240 individuals. This cohort already proved to be sufficiently large to detect single SNP effects as shown by Fehrmann et al. 

Another explanation for lack of correlation resides in the tissue in which mRNA expression was measured in both the IBD-PBMC and eQTL cohort. The Th17/IL23 pathway is highly specific to the Th17 cells, which have a very low frequency in peripheral blood. Therefore it is possible that the Th17/IL23 mRNA expression effects are diluted to undetectable levels in the heterogeneous cell populations of PBMCs and whole blood. To test this, we analyzed the probe-intensity values of the genes with and without eQTL SNPs in the microarray data of the eQTL cohort. It appeared that indeed the genes without eQTL SNPs have significantly lower expression compared to genes for which an eQTL SNPs could be detected (data not shown). It is also possible that peripheral Th17 cells are phenotypically and expression-wise not the same as mucosal Th17 cells, and that we have investigated the wrong cells.

With this finding in mind, we have to be careful when interpreting eQTL-data. Fu et al. found that eQTL effects can be highly tissue specific, which might result in omission of effects when expression is measured in the wrong tissue. eQTL data is usually based on whole blood mRNA expression and the pitfall is that this might result in omission of tissue specific effects. In other words, if mRNA expression is measured in the wrong tissue type, the correlation between the genetic risk variant and the mRNA expression of the true culprit gene cannot be picked up, and correlations with mRNA expression of other genes might falsely be interpreted as being relevant for disease pathogenesis. This might also be the case for the locus containing CCR6 and RNASET2. When analyzing eQTL data in whole blood this SNP has a very strong cis-eQTL effect on RNASET2, indicating that this might be the culprit gene. However, CCR6 is prioritized based on its role in the Th17/IL23 pathway in literature. Hence, prioritizing genes in risk loci based on eQTL data should always be combined with other tools like DAPPLE or GRAIL.

Our risk model includes all SNPs for which Th17/IL23 pathway genes have been prioritized. This inclusion strategy might have resulted in inclusion of SNPs in which the culprit gene is not involved in the Th17/IL23 pathway; and vice versa we might have omitted SNPs and culprit genes involved in the Th17/IL23 pathway with up until now unknown function. The answer to this question will only be definitely answered if causal variants and genes are discovered in all IBD associated loci.

Although at first perhaps surprising, it is very well possible that genetic risk variants do not
Correlation of genetic risk and mRNA expression in a Th17/IL23 pathway analysis in Inflammatory Bowel Disease

Contribute to disease pathogenesis through altered gene expression. What other mechanisms can account for this? One of the possibilities is an altered protein structure. This can be caused by rare, yet unidentified variants in exomes. Rivas et al. identified two such variants in the IL23R gene by targeted resequencing of IBD loci. Though the contribution of such rare variants to the missing heritability is probably relatively small. Another possibility is abrogated phosphorylation of genes, as recently shown for compound heterozygous missense mutations in the IL1 locus which caused IL-10 induced abrogated IL-10R1 phosphorylation. Furthermore, the ENCODE project has shown that regulatory elements of genes can be located at great distance of target genes. This implies that the culprit genes can be at a great distance from the associated loci, in which case we should look for trans-eQTL effects. In the current study we could not identify such trans-eQTL effects; this is likely due to lack of power.

In conclusion, our findings show limited influence of genetic risk variants on gene expression of genes involved in the Th17/IL23 pathway in PBMCs. Moreover prioritizing genes in associated loci based on their eQTL effects in whole blood should be interpreted with caution.

Supplementary Data

Supplementary data are available online:
Supplementary methods: http://links.lww.com/IBD/A427
Supplementary table 1: http://links.lww.com/IBD/A428
Supplementary table 2: http://links.lww.com/IBD/A430
Supplementary table 3: http://links.lww.com/IBD/A429
Supplementary table 4: http://links.lww.com/IBD/A431
Supplementary table 5: http://links.lww.com/IBD/A433
Supplementary figure 1: http://links.lww.com/IBD/A432

References


Part 2

Genetic risk factors for Inflammatory Bowel Disease in other diseases