CHAPTER 3

Systematic prioritization of candidate genes in disease loci identifies *TRAFD1* as a master regulator of IFN\(\gamma\) signalling in celiac disease

Adriaan van der Graaf\#, Maria Zorro\#, Annique Claringbould\', Urmo Vosa\', Raul Aguirre-Gamboa\', Chan Li\', Joram Mooiweer\', Isis Ricano-Ponce\', Zuzanna Borek\', Frits Koning\', Yvonne Kooy-Winkelaar\', Ludvig Sollid\', Shuo-Wang Qiao\', BIOS consortium, Vinod Kumar\', Yang Li\', Sebo Withoff\', Lude Franke\', Cisca Wijmenga\', Serena Sanna\', Iris Jonkers\'.

# These authors contributed equally
*Corresponding Authors

*BioRxiv (full version). Preprint first posted online on March 5, 2020; doi:https://doi.org/10.1101/2020.03.04.973487*
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease

Abstract

Background: Celiac disease (CeD) is a complex T cell–mediated enteropathy induced by gluten. Although genome-wide association studies have identified numerous genomic regions associated with CeD, it is difficult to accurately pinpoint which genes in these loci are most likely to cause CeD.

Results: We used four different in silico approaches – Mendelian Randomization inverse variance weighting, COLOC, LD overlap and DEPICT – to integrate information gathered from a large transcriptomics dataset. This identified 118 prioritized genes across 50 CeD-associated regions. Co-expression and pathway analysis of these genes indicated an association with adaptive and innate cytokine signalling and T cell activation pathways. 51 of these genes are targets of known drug compounds, suggesting that our methods can be used to pinpoint potential therapeutic targets. In addition, we detected 129 gene-combinations that were affected by our CeD-prioritized genes in trans. Notably, 40 of these trans-mediated genes appear to be under control of one master regulator, TRAFD1, and were found to be involved in IFNγ signalling and MHC I antigen processing/presentation. We then performed in vitro experiments that validated the role of TRAFD1 as an immune regulator acting in trans.

Conclusions: Our strategy has confirmed the role of adaptive immunity in CeD and revealed a genetic link between CeD and the IFNγ signalling and MHC I antigen processing pathways, both major players of immune activation and CeD pathogenesis.

Key words: Celiac disease, gene prioritization, expression quantitative trait loci (eQTL)
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFN\(\gamma\) signalling in celiac disease

Introduction

Celiac disease (CeD) is an auto-immune disease in which patients experience severe intestinal inflammation upon ingestion of gluten peptides. CeD has a large genetic component, with heritability estimated to be approximately 75%\(^1\). The largest CeD-impacting locus is the HLA region, which contributes approximately 40% of CeD heritability\(^2\). While the individual impacts of CeD-associated genes outside the HLA region are smaller, they jointly account for 60% of heritability. Previous genome-wide association studies (GWAS) have identified 42 non-HLA genomic loci associated with CeD\(^3\)\(^-\)\(^6\), but the biological mechanisms underlying the association at each locus and the genes involved in disease susceptibility are largely unknown. Yet identification of these non-HLA genetic components and an understanding of the molecular perturbations associated with them are necessary to finely understand CeD pathophysiology. Understanding the biological mechanisms of non-HLA CeD loci is difficult: only three of these loci point to single nucleotide polymorphisms (SNPs) located in protein-coding regions\(^3\). The other CeD-risk loci cannot be explained by missense mutations, making it necessary to look at other biological mechanisms such as gene expression to explain their role in CeD pathogenicity. Several studies have been performed to integrate expression quantitative trait loci (eQTLs) with CeD GWAS associations\(^4,7,8\), and several candidate genes, including UBASH3A, CD274, SH2B3 and STAT4\(^9\), have been pinpointed, implicating T cell receptor, NF\(\kappa\)B and interferon signalling pathways as biological pathways associated with CeD pathology. Unfortunately, these eQTL studies had limited sample sizes, which reduced their power to identify cis- and (especially) trans-eQTLs. Furthermore, previous attempts to integrate eQTLs have mostly annotated genomic loci based on catalogued eQTLs without formally testing the causality of the genes in the onset or exacerbation of CeD\(^8,10,11\).

Gene expression and GWAS data can also be integrated using methodologies that identify shared mechanisms between diseases. These methods can be roughly divided into three classes: variant colocalization methods, causal inference methods and co-expression methods. Colocalization methods consider the GWAS and eQTL summary statistics at a locus jointly and probabilistically test if the two signals are likely to be generated by the same causal variant\(^12\). Causal inference methods test if there is a causal relationship between expression changes and the disease, using genetic associations to remove any confounders\(^13,14\). Finally, co-expression methods do not use eQTL information, but rather test if there is significant co-expression between the genes that surround the GWAS locus\(^15\). Unfortunately, there is no current “gold standard” method for finding the causal gene behind a GWAS hit, as all the methods discussed here are subject to their respective assumptions, drawbacks and caveats. However, it is worthwhile to use all these methods in parallel to find the most relevant causal genes for CeD.

Here, we systematically applied all four methods to the latest meta-analysis results for CeD\(^5\) and coupled them with eQTL results from the Biobank Integrative Omics Study (BIOS) co-
hört, one of the largest cohorts for which there is genotype and RNA-seq expression data of peripheral blood mononuclear cells (schematic overview Fig. 1). We focused on 58 GWAS loci that showed significant association with CeD at $p < 5 \times 10^{-6}$. Our approach prioritized 118 genes in 50 loci and identified one gene, TRAFD1, as a master regulator of trans-effects. We then experimentally validated the role of TRAFD1-mediated genes using RNA-seq in a disease-relevant cell type. Our study yields novel insights into the genetics of CeD and is proof-of-concept for a systematic approach that can be applied to other complex diseases.

**Methods**

**Genotypes for eQTL analysis**

We used the BIOS cohort\textsuperscript{16} to map eQTLs in 3,746 individuals of European ancestry. The BIOS cohort is a collection of six cohorts: the Cohort on Diabetes and Atherosclerosis Maastricht\textsuperscript{17}, the Leiden Longevity Study\textsuperscript{18}, Lifelines DEEP\textsuperscript{19}, the Netherlands Twin Registry\textsuperscript{20}, the Prospective ALS Study Netherlands\textsuperscript{21} and the Rotterdam Study\textsuperscript{22}. As described in Vosa \textit{et al.}\textsuperscript{23}, each cohort was genotyped separately using different arrays. Genotypes were subsequently imputed to the HaploType Reference Consortium panel (HRC v1.0) on the Michigan imputation server\textsuperscript{24}.

We considered only biallelic SNPs with a minor allele frequency (MAF) > 0.01, a Hardy-Weinberg test $p$ value $> 10^{-6}$ and an imputation quality RSQR $> 0.8$. To remove related individuals, a genetic relationship matrix (GRM) was created using plink 1.9\textsuperscript{25} (command --make-grm-bin) on linkage disequilibrium (LD)–pruned genotypes (option: “--indep 50 5 2”). Pairs of individuals with a GRM value $> 0.1$ were considered related, and one individual was removed from each of these pairs. Population outliers were identified using a principal component analysis on the GRM, and we removed individuals who were more than 3 standard deviations from the means of principal component 1 or 2.

**Expression quantification**

We used the same procedure for RNA gene expression control and processing as described in Zhernakova \textit{et al.}\textsuperscript{16}. In brief, RNA was extracted from whole blood and paired-end sequenced using the Illumina HiSeq 2000 machine. Read alignment of RNA-seq reads was done using STAR (v2.3.0)\textsuperscript{26} using a reference genome with masked variants with MAF $< 0.01$ in the Genome of the Netherlands\textsuperscript{27}. Aligned reads were quantified using HTSeq\textsuperscript{28}. Samples were removed if they had fewer than 80% aligned reads, fewer than 85% exon-mapping reads, or if they had a median 3' bias larger than 70% or smaller than 45%. Unobserved expression confounders were removed following the procedure of Zhernakova \textit{et al.}\textsuperscript{16}, correcting the expression matrix for the first 25 principal components as well as 3' bias, 5'bias, GC content, intron base pair percentage and sex.
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease

eQTL analysis
After genotype and RNA-seq quality controls (QCs), 3,503 individuals, 19,960 transcripts and 7,838,327 autosomal SNPs remained for analyses. We performed genome-wide eQTL mapping for the transcripts using plink 1.925 with the --assoc command. We defined cis-eQTL variants as those located within ±1.5Mb of the transcript and trans-eQTLs as variants located outside these boundaries. To select variants that could explain the cis-eQTL signal of a gene, we used GCTA-COJO software29 v1.26. For this analysis, we required selected variants to reach a p-value threshold of 5 x 10^{-6} and included the BIOS cohort genotypes as LD reference. This identified 707 genes with at least one eQTL reaching this threshold, 357 of which had more than one conditionally independent eQTL variant.

CeD summary statistics associated regions and candidate genes
We used summary statistics from a CeD GWAS meta-analysis of 12,948 cases and 14,826 controls that analysed 127,855 variants identified using the Immunochip array5. SNP positions were lifted over to human genome build 37 using the UCSC liftover tool. We first identified lead associated variants in the CeD meta-analysis by performing p-value clumping: we used plink 1.925 to select variants at a p-value threshold of 5 x 10^{-6} and pruned variants in LD with these selected variants using standard plink settings (R^2 > 0.5, utilizing 1000 Genomes European sample LD patterns)25,30. We removed variants in an extended HLA region (chromosome 6, 25Mb to 37Mb) due to the complex long range LD structure in this region and because our main interest was in understanding the function of the non-HLA genetic component of CeD. We looked for candidate genes around the clumped variants as follows. First, we defined regions around every clumped variant by padding the clumped SNP position by 1Mb on both sides. We then joined all overlapping CeD-associated regions together and looked for gene transcripts that partly or fully overlapped with the associated regions. This approach identified 58 CeD-associated regions and 1,235 candidate genes that are potentially causal for CeD. Of note, the CeD-association windows were set to be smaller than the eQTL window so that eQTL associations would fully overlap the associated CeD GWAS peak even when a gene is on the edge of the CeD-associated region.

Gene prioritization using Mendelian Randomization–Inverse Variance Weighting (MR-IVW), COLOC, LD overlap and DEPICT
We prioritized CeD-associated genes using three eQTL-based methods – MR-IVW31, COLOC12 and LD overlap – and one co-regulation-based method, DEPICT15. For the MR-IVW method, we used the independent variants identified by GCTA-COJO as instrumental variables32,33 to test causal relationships between changes in gene expression and CeD. MR-IVW was only performed when there were three or more independent eQTLs available (164 genes). A gene was significant for the MR-IVW test if the causal estimates passed a Bonferroni threshold p-value of 3.0 x 10^{-4}. Heterogeneity of causal estimates was accounted for and corrected for using Weighted Median MR analysis and Cochran’s Q test34. For the COLOC method, we used the ‘coloc’ R package and considered a gene significant for the COLOC analysis if the posterior probability of shared variants (H4) was larger than 0.9. For the LD
overlap method, a gene was considered significant if there was high LD ($r^2 > 0.8$) between the top independent eQTL and the top CeD variant in the region. Finally, we applied DEPICT\textsuperscript{15} to the clumped CeD GWAS variants described in ‘CeD summary statistics associated regions and candidate genes’. Genes identified by the DEPICT analysis were considered significant if a False Discovery Rate (FDR) < 0.05 was found with DEPICT’s own permutation measure.

We scored each gene in the CeD-associated loci by considering each of the four prioritization methods. A gene was prioritized as ‘potentially causal’ in CeD pathology when one of the four methods was significant (one line of evidence). If multiple lines of evidence were significant, the gene was prioritized more highly than when only a single line of evidence was available.

To explore how the prioritized genes affect CeD risk, we gave each gene an effect direction based on the effect direction of the top variants in the eQTL and the CeD GWAS using the following algorithm:

1. If there was a concordant effect that was significant in the top variants of both the eQTLs and the GWAS, the direction of the concordant effect was chosen.
2. If there was a concordant effect, but no significance of the SNP in one of the datasets, we could not be sure of an effect direction, and a question mark was chosen. The only exception to this was if the MR-IVW was significant, when we chose the direction of the MR-IVW effect.
3. If there was a discordant effect between the top SNPs, and both were significant in both datasets, a question mark was chosen. The only exception to this was when the IVW was significant, when the IVW effect was chosen.
4. If there was a discordant effect and there was significance in only one of the GWAS from the eQTL top SNP, the GWAS direction was chosen.
5. If there was a discordant effect and there was significance in only one of the eQTL from the GWAS top SNP, the GWAS direction was chosen.
6. If there was otherwise a discordant effect, a question mark was chosen.

Each gene is given a mark: positive (‘+’), negative (‘-’) or unknown (‘?’). ‘+’ indicates that increased expression increases CeD risk. ‘-’ indicates that increased expression decreases CeD risk. ‘?’ indicates that it is unknown how the expression affects CeD risk.

**Co-regulation clustering**

The genes that have been prioritized may have some shared function in CeD pathology. To identify possible shared pathways, we performed co-regulation clustering analysis based on 1,588 normalized expression co-regulation principal components identified from RNA-seq information across multiple human tissues by Deelen \textit{et al.}\textsuperscript{35}. We performed pairwise Pearson correlation of our prioritized genes with these 1,588 principal components and derived a correlation Z score for each prioritized gene pair. We then performed hierarchical clustering of this Z score matrix using Ward distances and identified 4 clusters from the resulting dendrogram.
Trans eQTL and mediation analysis

238 autosomal genes that were not located in, but were associated with, a significant trans-eQTL variant \((p < 5 \times 10^{-8})\) in the CeD-associated regions were identified and used as potential targets for mediation by our associated genes in the CeD-associated loci (86 potential cis mediating genes). We first selected trans-eQTL genes that were co-expressed (Pearson \(r > 0.1\), 197 gene combinations) with prioritized genes, then performed mediation analysis by running the trans-eQTL association again using the expression of the cis-eQTL gene as a covariate. We defined a trans-mediated gene if, after mediation analysis, the change (increase or decrease) in the effect size of the top trans-eQTL variant was significant according to the statistical test described in Freedman and Schatzkin\(^{36}\). For this analysis, we used a Bonferroni-adjusted \(p\)-value of \(3.0 \times 10^{-4}\).

Cell type proportion and SH2B3 expression mediation analysis

To assess if the cis-eQTL effect of TRAFD1 was not a proxy for cell-type composition, we performed mediation analyses in a fashion similar to the trans mediation analysis above using cell proportions measured in a subset of individuals in the BIOS cohort. To ensure that there was no residual effect of SH2B3-expression on the mediating effect of TRAFD1, we corrected the original TRAFD1 expression levels for the expression levels of SH2B3, leaving TRAFD1 expression independent of SH2B3, and reran the mediation analysis.

Literature review. We performed a Reactome pathway\(^37\) analysis to determine the potential function of the prioritized genes. This was complemented with a literature search (research and review papers) in Pubmed. For the coding and non-coding genes for which no studies were found, Genecards (www.genecards.org) and Gene Network v2.0 datasets (www.genenetwork.nl)\(^{35}\) were used, respectively. Information regarding the potential druggability of the prioritized genes was obtained from DrugBank\(^{38}\), the pharmacogenetics database\(^{39}\) and a previous study that catalogued druggable genes\(^{40}\).

THP-1 culture. The cell line THP-1 (Sigma Aldrich, ECACC 88081201) was cultured in RPMI 1640 with L-glutamine and 25mM HEPES (Gibco, catalogue 52400-025), and supplemented with 10% fetal bovine serum (Gibco, catalogue 10270) and 1% penicillin/ streptomycin (Lonza, catalog DE17602E). The cells were passed twice per week at a density lower than 0.5 \(\times 10^6\) cells/ml in a humidified incubator at 5% \(\text{CO}_2\), 37°C.

siRNA treatment. THP-1 cells were plated at 0.6 \(\times 10^6\) cells/ml and transfected with 25 nM siRNA using Lipofectamine RNAimax transfection reagent (Invitrogen, catalogue 13788), according to the manufacturer’s protocol. Cells were treated with an siRNA to target TRAFD1 (Qiagen catalogue 1027416, sequence CCCAGCGACCCATTAACAAT) (Knockdown (KD)), and cells treated with transfection mix without siRNA (Wild type (WT)) or non-targeting control siRNA (scrambled (SCR)) (Qiagen catalog SI03650318, sequence undisclosed by company) were included as controls. All the treatments were performed in triplicate. 72 hours after
transfection, a small aliquot of cells was stained for Trypan Blue exclusion to determine cell viability and proliferation. The cells were stimulated with either LPS (10 ng/ml) from *E. coli* (Sigma catalogue 026:B6) or media alone (unstimulated) for 4h. Subsequently, the cells were centrifuged, and the cell pellets suspended in lysis buffer and stored at -80°C until used for RNA and protein isolation.

**qPCR.** The total RNA from THP-1 cells was extracted with the mirVana™ miRNA isolation kit (AMBION, catalogue AM1561) and subsequently converted to cDNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo scientific, catalogue K1631). qPCR was done using the Syber green mix (Bio-Rad, catalogue 172-5124) and run in a QuantStudio 7 Flex Real-Time system (Applied Biosystems, catalogue 448598). Primer sequences to determine KD levels of *TRAFD1* were 5’ GCTGTTAAAGAAGCATGAGGAGAC and 3’ TTGGAACCACATAGTTCCGTCCG. *GAPDH* was used as endogenous qPCR control with primers 5’ ATGGGGAAGGTGAAGGTCG and 3’ GGGGTCATTGATGGCAACAATA. Relative expression values of *TRAFD1* were normalized to the endogenous control *GAPDH* and calculated using the ΔΔCT method, then given as a percentage relative to SCR expression levels.

**Western blot (WB).** Cell pellets from THP-1 cells were suspended on ice-cold lysis buffer (PBS containing 2% SDS and complete protease inhibitor cocktail (Roche, catalog 11697498001)). Protein concentration of cell extracts was determined using the BCA protein kit (Pierce, catalog 23225). Proteins were separated on 10% SDS-polyacrylamide electrophoresis gel and transferred to a nitrocellulose membrane. After 1 hour of blocking with 5% fat-free milk in Tris-Tween-Buffer-Saline, the membranes were probed for 1 hour at room temperature with mouse mono-clonal TRAFD1 antibody 1:1000 (Invitrogen, catalog 8E6E7) or mouse mono-clonal anti-actin antibody 1:5000 (MP Biomedicals, catalog 08691001), followed by incubation with goat anti-mouse horseradish peroxidase–conjugated secondary antibodies 1:10000 (Jackson Immuno Research, catalog 115-035-003). After three 10-minute washes, the bands were detected by Lumi light WB substrate (Roche, catalogue 12015200001) in a Chemidoc MP imaging system (Bio-Rad) and quantified using Image Lab™ software (Bio-Rad). The band intensity of TRAFD1 was normalized to actin, and the TRAFD1 SCR control level was set as 100%.

**Statistical analysis for in vitro experiments in THP-1 cells.** The statistical analyses of proliferation, qPCR and WB were performed using Prism 5 software (GraphPad Software, Inc.). Results are presented as mean ± SEM from a representative experiment. Statistical differences were evaluated using a one-tailed *t*-test.

**RNA sequencing (RNA-seq) in THP-1 cells.** RNA from THP-1 cells was extracted with the mirVana™ miRNA isolation kit (AMBION, catalog AM1561). Prior to library preparation, extracted RNA was analysed on the Experion Stdsend RNA analysis kit (Bio-Rad, catalog 7007105). 1 μg of total RNA was used as input for library preparation using the quant seq 3’
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease

It can be difficult to determine if a set of genes is ‘on average’ more or less differentially expressed due to co-expression between the genes within the set. To mitigate this, we performed a permutation test that considers the median absolute T statistic calculated by DESeq²⁴ in the WT-SCR experiment as a null observation and compared this null observation with the SCR-KD experimental comparison. This allowed us to compare the expected differential expression of a set of genes, based on the WT-SCR comparison, with the observed differential expression of the same set of genes in the SCR-KD comparison, while still incorporating the co-expression structure of the data. To do this, we randomly selected a same-sized set of genes 500,000 times in each relevant experiment (WT-SCR or SCR-KD), and determined the observed median absolute T statistic. We calculated a ratio of how often the permuted value is higher than the observed value. For example, the observations can be that 1% of permuted gene sets are more differentially expressed in the WT-SCR experiment, while only 0.01% of permuted genes sets are more differentially expressed in the SCR-KD experiment. Finally, we divide these values by one another, (percentage SCR-KD)/(percentage WT-SCR), to calculate a fold increase in differential expression. In the example given above, this indicates that the KD is 100 times (0.01/1 = 100) more differentially expressed than expected.

Gene set permutation analysis. It can be difficult to determine if a set of genes is ‘on average’ more or less differentially expressed due to co-expression between the genes within the set. To mitigate this, we performed a permutation test that considers the median absolute T statistic calculated by DESeq²⁴ in the WT-SCR experiment as a null observation and compared this null observation with the SCR-KD experimental comparison. This allowed us to compare the expected differential expression of a set of genes, based on the WT-SCR comparison, with the observed differential expression of the same set of genes in the SCR-KD comparison, while still incorporating the co-expression structure of the data. To do this, we randomly selected a same-sized set of genes 500,000 times in each relevant experiment (WT-SCR or SCR-KD), and determined the observed median absolute T statistic. We calculated a ratio of how often the permuted value is higher than the observed value. For example, the observations can be that 1% of permuted gene sets are more differentially expressed in the WT-SCR experiment, while only 0.01% of permuted genes sets are more differentially expressed in the SCR-KD experiment. Finally, we divide these values by one another, (percentage SCR-KD)/(percentage WT-SCR), to calculate a fold increase in differential expression. In the example given above, this indicates that the KD is 100 times (0.01/1 = 100) more differentially expressed than expected.

Gene set permutation analysis. It can be difficult to determine if a set of genes is ‘on average’ more or less differentially expressed due to co-expression between the genes within the set. To mitigate this, we performed a permutation test that considers the median absolute T statistic calculated by DESeq²⁴ in the WT-SCR experiment as a null observation and compared this null observation with the SCR-KD experimental comparison. This allowed us to compare the expected differential expression of a set of genes, based on the WT-SCR comparison, with the observed differential expression of the same set of genes in the SCR-KD comparison, while still incorporating the co-expression structure of the data. To do this, we randomly selected a same-sized set of genes 500,000 times in each relevant experiment (WT-SCR or SCR-KD), and determined the observed median absolute T statistic. We calculated a ratio of how often the permuted value is higher than the observed value. For example, the observations can be that 1% of permuted gene sets are more differentially expressed in the WT-SCR experiment, while only 0.01% of permuted genes sets are more differentially expressed in the SCR-KD experiment. Finally, we divide these values by one another, (percentage SCR-KD)/(percentage WT-SCR), to calculate a fold increase in differential expression. In the example given above, this indicates that the KD is 100 times (0.01/1 = 100) more differentially expressed than expected.

Available RNA-seq datasets. Four available RNA-seq datasets were included to study the pattern of expression of prioritized genes. A brief description of each dataset is provided below.

Whole biopsy samples. Duodenal biopsies were obtained from 11 individuals (n=6 CeD pa-
tients and n=5 controls) who underwent upper gastrointestinal endoscopy. To identify DEGs between patients and controls, a filter of $|\log_2 \text{FC}| > 1$ and FDR $\leq 0.01$ was applied using the DESeq2 R package (Zorro MM, available in doi.org/10.1016/j.jaut.2020.102422).

**IELs.** CD8$^+$ TCR$\alpha\beta$ IELs cell lines were generated from intestinal biopsies and expanded for 12 days, as described previously$^{44}$. Cells were left unstimulated (controls) or treated for 3 hours with IFN$\gamma$ (300 ng/ml, Pbl Assay science, cat 11410-2), IL-15 (20 ng/ml, Biolegend, cat 570304) or IL-21 (3 ng/ml, Biolegend, cat 571204) (n=8 samples per condition). Differential expression analysis between unstimulated cells and cytokine-treated IELs was performed using the R package DESeq2. DEGs were defined as genes presenting a $|\log_2 \text{FC}| > 1$ and an FDR $\leq 0.01$ between untreated controls and cytokine-treated samples (Zorro MM, available in doi.org/10.1016/j.jaut.2020.102422).

**Gluten-specific (gs) CD4$^+$ T cells.** gsCD4$^+$ T cell lines were generated from intestinal biopsies and expanded for 2 weeks, as reported previously$^{45}$. Cells were stimulated for 3 hours with 2.5 $\mu$g/ml of anti-CD3 (Biolegend, catalog 317315) and anti-CD28 (Biolegend, catalog 302923) antibodies. Untreated cells were included as control. N=22 samples per condition. DEGs were extracted with the DESeq2 R package using the cut-off $|\log_2 \text{FC}| > 1$ and an FDR $\leq 0.01$ between unstimulated samples and control (Jonkers I, unpublished results).

**Caco-2 cells.** After 2 weeks of expansion in Transwells, the cells were treated with 60 ng/ml of IFN$\gamma$ (PeproTech) for 3 hours. Untreated cells were included as controls. RNA samples were extracted and further processed for RNA-seq. DEGs between control and stimulated cells were extracted with the DESeq2 R package using a cut-off of $|\log_2 \text{FC}| > 1$ and an FDR $\leq 0.01$ (Zorro MM, unpublished results. Chapter 2).

**Results**

**Gene prioritization identifies 118 likely causal CeD genes**

To identify genes that most likely play a role in CeD (prioritized genes), we combined a recent genome-wide association meta analysis$^{46}$ with (1) eQTLs derived from whole-blood transcriptomes of 3,502 Dutch individuals$^{16}$ and (2) a co-regulation matrix derived from expression data in multiple different tissues and 77,000 gene expression samples$^{15}$. We selected 1,258 genes that were within 1Mb of the 58 CeD-associated non-HLA variant regions ($p < 5x10^{-8}$) (see Methods), and prioritized the genes that are the most likely causally related to CeD using four different gene prioritization methods: MR-IVW$^{33}$, COLOC$^{12}$, LD overlap and DEPICT$^{15}$ (Fig. 1A) (Supplementary Table 1).

The first method we applied, MR-IVW, is a two-sample Mendelian Randomization approach called inverse variance weighting (see Methods). Our MR-IVW used summary statistics from two datasets: the eQTL and CeD GWAS. First, independent eQTLs at a locus were identified
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease.

(see Methods), then the effect sizes of the eQTL and the GWAS were combined to identify gene expression changes that are causal (or protective) for CeD\(^{14,33}\) (see Methods). We only applied this method to a subset of 164 prioritized genes for which at least three independent cis-eQTL variants (at \(p < 5 \times 10^{-6}\)) were identified (see Methods)\(^{32}\). We accounted for heterogeneity using the Q test and weighted median method and found that the effect sizes were very similar before and after correction (data not shown).

The second method, COLOC, is a variant colocalization test in which we used eQTL and CeD summary statistics for all the SNPs in a locus and Bayesian probability to infer whether the
eQTL and the CeD-association signals are likely to originate from the same causal variant\textsuperscript{12}. The third method, LD overlap, is a more classical annotation-type approach that prioritizes a gene if the top eQTL is in strong LD ($r^2 > 0.8$) with the variant most significantly associated with CeD in a locus. This and the COLOC method were applied to 707 genes for which at least one significant eQTL variant was found.

Finally, we used DEPICT\textsuperscript{15}, a gene-prioritization method based on co-regulation in expression datasets across multiple different tissues. DEPICT identifies enrichment for co-regulated genes from genes in a GWAS locus. In contrast to the other methods, DEPICT assessed the potential role of all 1,258 genes independently of the presence of an eQTL.

In total, 118 out of the 1,258 assessed genes were prioritized by at least one of the four methods. Of these 118 genes, 27 had two lines of evidence, 6 genes (CD226, NCF2, TRAFD1, HML3, COLCA1, CTSH) had three lines of evidence, and one gene (CSK) was supported by all four methods (Supplementary Table 2) (Fig. 1B-C). Overall, we identified potentially causal genes in 50 out of 58 CeD-associated regions.

The four different gene prioritization methods complement each other in different ways. DEPICT prioritized the most genes: 66 in total, 38 of them uniquely prioritized (38/66, 58% unique). One reason for this is that DEPICT is based on co-expression, not genetic background. Indeed, 16 genes prioritized by DEPICT do not have a significant eQTL associated with them. Overall, the most concordance was found between COLOC and LD overlap (29% and 25% unique genes, respectively) as these methods are the most similar, while MR-IVW uniquely prioritized a relatively large proportion of genes (9/19, 47% unique). Thus, each method helps prioritize genes with multiple lines of evidence, but also adds a unique set of genes based on the assumptions of the method.

To see if any of these genes could lead to therapeutic intervention in CeD, we searched for the CeD-associated genes in DrugBank and assessed their druggability potential following Finan \textit{et al.}\textsuperscript{47} (Supplementary Table 2). 45 of the 118 prioritized genes encode proteins that are targeted by an approved drug or a drug in development (Fig. 1C) (Supplementary Table 2). For example, drugs such as Natalizumab and Basiliximab that target the proteins encoded by ITGA4 and IL21R, respectively, are currently approved or under study for the treatment of immune-mediated diseases including rheumatoid arthritis\textsuperscript{48}, Crohn’s disease\textsuperscript{49} and multiple sclerosis\textsuperscript{50} or as an immune-suppressor to avoid kidney transplant rejection. An additional 16 genes encoded proteins that are similar to proteins targeted by already approved drugs\textsuperscript{40} (Supplementary Table 2).

\textbf{Co-expression patterns of cis-eQTL-prioritized loci reveal four functional clusters}

The biological function for the 118 prioritized genes and their role in CeD pathology is not always fully understood. We sought to infer biological function using a guilt-by-association
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease.

![Heatmap showing results of the Reactome gene set enrichment analysis](image)

**A** Heatmap showing Spearman correlations between gene expression patterns of each prioritized gene. Blue squares indicate negative correlation. Red squares indicate positive correlation. Both are shaded on a gradient scale according to the Z score. A dendrogram computed with Ward distances between the correlations is shown on top of the heatmap. Branches of the correlation. Red squares indicate positive correlation. Both are shaded on a gradient scale according to the Z score. A dendrogram computed with Ward distances between the correlations is shown on top of the heatmap. Branches of the correlation. Red squares indicate positive correlation. Both are shaded on a gradient scale according to the Z score. A dendrogram computed with Ward distances between the correlations is shown on top of the heatmap. Branches of the correlation. Red squares indicate positive correlation. Both are shaded on a gradient scale according to the Z score.

**B** Clusters

1. IFNγ signaling
2. IL-6 signaling
3. Costimulation by CD28 family
4. Chemokine receptor binding
   - G alpha signaling
   - Immune system
   - GPCR signaling
   - Signal transduction
   - Cytokine signaling

**C** Biopsy CTR
Biopsy CeD
IELs
IELs+IL21
IELs+IL15
IELs+IFNγ
gsCD4
gsCD4+aCD3/CD28

Scaled expression

**Fig. 2 Co-expression pattern of cis-eQTL prioritized genes reveals four functional clusters.** (A) Heatmap showing the Spearman correlations between gene expression patterns of each prioritized gene. Blue squares indicate negative correlation. Red squares indicate positive correlation. Both are shaded on a gradient scale according to the Z score. A dendrogram computed with Ward distances between the correlations is shown on top of the heatmap. Branches of the dendrogram are coloured differently to mark separate clusters. (B) Results of the Reactome gene set enrichment analysis of the genes belonging to each of the clusters identified in (A). Colour key denotes the significance (-log 10 adjusted p value) of each biological pathway. (C) Heatmaps depicting the scaled expression of prioritized genes belonging to the four clusters identified in (A) in three available RNA-seq datasets: intestinal biopsies from controls (CTR, n=5 samples) or CeD patients (CeD, n=6 samples); CD8+ TCRαβ intraepithelial lymphocytes (IELs) unstimulated or treated with IL-21, IL-15 or IFNγ for 3 hours (n=8 samples per condition) and gsCD4+ T cells unstimulated or treated with anti-CD3 (aCD3) and anti-CD28 for 3 hours (n=22 samples per condition). Clustering was performed using the “average” method in hclust().
co-regulation approach to identify clusters of shared molecular function (see Methods). We identified co-regulated genes by correlating our prioritized gene list in 1,588 principal components that were identified from the co-expression of 31,499 RNA-seq samples across multiple tissues (Fig. 2A). We then performed Reactome 2016 gene set enrichment analysis to investigate the biological processes enriched in each cluster (Supplementary Table 3 and data not shown).

We could not identify a specific biological process linked to our first co-regulation cluster. However, genes such as ULK3 (relevant for autophagy) and CSK (relevant to T cell receptor (TCR) signaling) are included in this co-regulation cluster. Our second cluster encompasses genes (e.g. STAT1, CD274 and IL12A) implicated in interferon gamma (IFNγ) signalling, interleukin (IL)-6 signalling and co-stimulation by CD28. Co-regulation cluster 3 contains genes (e.g. CD28, CTLA4 and ICOS) associated with co-stimulation by CD28, a process that is essential for modulating T cell–activation. Finally, co-regulation cluster 4 contains chemokine (e.g. CCR1, CCR2 and CCR3) and cytokine signalling genes (e.g. IL2RA, IL21 and IL18R1) (Fig. 2B). The biological processes overrepresented in these co-regulation clusters are essential for the activation and function of the adaptive and innate immune system, which confirms and extends previous findings that implicate both arms of the immune system in CeD disease pathogenesis. Of note, approximately 10% of the prioritized genes are long non-coding RNAs (lncRNAs) rather than protein-coding genes (Supplementary Table 1). Although little is known about the function of lncRNAs, their co-regulation pattern with the genes in clusters 2 and 4 suggests that they may be associated with cytokine/chemokine signalling (Fig. 2A, B). Moreover, by using Genenetwork, we found that the lncRNAs RP3-395M20.9, AC007278.2 and AC104820.2 may be involved in tumour necrosis factor (TNF) signalling, neutrophil degranulation and chemokine receptor signalling, respectively, implying a role for these uncharacterized lncRNAs in immune regulation in CeD.

CeD candidate genes operate in immune and intestinal epithelial cells
To complement our Reactome gene set enrichment analysis and dig deeper into the biological processes and cell types in which the prioritized genes may act, we analysed their expression profiles in available RNA-seq datasets from disease-relevant cell types including 1) small intestinal biopsies of active CeD patients and healthy controls, 2) intra-epithelial lymphocytes (IE-CTLs) stimulated with disease-relevant cytokines IL-21, IL-15 and IFNβ, and 3) gsCD4+ T cells stimulated with antiCD3-antiCD28, which mimics the disease-specific response to gluten peptides (Fig. 2C and data not shown). Here we noted that the genes grouped in co-regulation clusters 1 and 2 are highly expressed in small intestinal biopsies and IE-CTLs, which is in line with the IFNγ pathway enrichment seen in co-regulation cluster 2 (Fig. 2B). IFNγ is mainly produced by gsCD4+ T cells and IE-CTLs and is known to disrupt the integrity of the intestinal epithelial cells in CeD-associated villous atrophy. Within this cluster we also found genes specific to antigen-presenting cells (B cells, monocytes and dendritic cells) and epithelial cells such as IL12A and COLCA1, which are most expressed in small intestinal biopsies (Fig.
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease.

The genes in co-regulation clusters 3 and 4 are highly expressed in gsCD4+ T cells, especially after stimulation with antiCD3-antiCD28, indicating that these prioritized genes may be biologically relevant in the immediate T cell receptor response to gluten ingestion.

The gene expression pattern of the prioritized genes, when combined with information from our literature search, suggests that these genes may control general biological processes (e.g. apoptosis, gene regulation and cytoskeleton remodelling) as well as specific immune functions (e.g. cell adhesion, cell differentiation and TCR signalling) in diverse cell types (e.g. T cells, neutrophils, B cells, monocytes, epithelial cells) (Fig. 3 and data not shown). The non-HLA genetic loci associated to CeD thus seem to affect a complex network of cells and biological processes.

Mediation analysis uncovers TRAFD1 as a major trans-eQTL regulator

To further understand the potential regulatory function of the prioritized genes, we identified downstream regulatory effects by performing a trans-mediation analysis using a two-step approach (Methods) (Supplementary Fig. 1A). We first considered all genes with a trans-eQTL (p < 5x10^-8) located in any of the 58 CeD-associated regions, then performed a mediation analysis.
Chapter 3
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease

analysis by re-assessing the trans-eQTL effect after adjusting the expression levels for the expression of the prioritized gene(s) in the same locus (Fig. 4A).

Of the 497 possible prioritized gene–trans-eQTL gene combinations, we found 129 that exhibited significant mediation effects. These combinations map to 6 associated regions and represent 13 unique mediating cis-eQTL genes and 67 unique mediated trans-eQTL genes (Supplementary Table 4). Among all the associated regions, the CeD-associated region on chromosome 12 contained the largest number of both cis-mediating genes (N=5) and trans-mediated genes (N=60). In this region, TRAFD1 mediated more trans genes than all of the other regional cis-regulators and also had the highest mediation impact (average Z-score difference in effect size between mediated and unmediated analysis = 2.81) (Methods, Supplementary Fig. 1B and data not shown). Of note, the top eQTL variant of TRAFD1 is a missense variant in the nearby gene SH2B3. This missense variant has been associated to a number of complex traits, including blood cell types and platelets, and diseases. However, we found that cell-type composition did not affect the eQTL-association of TRAFD1 in our cohort (p > 0.044 for 24 different cell-type traits) (Methods and data not shown). To ensure that the mediated trans genes of TRAFD1 were not mediated by SH2B3, we corrected TRAFD1 expression levels for SH2B3 and re-ran the mediation analysis. Here we found that the mediating effect of TRAFD1 was still significant for all 40 genes found initially and that the median Z-score difference between mediated and unmediated was higher than that of SH2B3, although it was slightly attenuated compared to the original TRAFD1 signal (Supplementary Figure 1B and data not shown).
not shown). Based on these results, we conclude that TRAFD1 is a master regulator of gene expression changes in the associated region (Fig. 4B and data not shown).

Strikingly, three of the TRAFD1 trans-mediated genes – STAT1, CD274 and PDCD1LG2 – are also prioritized cis-genes in their respective loci (Fig. 4B). These results suggest that the trans-mediated TRAFD1-effects may have an additional additive effect in these CeD-associated loci.

TRAFD1 is a poorly characterized gene that has been suggested to act as a negative regulator of the NFκB pathway. To further elucidate the biological processes in which the 40 TRAFD1 trans-mediated genes could be involved, we performed a Reactome gene set enrichment analysis (data not shown). Here we found that IFNγ signalling, cytokine signalling and major histocompatibility complex class I (MHCI) antigen processing/ presentation are strongly enriched pathways, which points to a role for TRAFD1 and TRAFD1 trans-mediated genes in antigen presentation and immune response (Fig. 5A).

By looking into RNA-seq datasets from disease-relevant cell types, we noted that most TRAFD1 trans-mediated genes are upregulated in biopsies from patients with active CeD, and these genes include STAT1, CXCL10 and TAP1, which are essential for IFN response, chemotaxis and antigen processing, respectively (Fig. 5B). Moreover, most TRAFD1 trans-mediated genes exhibit an increase in expression in response to IFNγ in intestinal

---

**Fig. 5 TRAFD1 is a regulator of IFNγ signalling genes.** (A) Results of the Reactome gene set enrichment analysis of TRAFD1-mediated genes (n=40 genes). Colour code denotes the significance (-log 10 adjusted p value) of each biological pathway. (B) Unscaled heatmaps depicting the expression of these genes in RNA-seq datasets from different cell types: whole biopsies from controls (Ctr, n=5 samples) of CeD patients (CeD, n=6 samples); intra-epithelial lymphocytes (IELs) unstimulated or treated with IFNγ for 3 hours (n=8 samples per condition); and Caco-2 cells untreated or stimulated with IFNγ for 3 hours (n=8 samples per condition). Red indicates that a gene is differentially expressed (DE), blue indicates that a gene is not differentially expressed (non-DE) (FDR<0.01 and |log2(RPKM)|>1). Grey (none or unstimulated), pink (IFNγ), green (IFNγ) and yellow (antiCD3/antiCD28) colours indicate the type of stimulation (treatment).
epithelial cells (Caco-2) or IFNβ in IELs (Fig. 5B). However, antiCD3-antiCD28 stimulation in gsCD4+T cells resulted in both up- and downregulation of the TRAFD1 trans-mediated genes, implying that TRAFD1 trans-mediated genes respond more strongly to IFN signalling (IFNγ or IFNβ) than to TCR activation by anti-CD3/anti-CD28. Indeed, the enrichment of the 40 TRAFD1 trans-mediated genes in significantly differentially expressed genes in biopsies, IELs, epithelial cells and gsCD4+T cells was strongest in IELs and epithelial cells upon IFN signalling (data not shown). Overall, our results suggest that TRAFD1 and TRAFD1 trans-mediated genes modulate IFN signalling upon antigen presentation, possibly via regulation of NFκB, in CeD pathology.

**TRAFD1 KD affects immune-activation genes**

We performed a siRNA KD experiment on TRAFD1 to gain more insights into the biological function of this gene and to independently validate the TRAFD1 trans-mediated genes. We also evaluated the transcriptional changes of knocking down TRAFD1 in the monocyte-like cell line THP-1 under resting conditions (unstimulated) or in the presence of LPS, a known inducer of the NFκB pathway62.

After siRNA treatment, we observed no significant differences in cell viability or proliferation among the controls (WT and SCR) and the KD treatment (Supplementary Fig. 2A, B). However, as expected for the KD cell line, we noted a significant reduction in the expression of TRAFD1 compared to the controls in WB and qPCR analyses (Supplementary Fig. 2C-E). KD of TRAFD1 was also confirmed in the RNAseq data, with TRAFD1 expression levels reduced by 41% in unstimulated KD cells compared to unstimulated SCR cells (adjusted \( p = 0.004 \)) and by 34% in LPS-stimulated KD cells compared to LPS-stimulated SCR cells (not significant, data not shown). The reduced KD effect upon LPS stimulation is consistent with our expectation that TRAFD1 acts as negative regulator of the NFκB pathway, which is activated by several stimuli, including LPS62. Thus, the KD was successful and neither the transfection method nor a reduced expression of TRAFD1 had a toxic effect (Supplementary Fig. 2A-E).

Next, we tested if the 40 TRAFD1 trans-mediated genes were more differentially expressed than expected after LPS stimulation. To disentangle differential expression from the co-expression inherently present in a gene expression dataset, we devised a permutation scheme that compared the control (WT vs. SCR) observations with the KD (SCR vs. KD) observations (see Methods). This scheme takes into account the co-expression of a gene set, as this co-expression is present in both the control and the experimental observation. After performing 500,000 permutations of 41 genes (40 trans mediated genes and TRAFD1) in the LPS-stimulated comparison, the median test statistic in the control observations was observed 44.5 times more often than in the KD observations (0.089% for WT-SCR vs. 0.002% for SCR-KD, Supplementary Fig. 4). This indicates that the 40 trans-mediated genes and TRAFD1 are 44.5 times more differentially expressed than expected. We did not find
Fig 6. TRAFD1 knockdown affects immune activation and stress-related genes. (A) Heatmap showing the expression profile of the 500 shared DEGs identified in the knockdown experiments (see Methods and Supplementary Fig. 2 F, G). A dendrogram on the left of the heatmap depicts the strength of similarities based on Ward distance. (B, C) Violin plots showing the normalized gene expression of the genes belonging to the first and second cluster of DEGs identified in (A) in THP-1 cells under different experimental conditions (WT=untransfected, SCR=non-targeting siRNA, KD=siRNA targeting TRAFD1) and stimulations (LPS=lipopolysaccharide). (D, E) Results of Reactome gene set enrichment analysis of the genes within the first (D) and second cluster (E). Significance (−log 10 adjusted p value) of each biological pathway is indicated by the colour key.

Increased differential expression of the same gene set in the unstimulated condition (0.298% for WT-SCR vs. 0.240% for SCR-KD, Supplementary Fig. 4), indicating that TRAFD1 mainly regulates genes in an LPS-stimulated state.

To identify the role of TRAFD1 in immune cells and processes, we compared gene expression changes in the unstimulated condition versus the LPS stimulated condition for each treatment (WT, SCR or KD) separately (Supplementary Fig. 2F). Differential expression analysis showed that 353 genes were uniquely upregulated and 330 genes uniquely downregulated after TRAFD1 KD treatment (Supplementary Fig. 2G, H). We found no REACTOME gene set enrichment for these unique KD genes. We found 500 upregulated and 433 downregulated genes that were differentially expressed in all three treatments upon LPS stimulation (Fig. 6A, Supplementary Fig. 2G, H). Upregulation (or downregulation) after LPS stimulation was treatment-dependent, i.e. the differential expression identified was increased (or decreased) from WT to SCR to KD (Fig. 6A). We performed hierarchical clustering, this separated the two gene sets into two clusters: cluster 1 shows a decreased response of genes in the TRAFD1 KD group (LPS cluster 1, Fig. 6B and data not shown) and cluster 2 displays an increased expression in the TRAFD1 KD cells under unstimulated conditions that persists after LPS stimulation (LPS cluster 2, Fig. 6C and data not shown). REACTOME gene set enrichment analysis indicated that the genes in LPS cluster 1 are involved in immune-related processes (e.g. cytokine signalling, RIG/IMDA5 induction of IFN signalling and IFN signalling, Fig. 6D), whereas the genes in LPS cluster 2 are associated with the heat shock response, which has
been shown to be activated as a consequence of immune activation or immune response to stress\textsuperscript{63} (Fig. 6E and data not shown). Together, these results suggest that TRAFD1 is a regulator of immune activation and inflammation.

**Discussion**

In the present study we aimed to identify CeD candidate genes using four *in silico* methods (MR-IVW, COLOC, LD overlap and DEPICT) and whole blood transcriptomics data from a population-based cohort. While previous studies have used at least one of these methods\textsuperscript{3,5,6,11}, to our knowledge this is the first effort that integrates the four different statistical approaches. This systematic prioritization approach resulted in 118 prioritized causal genes, including 46 that are direct targets of an approved drug or of drugs under development for other complex diseases, including autoimmune diseases. The co-expression pattern within a large RNA-seq dataset from blood\textsuperscript{35} suggests these genes are involved in cytokine signalling in innate and adaptive cells as well as in T cell activation pathways. We also identified TRAFD1 to be trans-regulator of 40 genes, with a strong enrichment in IFNγ signalling and MHC I antigen processing/presentation pathways, which are pivotal for the disease pathogenesis.

Among the cis-eQTL prioritized candidate genes, we found a co-regulated cluster of genes involved in T cell activation and co-stimulation. Within this co-regulation cluster (cluster 3), we found the THEMIS, IL2, CD28, CTLA4 and UBASH3A genes (Fig. 2), whose functions include T cell differentiation and activation and the TCR macromolecular complex (signalosome). These results highlight the key role of T cell activation in the pathogenesis of CeD\textsuperscript{64}.

Another co-regulation cluster grouped prioritized genes involved in cytokine and chemokine signalling events. For example, this group included CCR1 and CCR2, which control the activation and recruitment of monocytes, dendritic cells and neutrophils\textsuperscript{65}, and IL-21, which provides proliferation and survival signals to B cells\textsuperscript{66}. Deregulation in these genes could worsen the intestinal inflammation. We also found genes in this cluster, e.g. IL18RAP and IL18R1, that encode for cytokine receptors. IL18RAP and IL18R1 encode the IL-18 receptor, which is broadly expressed in various cell types and exhibits pleiotropic effects\textsuperscript{67}. This cluster also contains transcription factors genes, e.g. IRF4, ETS1 and REL. IRF4 and ETS1 are essential for T helper 1 (Th1) differentiation\textsuperscript{68,69}, while REL is a key regulator of NFκB signaling\textsuperscript{70}. This result is thus consistent with the novel genetic association reported between NFκB and CeD by Ricano-Ponce at al\textsuperscript{5}. Moreover, CeD patients show a persistent activation of the NFκB pathway in the intestinal mucosa\textsuperscript{71} as well as a significant increase in the methylation level of 8 genes that belong to this pathway\textsuperscript{72}. Thus, these results indicate that CeD patients present with a major defect in the NFκB signalling complex.

For practical reasons, most prioritization studies have been focused on incorporating cis-eQTLs\textsuperscript{73} and have mostly ignored trans-eQTLs, thus potentially missing long-distance co-
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease

regulated interactions. In our study, we took advantage of a large transcriptomics cohort to run a trans mediation analysis for CeD loci. One of the most remarkable findings of this approach was that 40 trans-mediated genes were found to be controlled by a single gene: TRAFD1. These 40 genes are enriched for IFNγ and MHC I antigen processing/presentation signalling pathways. Interestingly, gsCD4+ T cells exhibit a Th1 profile and produce a large amount of IFNγ, one of the most predominant cytokines in CeD. Some of the most striking effects of IFNγ include induction of apoptosis in intestinal epithelial cells, alteration of intestinal permeability and activation of monocytes and dendritic cells, which may act as antigen-presenting cells for gsCD4+ T cells.

TRAFD1 is thought to be a regulator of the NFκB signalling pathway, suggesting that CeD-risk SNPs may modulate the NFκB complex via both cis and trans regulatory mechanisms. Our results also point to a role for TRAFD1 in response to IFNγ; however, IFNγ does not typically activate the NFκB signaling and the IFNγ locus is not associated with CeD. Thus, TRAFD1 may activate the production of other cytokines, which in turn activate the NFκB complex.

IELs, which are the effector cells in CeD, have not thus far been genetically associated with the disease. However, given that MHC I antigen presentation presentation/processing are essential for IEL activation and the striking activation of the 40 trans mediated genes in IELs upon IFN stimulation, we propose that the IELs are also genetically linked to the disease through the action of TRAFD1.

Despite the approaches implemented in our study to uncover the novel gene interactions and biological pathways that may underlie the disease, a major drawback is the limited genome coverage of the CeD summary statistics used in this study. These were derived from a GWAS that used the Immunochip platform, a genotyping platform that only measures genotypes in regions known to be associated with immune function. We thus acknowledge that our current interpretation of CeD loci is biased toward immune-related mechanisms. Only when comprehensive whole-genome CeD association analyses have been performed will we have an unbiased understanding of the disease pathophysiology.

In our gene prioritization we observed that the different statistical gene prioritization methods applied to our data prioritized both concordant and unique genes. Therefore, we propose that investigators incorporate multiple methodologically orthogonal gene prioritization methods to identify a more comprehensive set of causal genes for a given disease. Here, we use two different (orthogonal) expression datasets (BIOS and DEPICT) and three prioritization methods using the same underlying data: MR-IVW, LD-overlap and COLOC. While we believe that the genes in this study represent the most robustly prioritized genes for CeD to date, it is difficult to validate if all the prioritized genes are truly causal based on statistical methodology alone. Functional validation of these genes in disease context is needed to rule out false positives.
The functional validation of TRAFD1 in the siRNA KD experiment in THP-1 cells does establish that this gene regulates the trans-mediated network identified by our eQTL and statistical analysis. Still, the effects of the SCR control and the transfection itself may have obscured some specific TRAFD1-mediated effects. Moreover, the CeD-associated effects of TRAFD1 may not be most pronounced in monocytes or upon LPS-stimulation. Indeed, context- and cell-type-specific effects of CeD-associated genetic variation may hamper the identification of the downstream effects of the prioritized cis- and trans-genes.

In conclusion, this study provides a framework for predicting candidate genes and their function using a systematic in silico approach that could be extended to other complex diseases. Using this approach, we not only confirmed previous association between adaptive cells (gscD4+ T cells and B cells) and CeD, we also unveiled a link between specific genes that may contribute to the disease via innate immune cells, epithelial cells and IELs. Finally, we found a gene network controlled by TRAFD1 that is part of two major pathways of immune activation, IFNγ signalling and MHC I antigen processing.

Acknowledgements.
We would like to thank Bana Jabri for providing IEL cell lines, Morris Swertz for data storage and cluster facilities and Kate McIntyre for editing the manuscript. This work was supported by an ERC Advanced grant [FP/2007-2013/ERC grant 2012-322698] and an NWO Spinoza prize grant [NWO SPI 92-266] to C.W. I.J is supported by a Rosalind Franklin Fellowship from the University of Groningen and an NWO VIDI grant [016.171.047].

Data Availability
Summary statistics of the CeD GWAS are available from the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega/studies/EGAS00001003805) under accession number EGAS00001003805. The individual-level data of the BIOS cohort is available upon request from https://www.bbmri.nl/acquisition-use-analyze/bios.
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease

References

1. Kuja-Halkola, R. et al. Heritability of non-HLA genetics in coeliac disease: A population-based study in 107 000 twins. Gut. 65, 1793–1798. (2016).
2. Bevan, S. et al. Contribution of the MHC region to the familial risk of coeliac disease. J. Med. Genet. 36, 687–690 (1999).
3. Trynka, G. et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. 43, 1193–1201 (2011).
4. Dubois, P. C. A. et al. Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet. 42, 295–302 (2010).
5. Ricaño-Ponce, I. et al. Systematic annotation of celiac disease loci refines pathological pathways and suggests a genetic explanation for increased interferon-gamma levels. Hum. Mol. Genet. 24, 397–409 (2015).
6. Ricaño-Ponce, I. et al. Refined mapping of autoimmune disease associated genetic variants with gene expression suggests an important role for non-coding RNAs. J. Autoimmun. 68, 62–74 (2016).
7. Kumar, V. et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. 7, (2011).
8. Jonkers, I. H. & Wijmenga, C. Context-specific effects of genetic variants associated with autoimmune disease. Hum. Mol. Genet. 26, R85–R192 (2017).
9. Fernandez-Jimenez, N. & Bilbao, J. R. Mendelian randomization analysis of celiac GWAS reveals a blood expression signature with diagnostic potential in absence of gluten consumption. Hum. Mol. Genet. 28, 3037–3042 (2019).
10. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).
11. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet. Epidemiol. 37, 658–665 (2013).
12. Graaf, A. van der et al. A novel mendelian randomization method identifies causal relationships between gene expression and low-density lipoprotein cholesterol levels. bioRxiv 671537 (2019).
13. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. 6, 5890 (2015).
14. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. 49, 139–145 (2017).
15. van Greevenbroek, M. M. J. et al. The cross-sectional association between insulin resistance and circulating complement C3 is partly explained by plasma alanine aminotransferase, independent of central obesity and general inflammation (the CODAM study). Eur. J. Clin. Invest. 41, 372–379 (2011).
16. Deelen, J. et al. Employing biomarkers of healthy ageing for leveraging genetic studies into human longevity. Exp. Gerontol. 82, 166–174 (2016).
17. Tigchelaar, E. F. et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: Study design and baseline characteristics. BMJ Open 5, e006772 (2015).
18. Lin, B. D. et al. The genetic overlap between hair and eye color. Twin Res. Hum. Genet. 19, 595–599 (2016).
19. Huisman, M. H. B. et al. Population-based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J. Neurol. Neurosurg. Psychiatry 82, 1165–1170 (2011).
20. Hofman, A. et al. The Rotterdam Study: 2016 objectives and design update. Eur. J. Epidemiol. 30, 661–708 (2015).
21. Võsa, U. et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. bioRxiv 18, 447367 (2018).
22. Das, S. et al. Next-generation genotype imputation service and methods. Nat. Genet. 48, 1284–1287 (2016).
23. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
24. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
25. Boomsma, D. I. et al. The genome of the Netherlands: design, and project goals. Eur. J. Hum. Genet. 22, 221–227 (2014).
26. Anders, S., Pyl, P. T. & Huber, W. HTSeq–A Python framework to work with high-throughput
sequencing data. Bioinformatics 31, 166–169 (2015).

29. Yang, J., Lee, S. H., Goddard, M. E. & Vis sinks, P. M. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011).

30. Auton, A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).

31. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur. J. Epidemiol. 32, 377–389 (2017).

32. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, 369–375 (2012).

33. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet. Epidemiol. 37, 658–665 (2013).

34. Bowden, J., Hemani, G. & Davey Smith, G. Invited commentary: Detecting individual and global horizontal pleiotropy in mendelian randomization—a job for the humble heterogeneity statistic? Am. J. Epidemiol. 187, 2681–2685 (2018).

35. Deelen, P. et al. Improving the diagnostic yield of exome-sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. Nat. Commun. 10, 2837 (2019).

36. Freedman, L. S. & Schatzkin, A. Sample size for studying intermediate endpoints within intervention trials or observational studies. Am. J. Epidemiol. 136, 1148–1159 (1992).

37. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128 (2013).

38. Wishart, D. S. et al. DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Res. 46, D1074–D1082 (2018).

39. Whirl-Carrillo, M. et al. Pharmacogenomics knowledge for personalized medicine. Clin. Pharmacol. Ther. 92, 414–417 (2012).

40. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. Sci. Transl. Med. 9, eaag1166 (2017).

41. Kim, D., Langmead, B. & Salzberg, S. L. HI-SAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357 (2015).

42. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

43. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010).

44. Jabri, B. et al. Selective expansion of intraepithelial lymphocytes expressing the HLA-E-specific natural killer receptor CD94 in celiac disease. Gastroenterology 118, 867–879 (2000).

45. Petersen, J. et al. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. Nat. Struct. Mol. Biol. 21, 480–488 (2014).

46. Ricaño-Ponce, I. et al. Immunochip meta-analysis in European and Argentinian populations identifies two novel genetic loci associated with celiac disease. Eur. J. Hum. Genet. 1–11 (2019).

47. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. Sci. Transl. Med. 9, eaag1166 (2017).

48. Chiu, Y. G. & Ritchlin, C. T. Denosumab: targeting the RANKL pathway to treat rheumatoid arthritis. Expert Opin. Biol. Ther. 17, 119–128 (2017).

49. Lynch, J. P., Metz, D. C., Rutgeerts, P., Vermeire, S. & Assche, G. Van. Biological therapies for inflammatory bowel diseases. Gastroenterology 136, 1182–1197 (2009).

50. Baldassari, L. E. & Rose, J. W. Daclizumab: Development, clinical trials, and practical aspects of use in multiple sclerosis. Neurotherapeutics 14, 842–858 (2017).

51. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128 (2013).

52. Goruppi, S. et al. The ULK3 kinase is critical for convergent control of cancer-associated fibroblast activation by CSL and GLI. Cell Rep. 20, 2468–2479 (2017).

53. Yang, T. et al. Knockdown of C-terminal Src kinase by siRNA-mediated RNA interference augments T cell receptor signaling in mature T cells. Eur. J. Immunol. 34, 2191–2199 (2004).

54. Abadie, V., Discenzo, V. & Jabri, B. Intraepithelial lymphocytes in celiac disease immunopathology. Semin. Immunopathol. 34, 551-66 (2012).

55. Nilsen, E. M. et al. Gluten specific, HLA-DQ restricted T cells from coelic mucosa produce cytokines with Th1 or Th0 profile dominated by interferon γ. Gut 37, 766–776 (1995).

56. Wapenaar, M. C. et al. The interferon gamma gene in celiac disease: Augmented expression correlates with tissue damage but no evidence for genetic susceptibility. J. Autoimmun. 23,
Systematic prioritization of candidate genes in disease loci identifies *TRAFO1* as a master regulator of IFNg signalling in celiac disease

1. Linterman, M. A. *et al.* IL-21 acts directly on B cells to regulate Bcl-6 expression and germinal center responses. J. Exp. Med. 15, 353-63 (2010).
2. Dinarello, C. A., Novick, D., Kim, S. & Kaplan, G. Interleukin-18 and IL-18 binding protein. Front. Immunol. 4, 289 (2013).
3. Mahnke, J. *et al.* Interferon regulatory factor 4 controls Th1 cell effector function and metabolism. Sci. Rep. 6, 35521 (2016).
4. Grenningloh, R., Bok, Y. K. & Ho, I. C. Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses. J. Exp. Med. 201, 615–626 (2005).
5. Liu, T., Zhang, L., Joo, D. & Sun, S. C. NF-κB signaling in inflammation. Signal Transduction and Targeted Therapy, 2 (2017).
6. Maiuri, M. C. *et al.* Nuclear factor κB is activated in small intestinal mucosa of celiac patients. J. Mol. Med. 81, 373–379 (2003).
7. Fernandez-jimenez, N. *et al.* Coregulation and modulation of NFκB-related genes in celiac disease: Uncovered aspects of gut mucosal inflammation. Hum. Mol. Genet. 23, 1298–1310 (2014).
8. Brynedal, B. *et al.* Large-Scale trans-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. Am. J. Hum. Genet. 100, 581–591 (2017).
9. Beitnes, A. C. R. *et al.* Rapid accumulation of CD14+CD11c+ dendritic cells in gut mucosa of celiac disease after in vivo gluten challenge. PLoS One. 7, e33556 (2012).
10. Deb, A., Haque, S. J., Mogensen, T., Silverman, R. H. & Williams, B. R. G. RNA-dependent protein kinase pkr is required for activation of NF-κB by IFN-γ in a STAT1-independent pathway. J. Immunol. 166, 6170-80 (2001).
Supplementary data

**Supplementary Fig. 1** Mediation effect on trans genes for all prioritized genes in the TRAFD1 region on chromosome 12. (A) Three boxes with the eQTL association curves of TRAFD1, SERPING1 and SERPING1 after mediation with TRAFD1. (B) Scatter plot indicating the absolute Z indicating the absolute Z difference between unmediated and mediated trans associations upon mediation (y axis) by the prioritized cis genes shown on the x axis and when correcting TRAFD1 expression for the expression of SH2B3 (‘TRAFD1 – SH2B3’).

**Supplementary Fig. 2** TRAFD1 knockdown validation. Cell viability (A) and proliferation (B) of THP-1 cells that were left untransfected (WT) or transfected with non-targeting siRNA (SCR) or siRNA targeting TRAFD1 (KD) for 72 hours. Protein and mRNA levels of TRAFD1 were determined by WB (C, D) and qPCR (E). Bars indicate mean ± SEM. Data are representative of three different experiments. Statistical differences were calculated with a one-sided t-test by using the SCR as 100% reference. p-value ≤ 0.0001 (****). (F) The differential expression analysis approach. Here we compared the gene expression between unstimulated samples and their respective LPS-stimulated samples to identify DEGs that respond to stimulation (|log2 FC| >1 and FDR ≤ 0.01). We then identified unique or shared DEGs responding to the stimulation between treatments (WT, SCR or KD), which are shown in two separate Venn diagrams: one for upregulated genes (G) and one for downregulated genes (H).
Systematic prioritization of candidate genes in disease loci identifies **TRAFD1** as a master regulator of IFNγ signalling in celiac disease

**Supplementary Fig. 3 Expression pattern of TRAFD1-mediated genes upon TRAFD1 knockdown.** Heatmap showing the pattern of gene expression of **TRAFD1** and of the 40 genes it mediates, scaled by row (see details in Methods and Fig. 5). Expression is shown in different treatments and stimulations as indicated by coloured bars on top of the heatmap.
Systematic prioritization of candidate genes in disease loci identifies TrafD1 as a master regulator of IFNγ signalling in celiac disease

Supplementary Fig. 4 DEGs upon TrafD1 knockdown are enriched in TrafD1-mediated genes. Here we compare the differential expression of 41 genes found in the trans mediation analysis of TrafD1 (40 trans-mediated genes and TrafD1) with the differential expression of 41 other randomly chosen genes. The histograms (blue) show the distribution of the median absolute T statistic of DEseq of 41 randomly chosen genes, when 500,000 sets are randomly chosen, compared to the observed value for the 41 genes that are from the trans-mediation analysis (red horizontal line). We compare the results of the control experiment (WT-SCR) in panels A and C with the results of the knockdown experiment (SCR-KD) in panels B and D. The fold differences between the control experiments and the knockdown experiments show how much more than expected the 41 trans-mediated genes are differentially expressed in the knockdown compared to the control.

| Region | Gene_name | MLD | n_eqtl | MR_Lw_eqtl | MR_Lw_pval | LD_overlap_sum | Coloc_h4 | Depict_fdr | L_evidence |
|--------|-----------|-----|--------|------------|-------------|----------------|----------|------------|------------|
| 15:74115895-76115895 | CSK | - | 3 | -0.5099 | 9.37e-09 | cojo_snp:0.92 | 1.00 | Yes | 4/4 |
| 1:182300050-184546061 | NCF2 | - | 3 | -0.7808 | 4.03e-14 | cojo_snp:1.00 | 0.00 | Yes | 3/4 |
| 18:66543688-68543688 | CD22 | + | 3 | 0.1694 | 4.69e-04 | top_snp:0.92 | 0.96 | Yes | 3/4 |
| 11:110196858-112196858 | COLCA1 | + | 4 | 0.0659 | 3.38e-05 | cojo_snp:1.00 | 1.00 | NA | 3/3 |
| 12:110715197-113906415 | TrafD1 | - | 2 | NA | NA | top_snp:1.00 | 1.00 | Yes | 3/3 |
| 15:78234957-80234957 | CTSH | + | 4 | 0.2165 | 2.51e-10 | top_snp:1.00 | 1.00 | NA | 3/3 |
| 20:29150077-31150077 | HM13 | + | 1 | NA | NA | top_snp:1.00 | 1.00 | Yes | 3/3 |
| 1:1505713-3539006 | FAM213B | - | 5 | -0.1326 | 3.35e-01 | cojo_snp:0.91 | 0.00 | Yes | 2/4 |
| 2:101966067-104125457 | IL18RAP | + | 3 | 0.1291 | 8.73e-10 | No LD | 0.00 | Yes | 2/4 |
| 2:101966067-104125457 | IL18R1 | + | 4 | 0.1725 | 1.83e-05 | No LD | 0.00 | Yes | 2/4 |
| 22:36630862-38633851 | RAC2 | + | 3 | 0.1097 | 4.47e-02 | cojo_snp:0.88 | 0.23 | Yes | 2/4 |
| 21:42855087-4638504 | UBA4H3A | + | 7 | 0.0144 | 9.10e-02 | cojo_snp:1.00 | 0.00 | Yes | 2/4 |
| 2:18767082-105765082 | APF3 | - | 3 | -0.1511 | 1.71e-04 | No LD | 0.91 | No | 2/4 |
| 18:66543688-68543688 | DOK6 | - | 1 | NA | NA | top_snp:0.94 | 0.95 | No | 2/3 |
| 15:74115895-76115895 | ULK3 | - | 1 | NA | NA | top_snp:0.83 | 1.00 | No | 2/3 |
| 6:0-1412741 | MYF4 | + | 1 | NA | NA | No LD | 0.98 | Yes | 2/3 |
| 16:755479-904667 | PARK7 | - | 2 | NA | NA | top_snp:0.93 | 0.94 | No | 2/3 |
| 2:42359275-44359275 | AC016735.2 | - | 4 | -0.1405 | 4.12e-03 | cojo_snp:0.94 | 0.97 | NA | 2/3 |
| 15:74115895-76115895 | MPR | + | 2 | NA | NA | top_snp:0.83 | 0.97 | No | 2/3 |
| 1:19968130-201951307 | Clorf106 | + | 2 | NA | NA | top_snp:0.89 | 0.99 | No | 2/3 |
| 1:19968130-201951307 | GP102 | + | 3 | 0.1543 | 1.71e-06 | No LD | 0.98 | NA | 2/3 |
| 9:0-1244657 | DOCK8 | 1 | 2 | NA | NA | cojo_snp:1.00 | 0.01 | Yes | 2/3 |
| 3:45174684-47543272 | F11P1 | + | 4 | 0.1069 | 8.67e-13 | cojo_snp:0.82 | 0.00 | NA | 2/3 |
Systematic prioritization of candidate genes in disease loci identifies \textit{TRAFD1} as a master regulator of IFN\(\gamma\) signalling in celiac disease
Supplementary Table 1. Prioritization of genes likely causal for celiac disease (CeD). This table contains all the genes in the prioritized CeD regions and their evidence for being causal to CeD. One gene per row is shown. Columns (in order): the human build 37 coordinates of the CeD region in which the gene is located (Region); the gene name according to the ENSEMBL GENES 96 database (human build 37) (Gene_name); the most likely effect direction (determined as described in Methods) (Effect_dir); the effect size (MR_ivw_effect) and p value (MR_ivw_pval) of the MR-IVW test; the summary of LD overlap (Ld_overlap_sum), with either the top eQTL variant (top_snp) or an independent eQTL variant (cojo_snp) with the r^2 linkage disequilibrium between the eQTL and the CeD top variant; the coloc posterior probability of causal variants being shared (Coloc_h4); if the gene passes DEPICTs own false discovery thresholds (Depict_fdr); and the lines of evidence that are significant compared to the lines of evidence that are available for a gene (L_evidence). Bold fields in any of the columns indicate that the prioritization method is significant according to our thresholds.

| Gene_name | Tier | Db_Target | Region | Effect_dir | MR_ivw_effect | MR_ivw_pval | Ld_overlap_sum | n_eqtl | Depict_fdr | L_evidence |
|-----------|------|-----------|--------|------------|---------------|-------------|----------------|--------|------------|------------|
| IL21      | 3A   | NA        | 12:110715197-113906415 | NA         | 0.59          | NA          | Yes            | 1/2    |            |            |
| PDXK      |      | NA        | 8:128264589-130264589  | NA         | 0.94          | NA          | Yes            | 1/2    |            |            |
| IL18RAP   | 3A   | NA        | 6:9808352-91978383     | NA         | 0.98          | NA          | No eQTL        | 1/2    |            |            |
| CTLA4     | 1    | Ipilimumab| 2:181007800-183135924  | NA         | 0.86          | NA          | No eQTL        | 1/2    |            |            |
| SMAD3     | 2    | NA        | 12:5511996-7511996     | NA         | 0.98          | NA          | No eQTL        | 1/2    |            |            |
| FASLG     | 3B   | NA        | 4:121974607-124561436  | NA         | 1.00          | NA          | No eQTL        | 1/2    |            |            |
| CTQTNF6   | 3A   | NA        | 2:100875901-192940451  | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| IL1RL1    | 3A   | NA        | 3:158682758-160793888  | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| CSK       | 1    | Dasatinib;Staurosporine;TG-100801;Fostamatinib | 12:110715197-113906415 | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| NCF2      |      | Dextromethorphan| 12:110715197-113906415 | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| PARK7     |      | Copper      | 12:110715197-113906415 | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| IL2RA     | 3A   | NA        | 12:5511996-7511996     | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| IL21R     | 3A   | NA        | 12:5511996-7511996     | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| MPI       | 2    | NA        | 12:5511996-7511996     | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
| PTPIN11   | 2    | Dodecane-Trimethylamine | 12:110715197-113906415 | NA         | 0.99          | NA          | No eQTL        | 1/2    |            |            |
Systematic prioritization of candidate genes in disease loci identifies **TRAFD1** as a master regulator of IFN\(\gamma\) signalling in celiac disease

**CXCR6** Tier_3B NA
**IL12A** Tier_3A NA
**HM13** NA Theophylline
**IL2** Tier_1 Pseudoephedrine;Cefazolin;SP4160;Girentuximab;Apremilast;Tapinarof;TG4010
**GPR183** Tier_3B NA
**CCR2** Tier_1 INCB3284;CCX915;Ptozalizumab
**STAT1** Tier_2 NA
**IL18R1** Tier_3A NA
**TNFRSF14** Tier_3A NA
**CD274** Tier_1 Atezolizumab;Durvalumab;Avelumab
**RAC2** NA Dextromethorphan
**SH2B3** NA Pazopanib
**CYP11A1** Tier_1 Aminogluthethimide
**BCL2L1** Tier_1 "4'-FLUORO-11'-BIPHENYL-4-CARBOXYLIC_ACID";Isosorbide;Gossypol
**PP1F** Tier_2 Ciclosporin;Proline;Triglyme;7-AMINO-4-METHYL-CHROMEN-2-ONE
**CCR3** Tier_2 NA
**CD226** Tier_3A NA
**GPR25** Tier_3B NA
**PLTP** Tier_3A NA
**CCR1** Tier_1 NA
**SLC22A4** Tier_1 NA
**SDSL** NA Pyridoxal phosphatase
**COX5A** NA Cholic_Acid;N-Formylmethionine
**PDCD1LG2** Tier_1 NA
**PTPN2** Tier_2 NA
**MMP9** Tier_1 Glutathione;Marinastat;Minocycline;Captopril;Glucosamine;Zinc.
**ITGAA** Tier_1 Natalizumab;ATL1102;CDP323;R1295;R411;Tinzapan;Vedolizumab
**MMEL1** Tier_3A NA
**CD28** Tier_1 NA
**TNFRSF9** Tier_1 Urelumab

**Supplementary Table 2.** This table contains all the prioritized cis genes in the CeD regions that are existing drug targets according to two different databases (DrugBank v5.1.4, and Finan et al. 2017). One gene per row is shown. Columns indicate (in order): the gene name according to the ENSEMBL GENES 96 database (human build 37) (**Gene_name**), the druggability tier based on Finan et al, with lower tiers making it more likely that the gene is druggable \(^4\), (Tier), and if the gene is a drug bank drug target (Db_target). Bold fields in any of the columns indicate that the prioritization method is significant according to our thresholds.

| Gene_name | Cluster_# | Gene_name | Cluster_# | Gene_name | Cluster_# | Gene_name | Cluster_# |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| ARHGAP21  | 1         | CD224     | 2         | CCDC116   | 2         | CCR3      | 4         |
| PIAS1     | 1         | TNFRSF11  | 4         | REL       | 4         | UBE2L3    | 2         |
| TNFRSF9   | 4         | CCR2      | 4         | PUS10     | 4         | COLCA1    | 2         |
| RGS1      | 4         | GPR18     | 4         | C1orf106  | 4         | SLC22A4   | 2         |
| PLTP      | 2         | IRF1      | 2         | CTLA4     | 3         | C5orf56   | 2         |
| MMP9      | 4         | RAC2      | 4         | ICOS      | 3         | PDCD1LG2  | 2         |
| HM13      | 2         | C1QTNF6   | 2         | FYCO1     | 2         | MIER1     | 2         |
| IL21R     | 1         | IL2RA     | 4         | CCR1      | 4         | TCTN1     | 2         |
| AAGAB     | 1         | UBAC2     | 2         | RTP3      | 2         | DOK6      | 2         |
| CSK       | 1         | ET51      | 4         | TAGAP     | 4         | TTC34     | 2         |
| CTSH      | 2         | TRAFD1    | 2         | SMAD3     | 2         | RP3-399M20.9 | 2      |
| DOCK8     | 4         | IRF4      | 4         | IL12A     | 2         | ADAM1A    | 2         |
| PP1F      | 2         | STAT4     | 4         | GPR183    | 4         | RP11-168O16.1 | 4      |
| IL2       | 3         | IL21      | 4         | UBE2E3    | 2         | FLT1P1    | 2         |
| POU2AF1   | 4         | PIANP     | 2         | GPR25     | 2         | ACO16735.2 | 2         |
| SH2B3     | 2         | SDSL      | 2         | BCL2L1    | 2         | ACO104820.2 | 2        |
| BACH2     | 4         | CYP11A1   | 4         | CXCR6     | 4         | ACO07278.2 | 2         |
| FRMD4B    | 4         | UK3       | 1         | RASGRPR1  | 4         | RP11-243J16.7 | 2      |
| ITGA4     | 4         | SCAMP2    | 1         | THEMIS    | 3         | MIL1F1P1  | 2         |
| STAT1     | 2         | MMEL1     | 2         | CBWD1     | 2         | AC115366.5 | 2         |
| ILIR1     | 4         | AFF3      | 2         | HECTD4    | 2         | RP3-399M20.8 | 2      |
| IL18R1    | 4         | SESN3     | 4         | AHSA2     | 2         | IQCJ-SCHIP1-JS1 | 2      |
Systematic prioritization of candidate genes in disease loci identifies \textit{TRAFD1} as a master regulator of IFN\(_\gamma\) signalling in celiac disease.

Supplementary Table 3. The 118 prioritized genes were assigned to a cluster based on a guilt-by-association co-regulation approach to find shared biological mechanisms. For each gene that was prioritized (Gene\_name), a cluster membership is given (Cluster\#).

| Gene\_name | Cluster# |
|------------|----------|
| IL18RAP    | 4        |
| ERRF1      | 2        |
| PARK7      | 2        |
| SMG7       | 2        |
| NCF2       | 4        |
| SLC33D1    | 2        |
| KIF21B     | 4        |
| FASLG      | 3        |
Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNg signalling in celiac disease.
1 Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
2 Department of Medicine, University of Chicago, Chicago, USA
3 Committee on Immunology, University of Chicago, Chicago, USA
4 School of Medicine and Surgery, University of Milano-Bicocca, Italy
5 Department of Gastroenterology and Hepatology, University Medical Center, Groningen, University of Groningen, Groningen, the Netherlands
6 Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, the Netherlands
7 Department of computational Biology for Individualised Infection Medicine, Centre for Individualised Infection Medicine, Helmholtz Centre for Infection Research, Hannover Medical School, Hannover, Germany
8 Department of Immunology, K.G. Jebsen Coeliac Disease Research Centre, University of Oslo, Oslo, Norway