

University of Groningen

## Potential impact of celiac disease genetic risk factors on T cell receptor signaling in gluten-specific CD4+ T cells

Bakker, Olivier B; Ramírez-Sánchez, Aarón D; Borek, Zuzanna A; de Klein, Niek; Li, Yang; Modderman, Rutger; Kooy-Winkelaar, Yvonne; Johannesen, Marie K; Matarese, Filomena; Martens, Joost H A

*Published in:*  
Scientific Reports

*DOI:*  
[10.1038/s41598-021-86612-5](https://doi.org/10.1038/s41598-021-86612-5)

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

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

*Publication date:*  
2021

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Bakker, O. B., Ramírez-Sánchez, A. D., Borek, Z. A., de Klein, N., Li, Y., Modderman, R., Kooy-Winkelaar, Y., Johannesen, M. K., Matarese, F., Martens, J. H. A., Kumar, V., van Bergen, J., Qiao, S-W., Lundin, K. E. A., Sollid, L. M., Koning, F., Wijmenga, C., Withoff, S., & Jonkers, I. H. (2021). Potential impact of celiac disease genetic risk factors on T cell receptor signaling in gluten-specific CD4+ T cells. *Scientific Reports*, *11*(1), 1-15. [9252]. <https://doi.org/10.1038/s41598-021-86612-5>

### **Copyright**

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

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

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



OPEN

## Potential impact of celiac disease genetic risk factors on T cell receptor signaling in gluten-specific CD4+ T cells

Olivier B. Bakker<sup>1,9</sup>, Aarón D. Ramírez-Sánchez<sup>1,9</sup>, Zuzanna A. Borek<sup>1,9</sup>, Niek de Klein<sup>1,9</sup>, Yang Li<sup>1,9</sup>, Rutger Modderman<sup>1</sup>, Yvonne Kooy-Winkelaar<sup>2</sup>, Marie K. Johannesen<sup>3,4</sup>, Filomena Matarese<sup>5</sup>, Joost H. A. Martens<sup>1,5</sup>, Vinod Kumar<sup>1,6,7</sup>, Jeroen van Bergen<sup>2</sup>, Shuo-Wang Qiao<sup>1,3,4</sup>, Knut E. A. Lundin<sup>1,3,8</sup>, Ludvig M. Sollid<sup>1,3,4</sup>, Frits Koning<sup>1,2</sup>, Cisca Wijmenga<sup>1,3</sup>, Sebo Withoff<sup>1</sup> & Iris H. Jonkers<sup>1</sup>✉

Celiac disease is an auto-immune disease in which an immune response to dietary gluten leads to inflammation and subsequent atrophy of small intestinal villi, causing severe bowel discomfort and malabsorption of nutrients. The major instigating factor for the immune response in celiac disease is the activation of gluten-specific CD4+ T cells expressing T cell receptors that recognize gluten peptides presented in the context of HLA-DQ2 and DQ8. Here we provide an in-depth characterization of 28 gluten-specific T cell clones. We assess their transcriptional and epigenetic response to T cell receptor stimulation and link this to genetic factors associated with celiac disease. Gluten-specific T cells have a distinct transcriptional profile that mostly resembles that of Th1 cells but also express cytokines characteristic of other types of T-helper cells. This transcriptional response appears not to be regulated by changes in chromatin state, but rather by early upregulation of transcription factors and non-coding RNAs that likely orchestrate the subsequent activation of genes that play a role in immune pathways. Finally, integration of chromatin and transcription factor binding profiles suggest that genes activated by T cell receptor stimulation of gluten-specific T cells may be impacted by genetic variation at several genetic loci associated with celiac disease.

In celiac disease (CeD), cereal-derived gluten peptides penetrate the small intestinal barrier, are subsequently modified by tissue trans-glutaminase 2 (TG2), then presented by HLA-DQ2- or HLA-DQ8-positive antigen-presenting cells (APCs) to gluten-specific CD4+ T-helper cells (gsTcells)<sup>1</sup>. This leads to robust activation of gsTcells that subsequently stimulate B cells to start producing auto antibodies to TG2 and deamidated gluten peptides<sup>1,2</sup> and activate CD8+ intraepithelial lymphocytes (IELs) to attack intestinal epithelial cells, leading to the villous atrophy that is characteristic of CeD. GsTcells are only found persistently in CeD patients<sup>3,4</sup> and can induce villous atrophy in patients upon gluten ingestion even after these individuals have been on a gluten-free diet for years<sup>5</sup>. Activation of gsTcells is thus central to CeD onset and pathology.

GsTcells have been shown to secrete many signaling molecules upon stimulation, including interleukin (IL)-2, IL-4, IL-6, IL-8, IL-10, IL-21, CD40LG, IFN $\gamma$  and TNF, and are often classified to be of type 1 helper class<sup>6-12</sup>. GsTcells uniquely express IL-21 and CXCL13, as well as several other markers characteristic of follicular and

<sup>1</sup>Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. <sup>2</sup>Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands. <sup>3</sup>K.G. Jebsen Coeliac Disease Research Centre, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. <sup>4</sup>Department of Immunology, Oslo University Hospital, Rikshospitalet, Oslo, Norway. <sup>5</sup>Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands. <sup>6</sup>Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, The Netherlands. <sup>7</sup>Nitte (Deemed to be University), Division of Infectious Diseases, Nitte University Centre for Science Education and Research (NUCSER), Paneer Campus, Deralakatte, Mangaluru 575018, India. <sup>8</sup>Department of Gastroenterology, Oslo University Hospital, Rikshospitalet, Oslo, Norway. <sup>9</sup>These authors contributed equally: Olivier B. Bakker, Aarón D. Ramírez-Sánchez and Zuzanna A. Borek. ✉email: i.h.jonkers@umcg.nl

regulatory T cells<sup>13</sup>. IL-21 and CXCL13, together with CD40LG and IL-4, play an important role in the interaction, differentiation and activation of T cells and plasma B cells<sup>13–16</sup>. Cytokines secreted by gsTcells are also important for activation and proliferation of IELs, in combination with IL-15, a cytokine important in CeD etiology that is produced by IELs<sup>17–20</sup>. GsTcells are thus central in the response to gluten peptides that leads to inflammation, anti-TG2 antibody production and villous atrophy in CeD.

To date, 43 genetic risk factors have been associated with CeD<sup>21–23</sup>, the most important being the HLA haplotypes HLA-DQ2 and -DQ8. While the role of HLA-DQ2 and -DQ8 in CeD is well defined<sup>24–26</sup>, the contribution of the non-HLA CeD risk-loci is mostly unclear. More than 95% of the single nucleotide polymorphisms (SNPs) associated with CeD are located in the non-coding genome and presumably deregulate genes important for CeD etiology<sup>27</sup>. Enrichment analysis of the CeD SNPs in regulatory regions suggests that CD4+ T cells are the major cell type affected by genetic risk factors<sup>28–30</sup>. Moreover, pathway and *cis*-eQTL analyses of genes in CeD loci suggest that they affect T cell receptor (TCR) signaling via alteration of expression of genes such as *UBASH3A*, *CD28* and *CSK*<sup>30–32</sup>. Overall, these observations confirm the importance of CD4+ T cell activation in CeD but do not delineate how CeD-associated SNPs affect gsTcells upon activation.

This knowledge gap is partially due to an incomplete understanding of the regulation of the response to stimulation in gsTcells. Recently, it was shown that the genetic risk loci associated with CeD are enriched with binding sites of specific transcription factors (TFs), including STAT4, STAT5A, STAT5B, T-BET, AP-1 subunit FOS and TFs from the NFκB signaling pathway<sup>29</sup>. Indeed, many of these TFs have been implicated in regulation of CD4+ T cell activation or in CeD<sup>12,33–36</sup>. However, the role of these TFs, as well as the dynamic transcriptional and epigenetic response in the activation of gsTcells, has not been described. Nor has the role of CeD-associated genetic variants in these dynamic transcriptional processes been explored in gsTcells.

Here, we set out to profile the transcriptomic and epigenetic response of gsTcells derived from CeD patients upon TCR-stimulation with anti-CD3 and anti-CD28 (aCD3/aCD28). This allowed us to identify the regulatory steps essential for the rapid and robust activation of cytokines important for CeD etiology and to prioritize which CeD-associated risk loci are related to the activation of gsTcells. Overall, we elucidate the dynamic events in gsTcells that can be induced by gluten peptides in CeD patients.

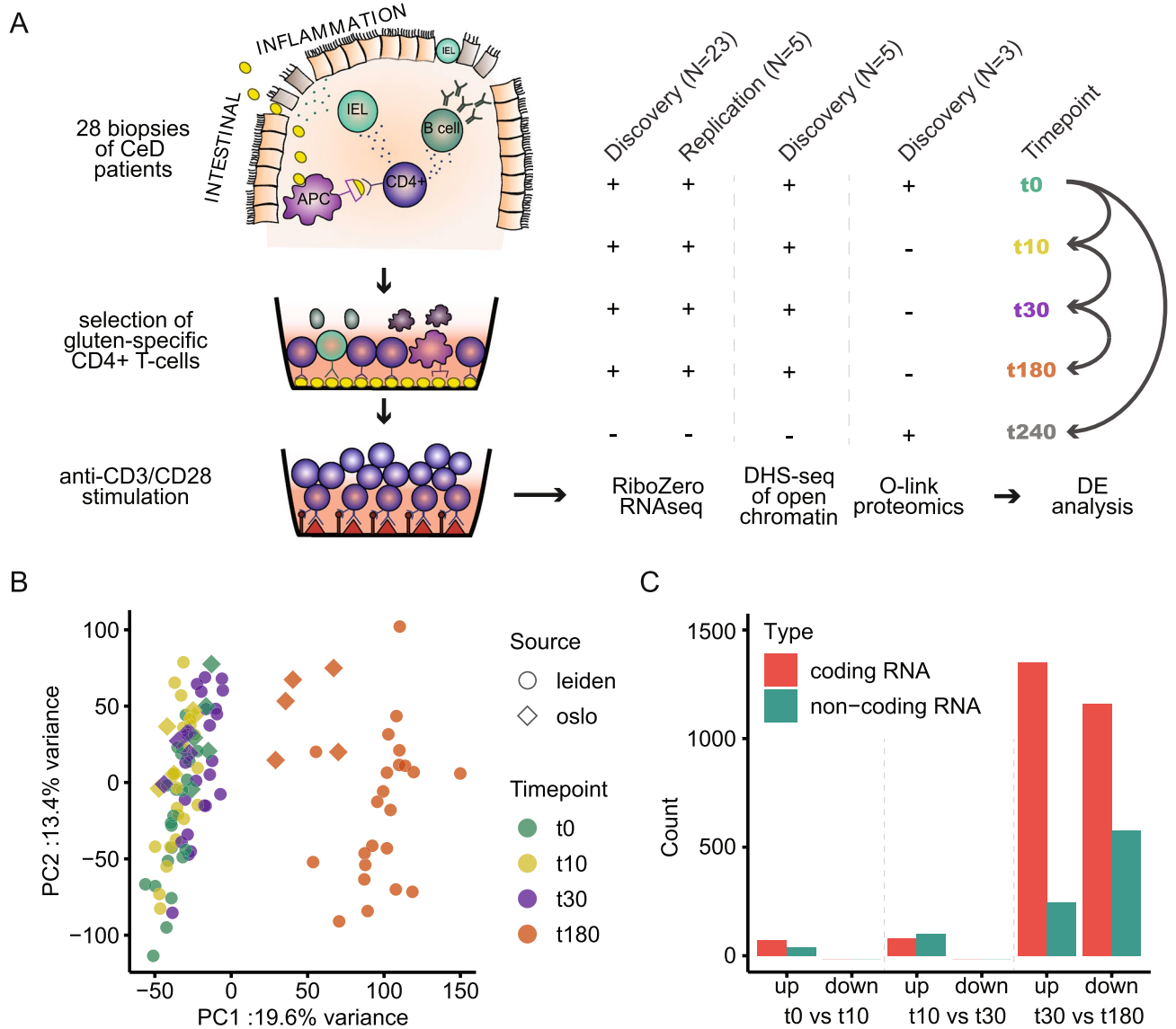
## Results

**Dynamic transcriptome changes in stimulated gsTcells.** To study the TCR-induced response of gsTcells, we opted for aCD3/aCD28 stimulation as a proxy for the interaction of gsTcell-TCR with gluten peptides presented by antigen-presenting cells in the context of HLA-DQ2 or -DQ8. Twenty-three CD4+ gsTcell clones isolated from biopsies from patients with active CeD were cultured and stimulated *in vitro* and used to perform transcriptomic (RNA-Seq; n = 23) and targeted proteomics analysis (n = 3) (discovery cohort, Suppl. Table 1). An additional five gsTcell clones were used for replication of the transcriptomic data and detection of open chromatin through DNase-Hyper-Sensitivity sequencing (DHS-sequencing) (Fig. 1A, Suppl. Table 1). The transcriptomic response of gsTcells to stimulation showed strong and consistent effects after 180 min relative to the earlier timepoints. Although there was considerable inter-clonal variation, the replication and discovery cohorts behaved very similarly (Fig. 1B). We used the discovery cohort to determine the dynamic transcriptional response of genes during the course of the stimulation, which showed clear distinctions between each time point (Suppl. Fig. 1). We performed differential expression (DE) analysis between consecutive timepoints to reveal the changes in gene expression over time. Between 0–10 min, 10–30 min and 30–180 min, 115, 182 and 3339 DE genes were identified, respectively (Fig. 1C, Suppl. Table 2). Finally, non-coding genes were found to be differentially expressed at all timepoints, but at 180 min the downregulated set was roughly twice as large as the upregulated set, which was not the case for the coding genes (Fig. 1C). GsTcell clones thus rapidly displayed strong and dynamic transcriptional changes after stimulation.

**DE genes cluster into response patterns with distinct functions.** To understand and categorize the activation of gsTcells, we clustered the DE genes using k-means clustering to identify temporal response patterns. This identified six distinct clusters that each represent a specific response (Fig. 2A, Suppl. Fig. 2, Suppl. Table 2). Genes in Clusters 1 (n = 366) and 2 (n = 162) were upregulated early (at 10 min). In contrast, cluster 3 genes (n = 1002) were upregulated after 30 and 180 min. Cluster 4 (n = 609) genes displayed an early decrease but recovered after 180 min. Cluster 5 (n = 588) genes responded similarly to cluster 1 genes, but their gene expression levels after 180 min were decreased compared to unstimulated expression levels. Finally, cluster 6 genes (n = 782) show a consistent late decrease in expression.

Gene set enrichment analysis using the Reactome database pinpointed enriched gene sets (false discovery rate (FDR) < 0.05) for clusters 2, 3, 4 and 6 (Fig. 2B) but found no enrichment in clusters 1 and 5. Cluster 2 is enriched for Nuclear Receptor transcription pathway genes (*NR4A1*, *NR4A2* and *NR4A3*) and IL10 signaling genes (*TNF* and *ICAM1*). Cluster 3 is predominantly associated with immune function, showing an enrichment of cytokine signaling pathway genes. Moreover, genes associated with transcriptional and translational processes are also enriched in cluster 3, consistent with an immune response that requires the production and secretion of cytokines and other signaling proteins. Cluster 4 is enriched in translational and mitochondrial response genes that are associated with a shift towards protein production. Finally, the downregulated genes in cluster 6 are enriched in genes involved p53-mediated regulation of caspases. Downregulation of cluster 6 genes may thus decrease apoptosis and cell death and favor proliferation, paving the way for a robust immune response by gsTcells that is mediated by the genes in cluster 3.

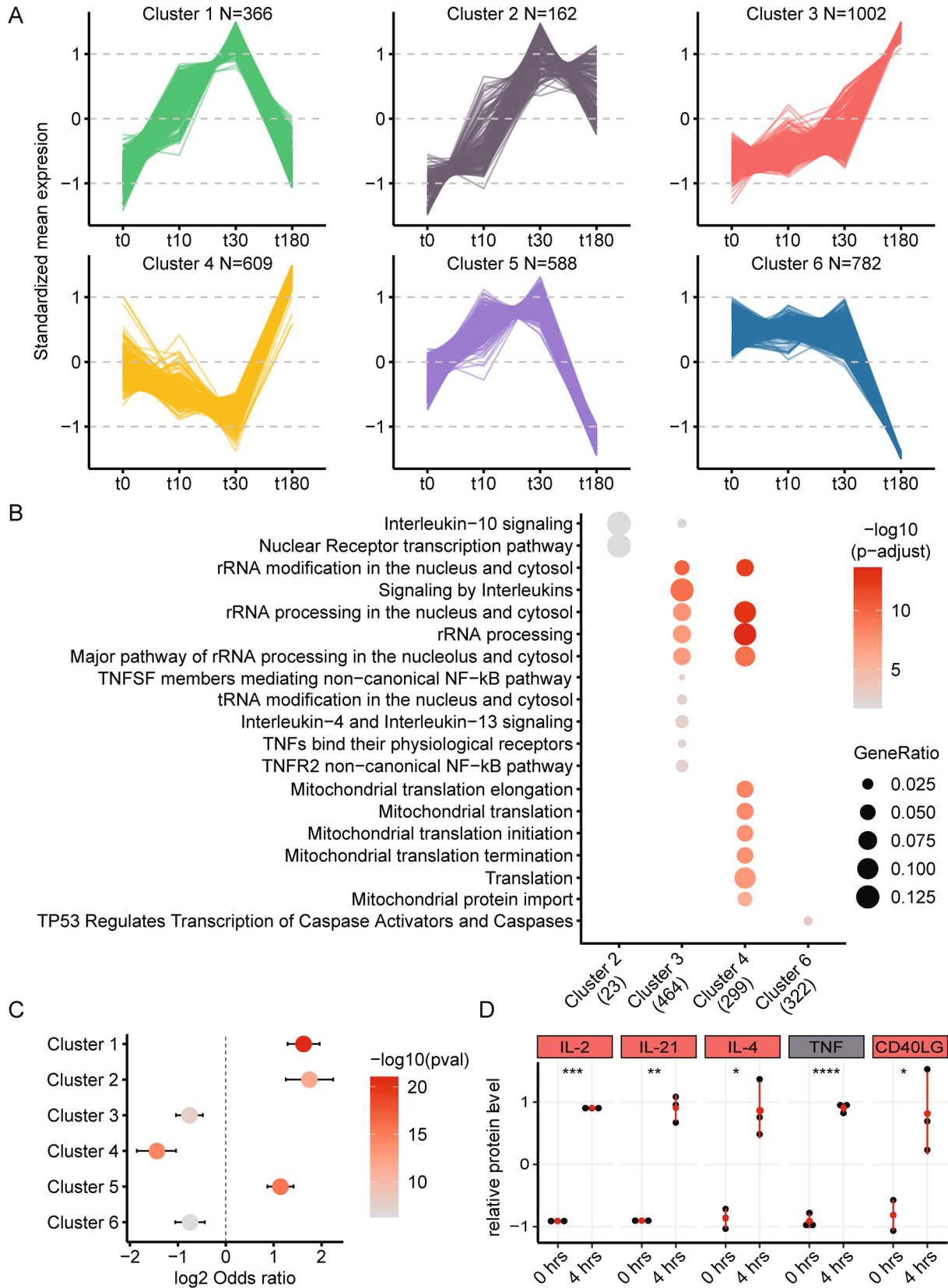
To ascertain how these pathways might be regulated, we investigated the relative enrichment of non-coding RNAs (ncRNAs) per cluster. The early-responding clusters 1, 2 and 5 are enriched for ncRNAs, implying that these RNAs play a role in regulating expression of genes at 180 min, when protein-coding genes are enriched



**Figure 1.** Stimulation of gluten-specific CD4+ T cells with aCD3/aCD28. **(A)** Experimental scheme of the discovery and replication cohort. The ‘+’ symbol indicates that a measurement is available at that timepoint. 28 gsTcell clones were isolated from CeD biopsies, 23 clones were used in discovery analysis and 5 were used for replication and DHS-sequencing. A final 3 clones from the discovery set were used for proteomic analysis. **(B)** PCA of the complete expression data of the discovery (circles) and replication (diamonds) cohorts. Each time point is indicated in a different color. **(C)** Differentially expressed genes identified by differential expression (DE) analysis between consecutive timepoints, plotted per comparison. Biotypes and direction of each DE gene are indicated.

(clusters 3, 4 and 6) (Fig. 2C). Several genes encoding TFs that mediate early immune and stress responses are also found in clusters 1 and 2. These TFs include all *EGR* TFs, *NR4A1*, *NR4A2*, *NR4A3*, *ATF3*, *FOS* and *FOSB*, of which the latter two are subunits of AP-1 (Suppl. Fig. 3A). Additionally, cluster 3 encompasses *REL* (encoding NFκB subunit c-REL), *NFκB1* (encoding NFκB subunit p50) and the NFκB inhibitory genes *NFκBIA*, *NFκBID* and *NFKIZ*, which suggests an early activation of the NFκB pathway and a subsequent feedback loop after 180 min (Suppl. Fig. 3B)<sup>37,38</sup>. Thus, the action of several TFs that are either activated or transcribed soon after stimulation, possibly in conjunction with ncRNAs, seems to mediate the strong response of genes after 180 min.

To confirm that transcriptional changes lead to secretion of cytokines, we measured a panel of 92 proteins in the gsTcell culture medium after 4 h of stimulation. Thirty of these proteins are encoded by genes differentially expressed during the stimulation of gsTcells (Suppl. Fig. 4A, Suppl. Table 3). We found that levels of IL-21, CD40LG, IL-2, IL-4 and TNF were significantly increased (nominal *p* value < 0.05) (Fig. 2D) in concert with increased expression of their corresponding genes (Suppl. Fig. 4B). These cytokines all have pro-inflammatory roles and contribute to activation and proliferation of other cell types, including B cells (IL-21, IL-4 and CD40LG) and other T cells like IELs (IL-2 and TNF)<sup>39,40</sup>.



◀**Figure 2.** Differentially expressed genes cluster into response patterns showing distinct functions. (A) Cluster analysis identified 6 robust clusters encompassing the response profiles of all 3509 differentially expressed (DE) genes. Numbers of genes per cluster are shown. Y-axis shows the mean expression of the genes at each respective time point, centered to mean zero and scaled to standard deviation one. (B) Reactome gene set enrichment analysis shows enriched pathways for 4 out of the 6 clusters. (C) Enrichment analysis by Fisher's exact test of non-coding RNAs per cluster. Significance is shown in shades of red. Log<sub>2</sub> odds ratios are plotted on the x-axis and indicate enrichment or depletion. Error bars indicate 0.95 confidence intervals. (D) Scaled relative protein levels (Olink) in the unstimulated condition and after 4 h of stimulation for the three independent experiments (black dots). Box colors above the dotplots indicate the cluster in which the DE genes are found. Mean, minimum and maximum relative protein levels are indicated in red. Nominal significance is indicated by asterisks (\**p* value < 0.05, \*\**p* value < 0.01, \*\*\**p* value < 0.001 and \*\*\*\**p* value < 0.0001).

In summary, the distinct and dynamic transcriptional changes we observe represent a robust translational immune response to TCR activation that leads to the secretion of several cytokines within 4 h.

### Transcriptional changes identified in gsTcells upon activation are similar to those in other T cells.

To examine the specificity of the gsTcell response, we compared the DE genes of gsTcells (between 0 and 180 min) to those of naïve CD4<sup>+</sup> T cells stimulated with aCD3/aCD28 [DICE consortium (n = 90)]<sup>41</sup>. A large proportion (57%) of the DE effects we observed in gsTcells were also found in the DICE consortium data (96% directional concordance; Fig. 3A, Suppl. Tables 4, 5). Directionally concordant genes (Q1 and Q3 in Fig. 3B) showed an enrichment for genes involved in interleukin signaling and rRNA processing (Suppl. Fig. 5A), as we had observed in cluster 3 (Fig. 2B), suggesting this to be a general response of CD4<sup>+</sup> T cells to stimulation. Genes uniquely activated in gsTcells include chemokines such as *CCL1*, *CXCL1* and *CCL4L1*, which encode for peptides that bind the receptors CCR8, CXCR2 and CCR5 and that can mediate recruitment of immune cells, including Type 2 innate lymphocyte cells, neutrophils and activated CD8<sup>+</sup> T cells, respectively (Fig. 3C, Suppl. Fig. 6A)<sup>42</sup>. In addition, the cytokine-encoding genes *IL5*, *IL9*, *IL19*, *IL17F* and *IL26*, as well as *RORC* (encoding for TF RORγt), are uniquely differentially expressed in gsTcells, albeit at low levels, and each points to a different subset classification of T helper cells for gsTcells<sup>43</sup>.

To determine which T helper subset the gsTcells most resemble, we compared the expression patterns of the main cytokines associated with Th1 (*IFNG*, *TNF*), Th2 (*IL4*, *IL5*, *IL13*), Th17 (*IL17A*), Treg (*IL10*, *TGFB1*), T follicular helper (*IL21*) and Th9 (*IL9*) cells (Suppl. Fig. 6B)<sup>6–13,43,44</sup>. We observed that cytokines characteristic for several subsets are strongly upregulated upon stimulation, with Th1 cytokines most strongly expressed. GsTcells are therefore most similar to Th1 cells but also share characteristics with multiple Th subsets.

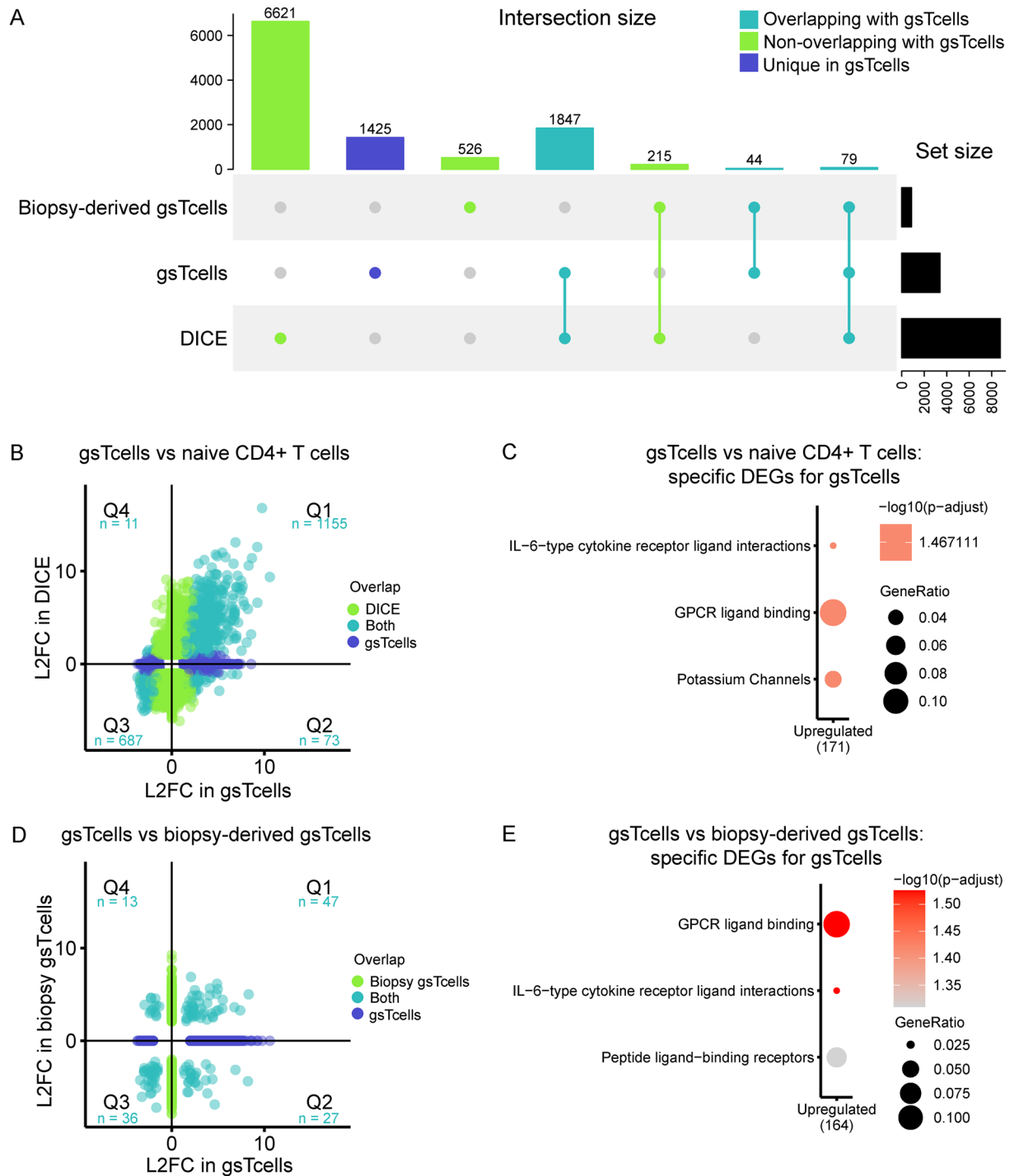
Overall, the strong concordance between the TCR-induced response profile of naïve CD4<sup>+</sup> T cells and gsTcells (Fig. 3A, B) suggests that the TCR response is largely shared in CD4<sup>+</sup> T cells. However, we also observed 1371 unique DE genes that were not observed in the naïve CD4<sup>+</sup> T cells or biopsy-derived gsTcells. This set of genes was mainly upregulated (1047/1371) and was enriched for G protein-coupled receptor ligand binding, IL-6-type cytokine receptor ligand interaction and peptide ligand-binding receptors (Fig. 3E). Nonetheless, when compared to naïve CD4<sup>+</sup> T cells, in vitro-cultured gsTcells show a distinct response profile on top of the shared TCR response profile that includes a diverse set of cytokines and chemokines that are comparable to that of Th1, Tfh and Th2 cells, which is consistent with previous evidence<sup>6–13,43,44</sup>.

### The expression profile of in vitro-cultured gsTcells is similar but not identical to the expression profile of gsTcells isolated directly from biopsies.

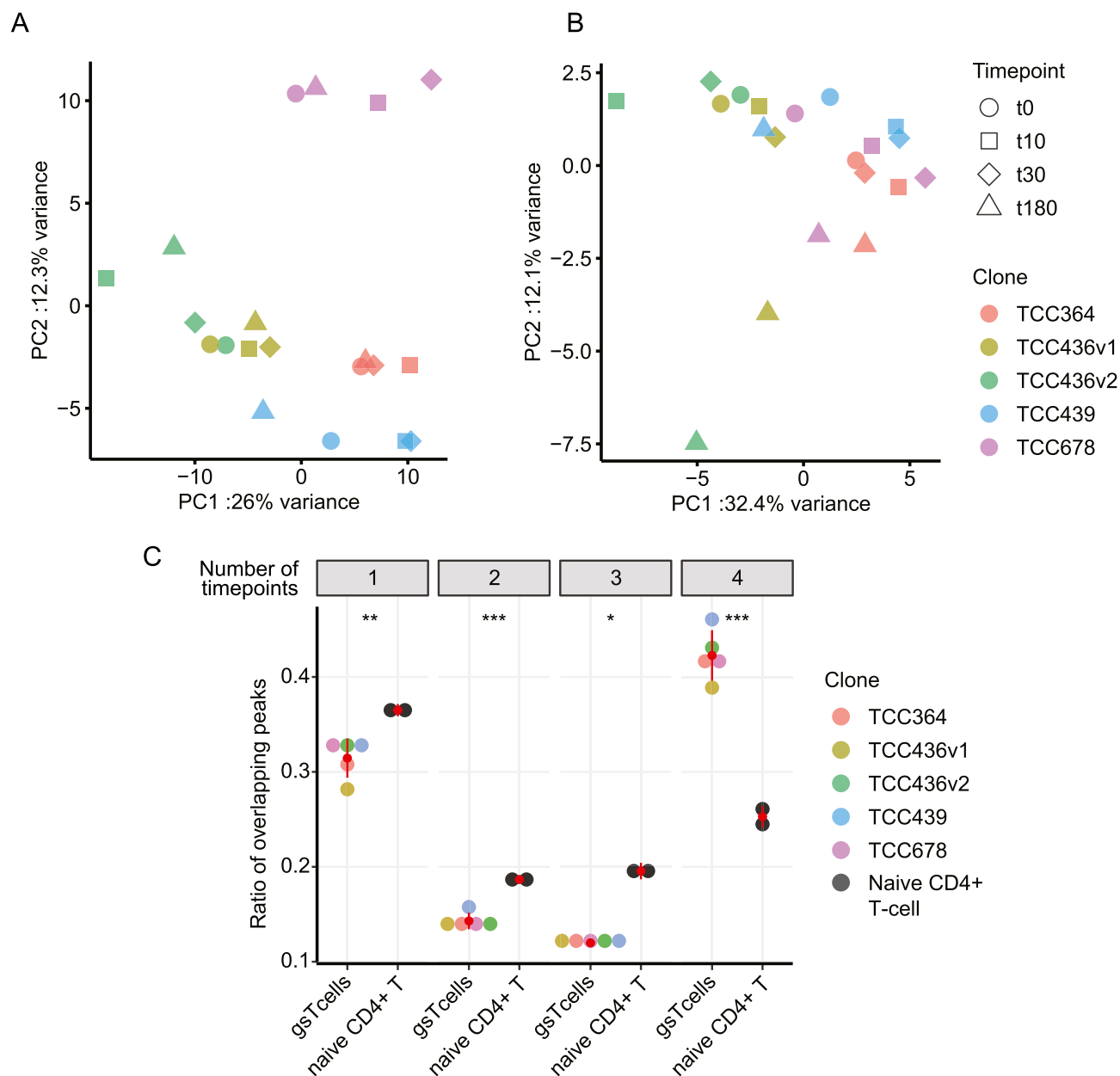
Next, we compared the transcriptomic response of our cultured gsTcells with gsTcells taken directly from CeD biopsies 6 days after an in vivo gluten challenge<sup>13</sup>. These biopsy-derived gsTcells were obtained from Christophersen et al.<sup>13</sup> and contained tetramer+ CD4<sup>+</sup> T cells and tetramer- CD4<sup>+</sup> T cells from CeD patients (n = 5) and gut CD4<sup>+</sup> T cells from healthy controls (n = 4). We found that 123 out of 3395 DE genes in gsTcells are shared with the 865 DE genes from biopsy-derived gsTcells (*p* value = 0.0056; two-sided Fisher's exact test) (Fig. 3D, Suppl. Tables 4, 6). These overlapping genes were enriched for the 'TNFR2 non-canonical NFκB pathway' in Reactome and included immune genes such as *CD200*, *MAP4K3* & *PDCD1*, *IL21* (a key regulator in CeD) and *IL22* (a regulator of intestinal epithelial homeostasis<sup>45</sup>) (Suppl. Fig. 5B, Suppl. Table 7). The differences between the in vitro-cultured gsTcells and biopsy-derived gsTcells could partly be attributable to differences in sample size. Moreover, the in vitro-cultured gsTcells were stimulated for 3 h, and the DE genes therefore reflect early transcriptional changes upon stimulation, which may not reflect the activation state of in vivo biopsy-derived gsTcells. Finally, continuous in vitro culturing of the gsTcell clones in the presence of IL-2 and IL-15 is likely to introduce changes in the expression profile of the gsTcell clones compared to the biopsy-derived gsTcells. Thus, gsTcells show disease-relevant characteristics but are not identical to 'fresh' biopsy-derived gsTcells.

### Chromatin of gsTcells remains largely stable.

Next, we assessed if changes in chromatin state agreed with changes in gene expression upon stimulation of gsTcells. We observed that the chromatin state measured by DHS-seq remained mostly stable at all four timepoints when assessed genome-wide, and most of the variation in open chromatin was explained by which gsTcell clone the data was derived from, and not by stimulation effects (Fig. 4A, Suppl. Table 8). Nonetheless, when investigating the open chromatin sites in a 5 kb window around the 3509 DE genes, we observed some variation between t180 and the other timepoints (Fig. 4B). However, differential peak-calling analysis did not find any sites with a log<sub>2</sub> fold change > 1. These results are probably due to the large variation between the clones (Fig. 4A, B) and the modest sample size. We also investigated the overlap



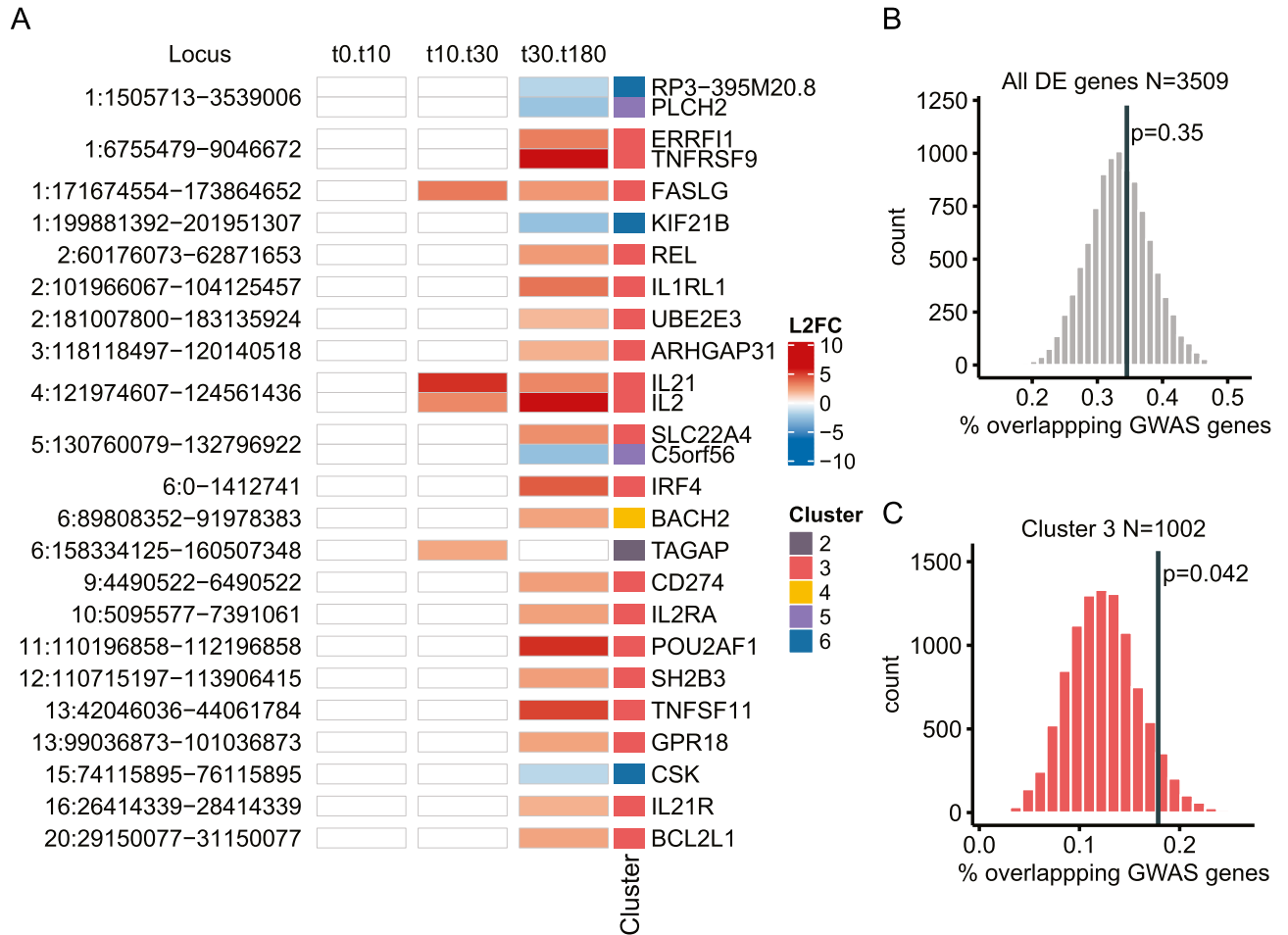
**Figure 3.** The transcriptional response of gsTcells shows differences and similarities with other T cells. **(A)** Upset plot comparing significant DE genes between stimulated naïve CD4+ T cells (DICE) (t0 vs t240, FDR < 0.05 and absolute log<sub>2</sub>FC > 1)<sup>41</sup>, gsTcells (t0 vs t180, FDR < 0.05 and absolute log<sub>2</sub>FC > 1) and biopsy-derived gsTcells (gsTcells vs CD4+ T cells in cases/controls, FDR < 0.05 and absolute log<sub>2</sub>FC > 2)<sup>13</sup>. At the top, the size of the intersecting sets with gsTcells are indicated in light blue, non-overlapping genes with gsTcells are shown in light green and genes unique to gsTcells are shown in dark blue. The right barplot shows the total number of DE genes per dataset. **(B)** Scatterplot of log<sub>2</sub>FC of DE genes between DICE (y-axis) and gsTcells (x-axis). Numbers in light blue indicate the number of genes in each quadrant that are significant in both analyses. **(C)** Gene set enrichment analysis done using Reactome of DE genes unique for gsTcells as compared to DICE naïve CD4+ T cells (adjusted *p* value < 0.05, absolute log<sub>2</sub>FC > 2). At the bottom is the direction of expression of the DE genes in gsTcells. Numbers in brackets indicate the number of DE genes present in all enriched pathways. Dot size indicates the ratio of the number of genes present in the gene set and the total gene set used in each pathway. **(D)** Comparison as in **(B)** for DE effects in CeD biopsy-derived gsTcells (y-axis). **(E)** Gene set enrichment analysis as in **(C)** for genes unique to gsTcells compared to CeD biopsy-derived gsTcells and DICE naïve CD4+ T cells (adjusted *p* value < 0.05, absolute log<sub>2</sub>FC > 2).



**Figure 4.** Open chromatin changes are minimal in gsTcells upon stimulation. **(A)** PC analysis of all peaks in each individual gsTcell clone in the discovery cohort. Timepoints indicated by shapes, clones by colors. **(B)** As in **(A)**, but only using peaks within a 5 KB window around the transcription start sites of the 3509 DE genes in the PCA. **(C)** Comparison of the relative changes in the individual gsTcells and naïve CD4+ T cells of two individuals between all timepoints. Overlaps between all timepoints were calculated using the multiinter Bedtools function, and the relative number of overlaps is plotted for each. Peaks that are present at all timepoints are represented in the '4' category. Peaks unique to one time point are represented in the '1' category. Clones are indicated with the same colors as in **(A)** and **(B)**. Nominal significance indicated with asterisks (\* $p$  value < 0.05, \*\* $p$  value < 0.01 and \*\*\* $p$  value < 0.001).

between peaks in each clone and compared that to the overlap between ATAC-seq peaks in naïve CD4+ T cells in a similar stimulation experiment<sup>46</sup>. Overall, the naïve CD4+ T cells showed more stimulation-specific effects, with 25% of peaks being shared at all four timepoints, as opposed to gsTcells, with 45% of the peaks are shared (Fig. 4C). Moreover, TF binding site enrichment analysis on peaks uniquely present in unstimulated gsTcells, but not seen in naïve CD4+ T cells under unstimulated conditions, showed that AP-1 binding sites are enriched in gsTcells even prior to stimulation (HOMER,  $p$  value =  $1e-16$ ). Altogether, this suggests that the early upregulation of gene expression upon activation in gsTcells is mostly independent of chromatin state and instead driven by the interplay of TFs and ncRNAs.



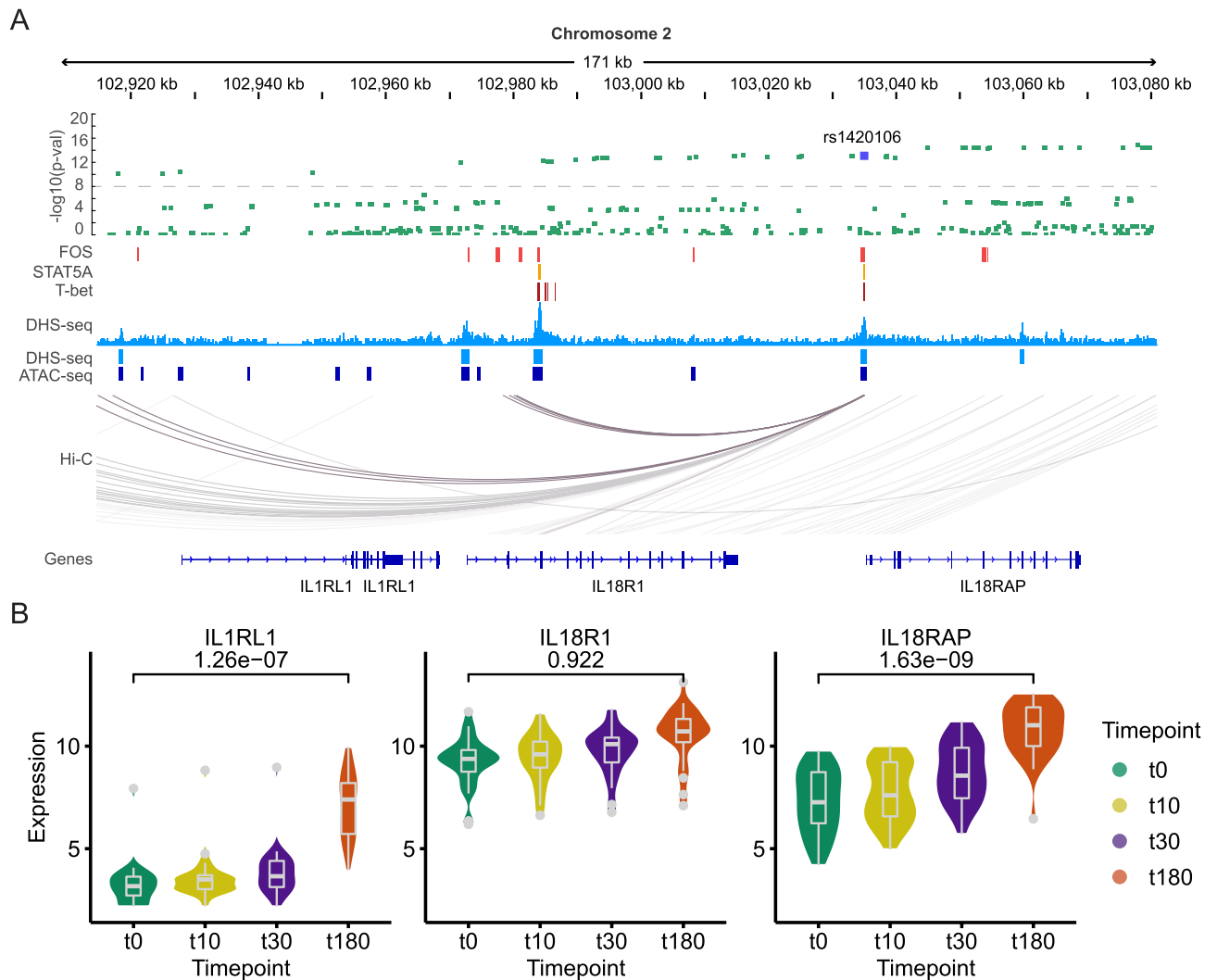


**Figure 5.** DE genes show enrichment for CeD loci. (A) Genes prioritized from CeD loci<sup>30</sup> were overlapped with DE genes and plotted per CeD locus. Log<sub>2</sub> fold change between timepoints is indicated, and clusters are depicted with colored boxes (right). (B) Enrichment of DE genes over the null distribution (histogram) in CeD loci ( $\pm 125$  kb window around start and end of gene) using GREA<sup>19</sup> for all DE genes ( $n = 3509$ ). X-axis indicates the number of genes that overlap with CeD loci as a percentage. The histogram shows the null distribution based on 10,000 permuted gene-sets. The black line indicates the value of the true gene-set. Nominal  $p$  values are indicated. (C) As in (B), but only for genes in cluster 3 ( $n = 1002$ ).

**DE genes show enrichment for CeD loci and CeD-associated genes.** To ascertain if the DE genes we found are affected by the genetic background associated with CeD, we integrated our transcriptional data with 118 *cis*-genes for CeD identified by a previous gene prioritization effort<sup>30</sup> (Fig. 5A). We observed that 26 of the prioritized genes are DE in the gsTcells. Of these 26 genes, 18 belong to cluster 3, in which immune response pathways are overrepresented, consistent with the mostly T cell-based GWAS signal of CeD<sup>23</sup>. Of particular interest are *IL21* and other cytokine-encoding genes (*IL2*, *TNFSF11* and *FASLG*), several cytokine receptor genes (*IL21R*, *IL1RL1*, *TNFRSF9* and *IL2RA*) and genes for TFs that can play a role in immune response (*REL*, *BACH2* and *IRF4*). Interestingly, *UBE2E3*, *CSK* and *SLC22A4* are DE in gsTcells but not in naïve CD4+ T cells (Fig. 3B), which implies that these genes may have a relatively specialized function in gsTcells.

We subsequently checked if the DE genes were overrepresented in CeD loci<sup>23</sup> using genomic region enrichment analysis (GREA)<sup>19</sup> (Fig. 5 B,C; Suppl. Fig. 8). GREA operates by comparing the overlap of DE genes in a window of  $\pm 125$  kb around a CeD locus to the overlap found randomly in a permutation-based null distribution. We found no significant enrichment for genes located in CeD loci when assessing all 3509 DE genes (Fig. 5B). However, when we assessed the enrichment per cluster (Suppl. Fig. 8), cluster 3 showed a nominally significant enrichment for genes located in CeD loci ( $p$  value = 0.042; Fig. 5C), with around 18% of cluster 3 genes being located near a CeD locus. Thus, the genes in cluster 3 may be affected by the genetic predisposition for CeD.

Next, we assessed whether any of the TFs that were DE in gsTcells were enriched to bind in CeD loci using Regulatory trait Locus Intersection (RELI) in conjunction with the provided database of ChIP-seq data for 389 TF-cell type pairs<sup>29</sup>. In total, 98 TF-cell type pairs showed significant enrichment for binding in CeD loci (FDR < 0.05) (Suppl. Table 9). Of these TFs, FOS, STAT5A and T-BET (encoded by the gene *TBX21*) were of particular interest as their genes showed a DE effect in gsTcells and the ChIP-seq data used in RELI was derived from CD4+ T cells. *FOS* falls in cluster 1 and showed an early but transient response to stimulation (Suppl. Fig. 9A), whereas *STAT5A* and *TBX21* showed a later response corresponding to the cluster 3 profile (Suppl. Fig. 9B and



**Figure 6.** Rs1420106 in the *IL1RL1/IL18R1/IL18RAP* locus overlaps with open chromatin and TF binding sites in gsTcells. **(A)** Overview of the *IL1RL1/IL18R1/IL18RAP* locus. From top to bottom: summary statistics of the CeD GWAS meta-analysis<sup>22</sup> (green); TF binding sites based on Chip-seq of *FOS* (red), *STAT5A* (orange) and *T-BET/TBX21*<sup>29</sup> (brown); DHS profile of unstimulated gsTcells with peak calls depicted below (light blue); peaks from naïve CD4+ T cell ATAC-seq<sup>46</sup> (dark blue); capture Hi-C data of activated CD4+ T cells depicting the 3D interactions between the highlighted region and other locations in the locus<sup>82</sup> (grey) and the gene annotations. The prioritized SNP rs1420106 is indicated in dark blue. **(B)** Gene expression pattern of genes in the locus. Y-axis represents the VST-normalized expression data. Black line and number indicate the adjusted *p* value (DeSeq2) of the DE effect between the t0 and t180 timepoints. Blood eQTL *p* values of SNP rs1420106 for the indicated genes are: *IL18RAP*, 3.27e-310; *IL1RL1*, 1.95e-144 and *IL18R1*, 1.63e-185 (eQTLgen.org<sup>50</sup>).

C, respectively). *FOS* is a well-known early response immune and stress response TF that can heterodimerize with the JUN or ATF TF families. Indeed, *ATF3* is also DE and has a very similar transcriptional profile to *FOS* (Suppl. Fig. 3A). *STAT5A* is required for IL2 signaling in CD4+ T cells<sup>47</sup>, and T-BET is the major TF for Th1 differentiation, expression of IFN $\gamma$  and other Th1-specific cytokines<sup>48</sup>. Thus, these TFs all have prominent roles in CD4+ T cell differentiation and activation.

To explore if any of the SNPs in the 43 CeD loci from the latest CeD meta-analysis<sup>23</sup> could directly affect gene regulation in gsTcells, we overlapped the TF binding sites of *FOS*, *STAT5A* and T-BET with the open chromatin regions, CeD GWAS summary statistics<sup>32</sup>, DE information, ATAC-seq of naïve CD4+ T cells<sup>46</sup> and capture Hi-C data of activated CD4+ T cells (Fig. 6). We found one locus of interest near the *IL18RAP/IL1RL1* genes. In this locus, one SNP (rs1420106, GWAS *p* value 8.7e-14) located in the promoter of *IL18RAP* overlapped with all three of the enriched TF binding sites, a DHS peak from gsTcells and an ATAC-seq peak from naïve CD4+ T cells (Fig. 6A). We assessed if this locus could act as an enhancer using publicly available capture Hi-C data from activated CD4+ T cells<sup>49</sup> and found interactions with the promoters of *IL1RL1* and *IL18R1* (Fig. 6A). Moreover, rs1420106 strongly affected the expression of *IL18RAP*, *IL1RL1* and *IL18R1* (eQTL *p* values 3.27e-310, 1.95e-144 and 1.63e-185, respectively) in whole blood (eQTLgen database<sup>50</sup>). Finally, both *IL1RL1* and *IL18RAP* were significantly upregulated in gsTcells at t180 compared to t0 (Fig. 6B). Together, this suggests that rs1420106 might

have a role in the activation response of gsTcells by modifying the expression of *IL18RAP*, *IL1RL1* and *IL18R1*. Similarly, we identified three more CeD loci (containing genes *BACH2*, *IL21-IL2* and *TAGAP*) that showed overlap of SNPs associated with CeD with DHS sites found in gsTcells that also bind the TFs FOS, STAT5A and/or T-BET (Suppl. Figs. 10–12).

Thus, CeD-associated genetics may play a complex role in gsTcells during activation of these important cells in CeD pathology.

## Discussion

In this study, we characterized gsTcells, one of the key players in CeD pathogenesis, by profiling transcriptomic and epigenetic changes during the early response to aCD3/aCD28 activation. We pinpointed pathways and TFs that may regulate these cells and confirmed that gsTcells are not restricted to any specific class of Th cells, but rather express cytokines, chemokines and TFs that are characteristic of Th1 but also of Tfh and Th2 subsets. We also identified *CCL1*, *CXCL1* and *CCL4L1* as unique DE genes that had not previously been shown to be expressed by CD4+ T cells in the context of CeD. Finally, we showed that the early response of gsTcells to stimulation is regulated by rapid upregulation of many TFs and ncRNAs in combination with the activation of JNK/AP-1 and NFκB TFs, whereas changes in chromatin were minor. Overall, our results provide an in-depth analysis of the molecular pathways that are activated in gsTcells upon TCR activation.

Our study illustrates that gsTcells express and secrete cytokines that can be associated to various subsets of Th cells, most prominently with the Th1 (*IFNG* and *TNF*), Th2 (*IL4*, *IL5*, *IL13*) and Tfh (*IL21*) subsets, which is in agreement with previous studies<sup>6–13,43,44</sup>. Based on their unique cytokine profile, gsTcells may exert multiple functions in CeD pathogenesis. Firstly, IFNγ is important for eliciting a strong response to foreign antigens such as the gluten peptides in CeD<sup>51</sup>. Moreover, IFNγ can also directly affect the integrity of the intestinal barrier<sup>52</sup>. Secondly, Th2-associated cytokines and IL-21 produced by Tfh cells are important for plasma cell differentiation, B cell activation and autoantibody production<sup>43</sup>. Finally, IL-21, IL-2 and TNF have been shown to activate CD8+ T cells and IELs in the gut, which thereby become “licensed to kill” epithelial cells, leading to the villus atrophy in CeD<sup>12,19,20,53</sup>. Thus, cytokines derived from gsTcells may play a role in several distinct disease mechanisms.

Our comprehensive analysis of the regulatory mechanisms that drive gene expression in activation of gsTcells agrees in part with the results of a study where the authors analyzed biopsy-derived gsTcells (Christophersen et al.<sup>13</sup>) but also uncovered differences. These differences might be due to differences in sample size and experimental design between the two studies but may also reflect the fact that the gsTcells used in our study have been expanded in vitro in the presence of IL-2 and IL-15. Nonetheless, the unique expression of genes that are not DE in activated naïve CD4+ T cells and the enrichment of genes that overlap with genes specific to biopsy-derived gsTcells suggest that in vitro-cultured gsTcells are a unique and appropriate model to delineate the dynamic transcription and regulation of gsTcells.

Similarly, comparison with other datasets such as the DICE data (Fig. 3A, B) and CeD-patient derived CD4+ T cells from PBMCs described by Quinn et al. (data not shown), validate our findings that the gsTcells have a similar but distinct expression profile and that important cytokines and other genes are associated to CeD etiology, such as *IFNG*, *IL21*, *IL17A* and *IL4*<sup>54</sup>, are differentially expressed. Still, the unique expression profile of gsTcells relative to naïve or blood-derived CD4+ T cells may be lost if the comparison with other antigen-specific T cells or memory CD4+ T cells activated with similar stimuli and timepoints.

Based on the expression pattern of the DE genes, we identified six major gene clusters with distinct dynamic responses and functions. Early responding genes were represented by clusters 1, 2 and 5 and contained multiple TFs, some cytokines and a disproportionate number of ncRNAs, implying that these genes regulate the response at later time points. An earlier study also observed a ncRNA response during the activation of lymphocytes, which supports the idea that ncRNAs are key in the development and activation of CD4+ T cell<sup>55</sup>. Some ncRNAs that are DE in gsTcells have also been implicated in immune activation, inflammation and proliferation in other studies, including *LINC00174*, *AF131217.1*, *TINCR* and *LINC00342*, which were all found in cluster 1<sup>56–60</sup>. Thus, ncRNAs seem to play a pivotal role in the immune response by changing the expression of specific genes in gsTcells.

The gsTcells we studied showed a stable open chromatin profile genome-wide, with only minor changes upon stimulation near the transcription start sites of DE genes. However, we cannot exclude the possibility that the stability we observe is a consequence of culturing gsTcells in the presence of cytokines to induce expansion<sup>61</sup>. In contrast to the stable open chromatin profiles in gsTcells, the transient expression of specific TFs and ncRNAs, in concert with the activation of common signaling pathways like JNK/AP-1 and NFκB, may be the source of the unique expression profile observed in the gsTcells.

Several DE genes are located in CeD-associated loci, and we found a subtle enrichment for genes of immune cluster 3 in CeD loci. However, the largest CeD association study to date was performed using the Immunochip platform, which is enriched for known immune regions, and thus we may have missed enrichment in other functional clusters. Nonetheless, 18% of the genes in cluster 3 are located in CeD loci, highlighting that TCR-mediated T cell activation, particularly in gsTcells, may be affected by CeD-associated SNPs.

To ascertain the role of CeD-associated SNPs in regulating gene expression in gsTcells, we integrated multiple publicly available functional data layers with the gene expression and DHS regions of the gsTcells and CeD-associated SNPs. While this integration provides suggestive evidence that the prioritized SNPs have regulatory potential in gsTcells, we cannot directly confirm that the genetic effect has a regulatory role. This would require an eQTL analysis with primary gsTcells derived from biopsies or functional validation by targeting these candidate regulatory elements, both of which are beyond the scope of this study. Despite these challenges, we provide evidence for potential genetic interference of CeD-associated SNPs in gsTcells in several loci and pinpoint several SNPs and regulatory regions within the genome that are the most likely candidates to cause this interference.

In summary, we present an in-depth characterization of early transcriptional dynamics of gsTcells in response to TCR activation. We highlight that this transcriptional response is most likely regulated by TFs and ncRNAs rather than large changes in chromatin state. Finally, we prioritize several CeD-associated genetic loci that may impact the TCR-activation in gsTcells directly.

## Methods

**Obtaining T cell clones from biopsies of CeD patients.** Gluten-specific CD4+ T cells were isolated from CeD patient small intestinal biopsies, as described previously<sup>25,62–65</sup>. Patients were diagnosed with CeD with small bowel biopsy confirmation and included at Leiden University Medical Center (LUMC), the Netherlands (n = 18), and the Riks Hospital in Oslo (RHO), Norway (n = 4), from which 23 and 5 gluten-specific T cell lines were isolated, respectively (Suppl. Table 1). Briefly, written informed consent was given by all patients and the small intestine biopsies from CeD patients were cultured with a mixture of gluten and TG2-treated gluten (deamidation) for 5 days. To expand the T cells, IL-2 (20 Cetus units/ml; Novartis, Arnhem, the Netherlands) and IL-15 (10 ng/ml; R&D systems, Abingdon, UK) were added. Subsequently, irradiated allogeneic peripheral blood mononuclear cells in the presence of phytohemagglutinin (1 µg/ml; Remel Inc. Lenexa, USA), IL-2 (20 Cetus units/ml) and IL-15 (10 ng/ml) were mixed with T cells for re-stimulation<sup>25</sup>. The resulting T cell clones were tested for reactivity against gluten digested by pepsin and trypsin and TG2-treated in proliferation assay. The pepsin/trypsin digest of gluten was prepared as described by Van de Wal et al.<sup>63</sup>. For deamidation, the pepsin/trypsin-digested gluten (500 mg/ml) was incubated with 100 mg/ml of guinea pig tTG (T-5398; Sigma, St. Louis, MO) at 37 °C for 2 h in PBS with 1 mM CaCl<sub>2</sub> and subsequently used in T cell proliferation assays. Proliferation assays were conducted as described by Van de Wal et al.<sup>63</sup>, for samples from LUMC, and Molberg et al.<sup>65</sup>, for samples from RHO. Gluten-specific lines were cloned by limiting dilution and expanded again by re-stimulation at 1- to 3-week intervals<sup>25</sup>. Clones were stored in liquid nitrogen. All methods were performed in accordance with relevant guidelines and regulations.

**Stimulation of T cell clones.** In all, 28 gluten-specific T cell clones were stimulated in 6-well plates coated overnight with anti-CD3 (2.5 µg/ml; Biolegend, San Diego, CA, USA) and anti-CD28 (2.5 µg/ml; Biolegend) or PBS (negative control) for 0, 10, 30 and 180 min. At each timepoint, cells were harvested for RNA isolation. Cell culture medium was harvested after 240 min for proteomic analysis.

**RNA isolation and library preparation of stimulated gsTcells.** GsTcells were harvested at each time point, washed with PBS and resuspended in a lysis buffer (Ambion, Life Technologies, Carlsbad, CA, USA). RNA was extracted with the mirVana RNA isolation kit (Ambion) according to the manufacturer's instructions. The quantity and quality of RNA was determined by Bioanalyzer (Agilent technologies, Santa Clara, CA, USA). The sequencing libraries were prepared from 1 µg of total RNA using the TruSeq Stranded Total RNA with Ribozero Globin kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Sequencing was done with the Illumina HiSeq2500 (Illumina).

**DNase I hypersensitivity sequencing and analysis.** Standard protocols for nuclei isolation, DNase I (Roche #04716728001) treatment and library preparation for DNase I hypersensitivity sequencing generated within the Blueprint consortium were followed. Protocol details<sup>66</sup> can be found at: [http://www.blueprint-epigenome.eu/UserFiles/file/Protocols/Blueprint\\_DNaseI\\_Protocol.pdf](http://www.blueprint-epigenome.eu/UserFiles/file/Protocols/Blueprint_DNaseI_Protocol.pdf). All samples were sequenced to a sequencing depth of approximately 50–60 million 50 bp single-end reads.

**Protein analysis.** Supernatants from unstimulated and 4 h-stimulated gsTcell cell cultures were taken and analyzed with the Immuno-Oncology panel of Olink (<http://www.olink.com/products/immuno-oncology>). Data was analyzed by subtracting relative log<sub>2</sub> protein levels in a blank medium control from the relative log<sub>2</sub> protein levels in the supernatants, followed by a two-sided *t*-test to measure significant change between supernatants of unstimulated and 4 h-stimulated gsTcells.

**Statistical methods.** Statistical analyses were performed in R (version 3.6.3)<sup>67</sup> unless otherwise specified. Visualization of results was done using the R package ggplot2 (version 3.3.0)<sup>68</sup>.

**RNA-seq quantification.** Before alignment, the reverse complement of the fastQ sequences were taken using the FASTX-Toolkit<sup>69</sup>. Alignment was done using Hisat2 (version 2.0.4)<sup>70</sup> against the forward strand, with default alignment parameters. The reference genome index was made using the Hisat2-build indexer and 1000 genomes reference genome version GRCh37 v75 with default parameters. For the samples that had paired-end data, only the first mate file was used for alignment. Reads mapping to multiple positions were removed. The genes were quantified using HTSeq (version 0.6.1.p1)<sup>71</sup> with options -m union, -t exon, -stranded yes and other options on default.

**DE analysis.** The raw count matrix, containing 63,682 genes and 112 samples (92 samples from the Leiden cohort and 20 from the Oslo cohort), was first filtered to remove any non-expressed genes by selecting only genes that had at least 1 read in 20 samples. This resulted in 29,772 genes to be tested for DE effects. Samples from the Leiden and Oslo cohorts were then split, and the DE effects assessed separately.

DE effects were quantified using the R package DESeq2 (version 1.26.0)<sup>72</sup>, including RNAseq batch and sex as covariates for the Leiden samples. No covariates were included for the Oslo samples because no sex information

was available and all samples had been sequenced in the same batch. DE effects were then mapped between the t0 and t10, t10 and t30, and t30 and t180 timepoints. DE effects in the Leiden cohort were filtered on having an absolute log<sub>2</sub> fold change (log<sub>2</sub>FC) of at least 1 and an FDR < 0.05. Oslo samples were used as the replication cohort, and comparisons between the two were made using unfiltered Oslo data. In total, we identified 3509 unique DE genes in the Leiden cohort. These genes were used for interpretation and downstream analysis. PC analyses were performed on the variance-stabilized count data.

**Clustering of DE genes into distinct response patterns.** DE genes were clustered into time patterns as follows. The gene expression matrix was VST-normalized using DESeq2 (version 1.26.0)<sup>72</sup>, after which the mean expression level for each gene was determined at each of the four time points. Each row was then centered to mean 0 and scaled to standard deviation of 1. The data was clustered using k-means clustering (k=6) on a Euclidean distance matrix using the R package TCseq (version 1.10.0)<sup>73</sup>. Cluster number was determined by assessing the stability of the clustering in terms of within-cluster sums of squares over 100 iterations of the clustering. We then determined that the optimal tradeoff between stability and informativeness of each cluster occurred with a cluster number of k=6. To verify the stability, we ran another 100 random k-means clustering runs using different parameters (nstart = 100, k = 6 and iter.max = 1000). This yielded 100 fully stable clusters that matched very well with the clustering definition maintained in the manuscript (98.3% of genes matched their cluster).

**Comparisons with DICE and biopsy-derived gsTcells data.** DE genes from primary naïve CD4+ T cells were retrieved from DICE (<https://dice-database.org/>)<sup>41</sup>. Briefly, naïve CD4+ T cells from healthy donors (n=91) were obtained from blood by FACS and stimulated using aCD3/aCD28 for 4 h. Biopsy-derived gsTcells were obtained from Christophersen et al.<sup>13</sup> and contained tetramer+ CD4+ T cells and tetramer- CD4+ T cells from CeD patients (n=5) and gut CD4+ T cells from healthy controls (n=4). First, we obtained all DE genes that were significant (adjusted *p* value < 0.05) and showed an absolute log<sub>2</sub>FC > 1. Next, we intersected all DE genes from each dataset to obtain those that were unique per dataset and those that were shared with gsTcells. We then obtained the overlapped DE genes of gsTcells between DICE (n = 1926) and CeD biopsies (n = 144) to evaluate the concordance of those genes using the log<sub>2</sub>FC. The overlapped genes were divided in four quadrants (Q1–Q4). Q1 and Q3 included concordant DE genes that were upregulated and downregulated, respectively. Q2 and Q4 consisted of non-concordant DE genes, with Q2 being upregulated in gsTcells but not in reference dataset and Q4 being vice-versa.

**Gene set enrichment analysis.** Reactome pathways<sup>74</sup> were used to identify the pathways or biological processes that were enriched for each set of genes. This analysis was performed using the R package clusterProfiler (version 3.14.3)<sup>75</sup>. *p* values were adjusted using the Benjamini–Hochberg procedure to account for multiple testing.

**Quantification and peak calling of DHS sequencing.** DHS reads were aligned to hg19 reference genome using bwa (version 0.6.1-r104)<sup>76</sup> with default settings, after which duplicates were marked using bamUtil (version: 1.0.2)<sup>77</sup>. Alignments were then filtered to have a mapping quality of at least 30 and a primary alignment and to not be duplicated using Samtools (version 1.9)<sup>78</sup>. Peaks were then called using macs2 (version 2.2.6)<sup>79</sup> enabling -broad -nomodel -shift -125 -extsize 250. Peaks were considered at an FDR threshold < 0.05.

**Differential peak calling of DHS sequencing.** To identify differentially accessible sites between timepoints, consensus peaks were first defined using the R package DiffBind (version 2.14.0)<sup>80</sup>, after which raw read counts were determined for each consensus peak. Differentially accessible peaks were then quantified between t0 and t10, t10 and t30, and t30 and t180 using DESeq2 (version 1.26.0)<sup>72</sup>. PC analyses of DHS data were performed on the RPKM-normalized log<sub>10</sub>-transformed read counts for the consensus peaks. Overlap of peaks between the gsTcells or public datasets was determined with Bedtools multiinter -cluster. TF binding site enrichment was performed with Homer findMotifsGenome.pl with the merged regions of untreated gsTcells as background<sup>81</sup>.

**Enrichment of differentially expressed genes in celiac disease loci.** To test for enrichment of DE genes in CeD loci, we used the R package GREASE (<https://github.com/raguirreg/GREASE>, version 0.1.0)<sup>19</sup>. We defined CeD genes as genes within a 125 kb window of a CeD GWAS top SNP. We then generated 10,000 random gene-sets that matched the CeD gene-sets in size. The 10,000 random gene-sets were used to generate an empirical null distribution of the overlap between our DE gene-set per cluster and the random gene-sets. We then estimated the one-sided empirical *p* value of the enrichment for each cluster of DE genes.

**Ethical approval.** Biological material was obtained from celiac disease patients according to protocols approved by the regional ethics committees (Medical Ethical Committees of VU University Medical Center, Leiden University Medical Center and University of Oslo), and the individuals donating material gave their written informed consent.

### Data availability

All data (RNA-Seq count tables, DHS counts and peaks) required to reproduce this study have been provided as supplementary files. The raw RNA-Seq and DNase reads supporting this study are available upon request to

the authors as this is privacy sensitive. All code and scripts used to generate the results and figures are available on Github ([https://github.com/OlivierBakker/gluten\\_specific\\_tcells](https://github.com/OlivierBakker/gluten_specific_tcells)).

Received: 3 December 2020; Accepted: 18 March 2021

Published online: 29 April 2021

## References

- Lindfors, K. *et al.* Coeliac disease. *Nat. Rev. Dis. Prim.* **5**, 1–18 (2019).
- Jabri, B. & Sollid, L. M. T cells in celiac disease. *J. Immunol.* **198**, 3005–3014 (2017).
- Molberg, *et al.* Gliadin specific, HLA DQ2-restricted T cells are commonly found in small intestinal biopsies from coeliac disease patients, but not from controls. *Scand. J. Immunol.* **46**, 103–108 (1997).
- Christophersen, A. *et al.* Healthy HLA-DQ2.5+ subjects lack regulatory and memory T cells specific for immunodominant gluten epitopes of celiac disease. *J. Immunol.* **196**, 2819–2826 (2016).
- Risnes, L. F. *et al.* Disease-driving CD4+ T cell clonotypes persist for decades in celiac disease. *J. Clin. Investig.* **128**, 2642–2650 (2018).
- Gianfrani, C. *et al.* Gliadin-specific type 1 regulatory T cells from the intestinal mucosa of treated celiac patients inhibit pathogenic T cells. *J. Immunol.* **177**, 4178–4186 (2006).
- Bodd, M. *et al.* HLA-DQ2-restricted gluten-reactive T cells produce IL-21 but not IL-17 or IL-22. *Mucosal Immunol.* **3**, 594–601 (2010).
- Nilsen, E. M. *et al.* Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon  $\gamma$ . *Gut* **37**, 766–776 (1995).
- Brottveit, M. *et al.* Mucosal cytokine response after short-term gluten challenge in celiac disease and non-celiac gluten sensitivity. *Am. J. Gastroenterol.* **108**, 842–850 (2013).
- Lahat, N. *et al.* Cytokine profile in coeliac disease. *Scand. J. Immunol.* **49**, 441–447 (1999).
- Sjöberg, V. *et al.* Intestinal T-cell responses in celiac disease—impact of celiac disease associated bacteria. *PLoS ONE* **8**, e53414 (2013).
- Kooy-Winkelaar, Y. M. C. *et al.* CD4 T-cell cytokines synergize to induce proliferation of malignant & nonmalignant innate intraepithelial lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E980–E989 (2017).
- Christophersen, A. *et al.* Distinct phenotype of CD4+ T cells driving celiac disease identified in multiple autoimmune conditions. *Nat. Med.* **25**, 734–737 (2019).
- Moens, L. & Tangye, S. G. Cytokine-mediated regulation of plasma cell generation: IL-21 takes center stage. *Front. Immunol.* **5**, 1–13 (2014).
- Vazquez, M. I., Catalan-Dibene, J. & Zlotnik, A. B cells responses and cytokine production are regulated by their immune micro-environment. *Cytokine* **74**, 318–326 (2015).
- Mesin, L., Sollid, L. M. & Di Niro, R. The intestinal B-cell response in celiac disease. *Front. Immunol.* **3**, 1–12 (2012).
- Meresse, B., Korneychuk, N., Malamut, G. & Cerf-Bensussan, N. Interleukin-15, a master piece in the immunological jigsaw of celiac disease. *Dig. Dis.* **33**, 122–130 (2015).
- Korneychuk, N. *et al.* Interleukin 15 and CD4+ T cells cooperate to promote small intestinal enteropathy in response to dietary antigen. *Gastroenterology* **146**, 1017–1027 (2014).
- Zorro, M. M. *et al.* Tissue alarmins and adaptive cytokine induce dynamic and distinct transcriptional responses in tissue-resident intraepithelial cytotoxic T lymphocytes. *J. Autoimmun.* **108**, 102422 (2020).
- Ciszewski, C. *et al.* Identification of a  $\gamma c$  receptor antagonist that prevents reprogramming of human tissue-resident cytotoxic T cells by IL15 and IL21. *Gastroenterology* **158**, 625–637 (2020).
- Dubois, P. C. A. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
- 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).
- 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.* **28**, 313–323 (2020).
- Kim, C. Y., Quarsten, H., Bergseng, E., Khosla, C. & Sollid, L. M. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4175–4179 (2004).
- 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).
- Sollid, L. M., Qiao, S. W., Anderson, R. P., Gianfrani, C. & Koning, F. Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics* **64**, 455–460 (2012).
- Withoff, S., Li, Y., Jonkers, I. & Wijmenga, C. Understanding celiac disease by genomics. *Trends Genet.* **32**, 295–308 (2016).
- Farh, K.K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
- Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
- van der Graaf, A. *et al.* Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFN $\gamma$  signaling in celiac disease. *Front. Genet.* **11**, 1–16 (2021).
- Kumar, V., Wijmenga, C. & Xavier, R. J. Genetics of immune-mediated disorders: from genome-wide association to molecular mechanism. *Curr. Opin. Immunol.* **31**, 51–57 (2014).
- 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).
- Monteleone, G. *et al.* A failure of transforming growth factor- $\beta$ 1 negative regulation maintains sustained NF- $\kappa$ B activation in gut inflammation. *J. Biol. Chem.* **279**, 3925–3932 (2004).
- Fernandez-jimenez, N. *et al.* Coregulation and modulation of NF $\kappa$ B-related genes in celiac disease: uncovered aspects of gut mucosal inflammation. *Hum. Mol. Genet.* **23**, 1298–1310 (2014).
- Maiuri, M. C. *et al.* Nuclear factor  $\kappa$ B is activated in small intestinal mucosa of celiac patients. *J. Mol. Med.* **81**, 373–379 (2003).
- Li, P. *et al.* BATF-JUN is critical for IRF4-mediated transcription in T cells. *Nature* **490**, 543–546 (2012).
- Renner, F. & Schmitz, M. L. Autoregulatory feedback loops terminating the NF- $\kappa$ B response. *Trends Biochem. Sci.* **34**, 128–135 (2009).
- Shih, V. F. S. *et al.* Kinetic control of negative feedback regulators of NF- $\kappa$ B/RelA determines their pathogen- and cytokine-receptor signaling specificity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9619–9624 (2009).
- Crotty, S. A brief history of T cell help to B cells. *Nat. Rev. Immunol.* **15**, 185–189 (2015).
- Zhu, J. & Paul, W. E. Peripheral CD4+ T-cell differentiation regulated by networks of cytokines and transcription factors. *Immunol. Rev.* **238**, 247–262 (2010).
- Schmiedel, B. J. *et al.* Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715 (2018).

42. Hughes, C. E. & Nibbs, R. J. B. A guide to chemokines and their receptors. *FEBS J.* **285**, 2944–2971 (2018).
43. Raphael, I., Nalawade, S., Eagar, T. N. & Forsthuber, T. G. T cell subsets and their signature cytokines in autoimmune and inflammatory diseases. *Cytokine* **74**, 5–17 (2015).
44. Spolski, R. & Leonard, W. J. IL-21 and T follicular helper cells. *Int. Immunol.* **22**, 7–12 (2009).
45. Keir, M. E., Yi, T., Lu, T. T. & Ghilardi, N. The role of IL-22 in intestinal health and disease. *J. Exp. Med.* **217**, 1–9 (2020).
46. Qu, K. *et al.* Individuality and variation of personal regulomes in primary human T cells. *Cell Syst.* **1**, 51–61 (2015).
47. Farrar, M. A. & Owen, D. L. STAT5 and CD4+ T cell immunity. *Fl000Research* **6**, 1–10 (2017).
48. Zhang, Y., Zhang, Y., Gu, W. & Sun, B. Th1/Th2 cell differentiation and molecular signals. In *T helper cell differentiation and their function* (ed. Sun, B.) 15–44 (Springer, 2014). [https://doi.org/10.1007/978-94-017-9487-9\\_2](https://doi.org/10.1007/978-94-017-9487-9_2).
49. Burren, O. S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol.* **18**, 1–19 (2017).
50. Vösa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* <https://doi.org/10.1101/447367> (2018).
51. Schroder, K., Hertzog, P. J., Ravasi, T. & Hume, D. A. Interferon- $\gamma$ : an overview of signals, mechanisms and functions. *J. Leukoc. Biol.* **75**, 163–189 (2004).
52. Eriguchi, Y. *et al.* Essential role of IFN- $\gamma$  in T cell-associated intestinal inflammation. *JCI Insight* **3**, e121886 (2018).
53. Setty, M. *et al.* Distinct and synergistic contributions of epithelial stress and adaptive immunity to functions of intraepithelial killer cells and active celiac disease. *Gastroenterology* **149**, 681–691 (2015).
54. Quinn, E. M. *et al.* Transcriptome analysis of CD4+ T cells in coeliac disease reveals imprint of BACH2 and IFN1 regulation. *PLoS ONE* **10**, 1–25 (2015).
55. Xia, F. *et al.* Dynamic transcription of long non-coding RNA genes during CD4+ T cell development and activation. *PLoS ONE* **9**, 1–11 (2014).
56. Lu, Q., Meng, Q., Qi, M., Li, F. & Liu, B. Shear-sensitive lincRNA AF131217.1 inhibits inflammation in HUVECs via regulation of KLF4. *Hypertension* **73**, E25–E34 (2019).
57. Liu, J., Yang, T., Zhang, Y. & Wang, S. Promotion of BZW2 by LINC00174 through miR-4500 inhibition enhances proliferation and apoptosis evasion in laryngeal papilloma. *Cancer Cell Int.* **20**, 1–10 (2020).
58. Guo, J. *et al.* Role of linc00174/miR-138-5p (miR-150-5p)/FOSL2 feedback loop on regulating the blood-tumor barrier permeability. *Mol. Ther. Nucleic Acids* **18**, 1072–1090 (2019).
59. Liu, C., Xu, Y., Wu, X. & Zou, Q. Clinical significance of linc00342 expression in the peripheral blood lymphocytes of patients with chronic kidney disease. *Int. J. Nephrol. Renovasc. Dis.* **12**, 251–256 (2019).
60. Xu, S., Kong, D., Chen, Q., Ping, Y. & Pang, D. Oncogenic long noncoding RNA landscape in breast cancer. *Mol. Cancer* **16**, 1–15 (2017).
61. Hedfors, I. A. & Brinchmann, J. E. Long-term proliferation and survival of in vitro-activated T cells is dependent on interleukin-2 receptor signalling but not on the high-affinity IL-2R. *Scand. J. Immunol.* **58**, 522–532 (2003).
62. Vader, W. *et al.* The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* **122**, 1729–1737 (2002).
63. Van De Wal, Y. *et al.* Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10050–10054 (1998).
64. Kooy-Winkelaar, Y. & Koning, F. Isolation and cloning of gluten-specific T cells in celiac disease. In *Methods in Molecular Biology* 53–59 (2015). [https://doi.org/10.1007/978-1-4939-2839-2\\_6](https://doi.org/10.1007/978-1-4939-2839-2_6).
65. Molberg, Ø., McAdam, S. N., Lundin, K. E. A. & Sollid, L. M. Studies of gliadin-specific T-cells in celiac disease. In *Celiac Disease* 105–124 (Humana Press, 2000). <https://doi.org/10.1385/1-59259-082-9:105>.
66. Yi, G. *et al.* Chromatin-based classification of genetically heterogeneous AMLs into two distinct subtypes with diverse stemness phenotypes. *Cell Rep.* **26**, 1059–1069.e6 (2019).
67. R Core Team. *R: A Language and Environment for Statistical Computing* (2019).
68. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
69. Gordon, A., Hannon, G. J. *et al.* Fastx-toolkit. *FASTQ/A short-reads preprocessing tools (unpublished)* [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit) 5 (2010).
70. Kim, D., Langmead, B. & Salzberg, S. L. hisat2. *Nat. Methods* (2015).
71. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
72. Love, A. M., Anders, S., Huber, W. & Love, M. M. Package ‘DESeq2’ (2017).
73. Wu, M. & Gu, L. TCseq: time course sequencing data analysis. *R Packag. version 1* (2018).
74. Fabregat, A. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
75. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
76. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
77. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**, 918–925 (2015).
78. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
79. Gaspar, J. Improved peak-calling with MACS2. *bioRxiv* <https://doi.org/10.1101/496521> (2018).
80. Stark, R. *et al.* DiffBind: differential binding analysis of ChIP-Seq peak data. *R Packag. version 100*, 3–4 (2011).
81. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
82. Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 1–12 (2018).

## Acknowledgements

We thank Kate Mc Intyre for editing the manuscript. We would like to acknowledge Morris Swertz for facilitating data processing and storage. O.B.B. is supported by a Netherlands Organization for Scientific Research (NWO) VIDI Grant (No. 016.171.047) and an NWO VENI Grant (No. 863.13.011). A.D.R.S. is supported by CONACYT-I2T2 scholarship (No. 459192). S.W. and C.W. were supported by The Netherlands Organ-on-Chip Initiative, an NWO Gravitation project (024.003.001) funded by the Ministry of Education, Culture and Science of the government of the Netherlands. K.E.A.L., L.M.S., C.W., S.W.Q. and M.K.J. are supported by Stiftelsen KG Jebsen (Project SKGJ-MED-017). I.J. is supported by a Rosalind Franklin Fellowship from the University of Groningen and an NWO VIDI Grant (No. 016.171.047).

### Author contributions

Z.A.B., R.M., Y.K.W., M.K.J., S.W.Q., F.M., J.v.B. and I.H.J. performed wet-lab experiments. K.E.A.L., L.M.S. and F.K. provided samples. Z.A.B., I.H.J., A.D.R.S. and O.B.B. conceived and wrote the manuscript. A.D.R.S. and O.B.B. performed the statistical analyses. Z.A.B. and N.K. assisted with quantifying the RNAseq and DHS data. I.H.J., Y.L., J.H.A.M., V.K., K.E.A.L., L.M.S., F.K. S.W. and C.W. supervised and edited the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-86612-5>.

**Correspondence** and requests for materials should be addressed to I.H.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021