A Possible Mechanism behind Autoimmune Disorders Discovered By Genome-Wide Linkage and Association Analysis in Celiac Disease

Malin Östensson1, Caroline Montén2, Jonas Bacelis2, Audur H. Gudjonsdottir4, Svetlana Adamovic3, Johan Ek5, Henry Ascher6, Elisabet Pollak3, Henrik Arnell7, Lars Browaldh6, Daniel Agardh2, Jan Wahlström3, Staffan Nilsson1, Åsa Torinsson-Naluai9,9*

1 Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden, 2 Diabetes and Celiac Disease Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden, 3 Institute of Biomedicine, Department of Medical and Clinical Genetics, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden, 4 Queen Silvia Children’s Hospital, Sahlgrenska Academy at the University of Gothenburg, Department of Pediatrics, Gothenburg, Sweden, 5 Buskerud Central Hospital, Department of Pediatrics, Drammen, Norway, 6 Sahlgrenska Academy at the University of Gothenburg, Department of Public Health and Community Medicine, Unit of Social Medicine, Gothenburg, Sweden, 7 Department of Pediatric Gastroenterology, Hepatology and Nutrition, Karolinska University Hospital and Division of Pediatrics, CLINTEC, Karolinska Institutet, Stockholm, Sweden, 8 Department of Clinical Science and Education, Karolinska Institutet Sodersjukhuset, Stockholm, Sweden, 9 Systems Biology Research Centre, Tumor Biology, School of Life Sciences University of Skövde, Skövde, Sweden

Abstract

Celiac disease is a common autoimmune disorder characterized by an intestinal inflammation triggered by gluten, a storage protein found in wheat, rye and barley. Similar to other autoimmune diseases such as type 1 diabetes, psoriasis and rheumatoid arthritis, celiac disease is the result of an immune response to self-antigens leading to tissue destruction and production of autoantibodies. Common diseases like celiac disease have a complex pattern of inheritance with inputs from both environmental as well as additive and non-additive genetic factors. In the past few years, Genome Wide Association Studies (GWAS) have been successful in finding genetic risk variants behind many common diseases and traits. To complement and add to the previous findings, we performed a GWAS including 206 trios from 97 nuclear Swedish and Norwegian families affected with celiac disease. By stratifying for HLA-DQ, we identified a new genome-wide significant risk locus covering the DUSP10 gene. To further investigate the associations from the GWAS we performed pathway analyses and two-locus interaction analyses. These analyses showed an over-representation of genes involved in type 2 diabetes and identified a set of candidate mechanisms and genes of which some were selected for mRNA expression analysis using small intestinal biopsies from 98 patients. Several genes were expressed differently in the small intestinal mucosa from patients with celiac autoimmunity compared to intestinal mucosa from control patients. From top-scoring regions we identified susceptibility genes in several categories: 1) polarity and epithelial cell functionality; 2) intestinal smooth muscle; 3) growth and energy homeostasis, including proline and glutamine metabolism; and finally 4) innate and adaptive immune system. These genes and pathways, including specific functions of DUSP10, together reveal a new potential biological mechanism that could influence the genesis of celiac disease, and possibly also other chronic disorders with an inflammatory component.

Citation: Östensson M, Montén C, Bacelis J, Gudjonsdottir AH, Adamovic S, et al. (2013) A Possible Mechanism behind Autoimmune Disorders Discovered By Genome-Wide Linkage and Association Analysis in Celiac Disease. PLoS ONE 8(8): e70174. doi:10.1371/journal.pone.0070174

Editor: Anna Carla Goldberg, Albert Einstein Institute for Research and Education, Brazil

Received March 18, 2013; Accepted June 14, 2013; Published August 2, 2013

Copyright: © 2013 Östensson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The principal funding for this study was provided through the regional agreement on medical training and clinical research (ALF) between Gothenburg County Council and Gothenburg University. We also thank Bengt Hre’s, Claes Groshinsky’s, Magnus Bergvall’s, Nilsson-Ehle’s, Tore Nilsson’s and Professor Nanna Svartz Foundations, Kungliga fysiografiska sällskapet in Lund, Frimurare barnhusdirektionen, Ruth and Richard Julin Foundation and the Swedish Society of Medicine. The Swedish Medical Research Council and the Celiac Disease Foundation supported family sample collections. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: asa.torinsson.naluai@gu.se

Introduction

Celiac disease (CD) is a common chronic disease and even though most often diagnosed in early childhood, it can present itself at any age. Most of the individuals with CD remain undiagnosed and an estimated 2% of the Swedish population is affected without having been diagnosed [1]. Ongoing disease will increase the overall risk for developing other chronic inflammatory diseases, neurological manifestations and malnutrition disorders. CD is the only autoimmune disorder where the actual genes responsible for the association in HLA are known (HLA-DQA1 and HLA-DQB1) [2]. In the past few years Genome Wide Association Studies (GWAS) have had tremendous success in identifying new genes, or gene regions, that influence common diseases. These studies use several hundreds of thousands of genetic markers (single nucleotide polymorphisms, SNPs) across all human chromosomes in order to pin down the chromosomal locations of genes, which could influence the disease.

A large joint effort has been done, not the least in CD, and 40 new CD-associated genetic regions marked by SNPs have been
Genotyping and Imputation

We included single nucleotide polymorphism (SNP) markers that had a call rate above 97%, which led to the exclusion of 1.3% of the Omni Express and 0.6% of the 660W-Quad SNP markers. Out of the 127,535,126 imputed genotypes, 88.3% had a posterior probability of over 0.95. Approximately 90% of the 944,512 SNP markers had a minor allele frequency of at least 0.01 after imputation.

Interaction Analyses

Since some markers just below genome-wide significance are still expected to be true findings, we wanted to try and separate these from the, in fact, true negative findings (those that show linkage and association close to genome-wide significance just by chance). In total, 603 SNP markers from 383 independent regions and their surrounding genes were identified by three inclusion criteria (Fig. 2 and Table S1). These genes were subsequently used for pathway and two-locus interaction analyses.

Two-locus interaction analysis. Two-locus interaction analysis, identified 582 SNP pairs with a p-value of less than 1.0×10^{-4} for the test comparing the model M_{6} of no association and the general two-locus model M_{6}. Out of these, 101 pairs from 87 regions deviated significantly (p<0.05) from a purely multiplicative model (M_{6}), which is the best fitting model when at least one of the SNP markers is false. Under the null hypothesis we expect to find 29 such pairs. The 101 pairs showed either epistasis (individuals carry both risk alleles) or evidence of heterogeneity (individuals carry either the one or the other risk allele from the two loci).

The results with a p-value <1.0×10^{-4} for epistasis and those with high p-value (>0.05), which represent pairs that did not show convincing deviation from the heterogeneity model are listed in Table 3 and 4. Several loci were in an epistatic relationship with HLA: rs4899272 (ACTN1), rs1073933 (COX7C), rs10482751 (TGFB2), rs571879 (APP), and rs7590305 (FABP1). Also, previously identified susceptibility loci for CD were involved in the area of epistasis.
| Chr | SNP | Genes | BP   | A1 | A2 | T/U | p-value | expTDT | T/DI (PLINK) exp TDT | TDI/GWAS catalog |
|-----|-----|-------|------|----|----|-----|---------|-------|---------------------|------------------|
| 1   | rs12743144 | PPP1R12B, SYT2, UBE2T | 127384568 | C   | T   | 90  | 61.44   | 0.31  | 4.34E-07            | 0.0095           |
| 1   | rs10886159 | EMA235, BAK1, TMEM137, BMI2 | 10886159 | C   | T   | 50  | 48.54   | 0.30  | 7.36E-07            | 0.10*            |
| 1   | rs160888894 | EAPF, SNK6, C4orf174 17 | 160888894 | C   | T   | 23  | 48.54   | 0.28  | 2.31E-07            | 0.22*            |
| 1    | rs113801444 | ST2, 2, ST2, P142  | 113801444 | C   | T   | 36  | 32.86   | 0.27  | 3.87E-06            | 0.22*            |
| 1    | rs1032355 | RG9MTD2, C4orf17, MTTP | 1032355 | G   | T   | 25  | 138.45  | 0.46  | 3.41E-07            | 0.0069           |
| 19   | rs4911642 | CCL8L2, PSI, TPTE2 22 | 4911642 | C   | T   | 38  | 85.28   | 0.41  | 7.30E-07            | 0.19*            |
| 20   | rs157640 | DOK5, 52847946 | 157640 | G   | T   | 73  | 138.45  | 0.46  | 3.14E-07            | 0.63*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 16  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
| 20   | rs12668824 | NAV1, 199861288 | 12668824 | C   | T   | 20  | 76.78   | 0.37  | 4.13E-06            | 0.26*            |
several interactions: rs4899272 (ACTN1), rs6741418 (STAT1, GLS), rs13096142 (CCR1,2,3,5), rs10197319 (ICOS, CTLA4) and rs870875 (CD247).

Pathway analysis. Biological functions clustered by Ingenuity Pathway Analysis (IPA) and Genetrail [13] are shown in Table 5, 6 and 7. Several clusters were significant after correction for multiple comparisons. The most significant network implicated by IPA included DUSP10 (Fig. 3 and Table 8). The second top network included the MHC complex (HLA) and the third top network included LPP, which is located within the most significantly non-HLA associated region identified in CD so far [3].

Gene Expression
Out of the 34 selected target genes, three were from the top associated SNPs (DUSP10, SVIL and PPP1R12B) and the remaining were genes identified from the two-locus and pathway analysis. Eight genes showed significant up- or down-regulation after correction for multiple testing using Bonferroni correction (Fig. 4). For the top associated genes, several transcript variants were tested (Table 9). For the PPP1R12B gene, Isoform c and d (transcript variants NM032103.2 and NM032104.2) also known as the small subunit (sm-M20) of myosin light chain phosphatase, show significant up-regulation in patients with CD autoimmunity compared to control patients. An additional ten genes showed nominally significant differences in expression (Expression Table 9).

Non-parametric Linkage (NPL)
The strongest linkage outside of HLA was detected in chromosome regions 5q23.2-q33.1, and 1q32.1. In total, thirteen regions with an NPL point wise p-value below 0.01 were detected (Fig. 5 and Table 10). In our previous linkage-scan, using almost the same set of families, we detected only one region (11q23-25) with a point wise p-value below 0.01 [14]. The reason for the improved results is mainly the almost perfect information content achieved by a dense set of highly successful SNP markers compared to a relatively sparse set of less successful microsatellite markers. Also in the NPL analysis, the PPP1R12B gene was located in one of the top regions (1q32.1).

Discussion
This study confirmed some previous GWAS findings and in addition, it established a new genome-wide significant region containing the DUSP10 gene. The top markers, rs12144971 and rs4240931 showed a substantial effect size in the HLA low-risk group with a transmitted versus non-transmitted allele ratio of 3.11 (Table 2).

DUSP10, TNF-α and Tissue Transglutaminase (TGM2)
The protein product of DUSP10 preferentially binds to the stress-activated p38 MAPK (mitogen-activated protein kinase) and plays an important role in regulating chemokine induction after infection by various pathogens [15], and in coordinating MAPK activity in response to oxidative stress [16]. In previous studies, both p38 MAPK and DUSP10 have been shown to activate TNF-α [17,18], of which one also demonstrates that TNF-α up-regulates TGM2 (the gene encoding the main autoantigen in CD [19]) in intestinal mucosa from untreated CD patients [17]. Whether this up-regulation of TGM2 is of importance for the immune response leading to formation of IgA-αTG and IgG-αTG autoantibodies, the serological markers for CD is still unresolved.
| Chr | SNP | Gene(s) | BP   | A1 | T   | U   | T/U chisq | p-value |
|-----|-----|---------|------|----|-----|-----|------------|---------|
| 1   | rs12144971 | DUSP10 | 220099108 | C   | T   | 26  | 35       | 0.74    |
| 1   | rs4240911 | DUSP10 | 220105678 | T   | C   | 26  | 31       | 1.03    |
| 10  | rs1247697 | SVIL   | 29901347  | C   | A   | 41  | 35       | 1.17    |
| 10  | rs12734338 | DUSP10 | 220139621 | G   | A   | 20  | 25       | 0.80    |
| 10  | rs11811613 | DUSP10 | 220122026 | G   | A   | 19  | 25       | 0.76    |
| 2   | rs13017044 | PRKCE  | 46086853  | A   | G   | 12  | 39       | 0.31    |
| 10  | rs11102146 | KCNA3  | 111007559 | C   | T   | 13  | 17       | 0.76    |
| 3   | rs3629249 | STAC   | 36329541  | A   | C   | 15  | 25       | 0.60    |
| 3   | rs1871350 | STAC   | 36348239  | C   | T   | 15  | 26       | 0.59    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 50  | 35       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 43  | 33       | 1.30    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 43  | 33       | 1.30    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
| 10  | rs10861397 | SORCS1 | 108678768 | G   | A   | 46  | 34       | 1.35    |
Pathway Analyses

In order to discover possible functional connections between DUSP10 and other genes, we analyzed genes surrounding the top 603 markers. A total of 845 genes were used in the analysis. Ingenuity pathway analysis (IPA) included DUSP10 within the most significant network. Also part of this network were GLS and RGS1, two genes previously identified within significant GWAS loci [3], as well as the insulin (INS) gene, and the immune regulatory nuclear factor kappa B (NF-kB) complex (Fig. 3 and Table 8). The second top network included the MHC complex (HLA) and also several genes within already identified GWAS loci: ACTN1, CD247, CCR5, ICOS and STAT1 [3]. In addition, both IPA and GeneTrail [13] identified T2D genes as the most significantly overrepresented gene cluster after correction for multiple testing (Table 5 and 6). Among this set of genes surrounding the 603 markers, many genes belonged to growth and nutrient signaling pathways, for example, INS, INSR, EGF, POMC, TIPRL and PRR5L. There were also related genes directly involved in energy metabolism; PDK1, COX7C, COQ3 and GLS.

Overlapping Results with Other GWAS Findings

Surprisingly, four out of six top loci identified by a GWAS for anorexia nervosa [20] and two out of three loci involved in plasma glucose levels in type 1 diabetic patients [21] were among our 603 and 35 best SNP markers respectively. One of the genes in anorexia, namely AKAP6, is also associated to fasting insulin-related traits as well as the autoimmune disease Ankylosing spondylitis [22]. Of the 40 identified regions in CD, seven regions overlap with our 603 SNP list (LPP, STAT4/GLS, RGS1, CCR1/CCR3, PUS10, ICOS/CTLA4 and CD247). Out of the 69 regions reported in the GWAS catalog for type 1 diabetes, eight overlap with the regions reported in this study and out of those eight, CTLA4/ICOS also overlap with the previously reported CD associations. We compared minor allele frequencies between the previous CD GWAS by Dubois et al. and our GWAS. In their top 42 associations, there was no SNP below a minor allele frequency of 0.08. In our top 42 associations, we identified five SNPs with a minor allele frequency below 0.06. This observation could just be a chance finding or perhaps an indication that rare variants are easier to discover using families. We also identified a relatively rare variant in the LPP gene region (rs17283813), with a minor allele frequency of 0.075. This SNP was not at all significant in the GWAS by Dubois et al. (Table S1).

Neither was there an association with the DUSP10 region in the GWAS by Dubois and co-workers. The associated markers in the DUSP10 region in our GWAS have a minor allele frequency around 0.5 and are hence very common in the population. It is difficult to say if this is a population specific effect or if DUSP10 could be detected in an HLA stratified population from another ethnicity. Interestingly, the DUSP10 region has also been identified as a risk factor for colon cancer by a meta-analysis of three GWAS from the UK. This is an indication that colon cancer and CD could share genetic risk factors.

Key Metabolic Regulators as well as the Top Associated gene PPP1R12B were Differently Expressed in CD Cases Compared to Controls

Another important finding was the difference between cases and controls and their gene expression patterns in the small intestine. Eight of the 34 candidate genes selected for quantitative measurements of gene expression, including PPP1R12B, PDK1, GLS, PRR5L and the INSR, showed significant up or down regulation of mRNA levels in cases compared to controls (Fig. 4).
Figure 2. Illustration of the three inclusion criteria used for pathway and interaction analyses. The first criteria of p-values less than $3.0 \times 10^{-2}$ in the linkage TDT analysis resulted in a total of 477 markers. The second criteria included a comparison of the results from this study with the results from the study by Dubois et al. [3]. We included 118 SNPs that had a simple score based on a combined p-value less than $5.0 \times 10^{-2}$ and in the same allelic direction in both datasets. The third criteria involved selecting markers with a large effect size. We included 65 markers which had a ratio of transmitted versus not transmitted (T/NT) alleles of over 5 or below 0.2, combined with a p-value of less than $2.0 \times 10^{-2}$.

Table 3. The top epistasis interaction results from the 101 two-locus interaction analysis.

| Snp 1   | Genes     | chr | Snp 2   | Genes     | chr | N  | P02 | P12 | Pm2 |
|---------|-----------|-----|---------|-----------|-----|----|-----|-----|-----|
| rs2187668 | HLA       | 6   | rs4899272 | ACTN1     | 14  | 95 | 4.0E-17 | 1.42E-13 | 4.0E-02 |
|         |           |     | rs204034  | SHISA9    | 16  | 94 | 1.3E-14 | 1.09E-12 | 5.0E-02 |
|         |           |     | rs571879  | APPL1     | 3   | 94 | 2.3E-15 | 5.21E-11 | 3.0E-02 |
| rs204999 | HLA       | 6   | rs1073933 | COX7C     | 5   | 94 | 9.9E-14 | 9.27E-12 | 3.0E-02 |
|         |           |     | rs11836636 | ATXNL3B   | 12  | 91 | 1.7E-12 | 8.15E-11 | 4.0E-02 |
| rs7745052 | FBXL4, C6orf168, USP45, COQ3, POU3F2, SFRS18 | 6   |         |           |     |    | 92 | 2.3E-05 | 1.79E-05 | 4.0E-02 |
| rs10749738 | FOXD3     | 1   | rs1373649 | BMPR1B    | 4   | 93 | 2.7E-05 | 1.78E-05 | 4.0E-02 |
| rs3860295 | RASSF5, IKBKE | 1   | rs13096142 | CCR5, CCR3, LTF, CCR2, CCR1 | 3   | 95 | 1.1E-05 | 6.48E-06 | 1.0E-02 |
| rs9396802  | KIF13A, NUP153, FAM8A1 | 6   | rs2194633 | NETO1     | 18  | 95 | 3.8E-06 | 6.82E-06 | 2.0E-02 |
| rs9296204  | MTCH1, P16 | 6   | rs4385459 | LY96, JPH1, GDAP1, TMEM70, TCEB1 | 8   | 95 | 2.8E-05 | 9.91E-06 | 3.0E-02 |
| rs9397928  | ARID1B*   | 6   | rs2415836 | FSCB*     | 14  | 93 | 2.8E-05 | 1.75E-05 | 3.0E-02 |
| rs1145212  | APOA5, ZNF259, BUD13 | 11  | rs10083673 | MYO5A     | 15  | 95 | 6.6E-05 | 1.77E-05 | 2.0E-02 |
| rs7756191  | DNAH8     | 6   | rs1108001 | NAV2, HATATIP2, DBX1, PRMT3 | 11  | 95 | 3.5E-05 | 2.60E-05 | 3.0E-03 |
| rs10197319 | ICO3, CTLA4 | 2   | rs882820  | SRL, TIPAP4 | 16  | 94 | 1.4E-05 | 3.03E-05 | 3.0E-05 |
| rs4899272  | ACTN1     | 14  | rs17703807 | C15orf41  | 15  | 83 | 2.9E-05 | 8.68E-05 | 1.0E-02 |

All SNP pairs which reached an interaction p-value of $P_{12} < 1.0 \times 10^{-2}$, in addition to $P_{m2} < 0.05$.

*closest known gene, located >500 kb from associated SNP.

$P_{02}$ – p-value for the test statistic comparing the models $M_0$ (no association) and the general model $M_G$.

$P_{12}$ – p-value for the test test comparing the models $M_R$ (heterogeneity) and the general model $M_G$.

$P_{m2}$ – p-value for the test comparing the models $M_M$ (multiplicative) and the general model $M_G$. 

doi:10.1371/journal.pone.0070174.g002
This could very well be a consequence of an ongoing inflammation or possibly also indicate an underlying metabolic difference. Glutamine is converted to glutamate by the enzyme glutaminase (GLS). In turn, glutamate can be converted to proline and subsequently catabolized by the enzyme proline dehydrogenase (GLS). In turn, glutamate can be converted to proline and is a critical regulator of insulin signaling pathways [28]. We could detect expression of both APPL1 and APPL2 in small intestinal biopsies and a significantly lower expression of APPL2 was detected in the CD autoimmunity cases as compared to controls (Fig. 4). Lower expression of APPL2 levels lead to enhanced adiponectin stimulated glucose uptake and fatty acid oxidation [29]. A SNP (rs10861406) included in the top 603 list was located upstream of APPL1, however the promotor of this gene was on the opposite side of a recombination hotspot and therefore not included in the gene list for pathway analyses. The most significant finding from our non-stratified linkage GWAS analysis was the association with the PPP1R12B gene region. *PPP1R12B* is involved in smooth muscle contractility and mediates binding to myosin [30]. Myosin light chain phosphatase from smooth muscle consists of a catalytic subunit (PP1c) and two non-catalytic subunits, M130 and M20. The two non-catalytic subunits are both encoded by the *PPP1R12B* gene. The M130 non-catalytic subunit (M130) mediates binding to myosin [30]. Myosin light chain phosphatase from smooth muscle consists of a catalytic subunit (PP1c) and two non-catalytic subunits, M130 and M20. The two non-catalytic subunits are both encoded by the *PPP1R12B* gene. The M130 non-catalytic subunit (M130) mediates binding to myosin [30]. Myosin light chain phosphatase from smooth muscle consists of a catalytic subunit (PP1c) and two non-catalytic subunits, M130 and M20. The two non-catalytic subunits are both encoded by the *PPP1R12B* gene. The M130 non-catalytic subunit (M130) mediates binding to myosin [30]. Myosin light chain phosphatase from smooth muscle consists of a catalytic subunit (PP1c) and two non-catalytic subunits, M130 and M20. The two non-catalytic subunits are both encoded by the *PPP1R12B* gene. The M130 non-catalytic subunit (M130) mediates binding to myosin [30]. Myosin light chain phosphatase from smooth muscle consists of a catalytic subunit (PP1c) and two non-catalytic subunits, M130 and M20. The two non-catalytic subunits are both encoded by the *PPP1R12B* gene. The M130 non-catalytic subunit (M130) mediates binding to myosin [30]. Myosin light chain phosphatase from smooth muscle consists of a catalytic subunit (PP1c) and two non-catalytic subunits, M130 and M20. The two non-catalytic subunits are both encoded by the *PPP1R12B* gene. The M130

### Table 4. The top heterogeneity results from the 101 two-locus interaction analysis.

| SNP1     | Genes          | chr | SNP2     | Genes          | chr | N  | P 02 | P 12 | PM2 |
|----------|----------------|-----|----------|----------------|-----|----|------|------|-----|
| rs4899272| ACTN1          | 14  | rs4820682| SRHD PS4 TFIP11| 22  | 95 | 7.1E-06 | 6.97E-02 | 2.0E-02 |
| rs4426448| DOK6           |     | rs70875  | CD247          | 1   | 94 | 9.4E-05 | 7.19E-02 | 3.0E-02 |
| rs482007 | PAEP           |     | rs87085  | CDRA2         | 8   | 94 | 4.1E-05 | 5.81E-01 | 5.0E-02 |
| rs571879 | APPL1, HES1X, IL17RD | 3    | rs4385459| LY96 JPH1 GDAP1 TMEM70 TCEB1 | 8   | 94 | 4.1E-05 | 5.81E-01 | 5.0E-02 |
| rs5790305| FABP1, THNSL2  | 2   | rs390495 | MICAL3         | 2   | 93 | 7.0E-05 | 9.09E-01 | 3.0E-03 |
| rs7745052| FBLX4, C6orf168, USP45, COQ3, POUSF2, 5F8S18 | 6    | rs4930144| IGF2AS TH MRPL23 TNNT3 SYT8 ASCL2 TNNI2 LSP1 IGF2 INS-IGF2 INS H19 | 11  | 50 | 1.9E-05 | 5.30E-01 | 3.0E-02 |
| rs10749738| FOXD3          | 1   | rs10498982| EPHA7*         | 6   | 93 | 2.0E-05 | 1.95E-01 | 4.0E-02 |
| rs2605393| STAC           |     | rs2605393| MICAL3         | 3   | 63 | 7.3E-05 | 4.37E-01 | 4.0E-02 |
| rs2187668| HLAQD          | 6   | rs11013804| KIAA1217      | 10  | 94 | 3.5E-14 | 8.40E-02 | 2.0E-02 |
| rs1676235| ESRRB ANGEL1, VASH1 |     | rs1676235| KIAA1217      | 14  | 43 | 2.0E-07 | 8.55E-02 | 3.0E-02 |
| rs958802 | KANK4 L1TD1, INADL | 1    | rs2194633| NETO1         | 18  | 95 | 1.9E-05 | 5.55E-01 | 3.0E-02 |
| rs2345981| KHRDS2         | 6   | rs6495130| RYR3          | 15  | 94 | 6.1E-05 | 1.58E-01 | 3.0E-02 |
| rs11940562| PCDH7*        | 4   | rs4905043| ITPK1 CHGA    | 14  | 44 | 4.6E-05 | 2.77E-01 | 2.0E-02 |
| rs4656538| POUSF1         | 1   | rs2187668| HLAQD         | 6   | 94 | 3.0E-13 | 1.19E-01 | 5.0E-02 |
| rs3860295| RASSF5 IKBE    | 1   | rs7046385| SM2            | 9   | 94 | 5.3E-05 | 1.07E-01 | 2.0E-02 |
| rs6741418| STAT1 GL5, STAT4 | 2    | rs10798004| C1orf25 C1orf26 | 8   | 97 | 7.2E-05 | 7.68E-02 | 4.0E-02 |
| rs1571812| VDLR           | 9   | rs1571812| VDLR          | 9   | 86 | 3.0E-05 | 9.19E-02 | 4.0E-02 |
| rs882820 | SRL TFAP4      | 16  | rs882820 | SRL TFAP4      | 16  | 87 | 4.2E-05 | 3.52E-01 | 6.0E-03 |
| rs1470379| VIM            | 10  | rs1470379| VIM           | 10  | 82 | 1.0E-05 | 3.70E-01 | 8.0E-03 |
| rs10946659| DCCD2 NRSN1 | 16  | rs10946659| DCCD2 NRSN1 | 9   | 78 | 1.9E-06 | 6.64E-01 | 9.0E-03 |
| rs10482751| TGFβ2         | 9   | rs10482751| TGFβ2 | 9   | 92 | 5.2E-05 | 1.86E-01 | 1.0E-02 |

All SNP pairs which reached an interaction p-value of P12 > 0.05, in addition to P02 < 0.05.

*closest gene located > 500 kb from associated SNP.
P02 – p-value for the test statistic comparing the models M0 (no association) and the general model MG.
P12 – p-value for the test statistic comparing the models MR (heterogeneity) and the general model MG.
PM2 – p-value for the test comparing the models M0 (heterogeneity) and the general model MG.

doi:10.1371/journal.pone.0070174.t004

*closest known gene located > 500 kb from associated SNP.*
The second most significant region in the HLA-stratified analysis after DUSP10 contains the SVIL gene. The product of this gene has been suggested to bind LPP [33]. In our two-locus interaction analysis, the LPP locus and a locus containing KIF13A was one of the 101 interaction pairs. KIF13A is a motor protein, which shuttles vesicles containing AP-1 and the mannnose-6-phosphate receptor [34]. KIF13A was significantly down-regulated in intestinal biopsies from CD patients in our gene expression analysis (Fig. 4). SVIL is associated with cell-focal adhesions (substrate contacts), which are important for rapidly moving cells such as for example immune cells but also for motility and polarity of intestinal epithelial cells. SVIL mRNA was down-regulated in our gene expression analysis, however, not significant after correction for multiple testing.

Proline and Glutamine Metabolism - Part of a “Danger Signal”

Amoebiasis was one of the nominally significant pathways in the GeneTrail analysis of genes surrounding the two-locus interaction SNPs (Table 7). Several of these genes were also present together with DUSP10 and the MHC class II genes in the two most significant IPA generated networks (marked in bold text in

Figure 3. Ingenuity network 1. The top network identified by the Ingenuity IPA software using genes surrounding all 603 most associated SNPs from the TDT analysis. Molecules in gray were present among the genes from our TDT analysis and molecules in white were added by the IPA software. The DUSP10 gene is marked in yellow.
doi:10.1371/journal.pone.0070174.g003
Table 5. Biological functions of genes surrounding the 603 top associated SNPs. Results from IPA.

| Function Annotation | p-value (Raw) | B-H p-value* | Molecules |
|---------------------|--------------|--------------|-----------|
| non-insulin-dependent diabetes mellitus | 0.0000057 | 0.025 | ABCBC, ADRA1B, ADRA1D, AGT, APOA5, ATP10A, B2CL1.1, CCR5, CD38, CDNAP2, FOXP1, FTO, HFE, HFE2, INS, KCNJ11, KIRREL3, KLF10, mir-154, mir-448, MTPP, PBX3, PIEZO2, PPARA, PPS3CA, PRDM10, RG55, VEGFA, ZMYM2 |
| quantity of metal | 0.0000082 | 0.025 | ABCBC, ADRA1B, AGP, PLP2, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CHGA, CX3CR1, CX5L3, DARC, DCC, DNL, EGF (includes EG:13645), FBXL5, FER1A, GNA14, GNB1, HFE, HFE2, IFG2, INS, INS, KCNJ11, LTF, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| incorporation of thymidine | 0.000010 | 0.025 | AGT, AKAP13, B2MP, CD40, EGF (includes EG:13645), IFG2, INS, INS, PRL, THBS2, TNSF13B, VEGFA, WT1 |
| quantity of Ca2+ | 0.000018 | 0.033 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CHGA, CX3CR1, CX5L3, DARC, DCC, DNL, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| eye development | 0.000022 | 0.033 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CHGA, CX3CR1, CX5L3, DARC, DCC, DNL, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| diabetes mellitus | 0.000027 | 0.034 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CHGA, CX3CR1, CX5L3, DARC, DCC, DNL, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| angiogenesis of bone | 0.000032 | 0.034 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CHGA, CX3CR1, CX5L3, DARC, DCC, DNL, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| quantity of bone | 0.000071 | 0.043 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CHGA, CX3CR1, CX5L3, DARC, DCC, DNL, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| development of head | 0.000069 | 0.043 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CHGA, CX3CR1, CX5L3, DARC, DCC, DNL, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| migration of cells | 0.000057 | 0.043 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CD40, CD99, CHGA, CMA1, CNTNAP2, CSF2RA, CTBP2, CTNNA2, CTSG, CX3CR1, CX5L3, DARC, DCC, DCL, DLX1, DNHLM3, DPH2, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| cell movement | 0.000073 | 0.043 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CD40, CD99, CHGA, CMA1, CNTNAP2, CSF2RA, CTBP2, CTNNA2, CTSG, CX3CR1, CX5L3, DARC, DCC, DCL, DLX1, DNHLM3, DPH2, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| apoptosis | 0.000069 | 0.043 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CD40, CD99, CHGA, CMA1, CNTNAP2, CSF2RA, CTBP2, CTNNA2, CTSG, CX3CR1, CX5L3, DARC, DCC, DCL, DLX1, DNHLM3, DPH2, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |
| quantity of leukocytes | 0.000076 | 0.043 | ABCBC, ADRA1B, AGP, ATP2B3, B2CL, B2MP, BTK, CMLG, CCR5, CD247, CD38, CD40, CD99, CHGA, CMA1, CNTNAP2, CSF2RA, CTBP2, CTNNA2, CTSG, CX3CR1, CX5L3, DARC, DCC, DCL, DLX1, DNHLM3, DPH2, EGF (includes EG:13645), FER1A, GNA14, GNB1, IFG2, INS, INS, KCNJ11, NTS, NUCB2, POMC, PRL, PRNP, PTGDR2, RGS1, RYR3, SELL, SOD1, TRPM8, TNXIP, VAV3, VEGFA |

A total of 823 genes surrounding the 603 top associated SNPs were put into the IPA software.

Surrounding genes were defined by either Grail (www.broadinstitute.org/mpg/grail/) or the Genome Browser (http://genome.ucsc.edu/). Gene families located in the same region were manually curated so that only one gene in each family remained in each region, based on a similar official gene symbol.

*Hochberg Y, Benjamini Y. Statistics in medicine 1990; 9:811–8.

doi:10.1371/journal.pone.0070174.0005

PLOS ONE | www.plosone.org 11 August 2013 | Volume 8 | Issue 8 | e70174
Table 6. Biological functions of genes surrounding the 603 top associated SNPs. Results from GeneTrail.

| Category         | rank | Subcategory                              | expected | observed | p-value (raw) | Genes                                                                 |
|------------------|------|------------------------------------------|----------|----------|---------------|----------------------------------------------------------------------|
| KEGG             | 1    | Type II diabetes mellitus                | 1.91     | 7        | 0.0026        | ABC2C8, CACNA1A, INS, INSR, KCNJ11, MAPK1, PRKCZ                       |
| KEGG             | 2    | Salivary secretion                       | 3.62     | 9        | 0.003         | ADRA1B, ADRA1D, AMY1B, ATP2B3, BST1, CALML6, CD38, CST2, RYR3         |
| KEGG             | 3    | Pathways in cancer                       | 13.35    | 23       | 0.007         | APP1L1, BCL2, BID, BMP2, CBL, CSF2RA, CTBP2, CTTNA2, DVL1, E2F3, EGF, FGF2, FH, ITGAV, LAM2, MAPK1, MIF, PTK2, RASSF5, STAT1, TECB1, TGF2, VEGFA |
| KEGG             | 4    | T cell receptor signaling pathway        | 4.40     | 10       | 0.012         | CA2B311, CBL, CD247, COS5, MAPK1, NCQ1, PDK1, PPP3CA, PRKCQ, VAV3     |
| KEGG             | 5    | TGF-beta signaling pathway               | 3.46     | 8        | 0.022         | BMP2, BMP8B1, DCN, ID4, MAPK1, NOG, TGF2B, THBS2                      |
| KEGG             | 6    | Cytokine-cytokine receptor interaction   | 10.79    | 18       | 0.022         | BMP2, BMP8B1, CCR5, CD40, CBL, CSF2RA, CX3CR1, CXCL13, EGF, IFNRA6, IL1R2, PRK1, PRL, TGF2, THBS, VEGFA                   |
| KEGG             | 7    | Arrhythmogenic right ventricular cardiomyopathy | 3.09    | 7        | 0.034         | ACTN1, CTTNA2, DMD, ITGA10, ITGAV, LAM2, SGCG                          |
| Gene Ontology    | 1    | negative regulation of phosphatase activity | 0.21    | 3        | 0.0006        | PPP2R4, TGF2B, TIPRL                                                  |
| Gene Ontology    | 2    | positive regulation of apoptosis         | 13.93    | 27       | 0.0008        | AGT, AKAP13, ARHGEF18, BCL2, BCL2L11, BCL2L13, BID, BIK, BMP2, BTK, CD38, HATAT2P1, IKBKE, ITGAV, MAGED1, MAPK1, MITCH1, PAVR, PPP2R4, PRUNE2, PVR, SOD1, TFAAP2, TGF2B, TAM1, VAV3, WTI |
| Gene Ontology    | 3    | regulation of phosphatase activity       | 0.53     | 4        | 0.0015        | BMP2, PPP2R4, TGF2B, TIPRL                                            |
| Gene Ontology    | 4    | glomerular epithelium development        | 0.08     | 2        | 0.0017        | BASP1, WTI                                                            |
| Gene Ontology    | 5    | vesicle                                  | 12.50    | 24       | 0.0017        | APP1L1, BGN, CD36, CTSG, CUZD1, CXC4, CYBA, DVL1, EGF, GRIA2, HFE, HPS4, LTF, NRSN1, PALM, RASSF9, SEC24A, SOD1, SYT1, SYT2, TGF2, TH, THBS2, VEGFA                 |
| Gene Ontology    | 6    | cellular defense response                | 2.38     | 8        | 0.0024        | CCR5, CD300C, CD5L, CX3CR1, DCDC2, LSP1, LY96, NCR2                   |
| Gene Ontology    | 7    | cytoplasmic vesicle                      | 12.17    | 23       | 0.0026        | BGN, CD36, CTSG, CUZD1, CXC4, CYBA, DVL1, EGF, GRIA2, HFE, HPS4, LTF, NRSN1, PALM, RASSF9, SEC24A, SOD1, SYT1, SYT2, TGF2, TH, THBS2, VEGFA |
| Gene Ontology    | 8    | phosphoinositide 3-kinase cascade        | 0.33     | 3        | 0.0033        | AGT, INS, TGF2B                                                       |
| Gene Ontology    | 9    | hindbrain development                    | 0.37     | 3        | 0.0048        | CTNN2A, MYO1D, SDF4                                                  |
| Gene Ontology    | 10   | regulation of neuronal synaptic plasticity | 0.37    | 3        | 0.0048        | NETO1, SHSA9, SYNGRI                                                 |
| Gene Ontology    | 11   | neuron projection membrane               | 0.12     | 2        | 0.0049        | CTNNAP2, SHSA9                                                       |
| Gene Ontology    | 12   | dopamine biosynthetic process            | 0.12     | 2        | 0.0049        | TGF2B, TH                                                            |
| Gene Ontology    | 13   | hydrogen peroxide biosynthetic process   | 0.12     | 2        | 0.0049        | CYBA, SOD1                                                           |
| Gene Ontology    | 14   | positive regulation of respiratory burst | 0.12     | 2        | 0.0049        | INS, INSR                                                            |
| Gene Ontology    | 15   | cardiac epithelial to mesenchymal transition | 0.12    | 2        | 0.0049        | TGF2B, TH                                                            |
| Gene Ontology    | 16   | enzyme activator activity                | 7.11     | 15       | 0.0051        | AGT, APOA5, ARHGEF18, BCL2L13, BMP2, EGF, MMP17, OPN1, PPRM1, PPP1R12B, PPP2R4, RGS5, RGS5, TBC1D15, VAV3                |
| Gene Ontology    | 17   | phosphoinositide 3-kinase cascade        | 0.33     | 3        | 0.0033        | AGT, INS, TGF2B                                                       |
| Gene Ontology    | 18   | hindbrain development                    | 0.37     | 3        | 0.0048        | CTNN1A2, MYO1D, SDF4                                                 |
| Gene Ontology    | 19   | regulation of neuronal synaptic plasticity | 0.37    | 3        | 0.0048        | NETO1, SHSA9, SYNGRI                                                 |
| Gene Ontology    | 20   | neuron projection membrane               | 0.12     | 2        | 0.0049        | CTNNAP2, SHSA9                                                       |
| Gene Ontology    | 21   | dopamine biosynthetic process            | 0.12     | 2        | 0.0049        | TGF2B, TH                                                            |
| Gene Ontology    | 22   | hydrogen peroxide biosynthetic process   | 0.12     | 2        | 0.0049        | CYBA, SOD1                                                           |
| Gene Ontology    | 23   | positive regulation of respiratory burst | 0.12     | 2        | 0.0049        | INS, INSR                                                            |
| Gene Ontology    | 24   | cardiac epithelial to mesenchymal transition | 0.12    | 2        | 0.0049        | TGF2B, TH                                                            |
| Gene Ontology    | 25   | enzyme activator activity                | 7.11     | 15       | 0.0051        | AGT, APOA5, ARHGEF18, BCL2L13, BMP2, EGF, MMP17, OPN1, PPRM1, PPP1R12B, PPP2R4, RGS5, RGS5, TBC1D15, VAV3                |
Table 6. Cont.

| Category                | rank | Subcategory                                           | expected | observed | p-value (raw) | Genes                                      |
|-------------------------|------|-------------------------------------------------------|----------|----------|---------------|--------------------------------------------|
| Genes                   | 18   | epidermal growth factor receptor signaling            | 0.74     | 4        | 0.0054        | AGT EGF NCRI2 SNX6                           |
| Genes                   | 19   | extracellular matrix                                  | 7.23     | 15       | 0.0060        | ASRPN BGN CMA1 CPXM2 CTSG DCN ECM2 LAMA2 LUM MPP23B OGN SOD1 TGF82 USH2A VEGFA |
| NIA human disease 1     | 1    | Diabetes Mellitus. Type 2                              | 10.49    | 22       | 0.0003*       | ABCC8 AGT AKAP10 APOA5 BLC CCR5 CD40 CMA1 CYBA FABP1 DFO INS INSR KNCN11 MTTP PRKZSell TH THBS2 TXNIP VEGFA |
| NIA human disease 2     | 2    | Hyperlipoproteinemias                                 | 0.26     | 3        | 0.0012        | APOA5 FABP1 PPARA                              |
| NIA human disease 3     | 3    | Diabetic Angiopathies                                 | 1.94     | 7        | 0.0024        | CD40 CYBA INS KNCN11 PPARA TXNIP VEGFA        |
| NIA human disease 4     | 4    | Postmortem Changes                                   | 0.10     | 2        | 0.0026        | DAOA TH2                                      |
| NIA human disease 5     | 5    | Disease Progress                                      | 7.77     | 16       | 0.0030        | AGT BLC2 CCR5 CD40 CMA1 CX3CR1 DCN EGF HFE KNCN11 PPARA PRKZSELL SOX1 VEGFA WT1 |
| NIA human disease 6     | 6    | Birth Weight                                          | 1.69     | 6        | 0.0054        | EGF EPHX1 DFO KNCN11 INS TH                  |
| NIA human disease 7     | 7    | Pathological Conditions. Signs and Symptoms           | 23.66    | 34       | 0.0073        | AGT APOA5 BLC2 CCR5 CD40 CMA1 CX3CR1 CYBA DAAO DCN DISC1 DMD EGF EPHX1 FCER1A FTO HFL HTR2C INS INSR KNCN11 LTTF MTTP PDXNA2 POMC PPARA PRKZSELL SOX1 TH THBS2 TXNIP VEGFA WT1 |
| NIA human disease 8     | 8    | Bronchiolitis. Viral                                  | 0.15     | 2        | 0.0075        | CCR5 CX3CR1                                  |
| NIA human disease 9     | 9    | Kidney Failure. Acute                                 | 0.15     | 2        | 0.0075        | CYBA WT1                                     |
| NIA human disease 10    | 10   | Diseases in Twins                                    | 0.46     | 3        | 0.0086        | DISC1 HFE PDXNA2                              |
| NIA human disease 11    | 11   | Coronary Artery Disease                               | 4.55     | 10       | 0.0127        | AGT APOA5 CD36 CD40 CMA1 CX3CR1 CYBA PPARA THBS2 VEGFA |
| NIA human disease 12    | 12   | Dyslexia                                             | 0.26     | 2        | 0.0233        | DYX1C1 KIA0319                                |
| NIA human disease 13    | 13   | Myocardial Infarction                                 | 6.64     | 12       | 0.0282        | AGT AKAP10 APOA5 CCR5 CTSG CX3CR1 HFE INSR MTTP THBS2 TNRF54 VEGFA |
| NIA human disease 14    | 14   | Nutritional and Metabolic Diseases                   | 18.45    | 26       | 0.0295        | ABC8 AGT AKAP10 APOA5 BTC CBLB CCR5 CD36 CMA1 CYBA DCN FABP1 FTO HTR2C INS INSR KNCN11 MTTP POMC PPARA PRKZSELL TH THBS2 TXNIP VEGFA |
| NIA human disease 15    | 15   | Overweight                                            | 0.31     | 2        | 0.0338        | APOA5 FTO                                    |

A total of 823 genes surrounding the 603 top associated SNPs were put into the GeneTrail software. Surrounding genes were defined by either Grail (http://www.broadinstitute.org/mpg/grail/) or the Genome Browser (http://genom.ucsc.edu/). Gene families located in the same region were manually curated so that only one gene in each family remained in each region, based on a similar official gene symbol. *Significant after multiple testing correction using FDR adjustment. (p corr-value = 0.032). Size of test set: 823 (768 known). Number of known ref. IDs: 44829 Kegg: Number of annotated genes in test set was 220. Number of annotated genes in ref set was 5405. Gene Ontology: Number of annotated genes in test set was 476. Number of annotated genes in ref set was 11580. NIA human genes sets: Number of annotated genes in test set was 76. Number of annotated genes in ref set was 1487. doi:10.1371/journal.pone.0070174.t006
as part of an immune evasion strategy [36]. Leishmania Major inhibits CD40-triggered p38 MAPK signaling to anti-leishmanial functions [35]. It has been suggested that regulates DUSP expression and activity, which in turn contribute signaling through p38 MAPK and ERK1/2 [35]. CD40 also responses to another parasite, Leishmania Major, by shared rs6065961, Table S1). CD40 has been shown to regulate immune encoding for the immune molecule CD40 (associated SNP Table 8). Another gene present in these networks was the gene

---

**Table 7. Biological functions of genes surrounding SNPs from the two-locus interaction. Results from GeneTrail.**

| Category               | rank | Subcategory                  | expected | observed | p-value (raw) | enrichment | Genes                                      |
|------------------------|------|------------------------------|----------|----------|---------------|------------|--------------------------------------------|
| KEGG                   | 1    | Amoebiasis                   | 1.01     | 4        | 0.0178        | up         | ACTN1 CTSG GNA14 TGFB2                    |
| KEGG                   | 2    | T cell receptor signalling   | 1.04     | 4        | 0.0196        | up         | CBLB CD247 ICOS VAV3                     |
| KEGG                   | 3    | PPAR signaling pathway       | 0.66     | 3        | 0.0282        | up         | APOAS CD36 FABP1                         |
| KEGG                   | 5    | Ubiquitin mediated proteolysis| 1.34     | 4        | 0.0438        | up         | CBLB KLHL9 TCEB1 UBR5                    |
| KEGG                   | 6    | Primary immunodeficiency     | 0.34     | 2        | 0.0441        | up         | BTK ICOS                                 |
| KEGG                   | 7    | Basal transcription factors  | 0.35     | 2        | 0.0465        | up         | GTF2B TAF7L                               |
| NIA human disease gene sets | 1   | Hyperlipoproteinemias        | 0.07     | 2        | 0.0017        | up         | APOAS FABP1                              |
| NIA human disease gene sets | 2   | Diabetes Mellitus Type 2     | 2.80     | 6        | 0.0493        | up         | APOAS CCR5 CD36 CMA1 FABP1 TH             |

A total of 187 genes from the interaction analysis were put into the GeneTrail software.

Surrounding genes were defined by either Grail (www.broadinstitute.org/mpg/grail/) or the Genome Browser (http://genome.ucsc.edu/). Gene families located in the same region were manually curated so that only one gene in each family remained in each region, based on a similar official gene symbol. Size of test set: 186 (173 known). Number of known ref. IDs: 44829. KEGG: number of annotated genes in test set: 52. Genes in reference set: 5405. NIA human disease gene sets: number of annotated genes in test set: 20. Genes in reference set: 1487.
doi:10.1371/journal.pone.0070174.t007

---

Table 8). Another gene present in these networks was the gene encoding for the immune molecule CD40 (associated SNP rs6065961, Table S1). CD40 has been shown to regulate immune responses to another parasite, Leishmania Major, by shared signaling through p38 MAPK and ERK1/2 [35]. CD40 also regulates DUSP expression and activity, which in turn contribute to anti-leishmanial functions [35]. It has been suggested that Leishmania Major inhibits CD40-triggered p38 MAPK signaling as part of an immune evasion strategy [36].

Another overrepresented category from GeneTrail was the extracellular matrix (ECM) (Table 6). Also, in the two most significant Ingenuity networks from the 603 marker analyses, ECM molecules and matrix metalloproteinases (MMPs) were included (Table 8). The ECM represents a major barrier to parasites like amoebas and leishmania. Parasites produce a wide variety of proteases to break down the ECM in order to access essential nutrients and invade host tissue [37]. A different situation when the ECM is degraded is during nutrient deprivation. In this way the ECM can provide energy for starving host cells. Just like gluten, the ECM has an unusually high proline content. MMPs are enzymes, which break down ECM making proline readily available as a nutritional source. Pandhare and co-workers have shown that energy or nutrient stress activates MMPs as well as the degradation of proline and furthermore demonstrated that, as the levels of glucose decreased to 1 mM and lower in the medium, intracellular proline increased almost 2-fold [38]. If gluten lingering in the intestine conveys a signal of ECM degradation (due to increased proline levels), several other mechanisms will most likely signal that there is food available at the same time (salivary secretion as one example is shown in Table 6). In this case, the immune system will rule out starvation as a possibility and the only other sensible option would be to search for an invasive intruder breaking down the ECM. The autoantigen in CD, TGM2, counteracts proteolysis and degradation of ECM by crosslinking ECM proteins [39]. If DUSP10 and PRR5 up-regulate TNF-α and subsequently TGM2 [17,18,25], in CD, the purpose may very well be for TGM2 to help prevent an apparent or illusory pathogenic invasion. It has also been shown that down-regulation of SVIL protects against ECM invasion by pathogens [40]. In our gene expression analysis SVIL was nominally significantly down-regulated in cases (Table 9).

When the body “senses” a pathogen disturbing energy balance or breaking down ECM, but there are no pathogenic antigens present, maybe there could be a risk that “self” antigens become our immune systems futile attempt to rid the perceived pathogen. In HLA-DQA1*02/05 and HLA-DQB1*02 carriers, peptides derived from TGM2 could constitute such “self” antigens. It is possible that individuals carrying other HLA molecules still respond to this “phantom pathogen” and that under these circumstances, various other antigens present in the intestine at the time could become triggers of other autoimmune diseases. If the expression or presence of an autoantigen, like TGM2, was stimulated by the disturbed proline/glutamine homeostasis, it can explain why symptoms in CD also disappear by withdrawal of gluten.

**Conclusion**

At least four major functional components together with gluten, all seem to play a role in forming an individual’s risk for CD:

1. polarity and epithelial cell functionality, e.g. nutrient/vesicle transport, proliferation and apoptosis, important for cell migration from the crypt to the shedding (apoptosis) at the apical villi.
2. intestinal smooth muscle, which is important for the movement of the bowel as well as the villi.
3. growth and energy homeostasis, which includes proline and glutamine metabolism, and finally
4. the innate and adaptive immune system.
### Table 8. The top four networks generated by the Ingenuity IPA software (allowing only direct connections between proteins/genes).

| Rank | Top functions | Ingenuity score | Number of focus molecules |
|------|---------------|----------------|--------------------------|
| 1    | Cell Morphology, Cellular Assembly and Organization, Hair and Skin Development and Function. Ingenuity Score: 155, 109 focus molecules. | 97 | 86 focus molecules. |
| 2    | Cell Signaling, Metabolism. Ingenuity Score: 74, 67 focus molecules. | 74 | 69 focus molecules. |
| 3    | Cellular Assembly and Organization, Cellular Function and Maintenance. Ingenuity Score: 67, 69 focus molecules. | 77 | 69 focus molecules. |
| 4    | Post-Translational Modification, Carbohydrate Metabolism, Lipid Metabolism. Ingenuity Score: 74, 67 focus molecules. | 74 | 67 focus molecules. |

The results of the network analysis included our genome-wide significant finding (DUSP10) within the top scoring network. P38 MAPK which interacts with DUSP10 is included in the second top network. The MHC class II complex is part of the second network. Genes within ours (P38 MAPK and DUSP10) and previously identified genome-wide significant regions are marked in italic, bold text. Only bold text show genes involved in amoebiasis. Underlined genes showed differences in our gene expression analysis (Table 9).

Rank: Top functions; Ingenuity score: Number of focus molecules; Molecules in Network.

doi:10.1371/journal.pone.0070174.t008
A slight dysfunction combining these categories together with gluten consumption would result in a metabolic imbalance which in turn could convey enough stress or “danger signal”, to trigger the immunological process and tissue destruction. A schematic illustration showing a rough outline of a possible disease model is presented in Figure 6.

In this study, we identified DUSP10 to be significantly associated with celiac disease. We also identified mechanisms, which we believe influence the risk of developing disease. Our data points towards genes that are involved in cancer as well as metabolic and cardiovascular diseases. Besides understanding how they work in celiac disease, our findings could also have consequences for these other common diseases.

Whole genome analysis allows for discovering completely unknown mechanisms behind disease. Even if the discovered genes and gene variants won’t be able to predict who will develop disease in the future, they can be used to identify the underlying molecular pathways that influence disease. These molecular pathways would then be valuable targets for drug intervention. Our data provides new insights and hypotheses to the research field of CD and autoimmunity. However, the functional variants behind associations as well as mechanisms causing differences in gene expression and if and how these are relevant for disease, remains to be identified.

**Materials and Methods**

**Ethics Statement**

The regional ethics board in Gothenburg approved this study and participants in the study gave written informed consent after being fully informed about the aim of the study. For all children in the study, parental written consent was obtained.

**Study Population**

A total of 106 families with multiple affected individuals, mostly nuclear families with an affected sib pair (ASP), were collected from Sweden and Norway. There were 403 subjects and 97 families with DNA to complete the analysis. A total of 226 of the family members had CD, including 20 parents. The makeup and selection process regarding the families has been described previously in detail [41].

Small-intestinal biopsies, for the gene expression analysis, were collected at four pediatric clinics in Sweden: Skåne University Hospital in Malmö, Sach’s Childrens’ Hospital and Karolinska...
Table 9. Results from gene expression analysis of 34 candidate genes.

| Gene Symbol | Assay Id | Gene | Fold Change | p-value | p-value corr. b | Selection criteria |
|-------------|----------|------|-------------|---------|-----------------|-------------------|
| ADCY9a      | Hs00181599_m1 | adenylate cyclase 9 | 1.58 | DOWN | 7.55E-06 | two-locus |
| APPL2a      | Hs00216855_m1 | adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper cont. 2 | 1.51 | DOWN | 2.15E-05 | two-locus |
| GLS         | Hs00221514_m1 | glutaminase | 1.46 | UP | 4.99E-06 | two-locus/IPA/previous |
| IRS1        | Hs00233154_m1 | insulin receptor | 1.15 | UP | 7.75E-04 | two-locus/IPA/previous |
| PDE18A      | Hs00134678_m1 | protein phosphatase 1, regulatory (inhibitor) subunit 12B | 1.15 | UP | 0.029 | top |
| PDK1        | Hs01561850_m1 | pyruvate dehydrogenase kinase, isozyme 1 | 1.30 | DOWN | 8.39E-05 | two-locus |
| PRK1        | Hs00223154_m1 | kinesin family member 13A | 1.22 | DOWN | 1.76E-04 | two-locus |
| PPP1R12B    | Hs00364073_m1 | protein phosphatase 1, regulatory (inhibitor) subunit 12B | 1.44 | DOWN | 2.03E-04 | two-locus/IPA/previous |
| PTP4A       | Hs00186620_m1 | regulator of G-protein signaling 1 | 1.11 | DOWN | 0.015 | IPA |
| RGS2        | Hs00364078_m1 | protein phosphatase 1, regulatory (inhibitor) subunit 12B | 1.08 | DOWN | 0.053 | two-locus/IPA/previous |
| DUSP10      | Hs00200527_m1 | dual specificity phosphatase 10 | 1.12 | UP | 0.704 | two-locus |
| IKBKE       | Hs01063858_m1 | inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase epsilon | 1.18 | UP | 0.014 | two-locus/IPA/previous |
| UNC5C       | Hs00931734_m1 | supervillin | 1.13 | DOWN | 0.026 | two-locus/IPA/previous |
| ARID1B      | Hs00368175_m1 | AT rich interactive domain 1B | 1.13 | DOWN | 0.024 | two-locus/IPA/previous |
| PKN2        | Hs00178944_m1 | protein kinase N2 | 1.14 | DOWN | 0.026 | two-locus/IPA/previous |
| ITPK1       | Hs00356546_m1 | inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase | 1.12 | UP | 0.024 | two-locus/IPA/previous |
| ITK         | Hs00397766_m1 | dipeptidyl-peptidase 10 | 1.75 | DOWN | 0.035 | two-locus/IPA/previous |
| RGS5        | Hs0031734_m1 | regulator of G-protein signaling 1 | 1.34 | DOWN | 0.053 | two-locus/IPA/previous |
| SVIL        | Hs00356656_m1 | regulator of G-protein signaling 1 | 1.34 | DOWN | 0.035 | two-locus/IPA/previous |
| MAGED1      | Hs00986269_m1 | melanoma antigen family D, 1 | 1.10 | DOWN | 0.053 | two-locus/IPA/previous |
| FOXD3       | Hs00255287_s1 | forkhead box D3 | 1.10 | DOWN | 0.053 | two-locus/IPA/previous |
| ITPK1-AS1   | Hs01053867_s1 | inositol 1,3,4-triphosphate 5/6 kinase Associated | 1.10 | DOWN | 0.053 | two-locus/IPA/previous |
| LPP         | Hs00353878_m1 | lipoprotein receptor | 1.10 | DOWN | 0.053 | two-locus/IPA/previous |
| CCR5        | Hs00168212_m1 | regulator of G-protein signaling 5 | 1.10 | DOWN | 0.047 | two-locus/IPA/previous |
| CD5L        | Hs00353257_m1 | fibronectin | 1.10 | DOWN | 0.053 | two-locus/IPA/previous |
| CTNNB1      | Hs00353623_m1 | chemokine (C-C motif) receptor 3 | 1.10 | DOWN | 0.053 | two-locus/IPA/previous |

Celiac Disease Genome-Wide Linkage and Association
Cont.

| Gene symbol | Assay id  | Gene Fold Change | p-value | p-value corr. | Selection criteria |
|-------------|----------|------------------|---------|---------------|-------------------|
| TIPRL       | Hs00295580_m1 | TIP41, TOR signaling pathway regulator-like (S. cerevisiae) | 1.01 | DOWN | 0.752 | genetrail |
| KHDRBS2     | Hs01061150_m1 | KH domain containing, RNA binding, signal transduction associated 2 | 1.06 | UP | 0.840 | two-locus |
| GTF2B       | Hs00976258_m1 | general transcription factor IIB | 1.03 | UP | 0.888 | IPA/genetrail |
| ACTN1       | Hs00998100_m1 | actinin, alpha 1 | 1.01 | DOWN | 0.914 | two-locus |
| DUSP10      | Hs04189838_m1 | dual specificity phosphatase 10 | No expression detected | top |

Expression (e.g. mRNA levels) of these genes was either up- or down-regulated in small intestinal biopsies from CD cases compared with control patients. Effect direction is presented for cases with control group as a reference. The selection column indicates if the gene was selected due to its presence in two-locus or pathway analyses. "Top" indicates top SNP in the present GWAS and "top previous" indicates that it was present in the GWAS by Dubois et al. All the gene assays (primers and probes) were predesigned and ordered from Life technologies (CA, USA).

Reference genes tested were: ACTB (Hs00357333_g1), B2M (Hs99999907_m1), EPCAM (Hs00158980_m1), GUSB (Hs99999908_m1), HPRT1 (Hs99999909_m1), MUC1 (Hs00159357_m1), PGK1 (Hs00999906_m1). For the results a combined value of ACTB, EPCAM, and PGK1 showed to be optimal when analysed by GeNorm and were selected as reference.

Gene not included in the gene list used for pathway analyses due to recombination between associated SNP and gene promotor: rs10861406. (APPL2), rs882820 (ADCY9). However, possible regulatory site could be close to associated SNP and influence gene expression.

bP-values corrected using Bonferroni correction.

dGene not included in the gene list used for pathway analyses due to recombination between associated SNP and gene promotor.

Celiac Disease Genome-Wide Linkage and Association

Gene Expression Analysis

We performed quantitative gene expression analysis using duodenal biopsies from CD autoimmunity patients and control patients. Biopsies were immediately put in RNAlater solution (Life Technologies, CA, USA). Total RNA was extracted using the miRNeasy Mini Kit (QIAGEN, Germany). RNA was converted to cDNA and quantitative PCR was run using TaqMan chemistry and the ABI7900 SDS instrument (Life Technologies, CA, USA).

Seven control genes were evaluated using GeNorm [45] (ACTB (Hs00357333_g1), B2M (Hs99999907_m1), EPCAM (Hs00158980_m1), GUSB (Hs99999908_m1), HPRT1 (Hs99999909_m1), MUC1 (Hs00159357_m1), PGK1 (Hs00999906_m1)) and the geometrical mean of ACTB, EPCAM, and PGK1 were selected as reference for the relative quantification analysis (Delta-Delta Ct method). A total of 34 expressed genes located close to some of the most significantly associated SNPs were evaluated (Table 9). The top associated genes from our Linkage GWAS (DUSP10, STL, PPP1R12B) were selected as well as several genes from the two-locus interaction analysis and pathway analyses including LPP which is the top associated from the GWAS by Dubois et al. Also the RGS genes (RGS1, 2 and 5) and GLS show genome-wide association in the study by Dubois et al. and are also present in our two-locus and pathway analyses.

Genotyping and Imputation

Samples were genotyped using two different SNP arrays, 211 samples with Human Omni Express and 192 samples with Human 660W-Quad (Illumina Inc, CA, USA). A total of 308,246 markers were available on both arrays and were therefore genotyped in the entire material. For the remaining 682,470 and for sporadic missing values we performed genotype imputation using the Impute 2 software [46], with the Hapmap 2 (rel. 24 Build 36) as a reference.

All individuals in the same family were located on the same plate. Quality control was first performed separately for the two arrays. SNP markers with less than 97% call rate in either of the two arrays were excluded.

Mendelian errors were detected by PLINK [47], 125,874 family-wise mendelian inconsistencies were set to missing (in each family the genotypes were set to missing for all subjects if there were any mendelian inconsistencies for a specific SNP).

Statistical Analysis

Linkage. For the linkage analysis only markers from both platforms were considered. From this set of 271,078 common SNP markers a LD pruned set of 105 539 SNPs were selected using PLINK. Parameters were a window size of 50 and $R^2$<0.5. The Decode genetic map as supplied by Illumina was used to run non-parametric linkage using Merlin version 1.1.2 [48] with the NPL
all method [49]. Marker allele frequencies were estimated from the founders.

Transmission Disequilibrium Test (TDT). Spielman et al introduced the Transmission Disequilibrium Test (TDT) in 1993 [50].

The imputation analysis provides us with (posterior) probabilities for each of the possible genotypes at each locus and to utilize all posterior probabilities, we performed an analysis where we use the expected values of the transmission counts. The test statistic will then have the following form,

\[
T_{\text{imp}} = \frac{(E[b] - E[c])^2}{E[b] + E[c]},
\]

(1)

\(T_{\text{imp}}\) has approximately the same distribution as the test statistic \(T\) in [50].

Stratified TDT. We implement a stratified TDT analysis where trios are split into a low-risk and a high-risk group based on the HLA genotype of the affected offspring. Children carrying the HLA-DQA1*02/05 risk allele and homozygous for the HLA-DQB1*02 risk allele (i.e. individuals carrying the DR3/DR3 or the DR3/DR7 haplotypes) were put in the “high-risk” group and the remaining children were put in the “low-risk” group. The rationale behind this is explained in the introduction and further information about this stratification can be found in our previous Linkage study [14]. A standard TDT analysis, with the 0.95 cut-off for imputation probabilities, was applied to each of these groups using PLINK [47].

Test for two-locus interaction. To examine possible interactions between marker variants, we used a pairwise test based on the one introduced by Kotti [51]. Consider two biallelic markers without linkage disequilibrium between their alleles. In the general model \(M_G\), the penetrance matrix has 9 parameters,

\[
\phi^G = \begin{bmatrix}
\phi_{00} & \phi_{01} & \phi_{02} \\
\phi_{10} & \phi_{11} & \phi_{12} \\
\phi_{20} & \phi_{21} & \phi_{22}
\end{bmatrix},
\]

Let \(n\) be the 3×3 matrix of genotype counts among the cases for the two markers, and let \(m\) be the corresponding matrix for the non-transmitted allele combinations. The likelihood for the models is

---

**Figure 5. NPL results.** Non-Parametric Linkage score displayed as \(-\log_{10}(p\text{-value})\) on the y-axis and chromosome 1–22 and X on the x-axis. doi:10.1371/journal.pone.0070174.g005

| chr | from(Mb) | to(Mb) | max NPL | p-value |
|-----|----------|--------|---------|---------|
| 6*  | 12.5     | 52.6   | 5.42    | 3.03E-08|
| 5   | 124.5    | 149.3  | 3.33    | 4.36E-04|
| 1   | 200.2    | 231.8  | 3.12    | 9.20E-04|
| 11  | 122.2    | 130.2  | 2.95    | 1.59E-03|
| 9   | 30.3     | 34.7   | 2.82    | 2.40E-03|
| 4   | 96.5     | 111.3  | 2.81    | 2.46E-03|
| 3   | 104.7    | 108.6  | 2.70    | 3.49E-03|
| 14  | 85.7     | 86.4   | 2.57    | 5.16E-03|
| 6   | 160.4    | 161.0  | 2.54    | 5.50E-03|
| 11  | 77.6     | 78.4   | 2.48    | 6.64E-03|
| 18  | 55.0     | 55.1   | 2.45    | 7.20E-03|
| 1   | 29.7     | 29.9   | 2.42    | 7.87E-03|
| 2   | 127.0    | 127.1  | 2.41    | 8.00E-03|
| 2   | 106.3    | 106.4  | 2.37    | 8.89E-03|

Regions showing significant linkage (the HLA region only) and putative linkage (nominal \(p<0.01\). Regions in the table are defined as the Megabase (Mb) interval showing a nominal \(p<0.01\). Neighbouring regions were merged if \(<15\ Mb pairwise distance\).

Max NPL – the maximum Z score across the region between the positions ‘from’ and ‘to’.

p-value – the p-value for the max NPL score.

*=The HLA region.

doi:10.1371/journal.pone.0070174.t010
For this analysis we use one affected subject from each family and markers were chosen based on the expected counts TDT (equation 1) and three different inclusion criteria:

1. P-value less than $3.0 \times 10^{-4}$.
2. P-value less than 0.01 in our analysis and with a p-value less than 0.05 in the GWAS by Dubois et al. [3] and if the product of these p-values were less than $5.0 \times 10^{-5}$ and the association were in the same allelic direction.
3. An allele transmission ratio of $<0.2$ or $>5$ combined with a p-value less than $2.0 \times 10^{-3}$.

We defined 383 regions using the inclusion criteria above (Fig. 2 and Table S1) a region consisted of a set of markers where the distance between adjacent markers was less than 100 kb. With these regions defined we analyzed all pairwise interactions using a Likelihood Ratio (LR) tests comparing the following four models:

- $N_{M0}$: None of the two loci is associated with CD,
- $N_{M0}$: Heterogeneity model [52], with penetrance

where $\alpha_i$ and $\beta_j$ are the penetrance factors for the genotypes $A_i$ and $B_j$ [53] respectively.

- $M_{G}$: Multiplicative model,
- $M_{G}$: the general model.

The restricted model used in [51] is the multiplicative model. We use the $M_{G}$-versus-$M_{M}$ test to filter out false positives, based on that if one or both of the SNPs were marginally significant by chance, then the joint distribution (penetrance) of these markers should follow a multiplicative model.

We have the likelihood ratio statistic

$$T_{jk} = -2 \log \frac{\max L_j}{\max L_j}$$

$T_{jk}$ will follow a $x^2$ distribution under the restricted model if $M_j$ is nested in $M_k$. The maximum likelihood estimates of the penetrance parameters and allele frequencies do not have a simple explicit expression, so to maximize the likelihoods we use the function `optim` in the statistical software R.

**Gene Selection**

Out of the 603 SNPs selected from the three inclusion criteria (Fig. 2 and Table S1), we were able to identify genes surrounding...
444 SNPs using GRAIL [54]. Grail uses known recombination hotspots in order to limit the region of interest surrounding each SNP marker. Genes around the remaining SNPs were identified with the Genome Browser [http://genome.ucsc.edu] and the 5 closest genes within 250 kb from the associated SNPs were included. In cases where there were no genes within this distance we included the closest gene.

Pathway Analysis
We analyzed connections between genes in different regions, using GeneTrail [13] and the Ingenuity Pathway Analysis (IPA) software (Ingenuity Inc., CA, USA). Within each associated region, all but one gene from the same gene family were removed. This was done in order not to amplify the significance of homologous gene clusters, i.e. chemokine receptor-, interferon- and histone-gene clusters.

URLs
PLINK [http://pngu.mgh.harvard.edu/purcell/plink/]
KEGG [www.genome.jp/kegg/]
Gene Ontology [www.genontology.org/]
GWAS catalog [http://www.genome.gov/gwastudies/]
GRAIL [http://www.broadinstitute.org/mpg/grail/]
SNAP [http://www.broadinstitute.org/mpg/snap/lpdlplot.php/]
GeneTrail [http://genetraill.bioinf.uni-sb.de/]

References
1. Mylrea A, Ivarsson A, Webb C, Danielson L, Hernell O, et al. (2009) Celiac disease revealed in 3% of Swedish 12-year-olds born during an epidemic. J Pediatr Gastroenterol Nutr 49: 170–176.
2. Soloid LM, Markussen G, Ekj J, Gjerde H, Vardal F, et al. (1989) Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J Exp Med 169: 345–350.
3. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 42: 295–302.
4. Festen EA, Gothey P, Green T, Boucher G, Beauchamp C, et al. (2011) A meta-analysis of genome-wide association scans identifies IL1RAP, PTTRN2, TAGAP, and PASU1 as shared risk loci for Crohn’s disease and celiac disease. PLoS Genet 7: e1001283.
5. Hunt KA, Zhermakova A, Turner G, Heap GA, Franke L, et al. (2008) Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 40: 395–402.
6. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, et al. (2007) A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PASU1 as shared risk loci for Crohn's disease and celiac disease. PLoS Genet 7: e1000249.
7. Kumar V, Wijnenga C, Witthoff S (2012) From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. Semin Immunopathol 34: 567–580.
8. Risch N, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. Genome Res 8: 1273–1288.
9. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516–1517.
10. Risch N, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. Genome Res 8: 1273–1288.
11. Nalini AT, Nilsson S, Gutjonsdottr AH, Louka AS, Ascher H, et al. (2001) Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. Genet Epidemiol 25: 827–829.
12. Zhermakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, et al. (2011) Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet 7: e1002405.
13. Dubois et al. [3].
14. Teng CH, Huang WN, Meng TC (2007) Several dual specificity phosphatases coordinate to control the magnitude and duration of JNK activation in signaling response to oxidative stress. J Biol Chem 282: 29395–29407.
15. Holmes B, Artinian N, Anderson L, Martin J, Masri J, et al. (2012) Protor-2 interacts with trinucleotripol to regulate mRNA stability during stress. Cell Signal 24: 309–315.
16. Luo LJ, Liu F, Wang XY, Dai FY, Dai XL, et al. (2012) An essential function for MKP5 in the formation of oxidized low density lipid-induced foam cells. Cell Signal 24: 1889–1898.
17. Dieiterich W, Ehns T, Bauer, M, Donner P, Volta U, et al. (1997) Identification of tissue transglutaminase as the autoantigen of celiac disease. Nat Med 3: 797–801.
18. Wang K, Zhang H, Blos CS, Duvvari V, Kaye W, et al. (2011) A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. Mol Psychiatry 16: 949–959.
19. Paterson AD, Waggott D, Boright AP, Hossein SM, Shen E, et al. (2010) A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. Diabetes 59: 330–339.
20. Lin Z, Beit JX, Shen M, Li Q, Liao Z, et al. (2012) A genome-wide association study in Han Chinese identifies new susceptibility loci for angiogenesis. Nat Genet 44: 77–83.
21. Liu W, Le A, Hancock C, Lane AN, Dang CV, et al. (2012) Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. Proc Natl Acad Sci U S A 109: 19883–19890.
22. D'Alessandro G, Calcagno E, Tartari S, Rizzardi M, Invernizzi RW, et al. (2011) Glutamate and glutathione interplay in a motor neuronal model of amyotrophic lateral sclerosis reveals altered energy metabolism. Neurobiol Dis 45: 346–355.
23. Holmes B, Artinian N, Anderson L, Martin J, Masri J, et al. (2012) Protor-2 interacts with trinucleotripol to regulate mRNA stability during stress. Cell Signal 24: 309–315.
24. Saito T, Jones CC, Huang S, Czech MP, Plich PF (2007) The interaction of Akt with APPL1 mediates apoptosis and glucose transport by c-Myc. Proc Natl Acad Sci U S A 104: 3530–3535.
25. Holmes B, Artinian N, Anderson L, Martin J, Masri J, et al. (2012) Protor-2 interacts with trinucleotripol to regulate mRNA stability during stress. Cell Signal 24: 309–315.
26. Saito T, Jones CC, Huang S, Czech MP, Plich PF (2007) The interaction of Akt with APPL1 mediates apoptosis and glucose transport by c-Myc. Proc Natl Acad Sci U S A 104: 3530–3535.
27. Holmes B, Artinian N, Anderson L, Martin J, Masri J, et al. (2012) Protor-2 interacts with trinucleotripol to regulate mRNA stability during stress. Cell Signal 24: 309–315.
28. Saito T, Jones CC, Huang S, Czech MP, Plich PF (2007) The interaction of Akt with APPL1 mediates apoptosis and glucose transport by c-Myc. Proc Natl Acad Sci U S A 104: 3530–3535.
29. Holmes B, Artinian N, Anderson L, Martin J, Masri J, et al. (2012) Protor-2 interacts with trinucleotripol to regulate mRNA stability during stress. Cell Signal 24: 309–315.
30. Ito M, Nakano T, Erdodi F, Hartshorne DJ (2004) Myosin phosphatase: a novel MAPK-regulated protein phosphatase that opposes Rap1 in the regulation of Mek1 activation. J Biol Chem 279: 31608–31615.
31. Holmes B, Artinian N, Anderson L, Martin J, Masri J, et al. (2012) Protor-2 interacts with trinucleotripol to regulate mRNA stability during stress. Cell Signal 24: 309–315.
32. Ito M, Nakano T, Erdodi F, Hartshorne DJ (2004) Myosin phosphatase: a novel MAPK-regulated protein phosphatase that opposes Rap1 in the regulation of Mek1 activation. J Biol Chem 279: 31608–31615.
31. Goto S, Bono H, Ogata H, Fujisuchi W, Nishihira T, et al. (1997) Organizing and computing metabolic pathway data in terms of binary relations. Pac Symp Biocomput: 175–186.
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29.
33. Takizawa N, Smith TC, Nebl T, Crowley JL, Palmieri SJ, et al. (2006) Supervillin modulation of focal adhesions involving TRIP6/ZRP-1. J Cell Biol 174: 447–458.
34. Nakagawa T, Setou M, Seog D, Ogasawara K, Dohmae N, et al. (2000) A novel motor, KIF13A, transports mannose-6-phosphate receptor to plasma membrane through direct interaction with AP-1 complex. Cell 103: 569–581.
35. Srivastava N, Sudan R, Saha B (2011) CD40-modulated dual-specificity phosphatases MAPK phosphatase (MKP)-1 and MKP-3 reciprocally regulate Leishmania major infection. J Immunol 186: 5863–5872.
36. Awasthi A, Mathur R, Khan A, Joshi BN, Jain N, et al. (2009) CD40 signaling is impaired in L. major-infected macrophages and is rescued by a p38MAPK activator establishing a host-protective memory T cell response. J Exp Med 197: 1037–1043.
37. Pina-Vazquez C, Reyes-Lopez M, Oriz-Estrada G, de la Garza M, Serrano-Luna J (2012) Host-parasite interaction: parasite-derived and -induced proteases that degrade human extracellular matrix. J Parasitol Res 2012: 748206.
38. Pandhare J, Donald SP, Cooper SK, Phang JM (2009) Regulation and function of proline oxidase under nutrient stress. J Cell Biochem 107: 759–768.
39. Mangala LS, Arun B, Sahin AA, Mehta K (2005) Tissue transglutaminase-induced alterations in extracellular matrix inhibit tumor invasion. Mol Biol Cell 20: 948–962.
40. Gudjonsdottir AH, Nilsson S, Ek J, Kristiansson B, Ascher H (2004) The risk of celiac disease in 107 families with at least two affected siblings. J Pediatr Gastroenterol Nutr 38: 338–342.
41. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5: e1000529.
42. Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506–516.
43. Otto S, Bickeboller H, Clerget-Darpoux F (2007) Strategy for detecting susceptibility genes with weak or no marginal effect. Hum Hered 63: 85–92.
44. Rice JP, Neuman RJ (2005) Two-locus models of disease. Genetic Epidemiology 9: 347–365.
45. Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46: 222–228.
46. Raychaudhuri S, Plenge RM, Rossin EJ, Ng ACY, Purcell SM, et al. (2009) Identifying Relationships among Genomic Disease Regions: Predicting Genes at Pathogenic SNP Associations and Rare Deletions. PLoS Genet 5.