
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU


HARVARD LIBRARY
Office for Scholarly Communication

# Meta-Analysis of Genome-Wide Association Studies in Celiac Disease and Rheumatoid Arthritis Identifies Fourteen Non-HLA Shared Loci

**Citation**
Zhernakova, Alexandra, Eli A. Stahl, Gosia Trynka, Soumya Raychaudhuri, Eleanora A. Festen, Lude Franke, Harm-Jan Westra, et al. 2011. Meta-Analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genetics 7(2): e1002004.

**Published Version**
doi:10.1371/journal.pgen.1002004

**Permanent link**
http://nrs.harvard.edu/urn-3:HUL.InstRepos:5360624

**Terms of Use**
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

## Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

OPEN ACCESS Freely available online

PLoS GENETICS

# Meta-Analysis of Genome-Wide Association Studies in Celiac Disease and Rheumatoid Arthritis Identifies Fourteen Non-HLA Shared Loci

**Alexandra Zhernakova[1,2,3], Eli A. Stahl[3,4], Gosia Trynka[5], Soumya Raychaudhuri[3,4,6,7], Eleanora A. Festen[5], Lude Franke[5,8], Harm-Jan Westra[5], Rudolf S. N. Fehrmann[5], Fina A. S. Kurreeman[1,3,4], Brian Thomson[4], Namrata Gupta[4], Jihane Romanos[5], Ross McManus[9], Anthony W. Ryan[9], Graham Turner[9], Elisabeth Brouwer[10], Marcel D. Posthumus[10], Elaine F. Remmers[11], Francesca Tucci[12], Rene Toes[1], Elvira Grandone[13], Maria Cristina Mazzilli[14], Anna Rybak[15], Bozena Cukrowska[16], Marieke J. H. Coenen[17], Timothy R. D. J. Radstake[18], Piet L. C. M. van Riel[18], Yonghong Li[19], Paul I. W. de Bakker[3,4,20,21], Peter K. Gregersen[22], Jane Worthington[23], Katherine A. Siminovitch[24], Lars Klareskog[25], Tom W. J. Huizinga[1], Cisca Wijmenga[5☯], Robert M. Plenge[3,4,7*☯]**

**1** Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands, **2** Complex Genetics Section, Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands, **3** Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **4** Broad Institute, Cambridge, Massachusetts, United States of America, **5** Genetics Department, University Medical Centre Groningen and University of Groningen, Groningen, The Netherlands, **6** Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **7** Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **8** Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom, **9** Department of Clinical Medicine and Institute of Molecular Medicine, Trinity Centre for Health Sciences, Trinity College, St James's Hospital, Dublin, Ireland, **10** Department of Rheumatology and Clinical Immunology, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands, **11** Genetics and Genomics Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, **12** European Laboratory for Food Induced Disease, University of Naples Federico II, Naples, Italy, **13** Unita' di Aterosclerosi e Trombosi, I.R.C.C.S Casa Sollievo della Sofferenza, S. Giovanni Rotondo, Foggia, Italy, **14** Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy, **15** Department of Gastroenterology, Hepatology, and Immunology, Children's Memorial Health Institute, Warsaw, Poland, **16** Department of Pathology, Children's Memorial Health Institute, Warsaw, Poland, **17** Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **18** Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **19** Celera, Alameda, California, United States of America, **20** Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **21** Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands, **22** The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, United States of America, **23** Arthritis Research Campaign–Epidemiology Unit, The University of Manchester, Manchester, United Kingdom, **24** Department of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Canada, **25** Rheumatology Unit, Department of Medicine, Karolinska Institutet at Karolinska University Hospital Solna, Stockholm, Sweden

## Abstract

Epidemiology and candidate gene studies indicate a shared genetic basis for celiac disease (CD) and rheumatoid arthritis (RA), but the extent of this sharing has not been systematically explored. Previous studies demonstrate that 6 of the established non-HLA CD and RA risk loci (out of 26 loci for each disease) are shared between both diseases. We hypothesized that there are additional shared risk alleles and that combining genome-wide association study (GWAS) data from each disease would increase power to identify these shared risk alleles. We performed a meta-analysis of two published GWAS on CD (4,533 cases and 10,750 controls) and RA (5,539 cases and 17,231 controls). After genotyping the top associated SNPs in 2,169 CD cases and 2,255 controls, and 2,845 RA cases and 4,944 controls, 8 additional SNPs demonstrated $P<5\times10^{-8}$ in a combined analysis of all 50,266 samples, including four SNPs that have not been previously confirmed in either disease: rs10892279 near the *DDX6* gene ($P_{combined}=1.2\times10^{-12}$), rs864537 near *CD247* ($P_{combined}=2.2\times10^{-11}$), rs2298428 near *UBE2L3* ($P_{combined}=2.5\times10^{-10}$), and rs11203203 near *UBASH3A* ($P_{combined}=1.1\times10^{-8}$). We also confirmed that 4 gene loci previously established in either CD or RA are associated with the other autoimmune disease at combined $P<5\times10^{-8}$ (*SH2B3*, *8q24*, *STAT4*, and *TRAF1-C5*). From the 14 shared gene loci, 7 SNPs showed a genome-wide significant effect on expression of one or more transcripts in the linkage disequilibrium (LD) block around the SNP. These associations implicate antigen presentation and T-cell activation as a shared mechanism of disease pathogenesis and underscore the utility of cross-disease meta-analysis for identification of genetic risk factors with pleiotropic effects between two clinically distinct diseases.

**Citation:** Zhernakova 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(2): e1002004. doi:10.1371/journal.pgen.1002004

**Editor:** Joshua M. Akey, University of Washington, United States of America

**Received** October 19, 2010; **Accepted** December 24, 2010; **Published** February 24, 2011

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

**Funding:** RMP was supported by grants from the NIH (R01-AR057108, R01-AR056768, U01-GM092691) and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. SR is supported by an NIH Career Development Award (1K08AR055688-01A1). FK is supported by a Marie Curie International Outgoing Fellowship for Career Development from the European Community's FP7 (Grant Agreement number PIOF-GA-2009-237280). The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278 from the National Center for Research Resources. The work presented is made possible by grants from the Coeliac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative, and partially funded by the Dutch Government (BSIK03009 to CW) and by the Netherlands Organisation for Scientific Research (NWO, VICI grant 918.66.620 to CW). LF received a Horizon Breakthrough grant from the Netherlands Genomics Initiative (93519031) and a VENI grant from NWO (ZonMW grant 916.10.135). AZ received a Rubicon grant from NWO (825.10.002). GT received a Ter Meulen Fund grant from the Royal Netherlands Academy of Arts and Sciences (KNAW). RMcM is supported by a Science Foundation Ireland Grant 09/IN.1/B2640. KAS is supported by a Canada Research Chair, Sherman Family Chair in Genomic Medicine, CIHR grants MOP79321 and IIN84042, and ORF grant RE01061. This work is partly supported by Coordination Theme 1 (Health) of the European Community's FP7, Grant Agreement number HEALTH-F2-2008-223404 (MASTERSWITCH project); by the EU FP6 supported AutoCure; and by the intramural program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** YL has employment and stock interest in Celera.

* E-mail: rplenge@partners.org

☯ These authors contributed equally to this work.

## Introduction

Autoimmune disorders, including rheumatoid arthritis (RA) and celiac disease (CD), affect about 5% of the population and have a complex genetic background. Family-based epidemiology studies suggest that there is a shared genetic basis between the two autoimmune diseases [1]. Recent genome-wide association studies (GWAS) have confirmed HLA and identified at least 26 other non-HLA genetic loci with common alleles associated to each disease (Table S1 and S2) [2,3]. The strongest genetic risk factor is the HLA locus [2,3], where different alleles confer risk of the two diseases. Six other risk loci outside of the HLA locus are shared between CD and RA and include *MMEL/TNFRSF14* [2,4], *REL* [2,5,6], *ICOS-CTLA4* [2,3,5,7], *IL2-IL21* [2,3,8,9,10], *TNFAIP3* [2,3,6,11], and *TAGAP* [2,3,8], (Chen et al, submitted) (Table 1 and Figure 1). These shared risk loci have emerged by simple cross-comparison across published studies, rather than a rigorous and systematic analysis of an integrated dataset. Because of the nature of these reports, it is unknown whether the other CD and RA risk alleles confer risk of both diseases. Moreover, it is unknown whether there are additional shared risk alleles that have not yet been discovered in any one disease.

A major challenge in identifying common alleles of modest effect is the sample size required to have sufficient power to obtain associations at a stringent level of statistical significance. Recent studies of height [12], lipids [13] and body mass index [14] have shown quite convincingly that very large sample sizes – more than 100,000 individuals – yield reproducible SNP associations for common alleles of modest effect size. For diseases such as CD and RA, which are relatively uncommon in the general population (prevalence ~0.5–1% for each disease), similar sized cohorts are difficult to ascertain. One solution to this problem is to combine two phenotypes to search for pleiotropic risk alleles. So far this approach has only been done for closely related phenotypes, such as the Crohn's disease and ulcerative colitis (together known as inflammatory bowel disease (IBD)) [15], or for medical traits that are known risk factors for disease (e.g., lipids and coronary artery disease, obesity and type 2 diabetes) [13,16].

Another challenge is how to interpret statistical significance of SNP associations in combined analysis of two clinically distinct phenotypes. In a GWAS of common variants for a single phenotype, most consider $P<5\times10^{-8}$ as statistically significant, as any SNP at random from the genome has the same probability of being associated with the phenotype and there are approximately 1 million uncorrelated common SNPs in the human genome [17]. However, this *P*-value threshold does not take into consideration that (a) many common SNPs, not just a single SNP, are associated with disease, and (b) the pleiotropy of risk alleles for related diseases should, in theory, increase the prior probability that an allele is a true-positive. In the case of autoimmunity, alleles often contribute to risk of more than one autoimmune disease [18]. Accordingly, a SNP with a confirmed association in one autoimmune disease has a higher prior probability of being associated with another autoimmune disease. This principle has been used to declare that SNPs are confirmed disease associations, if the SNP does not reach a stringent level of significance (e.g., $P<5\times10^{-8}$) in the other autoimmune disease [7,19]. Nonetheless, there are no formal criteria for assigning increased prior probabilities for SNPs across autoimmune diseases.

**Table 1.** Comparison of CD and RA risk alleles at seven shared risk loci.

| Locus | Locus name | Top CD SNP | Top RA SNP | LD between CD and RA SNPs (D′ and $r^2$) | | Comment |
|---|---|---|---|---|---|---|
| 1p36.3 | *MMEL1/TNFRSF14* | rs3890745 [2] | rs3748816 [3] | 1 | 0.93 | Same allele, same direction |
| 2p16.1 | *REL* | rs13003464 [2] | rs13031237 [5] | 0.11 | 0.01 | Different alleles |
| 2q33.2 | *ICOS/CTLA4* | rs4675374 [2] | rs3087243 [3] | 0.80 | 0.12 | Incomplete LD |
| 4q27 | *IL2/IL21* | rs13151961 [2] | rs6822844 [10] [3] | 1 | 0.90 | Same allele, same direction |
| 6p21 | *HLA* | rs2187668 (DQA1*0501-DQB1*0201 tag) [2] | rs6910071 (DRB1*0401 tag) [3] | 1 | 0.04 | Different alleles |
| 6q23.3 | *TNFAIP3* | rs2327832 [2] | rs6920220 [3] | 1 | 1 | Same allele, same direction |
| 6q25.3 | *TAGAP* | rs1738074 [2] | rs212389 (Chen et al, submitted) | 0.56 | 0.27 | Different alleles |

Columns Top SNP CD and Top SNP RA – best reported SNP in the locus, as indicated in the reference paper. Association is indicated to the same allele if the $r^2$ between CD and RA SNP is above 0.9.

doi:10.1371/journal.pgen.1002004.t001

## Author Summary

Celiac disease (CD) and rheumatoid arthritis (RA) are two autoimmune diseases characterized by distinct clinical features but increased co-occurrence in families and individuals. Genome-wide association studies (GWAS) performed in CD and RA have identified the HLA region and 26 non-HLA genetic risk loci in each disease. Of the 26 CD and 26 RA risk loci, previous studies have shown that six are shared between the two diseases. In this study we aimed to identify additional shared risk alleles and, in doing so, gain more insight into shared disease pathogenesis. We first empirically investigated the distribution of putative risk alleles from GWAS across both diseases (after removing known risk loci for both diseases). We found that CD risk alleles are non-randomly distributed in the RA GWAS (and vice versa), indicating that CD risk alleles have an increased prior probability of being associated with RA (and vice versa). Next, we performed a GWAS meta-analysis to search for shared risk alleles by combing the RA and CD GWAS, performing both directional and opposite allelic effect analyses, followed by replication testing in independent case-control datasets in both diseases. In addition to the already established six non-HLA shared risk loci, we observed statistically robust associations at eight SNPs, thereby increasing the number of shared non-HLA risk loci to fourteen. Finally, we used gene expression studies and pathway analysis tools to identify the plausible candidate genes in the fourteen associated loci. We observed remarkable overrepresentation of T-cell signaling molecules among the shared genes.


In the current study, we hypothesized that there are additional alleles that influence risk of both CD and RA in a pleiotropic manner. To increase power to detect these alleles, we combined two previously published GWAS of each disease, followed by replication in both CD and RA. We use our GWAS data to arrive at an empirical threshold for declaring SNPs as shared risk alleles for the two diseases. In doing so, we identified fourteen shared CD-RA risk alleles, which point to T-cell receptor signaling as a key shared pathway of disease pathogenesis.

## Results

### Comparing known risk alleles across diseases

We first aimed to investigate the status of established CD and RA loci across these two diseases using genotype data from published GWAS datasets of CD (4,533 cases, 10,750 controls) [2] and RA (5,539 autoantibody positive RA cases and 17,231 controls) [3] (See Materials and Methods for description of both cohorts). We considered only those reported loci with at least one risk allele associated at $P<5\times10^{-8}$ with confirmation in independent samples. There are 26 non-HLA loci from each disease that satisfy this stringent criterion, representing 46 distinct risk loci (Tables S1 and S2). We investigated the association of the 26 non-HLA CD SNPs in RA, and the 26 non-HLA RA SNPs in CD. Figure 1A and 1B show the OR and 95% CI of the 52 SNPs and the association statistics within the two diseases. Of the 26 CD SNPs, 11 are associated with risk of RA at $P<0.05$ (Table S1). Similarly, from 26 RA SNPs, 9 are associated with risk of CD at $P<0.05$ (Table S2). After excluding the six loci established in both diseases, this distribution remains non-random ($P<2\times10^{-4}$, Fisher's test), indicating additional sharing of risk loci between the two diseases.

### Comparing distribution of putative risk alleles across diseases

To provide additional evidence that there are shared risk alleles, we analyzed the distribution of moderately associated SNPs from the GWAS datasets (i.e., putative risk alleles) across the two autoimmune diseases. We investigated whether the subset of SNPs associated with CD at $P<0.001$ in the CD-GWAS are randomly distributed in the RA GWAS results, and vice-versa. After removing the established CD and RA risk loci, we performed association analysis on a set of independent SNPs for each disease. In CD, 70,520 SNPs remained after pruning SNPs in linkage disequilibrium (LD) (see Materials and Methods for details), of which 342 were associated with CD at $P<0.001$. In RA, 70,812 SNPs remained after LD-pruning, of which 282 were associated with RA at $P<0.001$. Using Fisher's test, we observed a non-random distribution of association with CD in the subset of $P<0.001$ RA GWAS SNPs, as well as a non-random distribution of association with RA in the subset of $P<0.001$ CD GWAS SNPs ($P<5\times10^{-5}$ for both diseases; see Figure 2 and Table S3A). Similar results were obtained when we used the Wilcoxon rank sum and Kolmogorov-Smirnov tests to analyze the distributions of SNP associations across diseases (Table S3B). From this analysis, we conclude that a SNP associated with risk of CD at $P<0.001$ has an increased prior probability of being associated with RA, and a SNP associated with risk of RA at $P<0.001$ has an increased prior probability of being associated with CD.

### GWAS and replication—same allele, same direction

While the analyses described above indicate that additional shared risk alleles remain to be discovered, these analyses do not identify which specific SNPs influence risk of both disease. To identify new shared risk alleles, we performed an inverse variance weighted meta-analysis [20] in which we assumed that the same allele confers risk of both diseases. A total of 472,854 SNPs outside the HLA (Chr6: 20–40 MB) overlapped between the two GWAS datasets and were included in the meta-analysis. We did not exclude the established CD and RA loci outside of the HLA region from the meta-analysis, as we considered the possibility that there may be novel risk alleles within these loci. The Q-Q plot of CD+RA meta-anlaysis P-values ($P_{\text{combined}}$) shows an enrichment of non-HLA associated SNPs in the tail of the distribution (Figure 3A), with no evidence for systematic bias across all SNPs ($\lambda_{GC} = 1.011$). A similar result was obtained after excluding known associated loci for both diseases (Figure 3A). The Manhattan plot indicates loci where significance increased in the combined cohort (Figure S1).

Sixty-five SNPs from 21 distinct genomic regions were associated with both CD and RA in the combined analysis with $P_{\text{combined}}<1\times10^{-5}$, and with disease-specific $P<0.01$ (Tables S4 and S5). Of these 21 loci, five are established in both diseases (*TNFAIP3*, *CTLA4/ICOS*, *IL2/IL21*, *REL* and *MMEL1/TNFRSF14*); five are established CD loci (*SH2B3*, *PTPN2*, *8q24.2*, *SOCS1*, *ICOSLG*); and four are established RA loci (*ANKRD55*, *STAT4*, *TRAF1/C5* and *PRKCQ*). The remaining 7 have not been previously confirmed in either disease (Table 2, Table S5).

To determine which of these loci are associated with both diseases – particularly those 7 loci not previously implicated in either disease and 9 loci established as risk alleles in either CD or RA alone – we selected from each of these 16 loci one most associated SNP for replication in additional 2,169 CD cases and 2,255 controls, and 2,845 autoantibody positive RA cases and 4,944 controls (see Materials and Methods for sample information). Five out of 16 SNPs were previously genotyped in samples

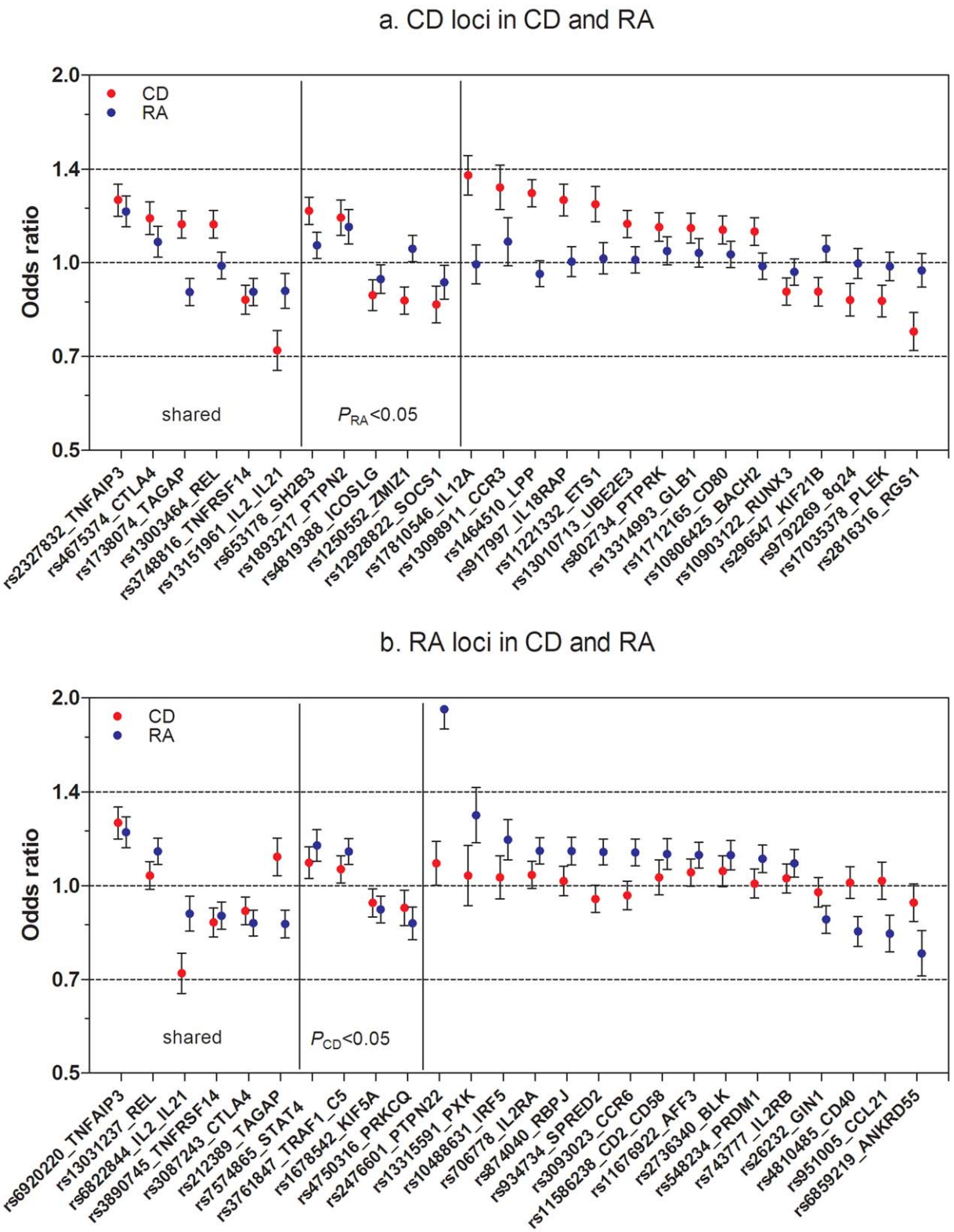


**Figure 1. Established CD and RA SNPs and their association across diseases.** (A) Known CD SNPs in RA. The figure represents OR and CI for the established CD SNPs ($p<5\times10^{-8}$, one SNP per locus) in RA meta-analysis (5,539 auto-anbitody positive cases and 17,231 controls). (B) Known RA SNPs in CD. The figure represents OR and CI for the established RA SNPs ($p<5\times10^{-8}$, one SNP per locus) in CD meta-analysis (4,533 cases and 10,750 controls). For the six shared loci established in both diseases, figure 1A includes the top CD SNP and figure 1B the top RA SNP. From six shared loci, three (*TNFRSF14*, *IL2/IL21* and *TNFAIP3*) are associated with same SNP or a good proxy ($r^2>0.9$) in both diseases; in other three loci – *CTLA4*, *REL* and *TAGAP* – the most associated SNPs in CD and RA are not in strong LD with each other ($r^2<0.3$), which is reflected in moderate association (*CTLA4*) or no association (*REL*) of these SNPs in the second disease. The *TAGAP* SNPs show association to opposite alleles in CD and RA.
doi:10.1371/journal.pgen.1002004.g001

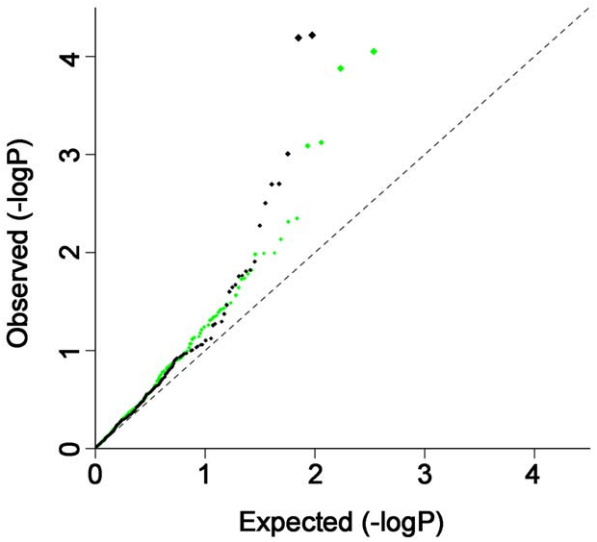


**Figure 2. QQ plot of CD associated SNPs in RA and RA associated SNPs in CD.** QQ plot of CD associated SNPs ($p<0.001$) in RA (green) and RA associated SNPs ($p<0.001$) in CD (black). The most strongly associated SNPs (after removing known risk loci) in one disease were further filtered for $P<0.001$, and the resulting LD-pruned SNP sets were then tested for their distribution of association in the other disease. The QQ-plots indicate excess sharing of moderately associated SNPs across CD and RA.
doi:10.1371/journal.pgen.1002004.g002

that overlapped with our replication samples [2,3], and are included here for completeness. Two SNPs – rs7283760 in the CD-established *ICOSLG* locus and rs2181622 in the RA-established *PRKCQ* locus – were not genotyped in the replication samples for technical reasons. We did not attempt replication of SNPs from the five established loci associated with risk of both CD and RA. We conducted association tests of the 14 SNPs in the replication and combined cohorts with inverse variance weighted meta-analysis, where we analyzed CD-only samples [replication ($P_{\text{CD-repl}}$) and GWAS+replication ($P_{\text{CD}}$)], RA-only samples [replication ($P_{\text{RA-repl}}$) and GWAS+replication ($P_{\text{RA}}$)], and RA+CD samples [all GWAS+replication samples together ($P_{\text{overall}}$)].

As shown in Table 2, of the 4 established CD risk SNPs, two replicated in the RA samples with $P_{\text{RA-repl}}<0.05$ and obtained $P_{\text{RA}}<0.001$ in all available RA case-control samples (*SH2B3* (12q24.1) and an intergenic region on 8q24.2, $P_{\text{RA}}=1.5\times10^{-5}$ and $9.1\times10^{-5}$ respectively). Similarly, of the 3 established RA risk SNPs tested in our study, two replicated in the CD samples with $P_{\text{CD-rep}}<0.05$ and obtained $P_{\text{CD}}<0.001$ in all available CD case-control samples (*STAT4* (2q32.3) and *TRAF1-C5* (9q33.2), $P_{CD}=9.7\times10^{-4}$ and $9.3\times10^{-4}$ respectively). All four of these SNPs have $P_{\text{overall}}<5\times10^{-8}$ in analysis of all 50,266 CD and RA samples.

Of the 7 SNPs not previously established as genome-wide significant in either CD or RA, four were significantly replicated in both diseases at $P_{\text{CD-repl}}<0.05$ and $P_{\text{RA-repl}}<0.05$, were associated to each disease with $P_{\text{CD}}<0.001$ and $P_{\text{RA}}<0.001$ and achieved $P_{\text{overall}}<5\times10^{-8}$ in the combined CD-RA cohort (*CD247* (1q24.2), *UBE2L3* (22q11.2), *DDX6* (11q23.3) and *UBASH3A* (21q22.3); see Table 2). The strongest signal in the combined analysis was observed from the *DDX6* locus (rs10892279, $P_{\text{overall}}=1.2\times10^{-12}$). This SNP achieved genome-wide significance $P_{\text{RA}}=1.1\times10^{-8}$ in the RA cohort alone, and $P_{\text{CD}}=2.0\times10^{-5}$ in the CD cohort. SNPs near *CD247* and *UBE2L3* were previously suggestively associated in both CD and RA [2,3]. The replication data presented here, together with the combined analysis of $P_{\text{overall}}<5\times10^{-8}$, demonstrate that these SNPs are indeed true positive associations for CD and RA. Of note, SNPs in the *UBE2L3* are also associated with risk of systemic lupus erythematosus [21] and Crohn's disease [22], and the *CD247* locus is associated with systemic sclerosis [23].

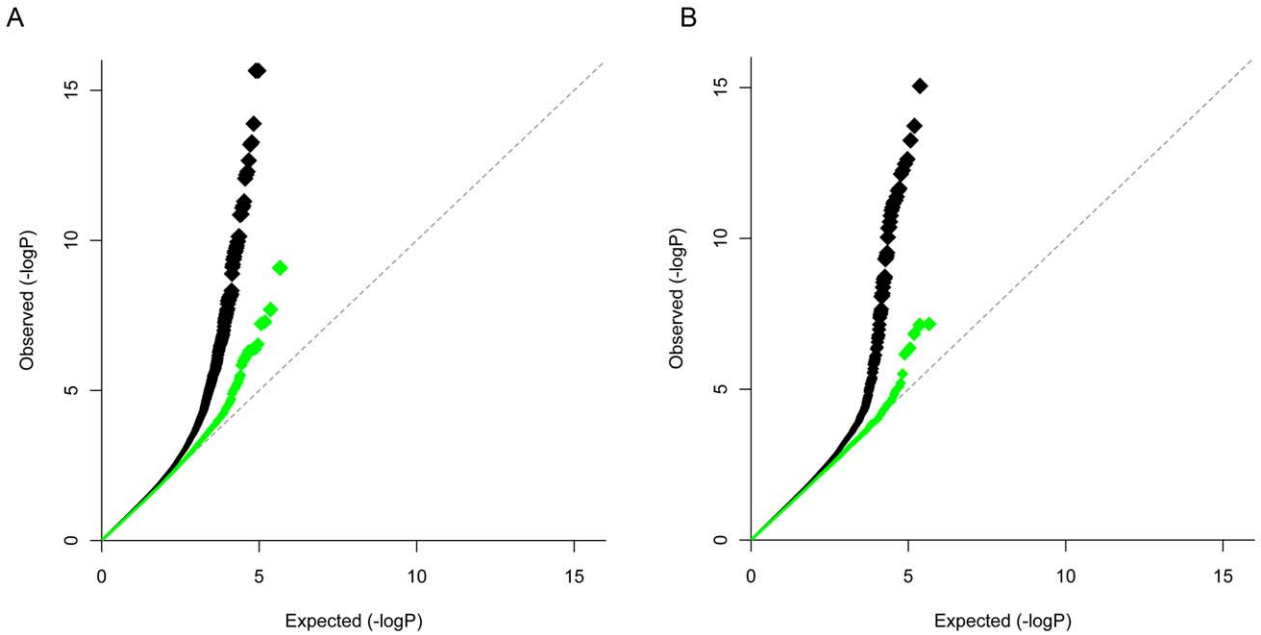


**Figure 3. QQ plot of CD-RA meta-analysis by directional method and opposite allelic effect.** CD-RA inverse variance weighted meta-analysis assuming allelic effects in the same direction in the two diseases (panel A) and opposite allelic effects (panel B). Black – all loci except the MHC region (chr. 6: 20–40 Mb). Green – all loci except MHC and established CD and RA regions (1 MB around previously validated SNPs excluded).
doi:10.1371/journal.pgen.1002004.g003

**Table 2.** CD-RA meta-analysis and replication, directional analysis.

| | | | | | | CD_GWAS | | CD_Repl | | CD_GWAS + Repl | | RA_GWAS | | RA_Repl | | RA_GWAS + Repl | | CD-RA GWAS+Repl | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ChrCHR** | **SNP** | **Min All** | **Locus** | **Status was** | **CD-RA P GWAS** | **CD P GWAS** | **CD OR GWAS** | **CD Repl*** | **CD OR Repl** | **CD P Meta** | **CD OR Meta** | **RA P GWAS** | **RA OR GWAS** | **RA Repl*** | **RA OR Repl** | **RA P Meta** | **RA OR Meta** | **CD-RA *P* META MAX [H]** | **Status now** |
| **Loci previously associated with celiac disease at $P<5\times10\text{-}8$ [G]** | | | | | | | | | | | | | | | | | | | |
| 12q24.1 | rs653178 | C | *SH2B3* | CD | 1.3E-11 | 8.2E-13 | 1.21 | **7.3E-09[A]** | 1.18 | **1.1E-18** | 1.19 | 8.8E-03 | 1.07 | **2.5E-04[C]** | 1.09 | **1.5E-05** | 1.08 | **3.0E-19** | CDRA |
| 8q24.2 | rs975730 | A | *8q24.2* | CD | 4.5E-06 | 1.6E-04 | 0.9 | **0.048[B]** | 0.93 | **4.6E-05** | 0.91 | 5.3E-03 | 0.93 | **2.3E-03[D]** | 0.91 | **9.1E-05** | 0.92 | **1.9E-08** | CDRA |
| 18p11.2 | rs1893217 | G | *PTPN2* | CD | 4.3E-10 | 1.8E-06 | 1.18 | 5.2E-05[A] | 1.17 | 7.9E-10 | 1.17 | 4.4E-05 | 1.14 | 0.075[C] | 1.05 | 1.2E-04 | 1.09 | 4.6E-12 | CD |
| 16p13,1 | rs243323 | G | *SOCS1* | CD | 6.6E-06 | 4.1E-04 | 0.9 | 0.013 [B] | 0.90 | 3.0E-05 | 0.9 | 3.9E-03 | 0.93 | 0.36[D] | 1.01 | 0.032 | 0.95 | 1.4E-05 | CD |
| **Loci previously associated with rheumatoid arthritis at $P<5\times10\text{-}8$ [G]** | | | | | | | | | | | | | | | | | | | |
| 9q33.2 | rs1953126 | T | *TRAF1* | RA | 2.8E-06 | 9.5E-03 | 1.08 | **0.017[B]** | 1.10 | **9.3E-04** | 1.08 | 7.2E-05 | 1.1 | **3.4E-06[D]** | 1.17 | **5.0E-09** | 1.13 | **4.2E-11** | CDRA |
| 2q32.3 | rs7574865 | T | *STAT4* | RA | 4.0E-08 | 7.7E-03 | 1.09 | **0.025[B]** | 1.10 | **9.7E-04** | 1.09 | 5.1E-07 | 1.16 | **9.7E-03[D]** | 1.10 | **5.8E-08** | 1.14 | **4.1E-10** | CDRA |
| 5q11.2 | rs1020388 | G | *ANKRD55* | RA | 7.7E-09 | 1.5E-04 | 0.9 | 0.32 [A] | 0.99 | 2.1E-03 | 0.94 | 1.3E-05 | 0.9 | NA | NA | NA | NA | 3.1E-07 | RA |
| **Loci previously not established in celiac disease or rheumatoid arthritis** | | | | | | | | | | | | | | | | | | | |
| 11q23.3 | rs10892279 | A | *DDX6* | none | 2.0E-07 | 5.3E-03 | 0.91 | **2.1E-04[B]** | 0.83 | **2.0E-05** | 0.89 | 7.4E-06 | 0.87 | **2.0E-04[D]** | 0.86 | **1.1E-08** | 0.87 | **1.2E-12** | CDRA |
| 1q24.2 | rs864537 | G | *CD247* | none | 2.6E-10 | 3.8E-07 | 0.87 | **0.046[A]** | 0.95 | **1.4E-06** | 0.91 | 8.9E-05 | 0.9 | **5.9E-03[C]** | 0.92 | **3.6E-06** | 0.91 | **2.2E-11** | CDRA |
| 22q11.2 | rs2298428 | T | *UBE2L3* | none | 6.8E-09 | 8.7E-07 | 1.18 | **0.02[A]** | 1.08 | **5.5E-07** | 1.13 | 8.2E-04 | 1.11 | **9.7E-03[C]** | 1.07 | **5.9E-05** | 1.09 | **2.5E-10** | CDRA |
| 21q22.3 | rs11203203 | A | *UBASH3A* | none | 3.1E-07 | 2.0E-03 | 1.1 | **0.036[B]** | 1.08 | **3.6E-04** | 1.10 | 4.2E-05 | 1.11 | **0.03[E]** | 1.08 | **7.8E-06** | 1.10 | **1.1E-08** | CDRA |
| 7p14.1 | rs11984075 | G | *ELMO1* | none | 3.1E-07 | 2.0E-03 | 1.15 | 0.052[B] | 1.12 | 5.2E-04 | 1.14 | 3.9E-05 | 1.19 | 0.07[F] | 1.09 | 2.7E-05 | 1.15 | 5.2E-08 | none |
| 12q14.1 | rs10876993 | C | *CDK4* | none | 9.9E-06 | 1.3E-03 | 0.91 | 0.062[B] | 0.93 | 4.0E-04 | 0.92 | 2.3E-03 | 0.92 | 0.19[F] | 0.97 | 2.4E-03 | 0.94 | 3.7E-06 | none |
| 10q26.1 | rs6585827 | A | *HTRA1* | none | 6.4E-07 | 9.9E-04 | 1.09 | 0.20[B] | 1.04 | 1.2E-03 | 1.08 | 1.90E-04 | 1.09 | 0.31[D] | 0.98 | 6.0E-03 | 1.06 | 2.7E-05 | none |

Directional meta-analysis of CD and RA. CD-RA GWAS meta-analysis, GWAS results for CD and RA, replication and combined meta-analysis results for loci that achieve $P<1\times10^{-5}$ in CD-RA meta-analysis and $P<0.01$ in GWAS in CD and RA risk (one SNP per locus). Listed are the rs ID for each SNP, the chromosome, one candidate gene in the region (selected by GRAIL analysis, expression analysis or manually based on immunological function; see text and Tables S8, S9, S10 and Table 4), minor alleles (column MinAll) (positive strand in HapMap release 22, major/minor based on frequency in CD GWAS controls). The association *P*-value and odds ratio (OR) with respect to minor allele are listed for CD and RA GWAS, CD-RA GWAS meta-analysis, replication analyses in each disease and meta-analysis of CD-RA GWAS and replication cohorts. Loci established in both diseases are not included in this table. A – replication for this SNP in CD was done in Dubois et al paper [2]; B – replication in current study with all available CD samples as described in Materials and Methods; C - replication for this SNP in RA was done in Stahl et al paper [3]; D - replication in current study with all available RA samples, as described in methods; E – replication in current study with RA replication cohort R1, R2, R3 and R5 (failed in R4); F – replication in current study with RA replication cohorts R1, R2, R3 and R4 (failed in R5) (see Table S11 for cohort description); G – *ICOSLG* (established CD locus) and *PRKCQ* (established RA locus) SNPs were not included to the replication and not mentioned in this table. H - The CD-RA P META MAX column shows the P-overall for SNP tested in replication in both CD and RA. For one SNP (*ANKRD55*) the replication results were available only in CD replication cohort, in this case the CD-RA P META MAX column shows the p-value for the meta-analysis of CD-RA GWAS and CD-replication. P-values are 2-tailed except for replication p-values in CD and RA, indicated by *.

doi:10.1371/journal.pgen.1002004.t002

### GWAS and replication—same allele, opposite direction

There is increasing evidence that alleles conferring risk of one autoimmune disease confer protection to another autoimmune disease [3,7,8,24,25,26]. We therefore performed an analysis of alleles that conferred risk in either CD or RA, but protection in the other disease (Figure 3B), followed by independent testing in our replication cohort. Nine loci were identified using the same criteria as above ($P_{combined}<1\times10^{-5}$, and disease-specific $P<0.01$; see Tables S6 and S7). The strongest shared signal from this analysis was at the *TAGAP* locus (6q25.3, rs212388 $P_{combined}=5.4\times10^{-12}$), an established risk locus in both CD and RA [2,3,8] (Chen, et al, submitted). Another locus that had an apparent opposite allelic effect was *REL* (2p16.1), although it shows a more complex pattern of association. From the three SNPs in the *REL* locus that were associated to both diseases with $P_{combined}<1\times10^{-5}$, and disease-specific $P<0.01$, two SNPs showed similar direction of association with CD and RA, whereas one SNP showed opposite directionality of association (Tables S4 and Table S6). Of the remaining SNPs, no single SNP replicated in both diseases at $P<0.05$ and achieved $P<5\times10^{-8}$ in an overall analysis of all data. We observed a trend of an association at the chromosome 2p23.1 (near the *LBH* gene) locus (rs7579944, $P_{CD}=9.7\times10^{-6}$ and $P_{RA}=2.3\times10^{-4}$ in the CD and RA cohorts, respectively; $P_{overall}=1.1\times10^{-8}$ in the combined analysis, but no formal replication in RA cohort ($P_{RA\text{-}repl}=0.13$)) (Table 3). Although these data strongly suggest that chromosome 2p23.1 is a shared CD-RA risk locus, additional replication will be required.

### Selecting most likely causal gene near associated SNP

We used two methods to identify the most likely causal gene in the region of the 14 shared non-HLA risk loci. First, we used a computational algorithm, GRAIL, which systematically searches for gene relationships across risk loci using PubMed abstracts [27]. In total, 14 shared loci contain 51 genes; 16 of these scored $P<0.1$ by GRAIL (Table S8). Second, we analyzed each shared SNP for evidence of cis-acting gene expression in peripheral blood cells derived from 1,469 individuals (Fehrmann et al, submitted). From 14 shared SNPs, 7 showed a significant (genome-wide FDR corrected $<0.05$) effect on expression of one or more transcripts in the LD block around the SNP (Table S9, Figure S2A-S2P). It is interesting to note that of the four novel SNP associations identified from this study, three show convincing effects on the expression of nearby genes, in particular rs864537-*CD247* ($P=3.5\times10^{-11}$), rs2298428-*UBE2L3* ($P=2.0\times10^{-99}$) and rs11203203-*UBASH3A* ($P=8.7\times10^{-10}$) (Table S9, Figure S2). Based on these two methods, 23 genes located in the 14 shared loci were selected as plausible candidates for shared CD-RA pathogenesis (Table 4 and Table S10).

## Discussion

In this study we demonstrate that there are 14 loci that contribute to risk of both RA and CD: 6 previously established risk loci and 8 loci identified in our study. Of the 8 new loci, 4 had not been associated previously with either disease at genome-wide significance (*CD247*, *UBE2L3*, *DDX6*, and *UBASH3A*) and 4 had been established in one but not the other autoimmune disease (*SH2B3*, *8q24.2*, *STAT4*, and *TRAF1-C5*). Our study represents the first systematic effort to compare the genetic basis of CD and RA in a very large sample set – more than 50,000 combined case-control samples – to identify risk alleles with pleiotropic effects on two clinically distinct autoimmune diseases.

To identify the shared risk loci, we performed two types of analyses. First, we compared the distribution of established and putative risk alleles across both autoimmune diseases. Both distributions were non-random, providing empirical evidence that the genetic basis of the two autoimmune diseases overlaps. Second, we combined GWAS data and performed independent replication to search for specific SNPs associated with both diseases. We performed the GWAS meta-analysis under a genetic model in which the same allele conferred risk of both autoimmune diseases, as well as a model in which the same allele conferred risk to one disease and protection from the other disease. Of the newly identified 8 shared risk alleles, all 8 confer same risk direction on both CD and RA.

Our study represents one of the first GWAS meta-analysis of clinically distinct but epidemiologically related diseases. This approach has appeal for diseases in which there is thought to be a shared genetic basis, as it adds power to detect alleles of modest effect size. A GWAS meta-analysis has been conducted on early onset inflammatory bowel disease (IBD), which include Crohn's disease (CrD) and ulcerative colitis (UC) [15]. CrD and UC are clinically similar diseases both affecting bowel, and often can not be distinguished between each other (presented as undifferentiated IBD), especially in children. In contrast to the IBD study, our GWAS meta-analysis combined phenotypes with different clinical presentations (enteropathy and inflammatory arthritis).

In combining GWAS data across clinically distinct phenotypes, an important question is how to interpret statistical significance and therefore how to declare a SNP as a confirmed association for each disease. In our study, we empirically demonstrated that SNPs associated with risk of either CD or RA have a higher probability of being associated with the other autoimmune – even if the SNP is not yet a confirmed association in either disease (Figure 2). We observed that a SNP associated with risk of CD at $P<0.001$ has an increased prior probability of being associated with RA, and a SNP associated with risk of RA at $P<0.001$ has an increased prior probability of being associated with CD. Based upon these analyses, we propose objective criteria for declaring a SNP as a shared CD – RA risk SNP in our study: it must achieve $P_{overall}<5\times10^{-8}$ in combined analysis of CD&RA, with the additional requirement of $P<0.05$ in an independent replication dataset and $P<0.001$ for each disease. Applying these criteria to our meta-analysis results we conclude that there are 14 non-HLA shared CD and RA risk loci (Table 1 and Table 2).

We applied two methods to select the most likely causal gene in the region of the 14 shared non-HLA risk loci, and in doing so gain insight into shared RA-CD pathogenesis: (1) a computational algorithm, GRAIL, which systematically searches for gene relationships across risk loci using PubMed abstracts [27] and (2) a dataset of cis-acting gene expression in peripheral blood cells derived from 1,469 individuals [27] (Fehrmann et al, submitted). Using these methods we prioritized 23 genes located in the 14 shared loci as plausible functional candidates. Interestingly, two out of four novel loci function in T-cell activation/signalling: *CD247*, which encodes for the zeta chain of the T-cell receptor-CD3 complex, and *UBASH3A*, which is a suppressor of T-cell receptor signaling, underscoring antigen presentation to T-cells as a critical shared mechanism of disease pathogenesis [28,29]. This observation is consistent with the known functions of several of the other shared RA-CD risk loci which were highlighted in GRAIL and expression analysis (*CTLA4*, *ICOS*, *TAGAP*, *SH2B3*, and *STAT4*). These genes are known to modulate T-cell activation and/or differentiation: *CTLA4* is a negative regulator of T-cell activation [30], *ICOS* is a T-cell co-stimulator molecule [31], *TAGAP* is up-regulated upon T-cell activation[32], *SH2B3 (LNK)* is an adaptor protein involved in T-cell activation [33], and *STAT4*

**Table 3.** CD-RA meta-analysis GWAS and replication—opposite allelic effect method.

| | | | | | | CD GWAS | | CD Repl | | CD GWAS + Repl | | RA GWAS | | RA Repl | | RA GWAS + Repl | | CD+RA GWAS+Repl | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CHR** | **SNP** | **Min All** | **locus** | **STATUS WAS** | **CD-RA GWAS** | **CD P GWAS** | **CD OR GWAS** | **CD Repl*** | **CD OR Repl** | **CD P Meta** | **CD OR Meta** | **RA P GWAS** | **RA OR GWAS** | **RA Repl*** | **RA OR Repl** | **RA P Meta** | **RA OR Meta** | **CD-RA P Meta MAX [F]** | **STATUS NOW** |
| **Loci previously associated with rheumatoid arthritis at $P<5\times10\text{-}8$** | | | | | | | | | | | | | | | | | | | |
| 2p14 | rs1876518 | T | *SPRED2* | RA | 2.2E-08 | 4.2E-03 | 0.93 | 0.18[A] | 1.04 | 5.4E-02 | 0.96 | 7.7E-07 | 1.13 | 1.4E-03[B] | 1.11 | 8.5E-09 | 1.12 | 1.9E-08 | RA |
| **Loci previously not established in celiac disease or rheumatoid arthritis [E]** | | | | | | | | | | | | | | | | | | | |
| 2p23.1 | rs7579944 | T | *LBH* | none | 3.2E-08 | 6.5E-05 | 1.12 | 0.025 [A] | 1.09 | 9.7E-06 | 1.11 | 1.3E-04 | 0.90 | 0.13 [B] | 0.96 | 2.2E-04 | 0.92 | 1.1E-08 | none |
| 1q23.1 | rs1772408 | A | *IFI16* | none | 5.0E-06 | 2.0E-04 | 1.15 | 0.016 [A] | 1.12 | 2.0E-05 | 1.14 | 4.5E-03 | 0.91 | 0.20[D] | 0.96 | 4.6E-03 | 0.93 | 8.4E-07 | none |
| 7q34 | rs4626515 | C | *JHDM1D* | none | 4.6E-07 | 3.2E-03 | 0.88 | 0.12 [A] | 1.08 | 0.070 | 0.93 | 4.1E-05 | 1.17 | 0.17 [B] | 1.06 | 7.9E-05 | 1.13 | 2.8E-05 | none |
| 11p15 | rs11043097 | C | *GALNTL4* | none | 2.0E-07 | 6.1E-05 | 1.15 | 0.45 [A] | 0.99 | 8.7E-04 | 1.10 | 7.5E-04 | 0.90 | 0.23[C] | 1.04 | 0.010 | 0.93 | 3.3E-05 | none |
| 14q13.2 | rs17103033 | C | *PPP2R3C* | none | 6.8E-06 | 2.8E-03 | 1.12 | 0.27 [A] | 0.97 | 0.032 | 1.07 | 7.6E-04 | 0.89 | 0.42 [D] | 0.99 | 3.8E-03 | 0.92 | 3.4E-04 | none |

Opposite allelic effect meta-analysis of CD and RA. CD-RA GWAS meta-analysis, GWAS results for CD and RA, replication and combined meta-analysis results for loci that achieve $P<1\times10\text{-}5$ in CD-RA meta-analysis and $P<0.01$ in GWAS in CD and RA risk (one SNP per loci). Listed are the rs ID for each SNP, the chromosomal location, one candidate gene in the region (selected by GRAIL analysis, expression analysis or manually based on immunological function; see text and Tables S8, S9, S10 and Table 4), minor alleles (column MinAll) (positive strand in HapMap release 22, major/minor based on frequency in CD GWAS controls). The association *P*-value and odds ratio (OR) with respect to minor allele are listed for CD and RA GWAS, CD-RA-meta analysis on GWAS, replication analyses in each disease and meta-analysis of CD-RA GWAS and replication cohorts. A – replication in current study with all available CD samples, as described in Materials and Methods; B - replication in current study with all available RA samples; C – replication in current study with RA replication cohort R1, R2 and R3 (failed in R4 and R5); D – replication in current study with RA replication cohorts R1, R2, R3 and R4 (failed in R5), see Table S11 for cohort description; E - The *TGFBR3* SNP fit the replication criteria but was not included to the replication step for technical reasons. F - The CD-RA P META MAX column shows the P-overall for SNP included to the replication in both CD and RA. Rs19126261 in *TGFBR3* locus was not included to the replication step; in this case the CD-RA P META MAX column shows the p-value for GWAS meta-analysis. P-values are 2-tailed except for replication p-values in CD and RA, indicated by *.

doi:10.1371/journal.pgen.1002004.t003

**Table 4.** Analysis of candidate genes within associated blocks.

| Shared SNP | Chr | Block_start | Block_end | GRAIL annotated genes in block | GRAIL selected genes, p<0.1 | Cis eQTL transcripts in LD block | Likely candidate gene for CDRA based on GRAIL and eQTL analysis | Evidence |
|---|---|---|---|---|---|---|---|---|
| rs3748816 | 1p36.3 | 2396198 | 2742203 | *PANK4, MMEL1, PLCH2, C1orf93, HES5, TNFRSF14* | *TNFRSF14* | *MMEL1, PLCH2, TNFRSF14, C1orf93* | *MMEL1, TNFRSF14, PLCH2, C1orf93* | GRAIL, eQTL |
| rs864537 | 1q24.2 | 165663466 | 165707466 | *CD247* | *CD247* | *CD247* | *CD247* | GRAIL, single gene |
| rs6706689 | 2p16.1 | 60920853 | 61033853 | *REL, PUS10* | *REL* | none | *REL* | GRAIL |
| rs7574865 | 2q32.3 | 191300239 | 191681739 | *STAT1, GLS, STAT4* | *STAT4, STAT1* | none | *STAT4, STAT1* | GRAIL |
| rs231804 | 2q33.2 | 204397239 | 204524239 | *ICOS, CTLA4* | *ICOS, CTLA4* | none | *ICOS, CTLA4* | GRAIL |
| rs13151961 | 4q27 | 123202845 | 123784345 | *IL2, IL21, ADAD1, KIAA1109* | *IL2, IL21* | none | *IL2, IL21* | GRAIL |
| rs6933404 | 6q23.3 | 137750000 | 138315000 | *OLIG3, TNFAIP3* | *TNFAIP3* | none | *TNFAIP3* | GRAIL |
| rs212388 | 6q25.3 | 159391579 | 159435579 | *RSPH3, EZR, TAGAP, FNDC1* | *TAGAP* | *TAGAP* | *TAGAP* | GRAIL, eQTL |
| rs975730 | 8q24.2 | 129129000 | 129812500 | *PVT1* | none | none | *PVT1* | single gene |
| rs1953126 | 9q33.3 | 122674767 | 123013267 | *PHF19, CEP110, GSN, TRAF1, RAB14, C5* | *TRAF1, C5* | *TRAF1, PHF19, C5,* | *TRAF1, PHF19, C5,* | GRAIL, eQTL |
| rs10892279 | 11q23.3 | 118080000 | 118252500 | *DDX6* | none | none | *DDX6* | single gene |
| rs653178 | 12q24.1 | 110322163 | 111513663 | *ATXN2, SH2B3, C12orf30, RPL6, ERP29, TRAFD1, PTPN11, ACAD10, BRAP, MAPKAPK5, C12orf51, ALDH2* | *TRAFD1, SH2B3* | *SH2B3/ATXN2, ALDH2* | *TRAFD1, SH2B3* | GRAIL, eQTL |
| rs11203203 | 21q22.3 | 42681500 | 42750500 | *UBASH3A, TMPRSS3* | *UBASH3A* | *UBASH3A, TMPRSS3* | *UBASH3A* | eQTL |
| rs2298428 | 22q11.2 | 19786946 | 20317446 | *CCDC116, UBE2L3, PI4KAP2, HIC2, LOC150223* | none | *UBE2L3* | *UBE2L3* | eQTL |

All chromosomal positions are based on NCBI build-36 coordinates. Grail selected genes column included genes with p-value P<0.1 as annotate by GRAIL, see Table S8. Cis eQTLs indicate genes that showed correlation in expression with associated SNP, see Table S9 and Figure S2.

doi:10.1371/journal.pgen.1002004.t004

is transcription factor important in differentiation of T helper cells [34].

How might these 14 loci influence risk of two clinically distinct autoimmune diseases? MHC class II alleles, the strongest risk factor in both diseases, are notably different between the two diseases: *HLA-DQ*A1* and **B1* alleles in CD and *HLA-DRB1* "shared epitope" alleles in RA. Under a model in which MHC class II molecules confer risk by preferentially presenting disease-specific antigens (gluten in CD, most likely citrullinated antigens in RA) to autoreactive T-cells, then disease specificity is determined in large part by the inheritance of specific HLA alleles and exposure to disease-specific antigens. Our genetic data extends this model to implicate downstream signaling events common to both diseases that may lead to altered T-cell activation and differentiation. Whether abnormal T-cell signaling occurs in the thymus (where autoreactive T-cells undergo negative selection), in the peripheral circulation (where autoreactive T-cells exert their effects), or in another manner remains to be determined.

There are several limitations of our study. First, we did not search for loci in which an allele contributes to risk of one autoimmune disease and an independent allele contributes to risk of the other autoimmune disease. The *REL* locus provides an example in which the risk alleles for the two autoimmune diseases appear distinct [2,5,6]. Second, our study is underpowered to detect shared risk alleles of more modest effect size, despite a combined sample size of >50,000 case-control samples. As more samples and SNPs are genotyped between these diseases, additional risk alleles will be discovered. Third, we did not attempt to fine-map the 26 established risk loci for both autoimmune diseases to determine if a single allele is responsible for risk in both autoimmune diseases. And fourth, we made no attempt to search for low-frequency or rare variants that are shared between RA and CD. Implementation of newer sequencing technologies will be required to search for rare risk variants.

In summary, our study adds four novel loci to established RA and CD risk loci (*CD247*, *UBE2L3*, *DDX6*, and *UBASH3A*). It also adds four loci previously established in one or the other disease to the list of shared CD-RA risk loci (*SH2B3*, *8q24.2*, *STAT4*, and *TRAF1-C5*). With six previously established CD-RA risk loci, there are now 14 shared CD-RA risk loci, out of 50 established loci for either of the two autoimmune diseases. We emphasize that these are conservative estimates of shared risk loci between the two diseases, as our study may be underpowered to detect common alleles of modest effect size, and we have not considered genetic models in which different alleles within one locus contribute to risk of the two diseases. In addition to the HLA associations, these shared risk loci clearly point to the critical role of antigen presentation via MHC class II molecules to the T-cell receptor, and subsequent activation and differentiation of T-cells in shared disease pathogenesis.

## Materials and Methods

### Ethics statement

Institutional review boards at each collection site approved the study, and all individuals gave their informed consent.

### Sample collection

**CD GWAS dataset.** CD case-control GWA study included 15,283 individuals (4533 cases, 10750 controls) from 5 populations: Finnish (FIN) (674 cases, 647 controls), Italian (IT) (541 cases, 497 controls), Dutch (NL, 876 cases, 803 controls), and two collections from UK population, UK1 (737 cases, 2596 controls) and UK3 (1922 cases, 1849 controls) (described in details in[2]).The genotyping of all cohorts except UK1 cases was done on Illumina platforms including 550K SNPs (either Illumina Hap550, or Illumina 610 or 670 Quad, or Illumina 1.2 M). The genotyping of UK1 cases (n = 737) was done on Illumina 317K arrays. The subset of SNPs successfully genotyped on Illumina 550 and Quadr platforms, but not on Hap300 platform (n = 196860) was further imputed in the UK1 dataset, using Plink and HapMap Phase 2 European CEU founders as a reference panel [35].

**RA GWAS dataset.** The RA meta-analysis includes 5,539 autoantibody positive RA cases and 17,231 controls of European ancestry as described previously [3]. This study comprises six GWAS case-control collections, genotyped on various platforms. The imputation was conducted on GWAS genotype data for each GWAS collection separately, using the IMPUTE software [36] and haplotype-phased HapMap Phase 2 European CEU founders as a reference panel. In total, 2.56 million SNPs were imputed. Identity by state (IBS) analysis was run on controls from both CD and RA GWAS datasets. The overlapping controls genotyped in both CD and RA datasets were excluded from the RA analysis.

**Replication cohorts.** The replication cohorts included 2,169 CD cases and 2,255 controls, and 2,845 antibody-positive RA cases and 4,944 controls. The CD replication cohorts included three case-control collections from Ireland, Italy and Poland; all collections were geographically matched and are described previously [2]. The five RA replication collections included (1) CCP or RF positive Dutch cases from Groningen and Nijmegen, together with geographically matched controls (Replication cohort 1, R1); (2) CCP positive white individuals from North America (Replication cohort 2, R2; this collection is called i2b2); (3) North American RF positive cases and controls matched on gender, age, and grandparental country of origin from the Genomics Collaborative Initiative (GCI, Replication cohort 3, R3); (4) CCP or RF positive Dutch cases and controls from Leiden University Medical Center (LUMC; Replication cohort 4, R4); and (5) CCP positive cases drawn from North American clinics and controls from the New York Cancer Project (together this collection is called NARAC-II), matched on ancestry informative markers data (Replication cohort 5, R5). All cohorts except i2b2 were described in detail in [3], whereas i2b2 is described in [37]. Summary information on these samples is presented in Table S11.

### Genotyping

Replication analysis of 15 SNPs was performed on the Sequenom iPlex platform in three centers – (1) Broad institute (all CD cases and controls, and RA replication cohorts R1 and R2); (2) Celera Diagnostics (Alameda California, USA; RA replication cohort R3 and R4); and (3) National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS, RA replication cohort R5). (See Table S11 for details). If the SNPs could not be designed into the iPLEX pool, then a proxy SNP was included. Information on the iPLEX design, proxies and cohorts genotyped in different centers is presented in Tables S11 and S12. We excluded SNPs in each replication collection if they were missing >10% genotype data, <1% MAF and $P_{HWE}<10^{-3}$. For 5 out of the 20 SNPs that satisfied the replication criteria in either the directional or opposite allelic effect analysis, replication results were already available for CD and RA samples from the studies Dubois et al [2] and Stahl et al[3], respectively. For these 5 SNPs, we included genotype data from all replication samples available in these studies.

### Data analysis

**GWAS meta-analysis.** The meta-analysis of CD and RA datasets was performed using an inverse variance-weighted method. Analysis was performed in *R* package as described

previously. [20,38] To detect associations of the same and opposite directions, two tests were performed. First, the directional meta-analysis was done; second, the direction of association was flipped in RA dataset and an opposite-allelic effect analysis was performed. In total, 477,662 SNPs either directly genotyped on the Illumina Hap550 platform, or genotyped on Hap300 and imputed in UK1 cases were included in the analysis; all SNPs overlapped with genotype or imputed SNPs from the RA dataset. In both diseases the genomic control corrected results were used for the meta-analysis.

**Replication analysis.** Replication and combined analyses were done with an inverse variance-weighted method. Combined (GWAS + replication) analysis within one disease (CD or RA) was done with a directional method. Replication association tests were one-tailed, for the same allele being risk or protective as in the GWAS meta-analysis. Combined analysis of all CD and RA samples was done for the same (directional or opposite) allelic effect as was estimated in GWAS meta-analysis.

**Distribution of risk alleles in GWAS.** For the analysis of distribution of risk alleles, we excluded SNPs located within 1 MB around each of the most associated SNPs (26 in each disease); for the MHC and PTPN22 loci, we excluded 20 Mb and 2 Mb, respectively (chr. 6: 20–40 Mb and chr. 1: 113–115 Mb). The pruning of SNPs in linkage disequilibrium (LD) was done by selecting SNPs to retain and then removing all SNPs with $r^2>0.1$ in the HapMap2 reference panel. Pairwise LD tables were generated from the HapMap2 release 24 phased haplotype data distributed with the IMPUTE software; $r^2$ values were calculated for all SNPs within 1 Mb of each other. For a given analysis, the most strongly associated SNPs (after removing 1 Mb around known associated SNPs) in one disease were retained. We also filtered for SNPs with $P<0.001$. The resulting LD-pruned SNP sets were then tested for non-random distributions of association in the other disease. Fisher's rule for combining *P*-values ($-2\sum_{i=1}^{n}\ln(P_i)\tilde{\chi}^2_{2n}$) ($-2\sum\ln(\mathrm{P})\sim\chi^2_{2n}$) was used to test the null hypothesis of a uniform distribution of *P*-values for association with a given disease. Kolmogorov-Smirnov and Wilcoxon rank sum tests were performed to test for overall difference and difference in location, respectively, of the distributions of *P*-values in a given disease, for SNPs with $P<0.001$ *versus* $P\geq0.001$ in the other disease. One-sided tests were conducted, with the alternative hypothesis that SNPs associated with one disease would also show evidence of association with the other.

**Gene expression.** The analysis of gene expression was done on PBMC of 1,469 individuals, as previously described [2]. In this dataset, we included SNPs with genotyping call-rate ≥95%, Hardy-Weinberg *P*-value ≥0.001, and MAF ≥5%. Expression data was quantile normalized, centered to the mean and scaled such that all probes had a standard deviation of 1. Principal component analysis was performed over the sample correlation matrix, in order to capture non-genetic variation. The variation described within the first 50 principal components was subsequently subtracted from the expression data as described by Fehrmann et al (Fehrmann et al, submitted). Effects were deemed *cis*-effects, when the mid-probe to SNP distance was ≤250 kb. False discovery rate was controlled at 5%, by comparing observed p-values with p-values obtained after permuting sample labels 100 times. The Fehrmann et al manuscript specifically investigated whether SNPs in the Illumina probe sets might explain the eQTL results: eQTL associated SNPs were checked for LD with SNPs from 1000 genomes pilot data located within probe sequences. Specifically for our study, we verified that none of the eQTL associated SNPs was in high LD ($r^2>0.1$) with any of the 1000 genome SNPs located within Illumina probe sequences.

## Supporting Information

**Figure S1** Manhattan plots for GWAS analysis. Manhattan plot for GWAS in CD (A), RA (B), CD-RA with directional meta-analysis (C) and CD-RA with opposite allelic effect (D). Figure E is a combined picture of the four analyses. Yellow – CD; blue – RA; green – CD-RA directional; purple – CD-RA opposite allele. In figure E, the 14 shared CD-RA loci are annotated. The HLA locus (chr6: 20–40 MB) is excluded from the analysis. In RA, three SNPs in *PTPN22* locus are associated at p<10-24 and therefore are excluded from the plot for the purpose of scale. The *PTPN22* SNPs excluded from RA plot are: rs2476601 $P(\mathrm{RA})=3.6\times10^{-68}$, rs2358994 $P(\mathrm{RA})=8.2\times10^{-31}$, and rs1230661 $P(\mathrm{RA})=6.1\times10^{-25}$.
(PDF)

**Figure S2** *Cis* eQTL genotype – expression correlation analyses in associated SNPs. Individual level gene expression data (residual variance after Transcriptional Components removed) from 1469 PAXgene samples. Spearman Rank Correlation coefficients and *P* values are shown for HT-12 and Ref. 8 data and for meta-analysis results. Right Y axis, average expression rank, is a measure of how strongly the tested probe is expressed amongst all probes in the dataset. Unannotated probes are manually plotted and localized to the following transcripts: Probe 2810202 – *PHF19*, Probe 6980470 – *TMPRSS3*, Probe 1230242 – *UBE2L3*.
(PDF)

**Table S1** Established CD SNPs and their association to RA. Candidate genes in the blocks are mentioned. Top *P*-value in CD is indicated as in the reference paper [2]. CD_*P* column indicates the p-value in CD-GWAS dataset (4,533 cases and 10,750 controls); RA_*P* column indicates the p-value in RA-GWAS dataset (5,539 cases and 17,231 controls). OR is given for the minor allele.
(DOC)

**Table S2** Established RA SNPs and their association to CD. Candidate genes in the blocks are mentioned. Top *P*-value in RA is indicated as in the reference papers [3,4,5,10,11,39,40,41,42,43]. CD_*P* column indicates the p-value in CD-GWAS dataset; RA_*P* column indicates the p-value in RA-GWAS dataset; OR is given for the minor allele. The CD results for RA SNPs that were not genotyped in CD GWAS (not present on Illumina Hap550 genotyping array), are either imputed or estimated from the best genotyped proxy SNP (indicated in column "genotyped/imputed"). When proxies were used, the $r^2$ with RA SNP is indicated in column '$r^2$ for proxy'. For imputed SNPs the imputation score is annotated in column 'Imp. score'. * - the perfect proxy rs13017599 was genotyped in the reference paper. + - the association in *TNFRSF14* does not reach $P<5\times10^{-8}$, however this locus was included as RA-established locus as it was implicated in several independent studies [3,39,41].
(DOC)

**Table S3** Distribution of CD associated SNPs ($P<0.001$) in RA dataset and RA associated SNPs ($P<0.001$) in CD datasets. 3a. Goodness of fit tests of no association, using Fisher's Rule for combining *P*-values. 3b Results of Kolmogorov-Smirnov and Wilcoxon rank sum tests for non-random distribution of associated SNPs across diseases. N – number of SNPs associated to CD and RA at p<0.001 after LD-pruning. Df – degrees of freedom.
(DOC)

**Table S4** SNPs associated to CD-RA with $P<1\times10$-5 and CD and RA with $P<0.01$; directional analysis. The table includes all SNPs associated to CD-RA in directional GWAS meta-analysis

with *P*(combined)<1×10-5 (column 'CD-RA_*P* dir') and *P*-value per diseases *P*<0.01 (columns 'CD_*P*' and 'RA_*P*'). OR is given for the reference (minor) allele.
(DOC)

**Table S5** CD-RA meta-analysis results, directional analysis – one top SNP per loci. One most associated SNP per locus is indicated. OR is given for the reference (minor) allele. Cut-off for the *P*-values is the same as in Table S4. SNPs are sorted for the *P*-value in CDRA meta-analysis. Loci established in both diseases are indicated in bold.
(DOC)

**Table S6** SNPs associated to CD-RA with *P*<1×10-5 and CD and RA with *P*<0.01; analysis of opposite allelic effect. The table includes all SNPs associated to CD-RA in GWAS meta-analysis of opposite alleleic effect with *P*(combined) *P*<1×10-5 (column 'CD-RA_*P*_opp') and *P*-value per diseases *P*<0.01 (columns CD_*P* and RA_*P*).
(DOC)

**Table S7** CDRA meta-analysis results, opposite allelic effect – one top SNP per loci. One most associated SNP per locus is indicated. OR is given for the reference (minor) allele. Cut-off for the *P*-values is the same as in Table S6. SNPs are sorted for the *P*-value in CD-RA meta-analysis. Loci established in both diseases are indicated in bold.
(DOC)

**Table S8** GRAIL analysis of associated loci. The *P*-value for each candidate gene is based on the number of relationships to other associated genes listed in the third column. GRAIL is available at http://www.broadinstitute.org/mpg/grail/.
(DOC)

**Table S9** Shared CD-RA risk variants correlated with *cis* gene expression. 'HT-12' comprise 1240 individuals with blood gene expression assayed using Illumina Human HT-12v3 arrays, 'Ref-8v2' comprise 229 individuals with blood gene expression assayed using Illumina Human-Ref-8v2 arrays. [A]Spearman rank correlation of genotype and residual variance in transcript expression. Meta-analysis eQTL *P* value shown if both datasets had identical probes. See Figure S2 for detailed results and Materials and Methods for sample information and references.
(DOC)

**Table S10** Characteristics of 23 candidate genes in shared loci.
(DOC)

**Table S11** Information on replication cohorts. Case-control collections included to the replication step in CD (top panel) and RA (bottom panel). For each collection, we list the source of controls, geographic origin, autoantibody status of RA cases, numbers of cases and controls, genotyping platform and genotyping center, and the strategy used to correct for case-control population stratification. See Materials and Methods for additional details. NIAMS – National Institute of Arthritis and Musculoskeletal and Skin Diseases.
(DOC)

**Table S12** Information on proxies used in three iPLEX pools. $R^2$ - $r^2$ between GWAS and SNP pools used in replication step (HapMap CEU, as calculated in SNAP (http://www.broadinstitute.org/mpg/snap/)). A – iPLEX pool used in Broad institute (BI) was used to genotype all replication cohorts for celiac disease and replication cohorts R1 and R2 in rheumatoid arthritis, as indicated in Table S11.
(DOC)

## Acknowledgments

We thank all individuals with celiac disease, rheumatoid arthritis, and control individuals for participating in this study. We thank Broad Institute, the genotyping facilities of Groningen, and WTCCC for help with genotyping and WTCCC consortium for the access to CD GWAS results. We thank all clinicians for sample collections, P. Dubois, P. Saavalainen, D. Barisani, M. T. Bardella, D. A. van Heel, M. Barbato, M. Bonamico, D. Kelleher, R. Greco, R. H. J. Houwen, V. M. Wolters, M. L. Mearin, and C. J. Mulder for establishing the CD cases collection. We thank J. J. Catanese, J. Smolonska, N. Patsopoulos, and R. B. Marques for technical contribution.

## Author Contributions

Conceived and designed the experiments: RM Plenge, C Wijmenga, A Zhernakova, EA Stahl, S Raychaudhuri. Performed the experiments: A Zhernakova, G Trynka, FAS Kurreeman, B Thomson, N Gupta, J Romanos, EF Remmers, Y Li. Analyzed the data: A Zhernakova, EA Stahl, RM Plenge, S Raychaudhuri, L Franke, RSN Fehrmann, EA Festen, FAS Kurreeman, PIW de Bakker, HJ Westra. Contributed reagents/materials/analysis tools: RM Plenge, C Wijmenga, A Zhernakova, R McManus, AW Ryan, G Turner, PK Gregersen, J Worthington, KA Siminovitch, L Klareskog, TWJ Huizinga, EF Remmers, F Tucci, R Toes, E Grandone, MC Mazzilli, A Rybak, B Cukrowska, E Brouwer, MD Posthumus, MJH Coenen, TRDJ Radstake, PLCM van Riel. Wrote the paper: RM Plenge, C Wijmenga, A Zhernakova, EA Stahl, S Raychaudhuri, FAS Kurreeman, G Trynka, R Toes, Y Li, TWJ Huizinga.

## References

1. Hemminki K, Li X, Sundquist K, Sundquist J (2009) Shared familial aggregation of susceptibility to autoimmune diseases. Arthritis Rheum 60: 2845–2847.
2. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 42: 295–302.
3. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 42: 508–514.
4. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, et al. (2009) Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat Genet 41: 1313–1318.
5. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, et al. (2009) REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 41: 820–823.
6. Trynka G, Zhernakova A, Romanos J, Franke L, Hunt K, et al. (2009) Coeliac disease associated risk variants in TNFAIP3 and REL implicate altered NF-{kappa}B signalling. Gut.
7. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, et al. (2008) Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 359: 2767–2777.
8. Hunt KA, Zhernakova 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.
9. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, et al. (2007) A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 39: 827–829.
10. Zhernakova A, Alizadeh BZ, Bevova M, van Leeuwen MA, Coenen MJ, et al. (2007) Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. Am J Hum Genet 81: 1284–1288.
11. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, et al. (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 39: 1477–1482.
12. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832–838.
13. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466: 707–713.

14. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet.
15. Imielinski M, Baldassano RN, Griffiths A, Russell RK, Annese V, et al. (2009) Common variants at five new loci associated with early-onset inflammatory bowel disease. Nat Genet 41: 1335–1340.
16. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
17. Pearson TA, Manolio TA (2008) How to interpret a genome-wide association study. JAMA 299: 1335–1344.
18. Zhernakova A, van Diemen CC, Wijmenga C (2009) Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet 10: 43–55.
19. Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, et al. (2009) Common and different genetic background for rheumatoid arthritis and coeliac disease. Hum Mol Genet 18: 4195–4203.
20. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, et al. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 17: R122–128.
21. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, et al. (2009) Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 41: 1234–1237.
22. Fransen K, Visschedijk MC, van Sommeren S, Fu JY, Franke L, et al. (2010) Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohn's disease. Hum Mol Genet 19: 3482–3488.
23. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, et al. (2010) Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet 42: 426–429.
24. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, et al. (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40: 955–962.
25. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 75: 330–337.
26. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, et al. (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 36: 337–338.
27. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, et al. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet 5: e1000534. doi:10.1371/journal.pgen.1000534.
28. Call ME, Wucherpfennig KW (2004) Molecular mechanisms for the assembly of the T cell receptor-CD3 complex. Mol Immunol 40: 1295–1305.
29. Tsygankov AY (2008) Multidomain STS/TULA proteins are novel cellular regulators. IUBMB Life 60: 224–231.
30. Egen JG, Kuhns MS, Allison JP (2002) CTLA-4: new insights into its biological function and use in tumor immunotherapy. Nat Immunol 3: 611–618.
31. Hutloff A, Dittrich AM, Beier KC, Eljaschewitsch B, Kraft R, et al. (1999) ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. Nature 397: 263–266.
32. Mao M, Biery MC, Kobayashi SV, Ward T, Schimmack G, et al. (2004) T lymphocyte activation gene identification by coregulated expression on DNA microarrays. Genomics 83: 989–999.
33. Li Y, He X, Schembri-King J, Jakes S, Hayashi J (2000) Cloning and characterization of human Lnk, an adaptor protein with pleckstrin homology and Src homology 2 domains that can inhibit T cell activation. J Immunol 164: 5199–5206.
34. Yao BB, Niu P, Surowy CS, Faltynek CR (1999) Direct interaction of STAT4 with the IL-12 receptor. Arch Biochem Biophys 368: 147–155.
35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
36. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39: 906–913.
37. Eyre S, Hinks A, Bowes J, Flynn E, Martin P, et al. (2010) Overlapping genetic susceptibility variants between three autoimmune disorders: rheumatoid arthritis, type 1 diabetes and coeliac disease. Arthritis Res Ther 12: R175.
38. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics 23: 1294–1296.
39. Barton A, Thomson W, Ke X, Eyre S, Hinks A, et al. (2008) Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. Nat Genet 40: 1156–1159.
40. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, et al. (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. N Engl J Med 357: 1199–1209.
41. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, et al. (2008) Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 40: 1216–1223.
42. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, et al. (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 357: 977–986.
43. Plant D, Flynn E, Mbarek H, Dieude P, Cornelis F, et al. (2010) Investigation of potential non-HLA rheumatoid arthritis susceptibility loci in a European cohort increases the evidence for nine markers. Ann Rheum Dis 69: 1548–1553.