scientific reports

Check for updates

OPEN

Single cell transcriptome sequencing indicates the cellular heterogeneity of small intestine tissue in celiac disease

Tian Shi[1,2], Yan Feng[1,2], Jin Ma[1,3], Weidong Liu[1,2], Na Li[1,2], Ting Li[1,2], Adilai Abudurexiti[1,2], Ailifeire Tuerxuntayi[1,2], Shenglong Xue[1,2] & Feng Gao[1,2]✉

Celiac disease (CeD) is an autoimmune small intestinal disease caused by gluten protein ingestion by genetically susceptible individuals. Genome-wide association studies and transcriptomic data have limited capacity to capture intercellular genetic variations. We aimed to construct a single cell transcriptome spectrum, analyze the immune microenvironment and cellular heterogeneity, discover disease-related specific genes and markers, and explore the pathogenesis of CeD. This study performed single cell RNA sequencing (scRNA-seq) on three small intestine biopsies from patients with CeD and three matched healthy Chinese controls. Immunohistochemistry (IHC) and quantitative polymerase chain reaction (qPCR) were used to validate potential diagnostic biomarkers of disease-differential genes. A total of 10 cell subpopulations were annotated, including three types of epithelial and stromal cells and seven types of immune cells. IHC revealed a pronounced overexpression of T cell disease-differential genes, *TRAT1*, *BCL11B*, and *ETS1* in intraepithelial lymphocytes in the CeD group. Further clinical validation using qPCR confirmed that *ETS1* ($P = 0.010$), *TRAT1* ($P < 0.001$), and *BCL11B* ($P = 0.036$) were enriched in the CeD small intestinal tissue. The *CD28/CTLA-4* pathway regulates the homeostasis of Treg cells. The IFITs family genes may serve as marker genes for antiviral specific $CD4^+$ T cell subsets. CeD-derived subsets of $CD8^+$ T cells frequently express genes associated with cytotoxicity, including *IFNG*, *GZMK*, *GZMH*, *GZMB*, *SH2D1A*, *PRF1*, and *NKG7*, as well as genes related to T cell exhaustion, such as *PDCD10*, *CTLA4*, *TIGIT*, *PDCD1*, and *DUSP4*. Inflammation and infection pathways were enriched in different cell populations. A single cell expression profile of CeD small intestinal tissue was successfully constructed using scRNA-seq in this study. New biomarkers for CeD-specific histopathology and potential therapeutic targets were discovered, and the biomarkers observed between inflammation and infection pathways were closely related to the onset of CeD.

Keywords Celiac disease, Single cell RNA sequencing, Immune microenvironment, Marker

Abbreviations

CeD	Celiac disease
Ctrl	Control
EEC	Enteroendocrine cell
FC	Fold change
GWAS	Genome-wide association studies
ILC	Innate lymphoid cell
MC	Mast cell
NK	Natural killer
PC	Plasma cell
PCA	Principal component analysis
scRNA-seq	Single cell RNA sequencing
TCR	T cell receptor

[1]Department of Gastroenterology, People's Hospital of Xinjiang Uygur Autonomous Region, No. 91, Tianchi Road, Tianshan District, Urumqi 830001, Xinjiang, China. [2]Xinjiang Clinical Research Center for Digestive Diseases, Urumqi, China. [3]Department of Pathology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China. ✉email: xjgf@sina.com

UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier

Celiac disease (CeD) is an autoimmune disorder affecting the small intestine. The disease manifests in genetically predisposed individuals upon exposure to gluten, and leads to an immune-mediated response[1]. The pathological characteristics of this disease predominantly include an elevation in the count of intraepithelial lymphocytes (exceeding 25 per 100 enterocytes) within the mucosa of the small intestine, crypt hyperplasia, and varying degrees of villous atrophy[2]. A systematic review of the global prevalence of CeD indicates that the seroprevalence is 1.4%, and the prevalence confirmed by biopsy results is 0.7%[3]. CeD occurs at any age, and two-thirds of affected individuals endure without being diagnosed[4]. The delayed diagnosis and treatment of CeD leads to a variety of complications for patients, increased risk of secondary autoimmune diseases, and even the occurrence of malignant tumors, thereby posing a serious threat to human health[5].

Major histocompatibility class II genes are the most crucial genetic factors in CeD. The carriage rate of human leukocyte antigen (HLA)-DQ2/DQ8 in the general population is 30–40%. Moreover, only approximately 3% of individuals carrying these genes develop CeD[6,7]. This indicates that genetic susceptibility cannot fully explain the occurrence of CeD and that multiple undetermined disease-contributing factors remain. These factors include non-HLA genes, non-gluten environmental factors, and immune factors.

Genome-wide association studies (GWAS) and transcriptomic analyses have been instrumental in elucidating the molecular mechanisms underlying CeD[8,9]. These bulk data represent the average expression levels of individual genes across a population of cells. However, this limits the ability to capture the heterogeneity of cells resulting from differences in genetic material among individual cells. Single cell RNA sequencing (scRNA-seq) is a robust tool for characterizing gene regulatory processes, functions, molecular profiles, and interactions of cell types in healthy tissues or organs, as well as under disease conditions[10,11]. Therefore, this study aimed to better understand the pathogenesis of CeD by revisiting and examining the disease process and by identifying diagnostic and therapeutic biomarkers. A droplet-based scRNA-seq platform (10x Genomics) was used to analyze single cell transcriptomic data from duodenal tissue samples from three patients with CeD and three healthy control individuals. To determine the pathogenesis of CeD at the cellular level, single cell transcriptomic profiles were constructed, and intercellular heterogeneity in CeD tissue samples was evaluated.

Thus, this study focuses on the intestinal mucosal immunological microenvironment in CeD, to elucidate the immunological landscapes and mechanisms leading to immune dysregulation. Potential diagnostic markers for differentially expressed genes (DEG) in CeD were validated using immunohistochemistry (IHC) and quantitative polymerase chain reaction (qPCR).

Methods

Patient selection and sample collection

Biopsy samples were collected from two female and one male patient at the Department of Gastroenterology of the People's Hospital of Xinjiang Uyghur Autonomous Region. All the patients were newly diagnosed with CeD and had not received any prior treatment. The diagnostic criteria for CeD were based on the 2017 World Gastroenterology Organization global guidelines on celiac disease[2]. Three healthy control individuals (Ctrl group) were recruited from the Health Examination Center of the People's Hospital of Xinjiang Uyghur Autonomous Region. These individuals were carefully selected to match the three patients with CeD in terms of age, sex, and ethnicity, thereby ensuring a controlled comparison. Controls were screened using serum-specific antibodises (anti-tissue transglutaminase antibody and anti-endomysial antibody) to exclude any CeD cases. The exclusion criteria included a history of malignant tumors, other autoimmune diseases, acute cardiovascular or cerebrovascular disease (within the last six months), history of gastrointestinal surgery, contraindications to endoscopic biopsy, and refusal to participate in the study.

The study adhered to the principles of the Declaration of Helsinki and approval was obtained from the Ethics Committee of the People's Hospital of Xinjiang Uygur Autonomous Region (Ethics No. KY2021042011).

Fresh tissue preparation and single cell isolation

In vitro biopsy specimens were washed twice with phosphate-buffered saline (PBS), stored at 4 ℃, transported at constant liquid temperature, and processed within 24 h. The tissue was transferred to a Petri dish and washed with Dulbecco's PBS to remove blood clots and necrotic tissue. Subsequently, the tissues were transferred to a centrifuge tube with digestion solution containing 0.05% trypsin-ethylenediaminetetraacetic acid (Gibco, USA), collagenase type II, type IV(Gibco, USA), and deoxyribonuclease I (Sigma, USA). The solution was then placed on a horizontal shaker in an oven, and was digested at 37℃ and 60 rpm for 1 h. The digested supernatant was transferred to a 40-μm cell strainer (BD, USA) along with the precipitate, and the filtrate was collected in a 50 mL centrifuge tube. After centrifugation, the supernatants were decanted, the cells were resuspended in red blood cell lysis buffer (Solarbio, China), left on ice for 3 min, and then centrifuged at 300 × *g* for 5 min at 4℃. Dead cells were cleaned using a Dead Cell Removal Kit (Miltenyi Biotec, Germany). Cell suspensions were washed twice with PBS containing 0.02% bovine serum albumin (BSA) at 300 × *g* for 5 min at 4 °C. Finally, the samples were resuspended in PBS containing 0.02% BSA and were passed through a 40-μm cell sieve to collect the cell suspension. Cell viability was measured using acridine orange and propidium iodide double staining (Logos Biosystems, Korea) with the LUNA-FL Automated Fluorescence Cell Counter. The sample had to meet the following sample criteria: cell diameter ≤ 40 μm, cell viability ≥ 90%, no visible cell clumping or impurities upon microscopic examination, minimum cell count of 1×10^5 cells/sample, and cell concentration range of 5×10^5 to 1.2×10^6 cells/mL. The prepared single cell suspension was then used for library preparation.

Library preparation and sequencing

The single cell suspension was loaded into the microfluidic chip port, and the 10x Genomics Chromium™ Single Cell Controller (10x Genomics, Pleasanton, CA) sequencing platform was utilized to generate gel bead-in emulsions (GEM). The GEMs were then transferred from the chip to PCR tubes, where a reverse transcription program was executed to synthesize the complementary DNA (cDNA). Subsequently, the cDNA was purified and amplified of the cDNA. The amplified cDNA products were subjected to fragmentation, end repair, A-tailing, size selection, and adapter ligation. Full sequencing adapters were added using PCR to yield the final sequencing library. To measure library concentration and analyze the distribution of library fragment sizes, we utilized a Qubit fluorometer and an Agilent 2100 Bioanalyzer were used, respectively. The required library concentration was ≥ 10 ng/μL, with fragment lengths predominantly ranging between 350 and 500 bp, and no significant irregular peaks were observed between 100 and 140 bp. Sequencing was performed on an Illumina NovaSeq 6000 system (Illumina, USA), utilizing a high-throughput mode of 2 × 150 bp paired-end reads, which yielded an average output ≥ 50,000 reads/cell. The resulting data were obtained in the FASTQ format.

Single cell data processing and cell subsets annotation

Cell Ranger (version 5.0.0) was used to process raw data. This analysis produced a raw matrix of unique molecular identifiers (UMIs), which was subsequently transformed into a Seurat object using the R package Seurat[12]. The following criteria was used to remove low-quality cells genes expressed in < 3 cells were removed; cells with gene expression > 6,000 were removed; cells with gene expression ≤ 200 were remove; and cells with a UMI proportion of ≥ 20% derived from mitochondrial genes were excluded. The average expression and standardized variance of genes in all cells were calculated, and the top 2,000 genes with the greatest changes were selected as representatives of all genes for principal component analysis (PCA) and dimensionality reduction. Harmony tools was used to eliminate batch effects[13]. The PCA method was used to calculate the similarity between cells, and the closer the samples were, the closer the expression trend of the cell genes. The Uniform Manifold Approximation and Projection (UMAP) method for dimension reduction analysis was used to calculate the similarity between samples and draw the clustering figure.

R software v3.5.1 was used to annotate the cell types, and the software compared the gene expression profiles of each cell with the reference database to annotate the cells as the most similar type. Cells within a cluster could be annotated as multiple cell types, and the cell type with the highest proportion was designated as the cell type of the cluster. Marker genes were selected based on their differential expression patterns across various clusters. The Wilcoxon rank-sum test was used by default for marker genes, and the Bonferroni method was used for *P*-value correction. Another important statistical indicator for marker genes is the log-fold change (FC), which reflects the strength of differential expression signals. Moreover, the cell types annotated by the software were for reference only; hence, subsequent steps considered marker genes and cell surface markers reported in other analyses of human intestinal or immune cells as literature-derived markers. The expression levels of these markers in each identified cell cluster were visualized, thereby facilitating manual verification of cell identities. All marker genes were ranked from highest to lowest according to their avg_logFC values, and the top 10 genes in terms of differential FC for each cluster were selected as characteristic markers for each cell type.

Differentially expressed genes and enrichment analysis

The DEGs and FC between the CeD and Ctrl groups were calculated using the Wilcoxon rank-sum test, and the *P*-values were adjusted using the Bonferroni correction method. Cluster Profiler software (v3.10.1) was used to analyze the enrichment of DEGs in Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Fisher's exact test was used to calculate the significance levels of the pathways with significant gene enrichment.

Immunohistochemistry analysis

The tissue samples were fixed in formalin and embedded in paraffin for further processing. Sections with a thickness of 4 μm were cut from the paraffin-embedded tissues. These sections underwent IHC staining to assess the expression patterns of candidate antibodies in CeD and Ctrl tissues. Primary antibody incubation for immunostaining was conducted overnight at 4 °C. The primary antibodies included mouse-anti-CD3 (Immunoway, Suzhou, China), rabbit-anti-TRAT1 (Proteintech, Wuhan, China), rabbit-anti-BCL11B (ZENBIO, Chengdou, China), and rabbit-anti-ETS1 (Immunoway, Suzhou, China). Biotin-conjugated anti-rabbit or anti-mouse immunoglobulins from ZSGB-Bio (Beijing, China) served as the secondary antibodies for immunostaining. Hematoxylin and eosin staining was performed for counterstaining.

Quantitative reverse-transcription PCR

Total RNA was extracted using the RNeasy Mini Kit (Qiagen, Germany). For the reverse transcription process, the EasyScript First-Strand cDNA Synthesis SuperMix was used according to the manufacturer's protocol. The PCR was conducted using the ABI PRISM 7500 Sequence Detection System (Foster City, USA). Primers for the selected genes are listed in Table 1:

Statistical analysis

Data processing was conducted using R software v3.5.1. Statistical tools, methodologies, and significance thresholds for each analysis are delineated in the materials and methods section and elaborated upon in the figure legends. *P*-values < 0.05 were deemed statistically significant.

Gene	Forward primer	Reverse primer
BCL11B	CGGGCGATGCCAGAATAGAT	CATGGTCAGCCTGGGTGATG
TRAT1	TGACCACACCAGGGTTGATG	ACAGATTTCTCTGGTCGGGC
ETS1	TAGCTGGCTACACAGGCAGT	TCTCCTGGCCACCTCATCTG

Table 1. qPCR primers used in the current study. *qPCR* quantitative polymerase chain reaction.

Results

Single cell transcriptomic analysis identifies the spectrum of cell populations in the small intestinal tissue of patients with CeD

To investigate the diverse cellular components within the small intestinal tissues of patients with CeD, we used the 10x Genomics platform for scRNA-seq analysis. We constructed a single cell transcriptomic spectrum of the duodenal tissues from three individuals with CeD and three matched controls (Ctrl). After quality control, 21,080 cells with transcriptome data were retained (Fig. 1A). The detailed clinical characteristics of these individuals are shown in Table 2, and cell counts are provided in Table S1. To address batch effects among different samples, we used the Harmony13 tool to ensure a uniform blend of cells from multiple specimens. Subsequently, UMAP was used for dimensionality reduction and clustering, and 21 cellular subpopulations (Clusters 0–20) were identified. Based on classic cell marker genes and literature reviews, we revisited the annotation of the initially-identified 20 cellular subgroups, ultimately delineating 10 distinct cellular subpopulations. These consisted of three types of epithelial and stromal cells, together with seven types of immune cells. Cluster 19 was annotated as endothelial cells (*CLDN5* and *MYCT1*)[14]; Cluster 16 was annotated as enterocyte cells (*GP2*)[15]; Cluster 18 and 20 were annotated as enteroendocrine cells (EEC; *UCN3* and *KCNH6*); Cluster 13 was annotated as innate lymphoid cells (ILC; *RORC* and *IL23R*); Cluster 3 was annotated as mast cells (MC; *MS4A2*, and *TPSB2*); Cluster 14 was annotated as myeloid cells (*CD14* and *S100A9*); Clusters 8–10 were annotated as plasma cells (PC; *CCR10*, *DERL3*, and *MZB1*)[16]; Cluster 5 were annotated as natural killer cell (NK; *EOMES*, *GZMK*, and *GNLY*); Clusters 0, 1, 2, 4, 5, 6, 11, 12, 15, and 17 were annotated as T cells (*CD3D* and *CD3G*); and Cluster 7 was annotated as B cells (*MS4A1/CD20*, *BANK1*, and *TNFRSF13C*) (Fig. 1B). We conducted a detailed examination of the expression patterns of the selected canonical genes across the identified cell clusters using violin plots to verify the accuracy of the cellular identification (Fig. 1C). Each cell type was represented in both CeD and Ctrl tissue samples; however, there were variations in the proportions of different cell types, inter-subpopulation distinctions, and gene expression patterns, indicating cellular heterogeneity in the small intestinal tissues of patients with CeD. Of the cell types, immune cells constituted the largest group in the dataset (Fig. 1D). The Wilcoxon rank-sum test was used to identify cell type-specific transcripts that delineated each cellular cluster and to ascertain the top DEGs within each cell type (Fig. 1E). We focused on the pivotal role of immune cells in the pathogenesis and progression of CeD. Accordingly, we conducted individual cluster analyses on seven major immune cell populations (T cells, NKs, MCs, PCs, B cells, ILCs, and myeloid cells) to investigate cellular heterogeneity between different immune cell types.

Heterogeneity of T-cell compartments in Ctrl and CeD small intestine samples

We analyzed the single cell transcriptomic data from 15,260 T cells using unsupervised clustering to investigate the composition and characteristics of various T cell subpopulations. We identified 13 cellular subgroups (Clusters 0–12) that were annotated as the two main clusters of $CD4^+$ and $CD8^+$ T cells (Fig. 2A,B). T cells derived from individuals with CeD and the Ctrls aggregated in distinct areas, indicating considerable differences in their functional properties (Fig. 2C). Clusters 1, 3, 5, and 9 exhibited high *ICOS* and *CCR6* expression levels and were annotated as $CD4^+$ T cells. Cluster 5 showed high *FOXP3* expression levels and was annotated as Tregs cells (C4-CD4). Clusters 0, 2, 4, 6, 7, 8, 10, 11, and 12 highly expressed *CD8A*, and were annotated as $CD8^+$ T cells (Fig. 2D). We compared the DEGs in T cells between the CeD and Ctrl groups, identified 23 DEGs, and found 13 relatively highly expressed genes and 10 relatively low expressed genes ($FC > 1.5$, $P < 0.01$) in CeD tissues (Additional file: Table S2). Furthermore, these genes were enriched in antigen processing and presentation, and NK cell-mediated cytotoxicity (Fig. 2E).

We selected the DEGs, *ETS1*, *TRAT1*, and *BCL11B*, in T cells from scRNA-seq of the small intestine to clinically validate their potential as diagnostic markers of CeD. Immunohistochemical validation was performed on 10 normal small intestine paraffin sections and 10 CeD small intestine paraffin sections (Fig. 3A). CD3, a classic lymphocyte marker, was used to label lymphocytes within the epithelium as a reference for comparison between the groups. Compared with that in the Ctrl group, *ETS1*, *TRAT1*, and *BCL11B* were enriched in the intraepithelial lymphocytes of the small intestine tissue in the CeD group. Furthermore, we measured the mRNA expression levels of *ETS1*, *TRAT1*, and *BCL11B* using qPCR (Fig. 3B). Compared with that in the Ctrl group, the expression levels of *ETS1* ($P = 0.010$), *TRAT1* ($P < 0.001$), and *BCL11B* ($P = 0.036$) in the small intestine tissue of patients with CeD were increased, consistent with the results of scRNA-seq.

The interferon-induced protein with tetratricopeptide repeats (IFIT) gene family was markedly overexpressed in $CD4^+$T cell subsets

$CD4^+$ T cells are key participants in CeD pathogenesis. Therefore, we analyzed the functional status of $CD4^+$ T cells in CeD. Four different cell clusters (C1-CD4–C4-CD4) were identified (Fig. 4A). Among them, 5,097 (87.83%) $CD4^+$ T cells were derived from CeD and 706 (12.17%) $CD4^+$ T cells were derived from the Ctrl group. The number of cells in each $CD4^+$ T cell subset increased significantly in the CeD group (Fig. 4B,C). The C1_$CD4^+$

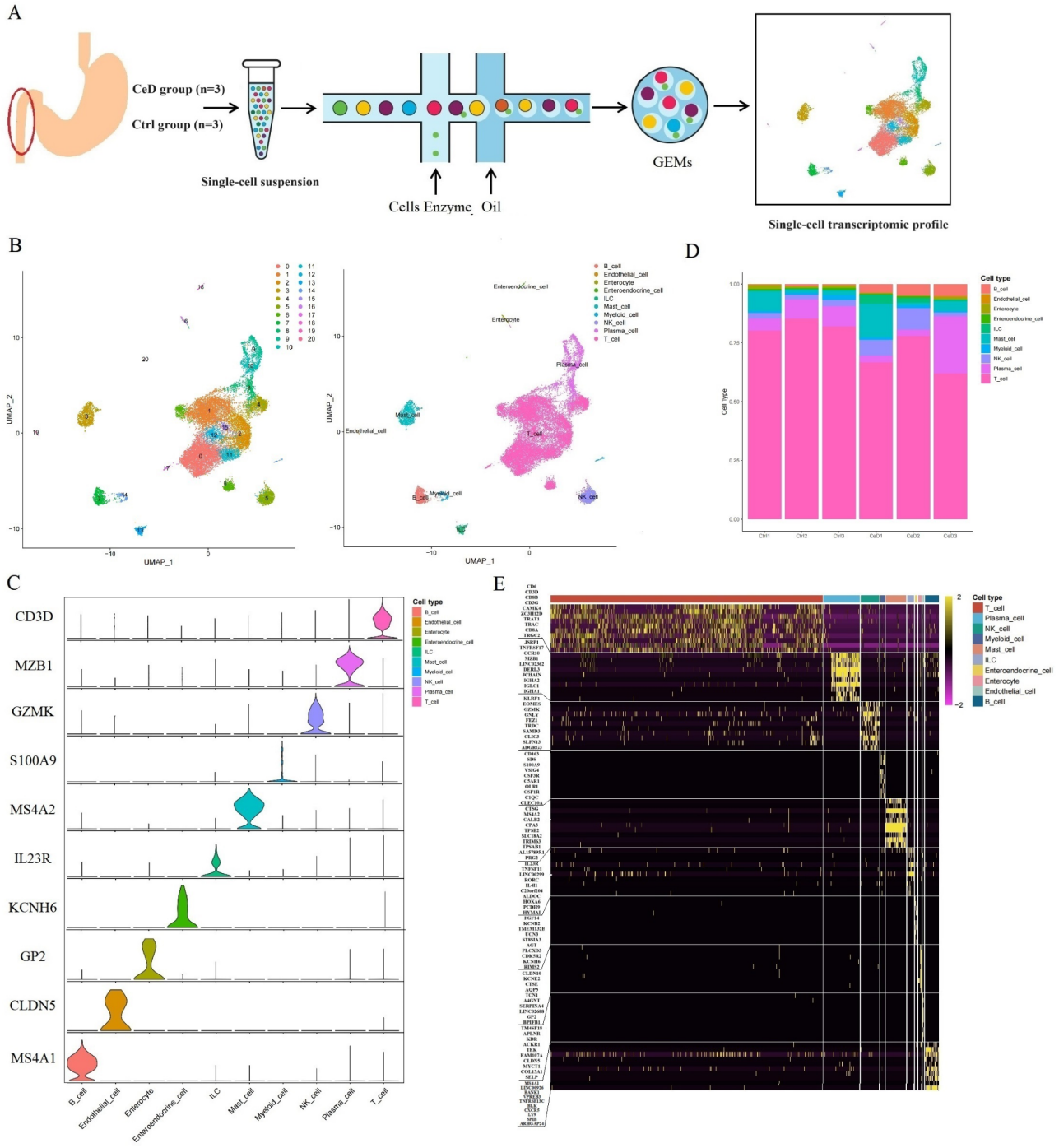

Fig. 1. Single cell expression profiling and cell typing in the small intestine of CeD and Ctrl samples. (**A**). Schematic representation of the analytical workflow for scRNA-seq in CeD. (**B**). UMAP clustering scatter plot delineating cell population distributions, encompassing 21,080 cells from Ctrl ($n=3$) and CeD ($n=3$) small intestinal tissue specimens. (**C**). Violin plots depicting the expression profiles of marker genes across identified cell types. (**D**). Barplot depicting the relative abundances of distinct cell populations within each sample. (**E**). Heatmap of differentially expressed genes of each cell type showing the top 10 genes.

T cell subset exhibited high expression of marker genes, such as *EGR1* and *IFNG*, and the C2_CD4$^+$ T cell subset primarily expressed marker genes, such as *GFI1* and *POU2F2*. The C3_CD4$^+$ T cell subset mainly expressed marker genes, such as *IFIT1*, *IFIT2*, and *IFIT3*. The *IFITs* gene family plays a crucial role in modulating innate immune responses and its members are important targets for robust antiviral activity. These genes are capable of inhibiting a wide array of viruses, thereby inducing cellular apoptosis and regulating immune responses[17]. The C4_CD4$^+$ T cell subset (Tregs cells) predominantly expressed genes such as *FOXP3*, *LAYN*, and *GEM* (Fig. 4D and Additional file: Table S3). GO enrichment analysis of the C3_CD4$^+$ T cell subgroups indicated the biological

Sample ID	Sex	Age (years)	Ethnic	IgA anti-tTG (CU)	Marsh score
CeD1	Female	56	Kazakh	4965.5	Marsh IIIc
Ctrl1	Female	52	Kazakh	1.9	–
CeD2	Female	40	Kazakh	2252.0	Marsh IIIc
Ctrl2	Female	38	Kazakh	1.9	–
CeD3	Male	39	Kazakh	4965.5	Marsh IIIc
Ctrl3	Male	39	Kazakh	1.9	–

Table 2. Clinical characteristics for enrolled patients.

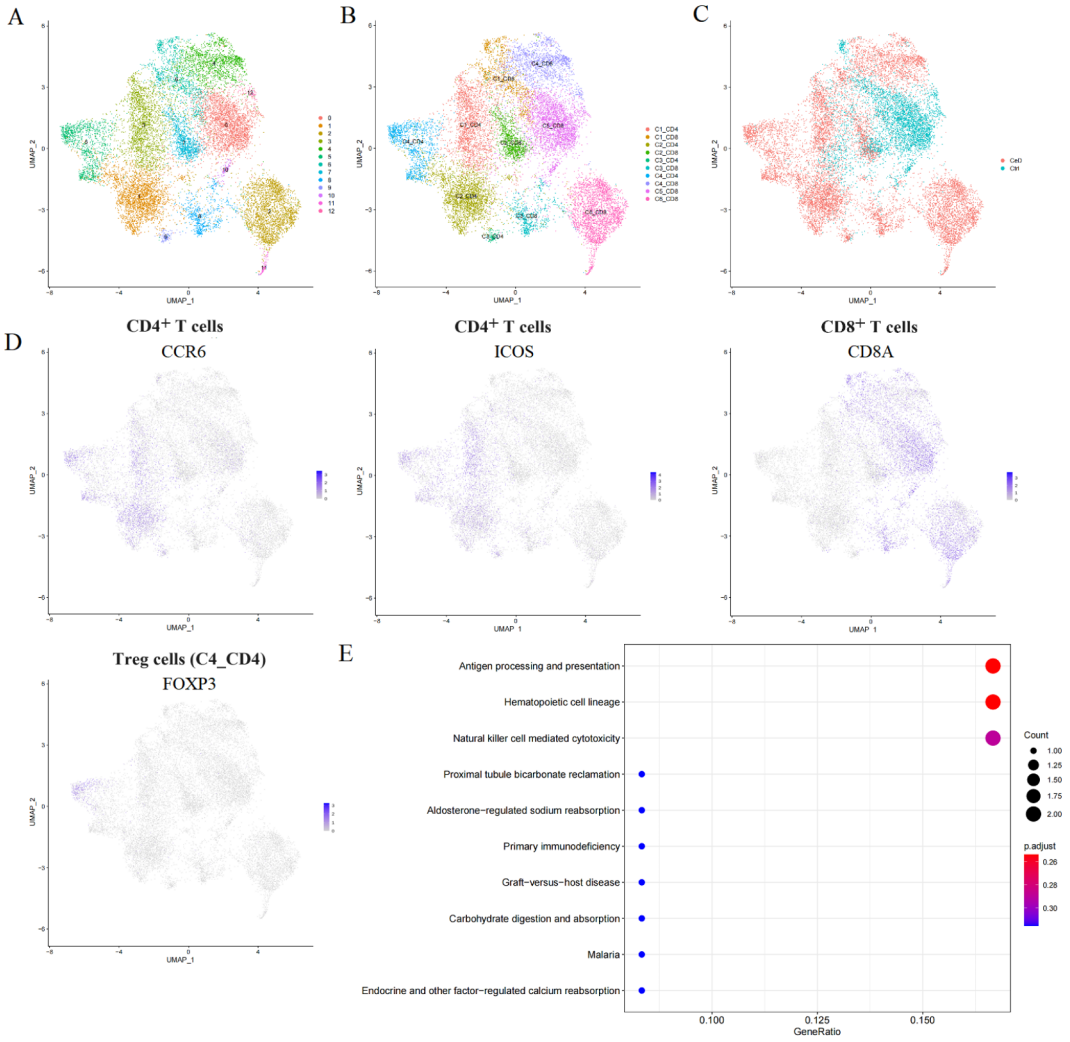

Fig. 2. Presents the single cell landscape of the T-cell compartment in the small intestine for both CeD and Ctrl samples. (**A**). UMAP visualization of T cell subsets. (**B**). UMAP visualization of $CD4^+$ T cell and $CD8^+$ T cell subsets. (**C**). UMAP visualization illustrating the distribution of T cell subsets between CeD and Ctrl groups. (**D**). Marker gene expression is depicted for $CD4^+$ T cells, $CD8^+$ T cells, and Treg cells, with darker colors indicating higher gene expression levels. **E**. The bubble plot displays KEGG analysis results for differentially expressed genes related to T cell diseases.

processes primarily involved in viral defense, which may be associated with activation due to exposure to gluten and/or viral infection or changes in the intestinal microbiome in CeD (Fig. 4E)[18]. Additionally, GO analysis of the C4_$CD4^+$ T cell subgroup (Treg cells) indicated their involvement in T cell activation, enriching a small cluster of genes related to co-stimulatory signals during T cell activation, such as *CTLA4*, *CD28*, and *TIGIT*. These phenotypes resembled the T cell activation observed in tetramer-positive T cells from CeD (Fig. 4F)[19].

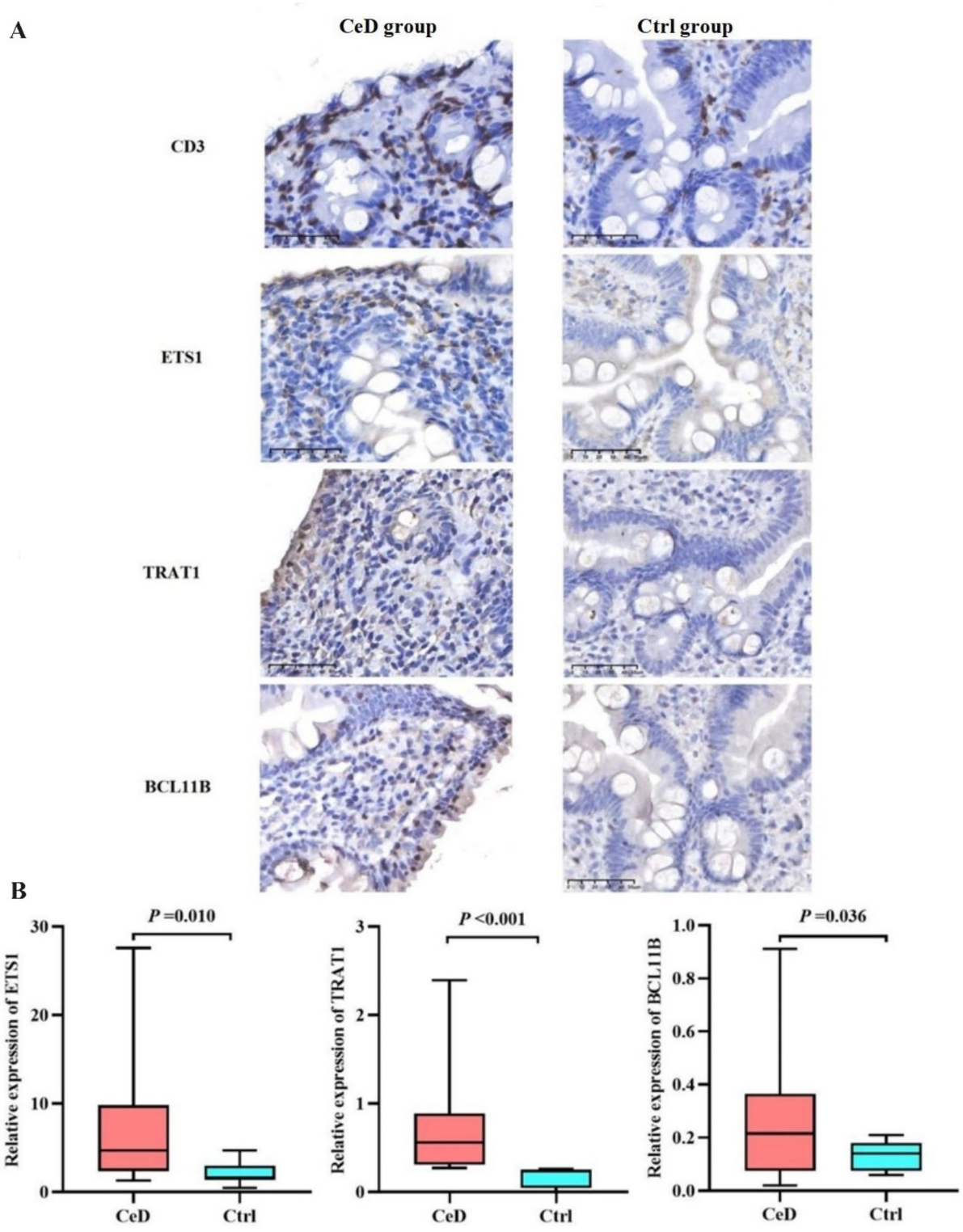

Fig. 3. Expression validation for marker genes. (**A**) Immunohistochemical staining of T cell disease differential genes in small intestine. Magnification: 40×. (**B**) The expression level of T cell disease differential genes *ETS1*, *TRAT1*, and *BCL11B* in small intestine between CeD and Ctrl groups.

The expression of cytotoxicity and T cell depletion-related genes was considerably upregulated in CD8$^+$ T cells from individuals with CeD

The 9457 CD8$^+$ T cells aggregated into six subgroups (C1_CD8–C6_CD8) (Fig. 5A). Of these, 5,610 (59.32%) CD8$^+$ T cells originated from patients with CeD, whereas 3,847 (40.62%) CD8$^+$ T cells originated from the Ctrl group. The majority of CD8$^+$ T cells in the Ctrl group were located within the C1_CD8 and C5_CD8 subgroups.

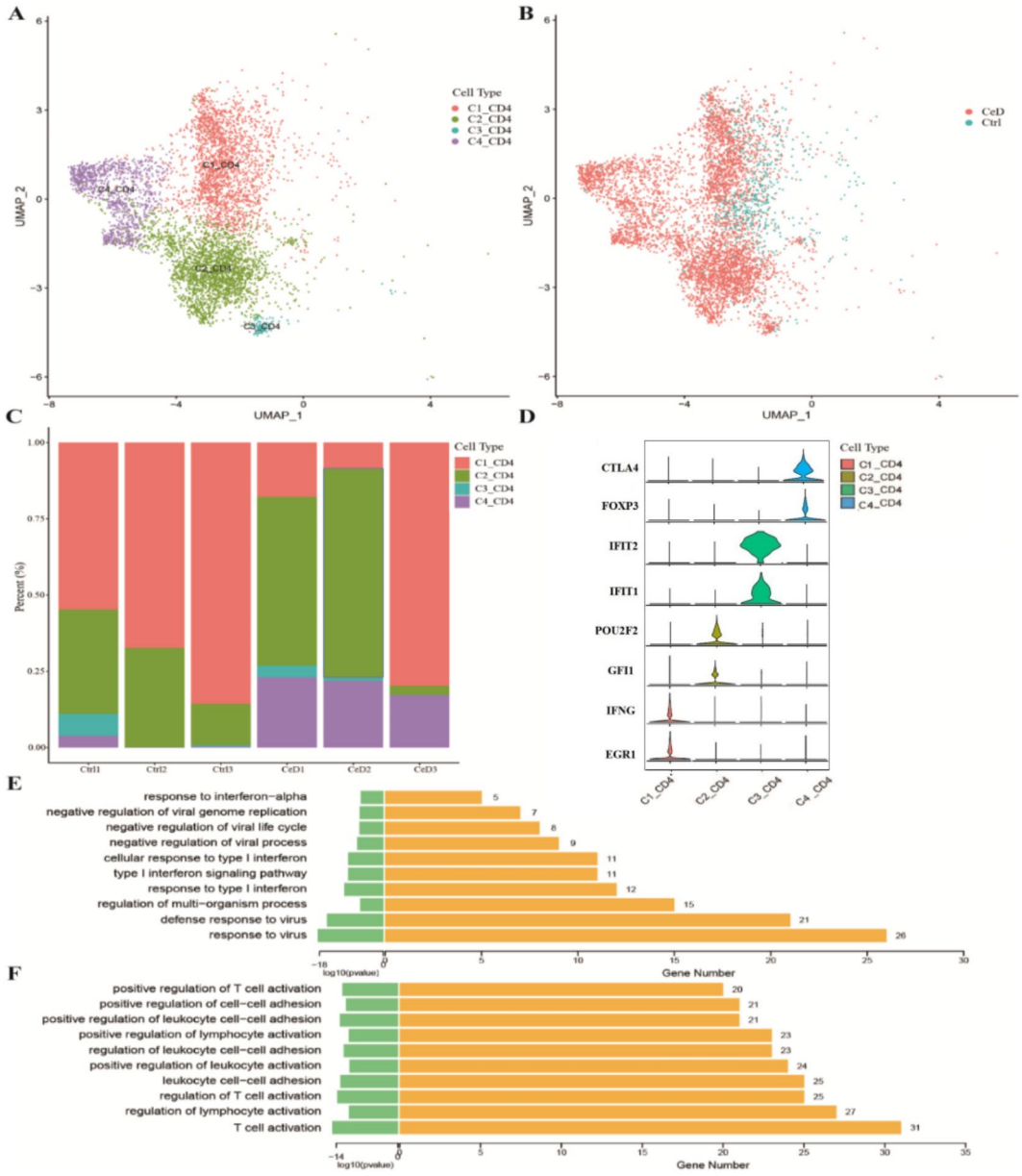

Fig. 4. The expression profile and cell typing of CD4⁺ T cells in small intestine tissue in CeD and Ctrl groups. (**A**) UMAP clustering scatter plot of CD4⁺ T cell subsets. (**B**) UMAP clustering scatter plot of the distribution of CD4⁺ T cell subsets among different groups. (**C**) Bar chart of CD4⁺ T cell subsets proportion in each sample. (**D**) Violin diagram of CD4⁺ T cell subsets marker gene. (**E**) GO analysis of C3_CD4⁺ T cell subsets. (**F**) GO analysis of C4_CD4⁺ T cell subsets.

In the CeD group, CD8⁺ T cells strongly dominated the C2_CD8, C3_CD8, C4_CD8, and C6_CD8 clusters (Fig. 5B,C). Differential expression analysis using the Wilcoxon rank-sum test (Fig. 5D; Table S4) indicated that the C1_CD8⁺ T cell subset exhibited differential expression of genes, such as *TNFSF9*, *CCL4L2*, *EGR1*, and *IFNG*. Of them, *EGR1* is involved in regulating the expression of pro-inflammatory factors[20], whereas *IFNG* (IFN-γ) is associated with the cytotoxic memory phenotype[21]. The C2_CD8⁺ T cell subset predominantly expressed major marker genes such as *AHNAK* and *SYNE2*. The C3_CD8⁺ T cell subset was distinguished by the upregulation of cytotoxicity-related genes, including *GZMK*, *GZMH*, *GZMB*, *SH2D1A*, *PRF1* (perforin), and *NKG7*. Concurrently, this subset expressed genes indicative of T cell exhaustion, including *PDCD10* and *RFF1*. The C4_CD8⁺ T cell subset principally expressed the marker genes, *IGLC2*, *TRDC*, and *NR4A1*, and extensively expressed members of the heat shock protein (HSP) family, including *HSPA1A*, *HSPA1B*, *HSPB1*, *HSPH1*, *HSPE1*, *HSPD1*, *HSPA4*, and *HSPA8*. C4_CD8⁺ T cell subpopulation marker genes are associated with several signaling pathways including "prion diseases", "mitogen-activated protein kinase (MAPK) signaling pathway", "antigen processing and presentation", "protein processing in the endoplasmic reticulum", and "longevity regulatory pathways" in the top terms of the KEGG pathway analysis (Fig. 5E). The C5_CD8⁺ T

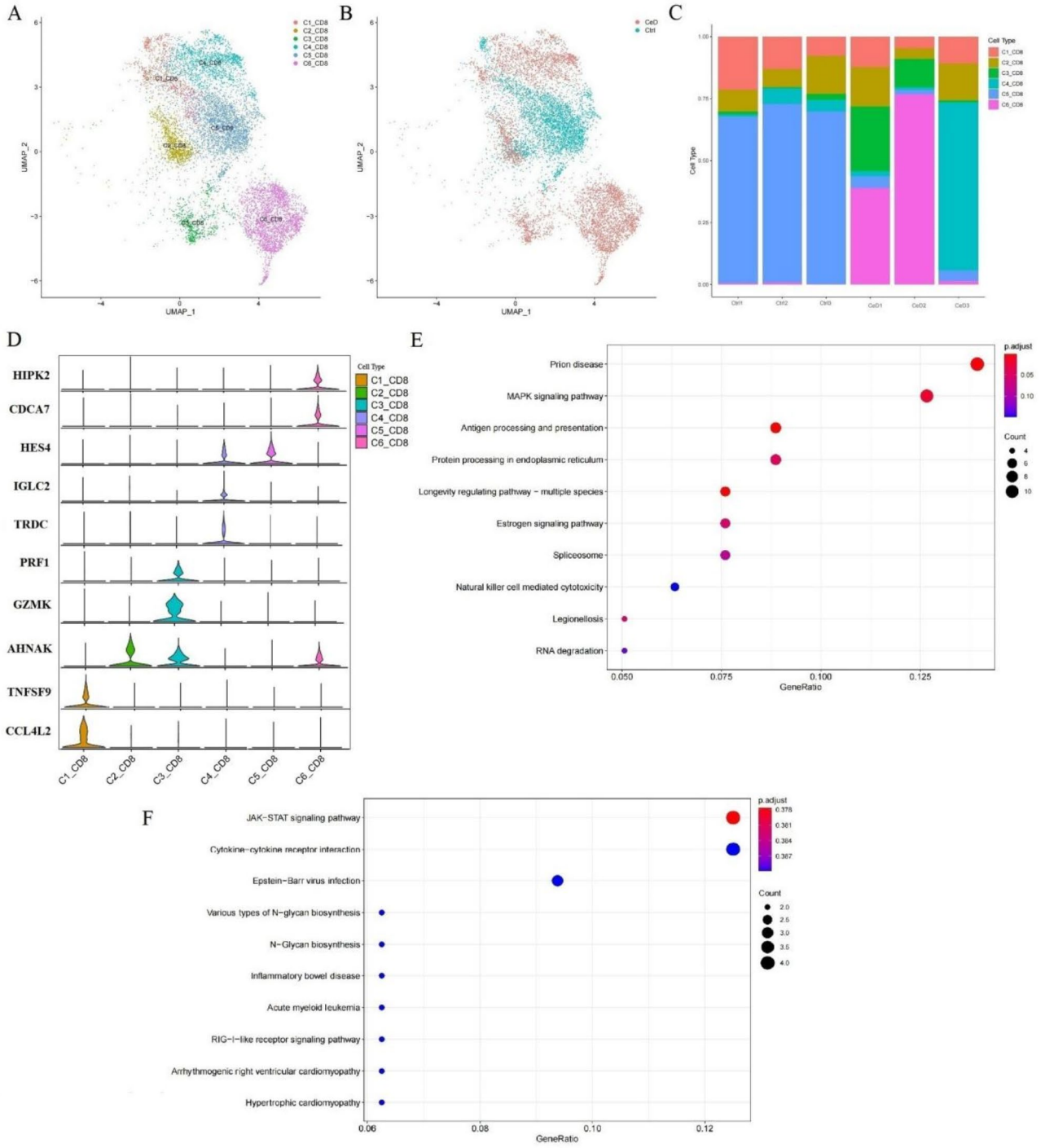

Fig. 5. The expression profile and cell typing of CD8$^+$ T cells in small intestine tissue in CeD and Ctrl groups. (**A**) UMAP clustering scatter plot of CD8$^+$ T cell subsets. (**B**) UMAP clustering scatter plot of the distribution of CD8$^+$ T cell subsets among different groups. (**C**) Bar chart of CD8$^+$ T cell subsets proportion in each sample. (**D**) Violin diagram of CD8$^+$ T cell subsets marker gene. (**E**) KEGG enrichment analysis of marker genes for the C4_CD8$^+$ T cell subsets. (**F**) KEGG enrichment analysis of marker genes for the C5_CD8$^+$ T cell subsets.

cell subset is characterized by the characteristic expression of the *IL-7R* gene, annotated as effector memory T cells, which play a vital role in regulating T cell development, differentiation, and survival[22]. The cells mainly participated in the activation of the JAK-STAT signaling pathway (Fig. 5F). Additionally, they expressed marker genes, such as *CYSTM1*, *PRXL2C*, and *HES4*. The C6_CD8$^+$ T cell subset primarily expressed marker genes (*CDCA7*, *HIPK2*, and *PRR5L)*, a cytotoxicity-related gene (*GZMB)*, and markers of T cell exhaustion (*TIGIT*, *PDCD1*, and *DUSP4)*. Table S4 lists the top 10 DEGs among the six CD8$^+$ T cell subsets.

Inflammation/infection pathways were closely related to the onset of CeD

We conducted functional enrichment analysis of DEGs across various cell types. Pathways related to infection and inflammation, including Epstein–Barr virus infection, Kaposi's sarcoma-associated herpesvirus infection, *Salmonella* infection, human cytomegalovirus infection, and pathogenic *Escherichia coli* infection, are widely present in diverse cell populations (NKs, MCs, ILCs, PCs, B cells, and myeloid cells) (Fig. 6).

Differential expression analysis of non-HLA susceptibility genes for CeD across various cell types

CeD arises from a confluence of genetic, immune, and environmental interactions. Although the carriage rate of HLA-DQ2/DQ8 among the general population is 30–40%, only approximately 3% progresses to CeD[23]. This emphasizes the involvement of additional genetic elements in the pathogenesis of this condition. To date, GWAS related to CeD have isolated 43 important non-HLA genetic risk factors[24]. To determine which cell types are most likely affected by CeD genetic determinants, we performed a cross-group (CeD vs. Ctrl) differential expression analysis on the 118 prioritized non-HLA CeD susceptibility genes from another study within various cell types[25,26] and crossed the DEGs between groups in the main cell populations. Within the set of 118 potential CeD risk genes, 20 exhibited significant differential expression in at least one cell type in our dataset, most notably T cells, B cells, MCs, NKs, ILCs, myeloid cells, EECs, and enterocytes. This indicates a cellular heterogeneity in the expression of CeD genetic risk factors and shows that these cell groups may be the main cell groups involved in CeD occurrence (Fig. 7). Non-HLA genes make a substantial contribution to the overall genetic predisposition to CeD[27], and our data assists in elucidating the complex genetics of CeD and its implications for the intestinal immune system.

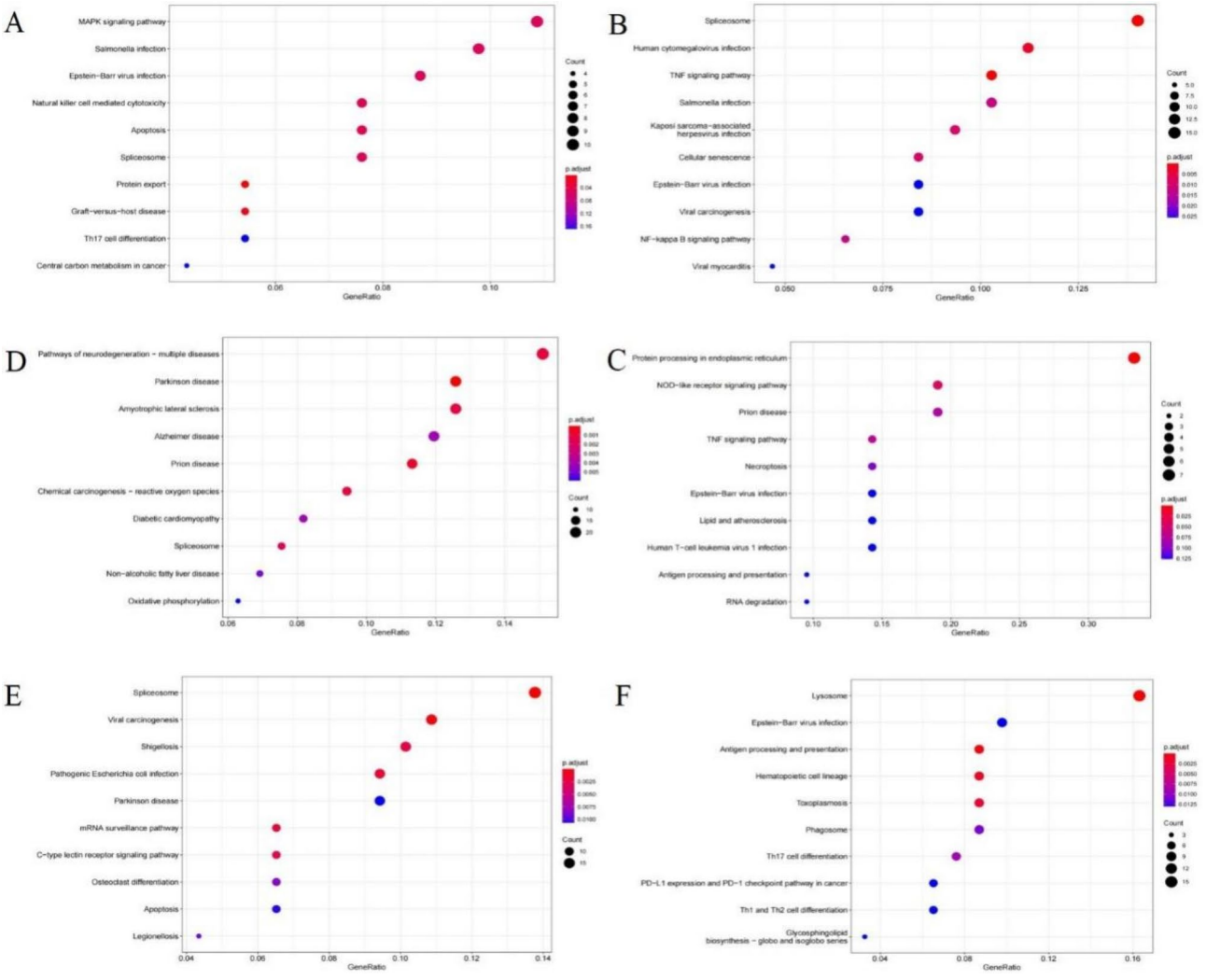

Fig. 6. Bubble chart of KEGG analysis for disease differential genes across various cell types. (**A**) KEGG Analysis Bubble Chart of DEGs in NK Cells. (**B**) KEGG Analysis Bubble Chart of DEGs in MC Cells. (**C**) KEGG Analysis Bubble Chart of DEGs in ILC Cells. (**D**) KEGG Analysis Bubble Chart of DEGs in PC Cells. (**E**.) KEGG Analysis Bubble Chart of DEGs in B Cells. (**F**) KEGG Analysis Bubble Chart of DEGs in myeloid cells.

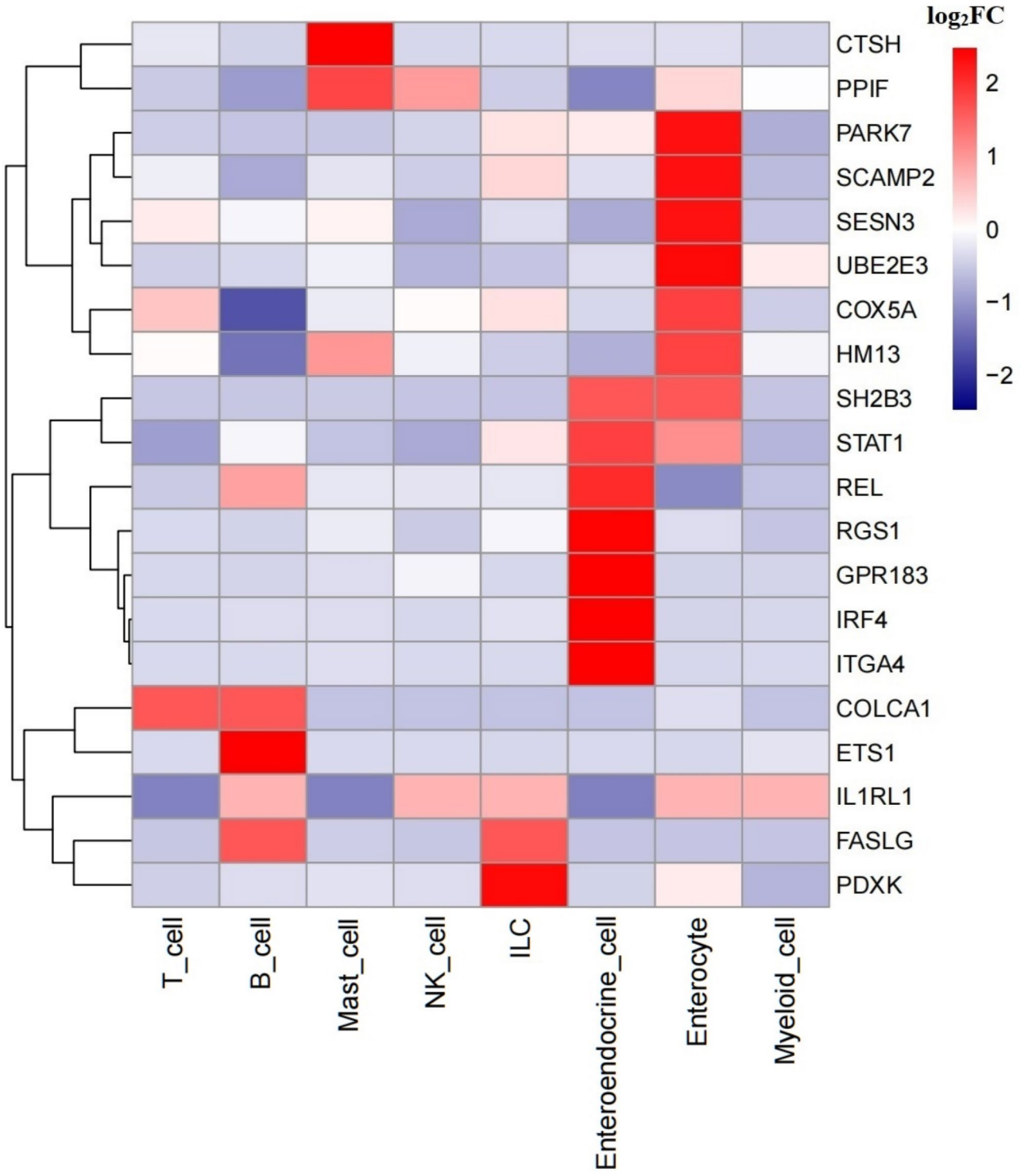

Fig. 7. Heat map of GWAS-related CeD non-HLA susceptibility genes and differentially expressed genes among different cell groups.

Discussion

CeD is a complex autoimmune disorder, leading to inflammation and villous atrophy confined to the small intestine, often affecting organ systems throughout the body[28,29]. This study presents, for the first time, novel molecular characteristics of Chinese adult patients with CeD at the single cell level. These characteristics are crucial for deciphering the heterogeneity of small intestine cells and potential disease mechanisms in CeD.

A study on CeD have focused on gluten-specific T cells[30]. In the current study, we found that the number of T cells in the CeD group was markedly increased. Substantial transcriptomic changes were observed in both $CD4^+$

and CD8$^+$ T cell populations, indicating that large subsets of both cell types were involved in the pathogenesis and progression of the disease, and that the subpopulations of T cells perform different functions. A subset of C3_CD4$^+$ T cells that significantly express *IFIT* family genes was identified. These genes are important targets for potent antiviral activity, and participate in the regulation of innate immune responses[17]. The GO enrichment analysis indicated that this subpopulation primarily plays a role in the biological processes associated with viral defense. Previous studies have not explored the major cell populations in which *IFIT* family genes regulate viral defense.

We annotated the C4_CD4$^+$ T cell subpopulation as Treg cells, and found a marked increase in Treg cell count in the CeD group. A considerable increase in the number of Foxp3$^+$ Treg cells in the intestinal mucosa of patients with active CeD has previously been reported[31,32]. In an in vitro gliadin challenge system, Foxp3$^+$ Treg cells undergo local expansion during gliadin provocation, consequently attenuating mucosal immune responses. In patients with active CeD, CD4$^+$CD25$^+$Foxp3$^+$ Treg cells from intestinal specimens demonstrated the capacity to modulate immune function in vitro[33]. However, during active CeD, these Treg cells fail to restrain small intestine mucosal inflammation, indicating a defect in their activation pathways for regulation[34]. Integrated GO enrichment analysis suggested that the C4_CD4$^+$ T cell subset was principally involved in T cell activation. A small cluster of gene sets pertinent to co-stimulatory signals during T cell activation were activated, including *CTLA4*, *CD28*, *TIGIT. CD28*, and *CTLA-4*. These are members of the immunoglobulin superfamily of receptors and are responsible for various aspects of T cell immune regulation, and directly influence the homeostasis of Treg cells[19,35]. Therefore, manipulating Treg cell homeostasis through intervention in the CD28/CTLA-4 pathway may be a potential therapeutic strategy for treating CeD.

Subsets of CD8$^+$ T cells derived from CeD cohorts frequently expressed genes associated with cytotoxicity and T cell exhaustion that play a significant role in the pathogenesis of CeD[35]. Furthermore, the C4_CD8$^+$ T cells subset broadly expressed members of the HSP family and significantly enriched in the MAPK signaling pathway. HSPs are highly conserved molecules that play important roles in immunomodulation[36]. Chladkova et al.[37] reported that the activation of p38 MAPK by gliadin fragments triggers dendritic cell transport to lymph nodes, leading to the production and expansion of differentiated gliadin-specific T cells. This facilitates the initiation of detrimental local and systemic immune responses and the development of CeD. According to the findings of this study, HSP-associated proteins may induce an immune response in the C4_CD8$^+$ T cell subset through the MAPK signaling pathway and may participate in the pathogenesis of CeD.

Many patients are suspected to have CeD based on the presentation of villous atrophy observed during endoscopic examination. However, many conditions cause villous atrophy, including food allergies, bacterial and parasitic intestinal infections, lymphocytic colitis, Crohn's disease, small intestinal bacterial overgrowth, common variable immunodeficiency, and non-steroidal anti-inflammatory drugs. Thus, identifying new, specific histopathological biomarkers is of great importance for improving the diagnostic rate of CeD and discovering potential therapeutic targets[38,39].

TRAT1 is a pivotal gene involved in regulating T cell receptors[40]. Xiao et al.[41] indicated a correlation between *TRAT1* expression levels and a spectrum of immune metrics, including stromal, immune, and ESTIMATE scores, as well as the extent of infiltration by various immune cell phenotypes, highlighting its contributory role in cytotoxic and regulatory responses. Our results suggest *TRAT1* utility as a potential diagnostic biomarker for this condition. *BCL11B* represents an indispensable transcriptional regulator pivotal for the lineage specification and maturation of T cells, with its dysregulation implicated in the occurrence of T cell tumors[42]. *ETS1* is a principal member of the ETS transcription factor family[43]. In a study associated with CeD GWAS, *ETS1* was identified as an immune-related gene primarily involved in the development of thymic T cells and was recognized as an important risk factor for CeD[44]. However, the *TRAT1*, *BCL11B*, and *ETS1* genes are mostly implicated in the development of cancer, with fewer studies focusing on their roles in CeD and other autoimmune diseases. As diagnostic markers for CeD, they need to be validated in a larger sample population, and the relationship between their expression levels in CeD and immune cells requires further exploration.

Infection is one of the environmental factors involved in the pathogenic mechanisms of CeD[45–47], and the relationship between infection factors and disease onset is typically complex and multidirectional. The current study found that Epstein–Barr, human cytomegalovirus, *Salmonella*, pathogenic *E. coli*, Kaposi sarcoma-associated herpes virus, and inflammation-related pathways were widely present in CeD-associated cells[48]. This could indicate that, in addition to activated gluten-specific T cells, a large population of other lymphocytes, are triggered by viral/bacterial stimuli as a potential contributing mechanism in the course of the disease.

This study had certain limitations. First, the low incidence of CeD, long sample collection time, and small sample size may have affected the stability of the results. Second, small intestinal tissue samples can only provide a limited number of tissue blocks. The limited number of captured cells resulted in fewer cell subsets. Finally, the study lacked longitudinal data analysis across the different pathological stages from pre-onset and early onset to post-treatment. This will be addressed in subsequent studies to comprehensively explore the development and outcome of the disease more comprehensively.

In this study, we successfully constructed a single cell transcriptome expression profile of CeD small intestine tissue using scRNA-seq. Moreover, we obtained tissue architecture information within each cell group through a separate clustering analysis of immune cell groups. In future studies, we intend to perform cell sorting of these specific subpopulations to validate their phenotypes and functionalities.

Data availability

The datasets generated and/or analysed during the current study are available in the the Gene Expression Bank (GEO). The names of the repository/repositories and accession number(s) can be found below: GSE282570. The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 30 July 2024; Accepted: 12 February 2025
Published online: 11 April 2025

References

1. Catassi, C., Verdu, E. F., Bai, J. C. & Lionetti, E. Coeliac disease. *Lancet* **399**, 2413–2426. https://doi.org/10.1016/s0140-6736(22)00794-2 (2022).
2. Bai, J. C. & Ciacci, C. World gastroenterology organisation global guidelines: Celiac disease February 2017. *J. Clin. Gastroenterol.* **51**, 755–768. https://doi.org/10.1097/mcg.0000000000000919 (2017).
3. Singh, P. et al. Global prevalence of celiac disease: Systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **16**, 823–36e2. https://doi.org/10.1016/j.cgh.2017.06.037 (2018).
4. Durazzo, M. et al. Extra-intestinal manifestations of celiac disease: What should we know in 2022? *J. Clin. Med.* **11,** 258. https://doi.org/10.3390/jcm11010258 (2022).
5. Kochhar, G. S., Singh, T., Gill, A. & Kirby, D. F. Celiac disease: Managing a multisystem disorder. *Cleve Clin. J. Med.* **83**, 217–227. https://doi.org/10.3949/ccjm.83a.14158 (2016).
6. Lebwohl, B., Sanders, D. S. & Green, P. H. R. Coeliac disease. *Lancet* **391**, 70–81. https://doi.org/10.1016/s0140-6736(17)31796-8 (2018).
7. Caio, G. et al. Celiac disease: A comprehensive current review. *BMC Med.* **17**, 142. https://doi.org/10.1186/s12916-019-1380-z (2019).
8. García-Santisteban, I., Romero-Garmendia, I., Cilleros-Portet, A., Bilbao, J. R. & Fernandez-Jimenez, N. Celiac disease susceptibility: The genome and beyond. *Int. Rev. Cell. Mol. Biol.* **358**, 1–45. https://doi.org/10.1016/bs.ircmb.2020.10.002 (2021).
9. Gnodi, E., Meneveri, R. & Barisani, D. Celiac disease: From genetics to epigenetics. *World J. Gastroenterol.* **28**, 449–463. https://doi.org/10.3748/wjg.v28.i4.449 (2022).
10. Jovic, D. et al. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin. Transl Med.* **12**, e694. https://doi.org/10.1002/ctm2.694 (2022).
11. Bigaeva, E., Uniken Venema, W. T. C., Weersma, R. K. & Festen, E. A. M. Understanding human gut diseases at single-cell resolution. *Hum. Mol. Genet.* **29**, R51–r58. https://doi.org/10.1093/hmg/ddaa130 (2020).
12. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420. https://doi.org/10.1038/nbt.4096 (2018).
13. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0 (2019).
14. Kabir, A. U. et al. Dual role of endothelial Myct1 in tumor angiogenesis and tumor immunity. *Sci. Transl Med.* **13,** eabb6731. https://doi.org/10.1126/scitranslmed.abb6731 (2021).
15. George, J. J. et al. Polycomb repressive complex 2 regulates genes necessary for intestinal Microfold cell (M cell) development. *Cell. Mol. Gastroenterol. Hepatol.* **12**, 873–889. https://doi.org/10.1016/j.jcmgh.2021.05.014 (2021).
16. da Silva, F. A. R. et al. Whole transcriptional analysis identifies markers of B, T and plasma cell signaling pathways in the mesenteric adipose tissue associated with Crohn's disease. *J. Transl Med.* **18**, 44. https://doi.org/10.1186/s12967-020-02220-3 (2020).
17. Zhou, X. et al. Interferon induced IFIT family genes in host antiviral defense. *Int. J. Biol. Sci.* **9**, 200–208. https://doi.org/10.7150/ijbs.5613 (2013).
18. Leonard, M. M. et al. Microbiome signatures of progression toward celiac disease onset in at-risk children in a longitudinal prospective cohort study. *Proc. Natl. Acad. Sci. U S A* **118,** e2020322118. https://doi.org/10.1073/pnas.2020322118 (2021).
19. Yao, Y., Wyrożżemski, Ł., Lundin, K. E. A., Sandve, G. K. & Qiao, S. W. Differential expression profile of gluten-specific T cells identified by single-cell RNA-seq. *PLoS One* **16**, e0258029. https://doi.org/10.1371/journal.pone.0258029 (2021).
20. Decker, E. L. et al. Early growth response proteins (EGR) and nuclear factors of activated T cells (NFAT) form heterodimers and regulate proinflammatory cytokine gene expression. *Nucleic Acids Res.* **31**, 911–921. https://doi.org/10.1093/nar/gkg186 (2003).
21. Vorkas, C. K. et al. Single-cell transcriptional profiling reveals signatures of helper, effector, and regulatory MAIT cells during homeostasis and activation. *J. Immunol.* **208**, 1042–1056. https://doi.org/10.4049/jimmunol.2100522 (2022).
22. Alkalay, M. J. Update on celiac disease. *Curr. Opin. Pediatr.* **32**, 654–660. https://doi.org/10.1097/mop.0000000000000936 (2020).
23. Brown, N. K., Guandalini, S., Semrad, C. & Kupfer, S. S. A clinician's guide to celiac disease HLA genetics. *Am. J. Gastroenterol.* **114**, 1587–1592. https://doi.org/10.14309/ajg.0000000000000310 (2019).
24. 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. https://doi.org/10.1038/s41431-019-0520-4 (2020).
25. van der Graaf, A. et al. Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signaling in celiac disease. *Front. Genet.* **11**, 562434. https://doi.org/10.3389/fgene.2020.562434 (2021).
26. Ramírez-Sánchez, A. D. et al. Single-cell RNA sequencing of peripheral blood mononuclear cells from pediatric coeliac disease patients suggests potential pre-seroconversion markers. *Front. Immunol.* **13**, 843086. https://doi.org/10.3389/fimmu.2022.843086 (2022).
27. Verdu, E. F. & Schuppan, D. Co-factors, microbes, and immunogenetics in celiac disease to guide novel approaches for diagnosis and treatment. *Gastroenterology* **161**, 1395–411e4. https://doi.org/10.1053/j.gastro.2021.08.016 (2021).
28. Lebwohl, B. & Rubio-Tapia, A. Epidemiology presentation, and diagnosis of celiac disease. *Gastroenterology* **160**, 63–75. https://doi.org/10.1053/j.gastro.2020 (2021). 06.098.
29. Therrien, A., Kelly, C. P. & Silvester, J. A. Celiac disease: Extraintestinal manifestations and associated conditions. *J. Clin. Gastroenterol.* **54**, 8–21. https://doi.org/10.1097/mcg.0000000000001267 (2020).
30. Christophersen, A. et al. Distinct phenotype of CD4(+) T cells driving celiac disease identified in multiple autoimmune conditions. *Nat. Med.* **25**, 734–737. https://doi.org/10.1038/s41591-019-0403-9 (2019).
31. Tiittanen, M., Westerholm-Ormio, M., Verkasalo, M., Savilahti, E. & Vaarala, O. Infiltration of forkhead box P3-expressing cells in small intestinal mucosa in coeliac disease but not in type 1 diabetes. *Clin. Exp. Immunol.* **152**, 498–507. https://doi.org/10.1111/j.1365-2249.2008.03662.x (2008).
32. Vorobjova, T. et al. Increased FOXP3 expression in small-bowel mucosa of children with coeliac disease and type I diabetes mellitus. *Scand. J. Gastroenterol.* **44**, 422–430. https://doi.org/10.1080/00365520802624177 (2009).
33. Zanzi, D. et al. IL-15 interferes with suppressive activity of intestinal regulatory T cells expanded in Celiac disease. *Am. J. Gastroenterol.* **106**, 1308–1317. https://doi.org/10.1038/ajg.2011.80 (2011).
34. Mazzarella, G. Effector and suppressor T cells in celiac disease. *World J. Gastroenterol.* **21**, 7349–7356. https://doi.org/10.3748/wjg.v21.i24.7349 (2015).
35. Rowshanravan, B., Halliday, N. & Sansom, D. M. CTLA-4: A moving target in immunotherapy. *Blood* **131**, 58–67. https://doi.org/10.1182/blood-2017-06-741033 (2018).
36. Kasperkiewicz, M. et al. Evidence for a role of autoantibodies to heat shock protein 60, 70, and 90 in patients with dermatitis herpetiformis. *Cell. Stress Chaperones* **19**, 837–843. https://doi.org/10.1007/s12192-014-0507-6 (2014).
37. Chladkova, B. et al. Gliadin fragments promote migration of dendritic cells. *J. Cell. Mol. Med.* **15**, 938–948. https://doi.org/10.1111/j.1582-4934.2010.01066.x (2011).

38. Dotsenko, V. et al. Transcriptomic analysis of intestine following administration of a transglutaminase 2 inhibitor to prevent gluten-induced intestinal damage in celiac disease. *Nat. Immunol.* **25** (7), 1218–1230. https://doi.org/10.1038/s41590-024-01867-0 (2024).
39. Dotsenko, V. et al. Genome-wide transcriptomic analysis of intestinal mucosa in celiac disease patients on a gluten-free diet and postgluten challenge. *Cell. Mol. Gastroenterol. Hepatol.* **11** (1), 13–32. https://doi.org/10.1016/j.jcmgh.2020.07.010 (2021).
40. Guo, Q. et al. The roles and mechanisms of TRAT1 in the progression of non-small cell lung cancer. *Curr. Med. Sci.* **42**, 1186–1200. https://doi.org/10.1007/s11596-022-2625-1 (2022).
41. Xiao, X. Y. et al. TRAT1 overexpression delays cancer progression and is associated with immune infiltration in lung adenocarcinoma. *Front. Oncol.* **12**, 960866. https://doi.org/10.3389/fonc.2022.960866 (2022).
42. Fang, H. et al. Expression pattern and diagnostic utility of BCL11B in mature T- and NK-cell neoplasms. *Pathology* **54**, 893–899. https://doi.org/10.1016/j.pathol.2022.04.012 (2022).
43. Ren, Y., Chen, B. & Zhang, M. Distinct prognostic and immunological roles of ETS1 and ETS2: A pan-cancer analysis. *Biomed. Res. Int.* **2023**, 4343350. https://doi.org/10.1155/2023/4343350 (2023).
44. Dubois, P. C. et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302. https://doi.org/10.1038/ng.543 (2010).
45. Brown, J. J., Jabri, B. & Dermody, T. S. A viral trigger for celiac disease. *PLoS Pathog.* **14**, e1007181. https://doi.org/10.1371/journal.ppat.1007181 (2018).
46. Barone, M. V. & Auricchio, S. A cumulative effect of food and viruses to trigger celiac disease (CD): A commentary on the recent literature. *Int. J. Mol. Sci.* **22,** 2027. https://doi.org/10.3390/ijms22042027 (2021).
47. Jiang, H. Y., Zhang, X., Zhou, Y. Y., Jiang, C. M. & Shi, Y. D. Infection, antibiotic exposure, and risk of celiac disease: A systematic review and meta-analysis. *J. Gastroenterol. Hepatol.* **35**, 557–566. https://doi.org/10.1111/jgh.14928 (2020).
48. Atlasy, N. et al. Single cell transcriptomic analysis of the immune cell compartment in the human small intestine and in Celiac disease. *Nat. Commun.* **13** (1), 4920. https://doi.org/10.1038/s41467-022-32691-5 (2022).

Acknowledgements
The authors thank the patients-participants at the People's Hospital of Xinjiang Uygur Autonomous Region for contributing intestinal biopsy samples.

Author contributions
Conceptualization, ST and GF; Writing orignal draft, ST and FY; data curation, FY and MJ; software, L-WD and LT; formal analysis, LN; methodology, TA and X-SL; project administration, AA; supervision, GF. MJ, L-WD, LN, LT, AA, TA, X-SL, and GF supervised and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding
This work was supported by Central Guidance of Local Science and Technology Development Special Fund Project (ZYYD2022A06), and Hospital Project of People's Hospital of Xinjiang Uygur Autonomous Region (20200406).

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
The Written Informed consent were signed by all participants and ethical approval was obtained from the Ethics Committee of the People's Hospital of Xinjiang Uyghur Autonomous Region (KY2021042011).

Consent for publication
All authors have agreed to publish this manuscript.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-90300-z.

Correspondence and requests for materials should be addressed to F.G.