UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN

SCIENZE DELLA RIPRODUZIONE E DELLO SVILUPPO

Understanding celiac disease at single cell resolution

Settore scientifico-disciplinare: MED/38

DOTTORANDO
Giuseppe Molinario

COORDINATORE
Prof. Adamo Pio D'adamo

SUPERVISORE DI TESI
Dott. Andrea Taddio

CO-SUPERVISORE DI TESI
Dott.ssa Luigina De Leo
Dott.ssa Sara Lega

ANNO ACCADEMICO 2024/2025

UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN

SCIENZE DELLA RIPRODUZIONE E DELLO SVILUPPO

Understanding celiac disease at single cell resolution

Settore scientifico-disciplinare: MED/38

DOTTORANDO
Giuseppe Molinario

COORDINATORE
Prof. Adamo Pio D'adamo

SUPERVISORE DI TESI
Dott. Andrea Taddio

CO-SUPERVISORE DI TESI
Dott.ssa Luigina De Leo
Dott.ssa Sara Lega

ANNO ACCADEMICO 2024/2025

CONTENTS

I. ABSTRACT

Celiac disease (CD) is one of the most common chronic lifelong diseases, with a prevalence of 1% in the general population. A triad of factors (genetic predisposition, gluten and environmental factors) is involved in the pathogenesis of celiac disease that is not yet fully understood. Recent discoveries made with single cell RNA sequencing (scRNA-seq) have significantly improved our understanding of the human intestine, both in normal conditions and also across the spectrum of gut diseases. To study celiac disease with a scRNA-seq approach could provide new insights into the pathogenesis of celiac disease. The aims of the project were: (1) To determine the optimal procedure of manipulation of intestinal biopsy samples from children with CD to generate single-cell RNA sequencing libraries; (2) generation of single-cell libraries from intestinal biopsies from children undergoing esophagogastroduodenoscopy for suspected celiac disease or other gastrointestinal disorders; (3) bioinformatic analysis of single cells data to identify all types of intestinal cells and identify differences between celiac patients and controls.

This study was performed using intestinal biopsy samples from children.

The GEXSCOPE kit (Singleron Biotechnologies) was used to generate the scRNA-seq libraries. The generated libraries were subsequently sequenced utilizing the Illumina platform. The resulting raw sequencing data were then processed and analyzed using the Celescope (a dedicated Linux-based analytical package) and Seurat (an RStudio software package) analytical tools.

For the first objective, a pilot study was conducted. Four children were enrolled and divided into two groups each one including 1 control and 1 CD. Intestinal samples were processed after storage in a preservation fluid (sCelLiVETM) for 3 days or immediately after collection. Single cell libraries were successfully generated from all patients. However, after sequencing, quality control analysis revealed:

- High mitochondrial RNA concentration and consequently low-quality scRNA-seq data in stored samples.
- Top-notch quality scRNA-seq data from fresh samples.

This indicates that sample processing time is a crucial factor in obtaining high-quality scRNA-seq data from intestinal biopsy samples. Therefore, the 2 samples processed after storage were excluded from the study. These two samples were excluded from

further studies, however they served as samples for the pilot study, so in principle they served to test the first hypothesis or to reach the achieve the first aim. So they were only excluded from the second part of the study.

For this project, single-cell libraries were generated from fresh intestinal biopsy samples from 5 celiac patients and 5 controls.

After sequencing and quality control step top-notch quality data from 70.000 single cells were included in the final dataset. The cluster analysis identified 21 cell types, which were found in all patients. A significant difference in proportion of different cell types was observed was observed between the groups: Children with CeD had higher number of immune cells and lower number of epithelial cells than controls. Finally, the cell-cell communication analysis between the various cell types was performed in both celiac patients and controls. This analysis revealed a high communication in the immune intestinal compartment of celiac patients, in comparison to controls in which cell communication appears to be homogeneous for all cell types.

II. RIASSUNTO

La celiachia (CD) è una delle malattie croniche permanenti più comuni, con una prevalenza dell'1% nella popolazione generale. Nella patogenesi della celiachia è coinvolta una triade di fattori (predisposizione genetica, glutine e fattori ambientali) che non è ancora del tutto compresa. Le recenti scoperte effettuate con il sequenziamento dell'RNA a singola cellula (scRNA-seq) hanno migliorato significativamente la nostra comprensione dell'intestino umano, sia in condizioni normali che nell'intero spettro delle malattie intestinali. Studiare la celiachia con un approccio scRNA-seq potrebbe fornire nuove informazioni sulla patogenesi della celiachia. Gli obiettivi del progetto erano: (1) Determinare la procedura ottimale di manipolazione di campioni di biopsia intestinale di bambini con CD per generare librerie di sequenziamento di RNA a singola cellula; (2) generazione di librerie a singola cellula da biopsie intestinali di bambini sottoposti a esofagogastroduodenoscopia per sospetta celiachia o altri disturbi gastrointestinali; (3) analisi bioinformatica dei dati delle singole cellule per identificare tutti i tipi di cellule intestinali e identificare le differenze tra pazienti celiaci e controlli.

Questo studio è stato eseguito utilizzando campioni di biopsia intestinale di bambini.

Per generare le librerie scRNA-seq è stato utilizzato il kit GEXSCOPE (Singleron Biotechnologies). Le librerie generate sono state successivamente sequenziate utilizzando la piattaforma Illumina. I dati grezzi di sequenziamento risultanti sono stati quindi elaborati e analizzati utilizzando gli strumenti analitici Celescope (un pacchetto analitico dedicato basato su Linux) e Seurat (un pacchetto software RStudio). Per il primo obiettivo è stato condotto uno studio pilota. Quattro bambini sono stati arruolati e divisi in due gruppi ciascuno comprendente 1 controllo e 1 CD. I campioni intestinali sono stati trattati dopo la conservazione in un fluido di conservazione (sCelLiVETM) per 3 giorni o immediatamente dopo la raccolta. Sono state generate con successo librerie di singole cellule da tutti i pazienti. Tuttavia, dopo il sequenziamento, l'analisi del controllo qualità ha rivelato:

- Elevata concentrazione di RNA mitocondriale e di conseguenza dati scRNA-seq di bassa qualità nei campioni conservati.
- Dati scRNA-seq di altissima qualità da campioni freschi.

Ciò indica che il tempo di elaborazione del campione è un fattore cruciale per ottenere dati scRNA-seq di alta qualità da campioni di biopsia intestinale. Pertanto, i 2 campioni

trattati dopo la conservazione sono stati esclusi dallo studio. Questi due campioni sono stati esclusi da ulteriori studi, tuttavia sono serviti come campioni per lo studio pilota, quindi in linea di principio sono serviti a testare la prima ipotesi o a raggiungere il primo obiettivo. Quindi sono stati esclusi solo dalla seconda parte dello studio.

Per questo progetto sono state generate librerie di singole cellule da campioni freschi di biopsia intestinale di 5 pazienti celiaci e 5 controlli.

Dopo la fase di sequenziamento e controllo qualità, nel set di dati finale sono stati inclusi dati di qualità di prim'ordine provenienti da 70.000 singole cellule. L'analisi dei cluster ha identificato 21 tipi di cellule, presenti in tutti i pazienti. È stata osservata una differenza significativa nella proporzione dei diversi tipi di cellule tra i gruppi: i bambini con CeD avevano un numero maggiore di cellule immunitarie e un numero inferiore di cellule epiteliali rispetto ai controlli. Infine, l'analisi della comunicazione cellula-cellula tra i vari tipi di cellule è stata eseguita sia su pazienti celiaci che su controlli. Questa analisi ha rivelato un'elevata comunicazione nel compartimento immunitario intestinale dei pazienti celiaci, rispetto ai controlli in cui la comunicazione cellulare sembra essere omogenea per tutti i tipi di cellule.

III. LIST OF FIGURES

1. INTRODUCTION

1.1 Celiac disease

Celiac disease (CD) is an autoimmune condition characterized by an abnormal reaction of the immune system to gluten ingestion. It is one of the most common autoimmune disorders and has a prevalence of 1-2% in the general population (Catassi et al. 2021). This pathology affects genetically predisposed individuals, carriers of the human leukocyte antigen (HLA) -DQ2 and/or -DQ8 haplotypes, in response to ingestion of gluten derived proteins, particularly gliadin, present in various cereals, including wheat, rye, barley, spelt and Khorasan (Kivelä et al. 2021; Lindfors et al. 2019). However, there are other genetic and/or environmental factors involved in this pathology, in fact only a small fraction of HLA-DQ2-positive and/or HLA-DQ8-positive subjects develop CD (3%) (Salazar, García-Cárdenas, and Paz-y-Miño 2017). Wheat gluten is a complex mixture of alcohol-soluble gliadin and alcohol-insoluble glutenin. These proteins contain a high percentage of proline and glutamine residues that give them resistance to proteolytic treatment by gastric and pancreatic enzymes (Sulic et al. 2015). Upon release into the gastrointestinal tract, gluten-derived gliadin peptides initiate immune system activation in individuals with celiac disease. These peptides can cross the intestinal epithelial barrier either between epithelial cells and/or by exploiting active transport mediated by IgA immunoglobulins (Hausch et al. 2002) and in the intestinal lamina propria are deamidated by the tissue transglutaminase 2 (TG2) enzyme. The deamidation process increases the affinity of peptides for HLA-DQ2/DQ8 molecules expressed on the membrane of antigen presenting cells (APCs), facilitating their interaction with gluten-specific CD4+ T cells. In patients with CD activation of CD4+ T cells involves the release of pro-inflammatory cytokines, including IFNγ and IL-21, which contribute to mucosal damage (Nisticò et al. 2006). This inflammatory process causes atrophy of the intestinal villi, the main characteristic of CD (Cardoso-Silva et al. 2019). Villous atrophy, together with chronic inflammation of the upper small intestine, can cause malabsorption of essential nutrients leading to symptoms ranging from anemia to bone fragility (Nisticò et al. 2006). The chronic inflammatory status of the duodenal mucosa can be restored to a normal pathophysiological state following a strict gluten-free diet (Figure 1).

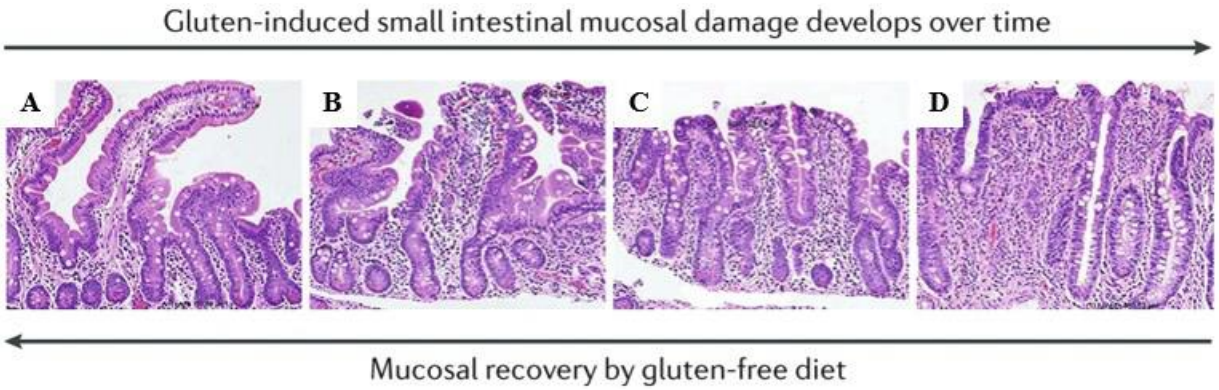


Figure 1: Small intestinal mucosa in celiac disease. *In celiac disease, the lesions of the intestinal mucosa induced by gluten ingestion evolve over time. Initially, the mucosa appears normal (A), but progresses to inflammation with hyperplasia of the crypts (B) and then to villous atrophy associated with hyperplasia of the crypts (C-D). The images are perpendicular sections of the intestinal mucosa obtained from biopsies of coeliac patients. The introduction of a strict gluten-free diet leads to the restoration of intestinal lesions and the recovery of the mucosa* (Lindfors et al. 2019).

1.1.1 Clinical manifestations

Celiac disease can occur at any age, from early childhood to late adulthood, and is characterized by two distinct peaks of onset. The first peak occurs shortly after the introduction of gluten during the first 2 years of life, while the second peak occurs in the second or third decade of life (Caio et al. 2019). The clinical manifestations of celiac disease are multiform, as it can be symptomatic (gastrointestinal and/or extraintestinal symptoms) or asymptomatic. The classic intestinal symptoms are represented by malnutrition, diarrhea, failure to thrive, abdominal pain and early distension in life (Barker and Liu 2008). Various extra-intestinal symptoms include microcytic anemia from iron deficiency or, more rarely, macrocytic anemia caused by folic acid and/or vitamin B12 deficiency (Balaban et al. 2019). Other extra-intestinal symptoms are also osteopenia and/ or osteoporosis, due to altered absorption of calcium and vitamin D3, growth retardation and short stature during infancy, tooth enamel abnormalities, aphthous stomatitis and hypertransaminemia. It is also possible to observe some neurological symptoms, such as headache, paresthesia, anxiety and depression. Clinical manifestations also include alterations in reproductive function, such as delayed or absent menstruation, repeated abortions, premature birth, early onset of menopause, and changes in reduced sperm count or motility. Surprisingly, some of these symptoms can be reversed at the beginning of a gluten-free diet (Caio et al. 2019) (Figure 2).

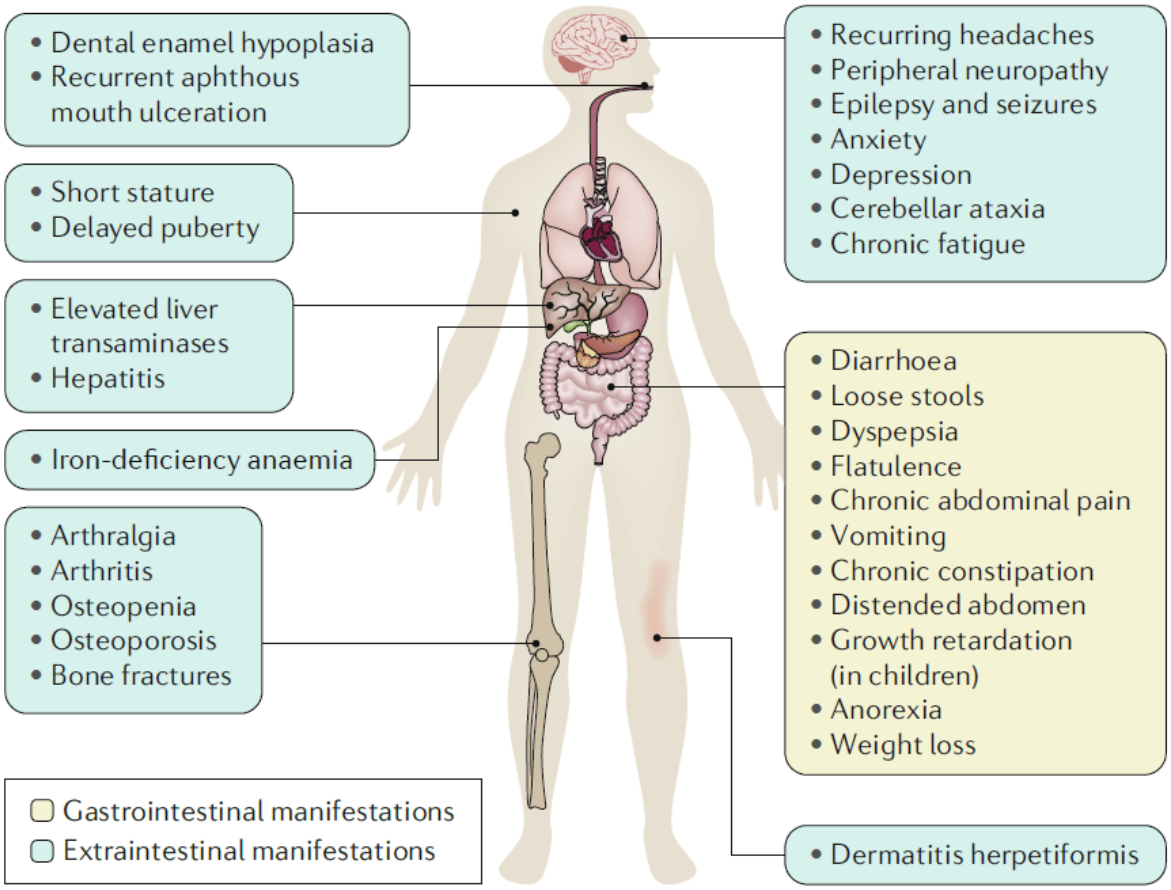


Figure 2: The clinical manifestations of CD. *In addition to the classical gastrointestinal symptoms, CD can manifest in various extraintestinal clinical presentations* (Lindfors et al. 2019).

1.1.2 Genetics

Celiac disease is one of the most common autoimmune diseases, about 3% of the world population is affected by this pathology. Several studies have confirmed the hypothesis that genetics plays an important role in the development of celiac disease. This disease develops in patients who have specific variants of the HLA class II genes. HLA is a highly polymorphic region of 4 Mb located on chromosome 6p21, which contains more than 200 genes and more than 3000 currently known alleles. The role of the HLA is to recognize foreign proteins, by measuring their recognition by the receptor of T lymphocytes. HLA class I molecules present peptides to CD8+ cytotoxic T cells, while HLA class II molecules present peptides to CD4+ helper T cells.

The reason why class II HLA is so important in celiac disease is related to its particular structure. This molecule is formed by two chains, α and β, glycosylated and covalently bound. This bond allows the formation of an antigen recognition pocket rich in positively charged residues. The presence of the positive residues means that the negative

residues formed by the deamination of glutamic residues in gluten can bind to HLA. The main HLA class II genes are DP, DQ and DR. The two variants of the *HLA* genes, *HLA-DQA1* and *HLA-DQB1*, encode the α and β chains, respectively, of the DQ2 and DQ8 heterodimeric proteins. These proteins are expressed on antigen-presenting cells and are strongly associated with celiac disease.

Some studies have shown that the human leukocyte antigen HLA-DQ2 (DQA1*05/ DQB1*02) is associated with most cases of celiac disease, unlike HLA-DQ8 (DQA1*0301/ DQB1*0302) which is present only in a minority of patients (Sollid et al. 2020). The presence of these haplotypes confers a genetic predisposition for celiac disease to approximately 30-40% of the global population. This indicates that while these molecules are necessary for disease development, they are not the sole determinant. Several studies have shown that celiac disease is a multifactorial disease that requires gluten ingestion, genetic predisposition (HLA DQ2 and/or DQ8) and a number of environmental factors not yet well identified (Figure 3). Trynka et al., (Trynka et al. 2011) have also identified through genome-wide studies the HLA locus and other DNA sequence variants that explain 50% of the heredity of celiac disease and the remaining 50% is unknown. This suggests that genome modifications, independent of the HLA DQ2 and/or DQ8, are involved in the pathogenesis of celiac disease and have not yet been studied.

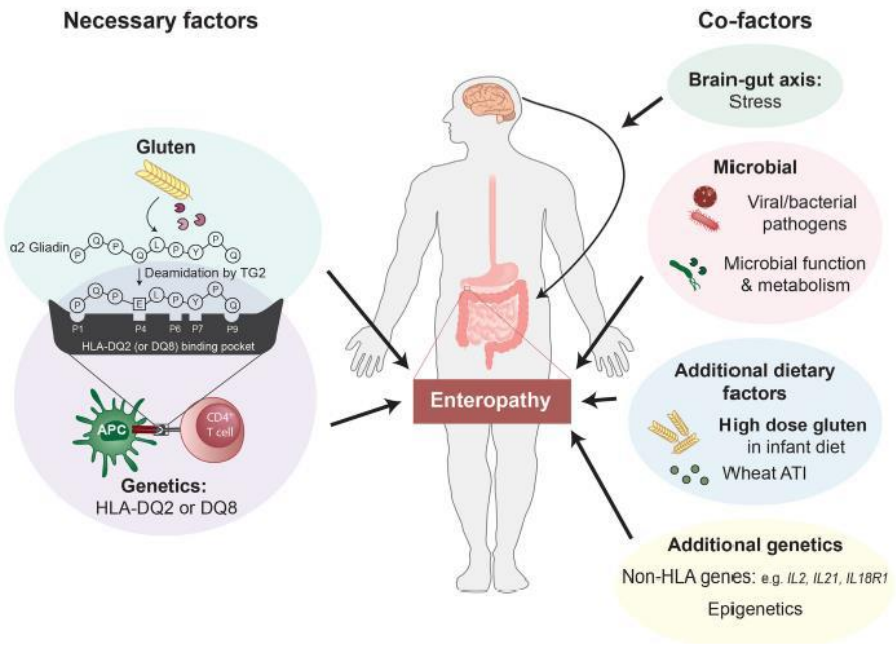


Figure 3: Main factors and co-factors triggering celiac disease (Verdu and Schuppan, 2021). *Among the main triggers of celiac disease are a genetic predisposition (HLA DQ2 and/ or DQ8), gluten ingestion but also other factors such as viral or bacterial pathogens.*

1.1.3 Diagnosis

The diagnosis of CD includes several tests that can prevent serological CD tests and/or observation of the morphology of the small intestinal mucosa.

1.1.3.1 Serological tests

The serological tests most used in the diagnosis of celiac disease are based on the search for anti-transglutaminase 2 antibodies (anti-TG2/anti-tTG Ab) and anti-endomysial antibodies of IgA class (EMA). The first-level screening test for celiac disease consists of testing for anti-tTG2 IgA Ab. Serum levels of anti-TG2 IgA exceeding the 7U/mL cut-off value in individuals on a regular gluten diet are a highly accurate indicator of active celiac disease and are strongly associated with the presence of villous atrophy upon small intestinal biopsy (Catassi et al. 2022). This screening has a sensitivity and specificity of more than 95%. In addition to this very accurate test, there is a second test that could be performed but it is less precise. This second test involves the detection of anti-TG2 IgG, which is carried out mainly in patients with IgA deficiency. Anti-TG2 antibodies can be determined using different methods, such as ELISA, fluorescent immunoassay, radioimmunoassay and chemiluminescence. When the anti-

TG2 IgA level is very high, a second assay is performed to detect anti-endomysial antibodies (EMA) in a separate sample.

EMAs are among the most specific serological markers for active celiac disease and are particularly useful in cases of doubtful positive for anti-TTG IgA or false positive for anti-TG2, such as those caused by transient infections (Maglio et al. 2011). The EMA, also of IgA class, have a specificity for celiac disease close to 100%, so if positive they greatly reinforce the diagnostic suspicion. The EMA test is an operator-dependent test because the presence of these antibodies is assessed by immunofluorescence. The EMAs recognize tissue transglutaminase (Salmi et al. 2006). This immunofluorescence-based test is performed on sections of human umbilical cord or monkey esophagus. Over the years, various kits have been developed to screen for anti-TG2 in serum; however, positive results generated by these kits require subsequent confirmation via a standard serum antibody assay (Salmi et al. 2014).

In addition to these main methods, there are also other less specific ones such as anti-gliadin antibodies (AGA) and anti-gliadin deaminated peptide (DP-AGA) (Hadjivassiliou et al. 2006). The determination of anti-DGP IgG can be used in patients with selective IgA deficiency and for children under 2 years (where anti-DGP IgG may appear before IgA-anti-TG2) (Catassi et al. 2022) .

1.1.3.2 Small intestinal biopsies

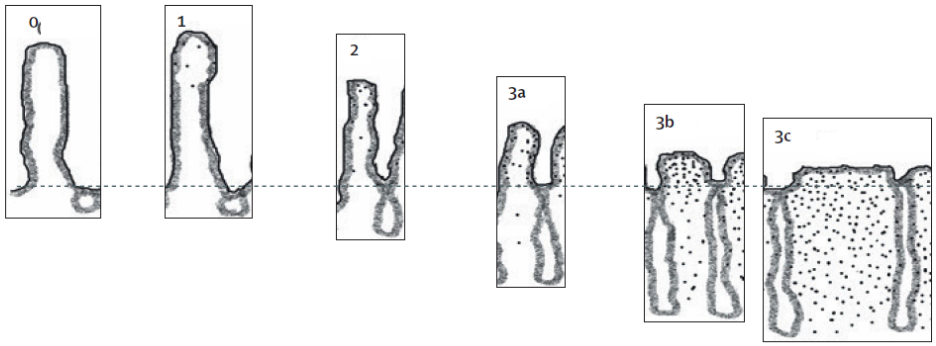
The execution of esophagogastroduodenoscopy with duodenal biopsy for histological examination at the duodenal level is fundamental for the diagnosis of celiac disease. This type of test should only be performed in case of a strong clinical suspicion and is carried out regardless of the results of serological tests.

Biopsies should preferably be taken during upper endoscopy from the bulb (at least 1 biopsy) and from the second or third portion of the duodenum (at least 4 biopsies) (McCarty et al. 2018).

During this examination, some macroscopic changes of the mucosa may already be visible, such as: absence of folds and edema, scalloping of the folds and micronodularity. These alterations have a very low sensitivity and this is why the biopsy examination should always be performed even in the absence of such signs (Korponay-Szabó et al. 2004).

The celiac disease presents precise histological features that are: atrophy of the mucosa of the small intestine, hypertrophy of the crypt and increase in intraepithelial lymphocytes (IELs). Intestinal morphometry is used as a reference method for accurate quantification of damage to the intestinal mucosa. This test involves the evaluation of certain factors such as: the count of IELs, the height of the villi, the depth of the crypt and the relationship between the height of the villi and the depth of the crypts. The histological classification is based on the Mars-Oberhuber (Catassi et al. 2022) (Figure 4):

- Marsh-Oberhuber grade 0: normal intestinal histology;
- Marsh-Oberhuber grade 1: infiltrative damage, is characterized by an increase in the number of IELs;
- Marsh-Oberhuber grade 2: hyperplastic damage, defined by the presence of crypt hyperplasia and an increase in the number of IELs;
- Marsh-Oberhuber grade 3: biopsy sample with any degree of villus atrophy, i.e. partial, subtotal and total (grades 3a, 3b and 3c, respectively).



	Marsh-Oberhuber grade					
	0	1	2	3a	3b	3c
Descriptor		Infiltrative	Hyperplastic	Partial villous atrophy	Subtotal villous atrophy	Total villous atrophy
Villous architecture	Normal	Normal	Normal	Shortened, blunt	Clearly atrophic	Complete loss
Crypt architecture	Normal	Normal	Enlarged	Enlarged	Enlarged	Severe hyperplasia
Intraepithelial lymphocytes	Normal (<25 per 100 enterocytes)	Increased	Increased	Increased	Increased	Increased
Vh:Cd ratio	≥3	≥3	≥2 to <3	≥1 to <2	≥0·5 to <1	<0·5

Figure 4: Small intestinal histology in CD. *The Marsh-Oberhuber scoring system is used to classify the level of damage to the small intestine caused by CD. Grade 0 indicates a normal small intestinal mucosa, whereas grades 1–3c classify the variable degree of damage.*

Using the Mars-Oberhuber classification as a basis, we attempted to simplify and standardize the work of anatomopathologists through a new version of the histological

classification proposed by Prof. Corazza and Dr. Villanacci (Corazza and Villanacci 2005). Specifically, the lesions that characterize Celiac disease have been divided into two main categories:

- Non-atrophic (grade A). In these cases, the lesions are characterized by a pathological increase in the number of intraepithelial lymphocytes, better recognized through the use of immunohistochemical techniques.
- Atrophic (grade B):
    - Grade B1: characterized by a villi/crypt ratio of less than 3:1, in which the villi are still identifiable.
    - Grade B2: where the villi are no longer identifiable.

1.1.3.3 Diagnostic criteria

Criteria for the diagnosis of celiac disease are provided by the ESPGHAN (European Society for Pediatric Gastroenterology, Hepatology, and Nutrition) guidelines.

The current guidelines envisage initially performing a serum assay of IgA and total IgA. If the level of IgA anti-TG antibodies in a person with normal total IgA is negative, the presence of celiac disease is unlikely and therefore the diagnostic process should be continued only in case of strong clinical suspicion. In case the anti-TTG IgA antibodies are positive, further tests must be performed to confirm the diagnosis of celiac disease. To confirm the disease, it is possible to test for anti-endomysial antibodies (EMA) and/or look for HLA DQ2-DQ8 alleles and also histological examination is always necessary, unless the antibody value is 10 times higher than normal. In the latter case, if the EMAs are also positive, the diagnosis can be confirmed without having to carry out a biopsy examination.

The current criteria divide the coeliac population into several classification categories:

- **Celiac disease**: symptomatic subject, with diagnostic investigations suggestive of disease (positive antibody, genetic and histological if performed);
- **Silent coeliac disease**: asymptomatic individual with positive antibody and histological evidence of mucosal damage;
- **Potential celiac disease**: symptomatic or not, with antibody positivity and predisposing genetics, without any noticeable morphological damage to the duodenal mucosa;

- **Latent celiac disease**: a person, symptomatic or not, with positive antibodies and predisposing genetics, without any noticeable morphological damage to the duodenal mucosa, who at some point in his life has presented a gluten-dependent enteropathy (Husby et al. 2012).

1.1.4 Treatment

To date, the only effective treatment for celiac disease is strict, lifelong adherence to a gluten-free diet (GFD). This "treatment" has shown success in alleviating intestinal symptoms but also extraintestinal, reducing level of auto antibodies and allowing the regrowth of intestinal villi (Caio et al. 2019). This diet is a diet that excludes harmful gluten peptides. Following this diet means giving up many foods that contain or are derived from wheat, rye, barley, and any hybrids of these grains. While this diet has positive effects from a clinical point of view, it also has negative effects from a psychological and social point of view. In recent years, efforts have been made to find other therapies for celiac disease, beyond the gluten-free diet, among which the use of probiotics to address dysbiosis has been explored (Catassi et al. 2022), but not with excellent results. Finally, possible therapeutic strategies such as treatment with monoclonal antibodies IL-15 (AMG 714), a selective oral inhibitor of TG-2 (ZED1227) or vaccination are currently being studied (Catassi et al. 2022).

1.1.5 Pathogenesis

Gluten ingestion triggers complex mechanisms of the disease (Figure 5). The most immunogenic gluten epitopes have been found to reside in α-gliadin and ω-gliadin (Sollid et al. 2020). Depending on their amino acid sequences, gliadin peptides can initiate small intestine damage in celiac patients either by directly activating the innate immune system or indirectly stimulating the adaptive immune response. Immunogenic peptides induce an adaptive immune response causing indirect damage while so-called "toxic" gliadine peptides activate innate immunity and cause direct damage (Sulic et al. 2015). Gluten has a high concentration of proline and for this reason the digestive enzymes are not able to degrade effectively the proteins of the gluten thus determining the presence of large peptides long up to 33 amino acids. Gliadin peptides can exploit both the transcellular and para-cellular pathways to enter the small intestine (Lebwohl, Sanders, and Green 2018). Gliadin molecules undergo a deamination process by the

tissue transglutaminase 2 (TG2) enzyme, this process occurs in both coeliac and healthy individuals. The effect of deamination is to convert residual glutamine into glutamic acid (Dieterich et al. 1997) which increases the affinity of gliadin peptide and the human leukocyte antigen (HLA) heterodimers-DQ2 or DQ8 expressed on antigen presenting cells. In the specific case of celiac disease, HLA-bound peptides are presented to CD4+ T cells that enter a phase of activation, clonal expansion and secretion of pro-inflammatory cytokines such as IFNγ and TNF-α (Tye-Din, Galipeau, and Agardh 2018). This process creates an inflammatory environment that causes damage to the mucosa of the small intestine characterized by increased IELs, hyperplasia of the crypts and atrophy of the villi. Gluten-specific T lymphocytes not only contribute to the inflammatory environment but also promote a cascade of events, which aim to produce autoantibodies against TG2. The development of enteropathy typical of celiac disease requires a complex interaction between adaptive and innate immunity, the latter involving the epithelial compartment (Abadie, Discepolo, and Jabri 2012). The cascade of events of the innate response in celiac disease is still not fully known. IELs are known to express the NKG2D and CD9/NKG2A killer T-cell receptors, which recognize stress-induced gene surface glycoproteins (MICA and MICB) and the protein HLA-E expressed on epithelial cells. IL-15 probably has an important role in the pathogenesis of celiac disease, since it shows a chronic hyper-expression both in the lamina propria and in the epithelium and is related to mucosal damage in patients (Di Sabatino et al. 2006). However, it is not known what happens at the epithelial level after ingestion of gluten. So far, the study of epithelial cells and pathogenesis is a complicated process but in recent years with the advent of new technologies such as new sequencing methodologies such as single cell RNA (scRNA-seq) sequencing new advances could be made.

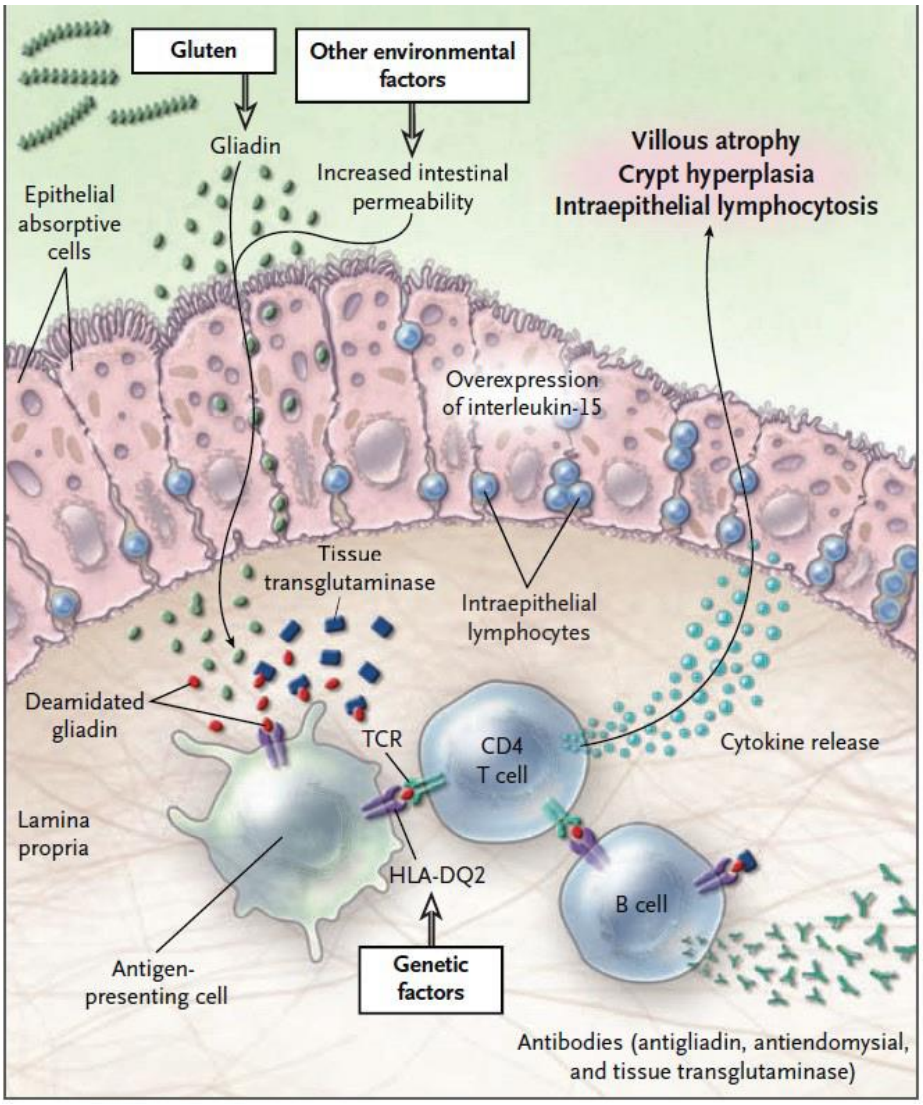


Figure 5: CD pathogenesis. *After ingestion of gluten, digestion produces large glutenderived gliadin peptides, which can cross the intestinal epithelial barrier and reach the lamina propria. Gliadin peptides cause changes in the epithelium through the innate immune system and, in the lamina proper, through the adaptive immune system. Within the epithelium, "toxic" gliadin peptides damage epithelial cells, leading to increased IL-15 secretion, which in turn activates IELs. These lymphocytes become cytotoxic and contribute to the apoptosis of enterocytes and direct epithelial damage. At the same time, the gliadin peptides enter the membrane, where they are deamidated by TG2. This modification increases their affinity for APCs expressing HLA-DQ2 or HLA-DQ8. In this way gliadin peptides are more efficiently presented to CD4+ T cells causing the production of pro-inflammatory cytokines (IFN-γ, TNFα, IL-21) generating tissue damage. This whole process involves villous atrophy and crypt hyperplasia, as well as activation and expansion of B cells, which produce specific antibodies to the deamidated gliadin peptides and TG2* (Green and Cellier 2007).

1.1.6 Human intestine

The intestine is a crucial organ for digestion, nutrient absorption, immune defense and hormone secretion (Wang et al. 2020). Having a complex structure and having multiple functions, in the human intestine there is a considerable amount of very different cell types that work in a coordinated way. Until recently, the understanding of the intestine

and its cellular composition was limited by the poor specificity of the avaible technologies. The advent of single-cell RNA sequencing (scRNA-seq) technologies has revolutionized this field by enabling the high-resolution profiling of transcriptional landscapes at the single-cell level, thereby offering unprecedented insights into the cellular diversity of complex tissues such as the intestine (Bigaeva et al. 2020; Shoaran et al. 2024). So far, many cell types have been identified and characterized, both epithelial ($EPCAM^+$) and immune cells (CD45+). The intestinal epithelium acts as a barrier between the host and the intestinal lumen and is characterized by rapid cell turnover. Intestinal epithelial cells (IECs) are specialized and polarized cells located in the gastrointestinal tract. Their main functions include the separation of intestinal lumen from underlying tissues, regulation of nutrient absorption and reception of signals from immune and mesenchymal cells and microbiota (Escudero-Hernández 2021). The intestinal barrier is organized in villi and crypts (Figure 6).

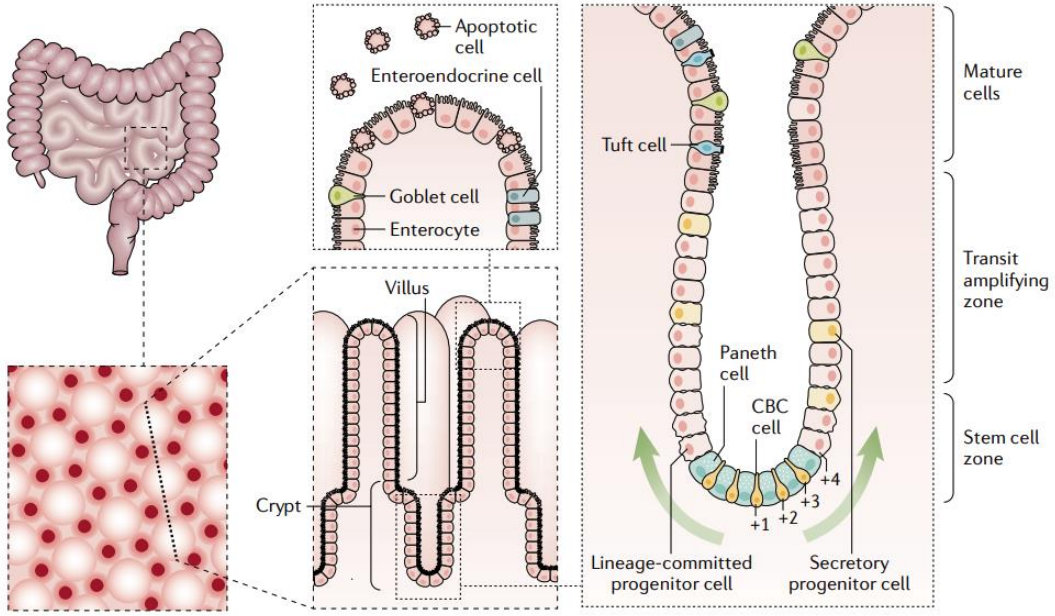


Figure 6: Crypt-villus structure. *The intestinal barrier is organized into villi and crypts. The crypt continuously produces new cells that differentiate and migrate to the villi. There are about 15 basal columnar cells (CBC) cells positioned from +1 to +3 at the base of the crypt (stem cell zone), surrounded by protective Paneth cells, and they continuously divide to give rise to transit amplifying cells (Tas). TAs are progenitors engaged in the lineage that in turn proliferate, contributing to rapid replacement of intestinal cells. Fully mature intestinal stem cells (ICEs) that have migrated to the tip of the villi are expelled and will undergo apoptosis, being simultaneously replaced by other cells that are migrating* (Gehart and Clevers 2019).

Villi are protrusions of the intestinal epithelium that have the function of increasing the surface area by improving the absorption capacity of the intestine. Crypts are tubular

invaginations that protect intestinal stem cells (ISCs) from potential damage of the digestive process through physical barriers and chemical protection (Gehart and Clevers 2019).

To date about 9 sub-types of intestinal epithelial cells (IEC) have been identified (Figure 7).

At the basis of intestinal crypts, there are stem cells involved in tissue homeostasis and repair (Wang et al. 2020). The main ISCs are basal columnar cells (CBC) of the crypt and are identified as Lgr5+ cells. These cells are capable of generating multiple lines, both absorbent and secretory cells (Escudero-Hernández 2021). Stem cells are surrounded by secreting Paneth cells and mesenchymal cells. Paneth cells are responsible for nurturing and supporting CBC through paracrine signals including epidermal growth factor (EGF), WNT and Notch ligands, and transforming growth factor (TGF)-β. Paneth cells are also able to produce antimicrobial products such as lysozyme, α-defensins and phospholipase A2, thus they also have a protective role towards CBC cells. Stem cells are in continuous division generating progenitor cells or transit amplifying (TA) cells that amplify, proliferate and become mature epithelial intestinal cells. The entire maturation process from stem cells to IECs prevents a cellular change that begins in the crypt and ends at the tip of the villus (Barker et al. 2007). The process of generating and maturing cells takes about four or five days to withstand different types of stress (mechanical, chemical and biological). The underlying mesenchymal cells provide structural support and secrete factors to regulate IECs' behavior and differentiation. scRNA-seq allowed the identification of subpopulations of major or differentiated stem cells, providing new markers and a deeper understanding of their transcriptional programs (Busslinger et al. 2021).

Enterocytes are the most abundant cells in the intestinal epithelium and their main function is nutrient absorption (they represent up to 80% and are responsible for the transport of ions, water, sugars, peptides and lipids). Single-cell analyses revealed functional differences in enterocytes along the proximal-distal axis of the intestine, in fact a different degree of nutrient absorption was demonstrated in the small and large intestines (Wang et al. 2020). In recent years, a transcriptional atlas of single-cell epithelium from the small intestine and the colon has been generated in healthy adults, and regional differences have been shown (Burclaff et al. 2022). Enterocytes that have

completed their differentiation process have been identified as mature enterocytes (FitzPatrick et al. 2025).

Goblet cells or caliciform cells are protective cells as they produce and secrete mucus which prevents microbial invasion. New potential markers for Goblet cells have been discovered that have made it possible to highlight a new subgroup, the caliciform TFF1+ cells showing distinct genes (Wang et al. 2020).

A new cell population, called INFLARE, has recently been identified. INFLARE cells are a metastatic population with a function similar to that of goblet cells, but are only present in inflammatory conditions (Oliver et al. 2024).

Cells of Paneth are located in the crypts and have a defensive role as Goblet cells. Single-cell studies suggest the presence of another type of cells called "Paneth-like" also in the large intestine, broadening our understanding of their distribution and function (Wang et al. 2020).

Tuft cells are a small cell population present in the intestine and play a detection and signaling role, especially in response to parasites and inflammation.

Finally, enteroendocrine cells (EEC) are a rare cell type and their function is to produce a wide range of hormones that can regulate digestion, appetite and intestinal motility.

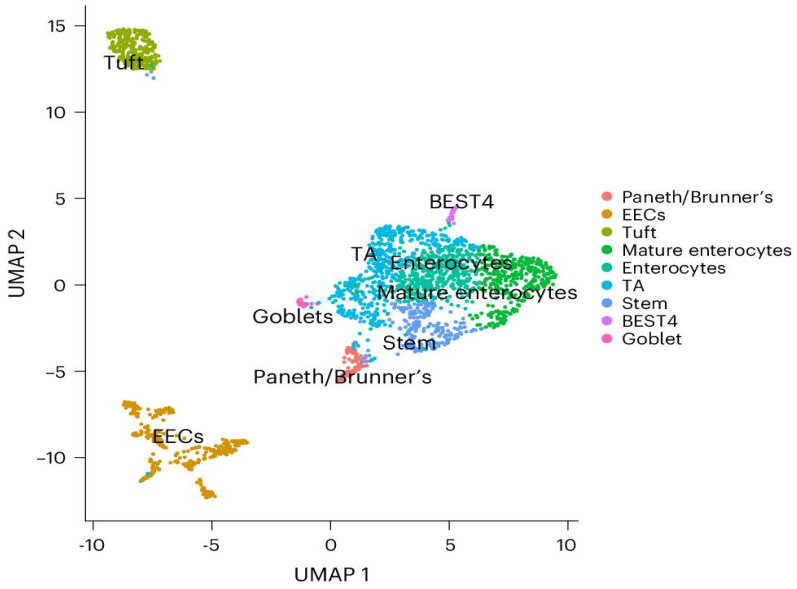


Figure 7: UMAP (Uniform Manifold Approximation and Projection) plot of small intestinal epithelial EPCAM$^+$ cells. *Clustering of epithelial cells after scRNA-seq analysis* (FitzPatrick et al. 2025)***.***

In addition to the epithelial component, the intestine hosts a large and complex population of immune cells (Figure 8). The scRNA-seq assays provided a very detailed view of the heterogeneity of immune cells residing in the fetus, both in disease and health conditions (Atlasy et al. 2022). As already known the main immune cells present in the intestine are the lymphocytes (T, B, ILCs), myeloid cells (macrophages and dendritic cells) and finally the plasmacellules. In particular, these studies have deepened the understanding of T cell populations (CD4+ and CD8+) and their heterogeneity. Analysis of scRNA-seq showed significant changes in CD4+ and CD8+ T cell transcripts in both inflammatory bowel disease and celiac disease. The same analysis also identified a population of natural intraepithelial lymphocytes (iNKT cells), which is less expressed in celiac disease (Atlasy et al. 2022). An interesting information has emerged from recent analyses indicating communication between T lymphocytes and INFLARE (Oliver et al. 2024). Myeloid cells represent a very heterogeneous cell population and consist of macrophages and dendritic cells mainly. Dendritic cells (DCs) are positioned to bind the antigens present in the intestinal lumen and then migrate to the mesenteric lymph nodes to initiate adaptive immune responses. The scRNA-seq divided the DCs in conventional dendritic cells (cDCs) and plasmacitoid dendritic cells (pDCs) and also characterized their different functions in inducing tolerance or pro-inflammatory immunity (Cerovic et al. 2014). Finally, to complete the population of intestinal immune cells there are the innate lymphoid cells (ILCs) acting as a first line of defence against pathogens and maintaining tissue homeostasis without the need for antigenic receptor rearrangement. ILCs are classified into three groups according to the transcription factors expressed and the cytokines produced. Innate lymphoid cells are:

1) **ILC1s**: Similar to T helper cells 1 (Th1), they produce IFN-γ and are important in the defense against intracellular pathogens.
2) **ILC2s**: Similar to T helper cells 2 (Th2), they produce cytokines such as IL-5 and IL-13, involved in the anthelmintic response and allergy.
3) **ILC3s**: Similar to T helper 17 (Th17) lymphocytes, they produce IL-17 and IL-22, essential for the protection of mucosal barriers against extracellular bacteria and fungi. ILC3s, in particular, are abundant in the intestine and play a critical role in maintaining epithelial barrier integrity and regulating commensal bacterial flora.

ILCs also interact directly with intestinal epithelial cells influencing the proliferation, differentiation and production of antimicrobial peptides.

Finally, in the intestine there are plasma cells. Plasma cells are differentiated immune cells derived from B lymphocytes specialized in the production and release of large amounts of antibodies. Their primary function is the production of Immunoglobulin A (IgA). These immunoglobulins are transported to the intestinal lumen where they create a protective barrier against pathogens and maintain the balance of the intestinal microbiota. This population is very bountiful and constitutes a fundamental component of mucosal immunity. An abnormal function or presence of plasma cells can contribute to the development of intestinal diseases.

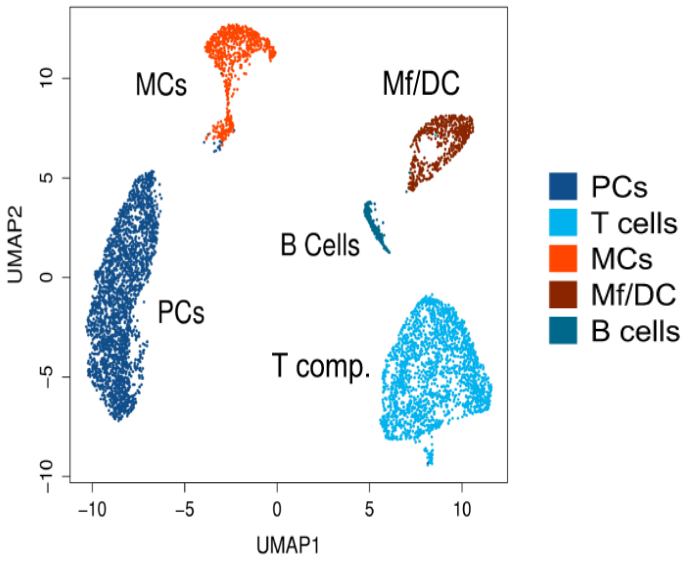


Figure 8: Single-cell CD45+ immune cells in human small intestine. *Clustering of immune cells after scRNA-seq analysis*(Atlasy et al. 2022)*. PC: Plasma Cells; MCs:Mast cell, Mf/DC: Macrophage/Dendritic cell.*

1.2 Single Cell transcriptomics

1.2.1 Introduction to next generation sequencing and Single Cell transcriptomics

Over the years, advances in molecular biology have radically transformed our ability to analyse nucleic acids, revolutionizing the way biological systems studies can be approached. From the discovery of DNA structure by Watson and Crick in 1953 (Watson and Crick 1953) to the sequencing of the first RNA molecule, the tRNA by Robert Holley (Holley et al. 1965), research has developed several increasingly sophisticated methods for sequencing DNA and/or RNA. The crucial turning point came in 1977 with the introduction of the Sanger chain termination method (Sanger, Nicklen, and Coulson 1977), which remained for years the reference point in the field of sequencing. However, in 2005 there was a major breakthrough with the commercialisation of Next Generation Sequencing (NGS) technology, or second-generation sequencing, marking a turning point in genomics (Margulies et al. 2005). The NGS technology enables massive and parallel sequencing of millions of DNA or RNA fragments while offering high productivity, sensitivity, accuracy and cost savings compared to the Sanger method (Mardis 2008). The NGS process involves several steps including fragmentation of genetic material, addition of known adapters (ligation), amplification and identification of bases via optical or electronic signals (Metzker 2010). The NGS is suitable for large-scale applications because it can run millions of reactions simultaneously. These advantages have made it an essential tool in many fields, including genetic diagnostics, oncology, evolutionary studies, metagenomics and drug discovery (Metzker 2010; Meyerson, Gabriel, and Getz 2010). This technology has found wide application in transcriptomics, the systematic study of all transcripts (mRNA) present in a cell or tissue. These technological advancements have driven the widespread adoption of bulk RNA sequencing (bulk RNA-seq), which analyses average gene expression across entire tissues or cell populations, providing a global overview of transcript abundance (Wang, Gerstein, and Snyder 2009). This analysis includes: mRNA extraction, cDNA back-transcription, amplification, library preparation, sequencing, and finally bioinformatics analysis for alignment and counting of gene mapped readings (Conesa et al. 2016). This process generates a matrix of counts from

which gene expression can be inferred. The bulk RNA-seq approach, based on cell populations, however has limitations in that it does not allow to distinguish the transcriptional contributions of single cells, thus providing only general information which may mask rare but potentially relevant cell sub-types (Trapnell 2015). This defect is a substantial limitation especially when analyzing complex or heterogeneous tissues, where even small cell populations can have an important clinical impact. To address these limitations, single-cell RNA sequencing (scRNA-seq) has been developed which allows the analysis of gene expression at the level of a single cell (Tang et al. 2009a). The advent of this technology has totally revolutionized the study of genomics by allowing to identify cellular states, map differentiation trajectories, to discover rare cell subtypes and better understand the molecular mechanisms underlying development and pathology (Papalexi and Satija 2018; Stuart and Satija 2019). The shift from bulk to single-cell analysis has thus opened a new era in molecular and translational biology, with profound implications for personalized medicine, biomedical research and understanding of cellular complexity in living systems.

1.2.2 Single Cell RNA Sequencing

Single-cell transcriptomics began with the development of single-cell qPCR (quantitative real-time polymerase chain reaction) (Bengtsson et al. 2008; Eberwine et al. 1992; Huang et al. 2014; Taniguchi, Kajiyama, and Kambara 2009; Warren et al. 2006). The first transcriptomes generated via single-cell RNA-sequencing (scRNA-seq) were published in 2009 (Tang et al. 2009b), only 2 years after the first applications of RNA-seq to bulk populations of cells.

The significance of scRNA-seq is related to its ability to assess transcriptional similarities and differences within a heterogeneous population. The first studies carried out with this method revealed levels of heterogeneity that had not been previously detected, for example in embryonic and immune cells (Deng et al. 2014; Jaitin et al. 2014; Yan et al. 2013). Thanks to this technology, the initial sample can be studied in ever greater depth. For example, the scRNA-seq allows us to find rare cell populations that are present within the tissue and that we would never have detected without this technology. Its ability to study the gene expression of each single cell allows us to divide cellular populations that until now had never been found or grouped under a single name. The scRNA-seq also allows to analyze if and how different populations

communicate with each other and therefore could also influence gene expression and/or activation of some pathways. This type of analysis can be crucial during the study of a disease because it could show communications between cell populations that may only occur in individuals affected by the disease, laying the groundwork for the discovery of new pharmaceutical targets. scRNA-seq also serves to identify lineage and developmental relationships among heterogeneous but related cellular states, which in many cases encompass embryonic development, cancer progression, the differentiation of myoblasts and lung epithelium, or the diversification of lymphocyte fate (Blakeley et al. 2015; Lönnberg et al. 2017; Petropoulos et al. 2016; Trapnell 2015; Treutlein et al. 2014; Venteicher et al. 2017). scRNA-seq is therefore the most advanced approach to unravel the heterogeneity and complexity of RNA transcripts able to reveal compositions of different cell types and functions within highly organized tissues/organs/organisms (Jovic et al. 2022).

The scRNA-seq workflow includes the following parts: single cell isolation, reverse transcription, cDNA amplification, single cell library, high throughput sequencing and data analysis (de Klerk and ’t Hoen 2015) (Figure 9).

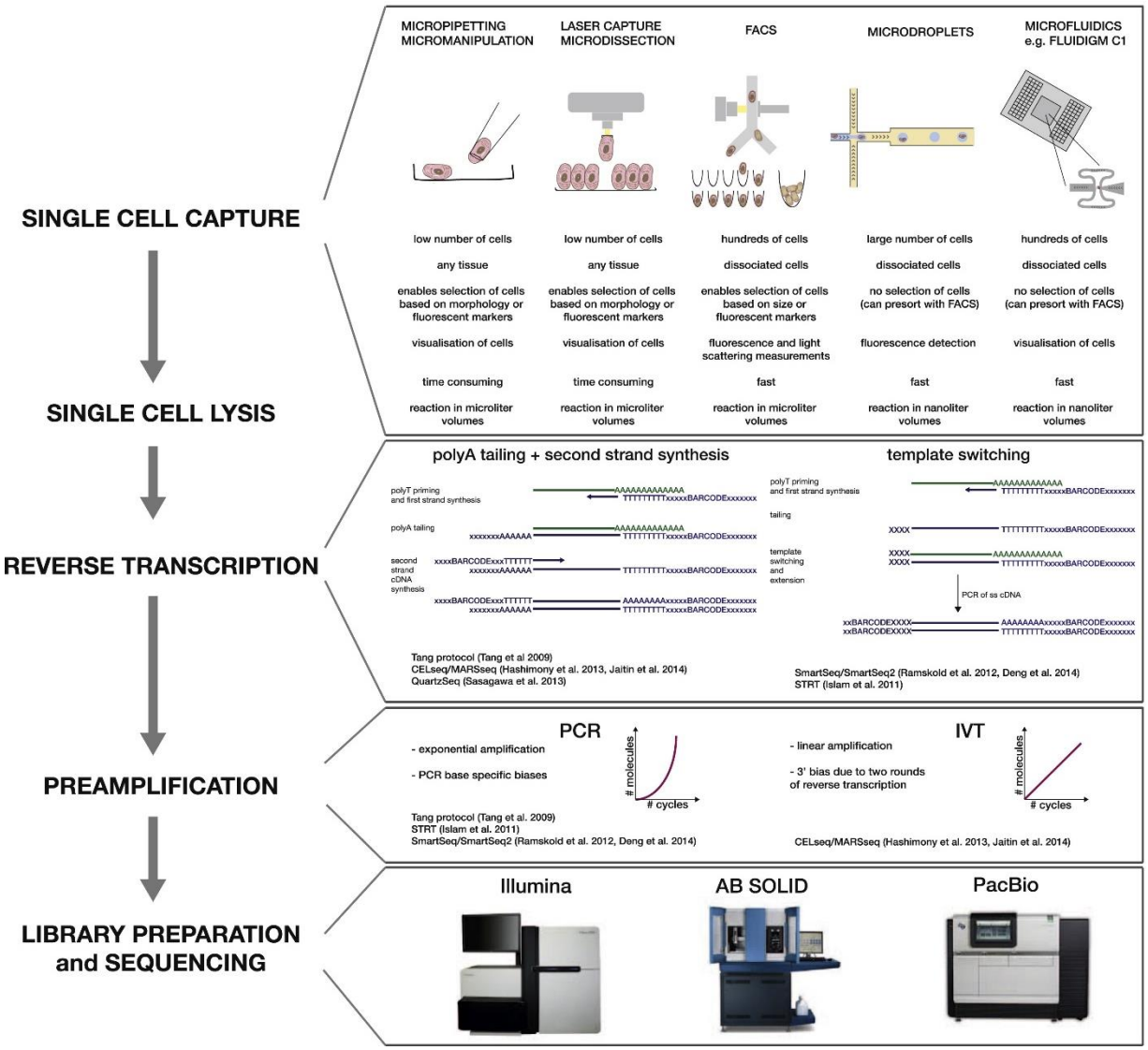


Figure 9: Single-Cell mRNA Sequencing Experiment. *Schematic of single-cell RNA-seq experiment showing each step and different experimental approaches* (Kolodziejczyk et al. 2015).

These steps are crucial and the generation of single cell libraries is much more complex than a normal bulk library. In particular the first step is crucial as it involves the isolation of single cells from the tissue. This step is the most delicate as the isolated cells must be alive and this is not always the case. There are many kits on the market that allow the isolation of single cells and depending on the kit the number of living cells can vary. Another factor affecting cell death is the time it takes for a sample to be collected and processed to isolate cells. It has been seen that when the sample is freshly processed, the number of living cells is higher. The quality of the final result will depend on this step. Normally, isolation is achieved by both mechanical and enzymatic dissociation. The next step is the lysis of these cells. The aim of this step is to release and capture the messenger RNA (mRNA) present in single cells. PolyT polymers are usually used

to capture the mRNA, and they bind to the polyadenylate tails of the mRNAs in order to avoid contamination by ribosomal RNA.

The mRNA captured by a back-transcriptase is reverse-transcribed into cDNA. According to the protocol used, the primers for the back-transcription can be associated with sequential adaptor sequences such as unique molecular identifiers (UMI) which can then distinguish the individual molecules. There are also cellular barcodes to preserve the origin of the cell from which each molecule is derived (Kivioja et al. 2011). Usually, the cDNA that is obtained from the back-transcription has a very low concentration so an amplification by PCR is carried out followed by another step of back-transcription. The amplified and retro-transcribed cDNA is then used to prepare the libraries which will then be sequenced. Several protocols have been developed over the years for creating scRNA-seq libraries that do not necessarily require the use of a specific machine. However, it should be remembered that if the loading of cells into single wells allowing the capture of individual molecules (Figure 10) is not carried out through dedicated microfluidic platforms this can result in an increase in the cost per reaction and to lower the sensitivity and accuracy of detection.

Figure 10: Microfluidic platforms for loading single cells. *The figure illustrates a dedicated microfluidic platform (left) used for the controlled loading of single cells into individual wells (right). The use of these microfluidic systems ensures precise positioning of single cells.*

In the last decade, however, the introduction of these platforms has radically transformed the field. Systems such as Chromium by 10x Genomics, ddSEQ by Bio-Rad, InDrop by 1CellBio and µEncapsulator by Dolomite Bio allow the encapsulation of thousands of single cells in droplets containing all the reagents needed for lysis, back transcription and barcoding, thus eliminating the need for cell sorting or microdissection (Klein et al. 2015; Macosko et al. 2015; Zheng et al. 2017). Despite the considerable advantages of scRNA-seq, there are some methodological limitations that must be

taken into account as they could influence the interpretation of the data. In particular, the limitations are:

- Lower Read Depth: In the case of scRNA-seq compared to bulk sequencing, this refers to the lower number of reads per cell. This limitation can result in the failure to detect transcripts expressed at low levels.
- Transcript Capture Bias (Poly-A): Most Illumina-based scRNA-seq platforms rely on capturing the poly-A tail of mRNAs. This approach may exclude quantification of certain lncRNAs and histones, introducing bias. This bias can occur due to mRNA cleavage events that render the poly-A tail unavailable for capture.
- Variation in Cellular Composition and Dissociation: The final representation of cell types in the dataset is closely related to cell survival during sample preparation. The complex and delicate tissue manipulation phase can cause severe cellular stress leading to cell death. Cell death is detected by the presence of mitochondrial RNA (mtRNA), and since the latter is an important cut-off during quality control, it could cause the loss of certain stress-sensitive cells or cell populations in the final dataset. This loss would therefore cause an underestimation of these specific populations.

The last step involves finally the analysis of the raw file data obtained by sequencing.
Single-Cell RNA Sequencing (scRNA-seq) involves the use of different sequencing platforms and technologies. We can identify two types of sequencers or sequencing technologies:

1) Sequencers based on DNBSEQTM (DNA Nanoball Sequencing) technology: These sequencers, developed by Complete Genomics (MGI), use the Nanoball DNA technology for high throughput sequencing. They are known for their accuracy and flexibility. Examples of sequencers: DNBSEQ-G400 and DNBSEQ-T7.
2) Sequencers based on the Illumina technology (Sequencing by Synthesis - SBS): The Illumina platforms are widely used for next-generation sequencing (NGS) and are compatible with many single cell library preparation solutions. Examples of sequencers: NovaSeq 6000 and NextSeq 2000.

Over the years there are more and more pipelines available to complete the analysis. About 20 different scRNA-seq protocols have been published to date (Kolodziejczyk et al. 2015). It is known, however, that some of these protocols provide full length

transcription data, while others are specific only to the 3' end. Recent meta-analyses have shown that all protocols are very accurate in determining the abundance of mRNA transcripts within a pool (Svensson et al. 2017; Ziegenhain et al. 2017). On the other hand, a sensitivity that is protocol dependent has been detected, in particular the minimum number of mRNAs required for a safe detection of gene expression varies from protocol to protocol. This information indicates that depending on the depth of sequencing required, one protocol will need to be chosen over another. The choice between bulk RNA-seq and scRNA-seq depends on several factors. Bulk RNA-seq remains the preferred choice when you want an average profile of the gene expression of a homogeneous tissue or when the number of samples is high and resources limited, as it is less expensive, easier to analyse and provides robust and reproducible results (Conesa et al. 2016) On the other hand, scRNA-seq is particularly indicated in contexts where strong cellular heterogeneity is assumed, such as in tumors, in the immune system, in embryonic development or in inflamed tissues, because it allows to decompose the complexity and identify specific cell subpopulations (Li and Wang 2021; Papalexi and Satija 2018). Both approaches are therefore complementary: the bulk RNA-seq can be used for a first exploration of transcriptional changes, while scRNA-seq allows more detailed characterization at the cellular level (Stuart and Satija 2019).

So, the scRNA-seq has therefore revolutionized almost every field of biology and medicine, providing insights that were previously impossible to achieve (Jovic et al. 2022).

2. AIM OF THE THESIS

Celiac disease is one of the most common chronic lifelong diseases, with a prevalence of 1% in the general population. A triad of factors (genetic predisposition, gluten and environmental factors) is involved in the pathogenesis of celiac disease that is not yet fully understood.

Recent discoveries made with single cell RNA sequencing (scRNA-seq) have significantly improved our understanding of the human intestine, both in normal conditions and also across the spectrum of gut diseases. To study celiac disease with a scRNA-seq approach could provide new insights into the pathogenesis of celiac disease.

This technology can allow for a more in-depth study of pathogenesis by examining the differences that may exist between celiac patients and controls. To date, however, few studies of this type have been conducted on the intestine and, in particular, on the pathogenesis of celiac disease. This could be due to the difficulty of handling the sample and, therefore, the low number of cells obtained.

The main objectives of this project are:

1) To determine the optimal procedure of manipulation of intestinal biopsy samples to generate single-cell RNA sequencing libraries.
2) Generation of single-cell libraries from intestinal biopsies from children undergoing esophagogastroduodenoscopy for suspected celiac disease or other gastrointestinal disorders.
3) Bioinformatic analysis of single cells data to identify all types of intestinal cells and identify differences between celiac patients and controls.

3. MATERIALS AND METHODS

3.1 Patients' cohort

After the approval of an independent ethics committee, intestinal biopsies were collected from 12 children at the IRCCS "Burro Garofolo" Institute for Maternal and Child Health in Trieste (Italy): 6 samples from celiac patients and 6 samples from control subjects.

For this study, there are two cohorts:

- First cohort: 4 patients: 2 patients with celiac disease (CD0-1) and 2 control subjects (CTRL0-1).
- Second cohort: 10 patients: 5 patients with celiac disease (CD1-5) and 5 control subjects (CTRL1-5).

Two patients from the first cohort are also present in the second cohort.

Celiac disease was diagnosed following ESPGHAN guidelines, presenting positive EMA, high anti-tTG values and a Marsh 1-3. For the same principles, the other patients were defined as control subjects. For all patients, ESPGHAN guidelines were also followed for sample collection, without exceeding the number of biopsies necessary for diagnosis.

3.2 Isolation

Following the collection of intestinal biopsies, epithelial cells were isolated using the following protocol:

1) Place 2-3 biopsies of the same district in a 50 mL Falcon with 5 mL HBSS w/o CaMg 1X.

WASHING:

1) Shake the 50 ml Falcon and remove the soil with a sterile Pasteur pipette. Add 5 mL of HBSS 1X for pouring and stir briefly. Repeat washing 2X (this wash is done to remove mucus)
2) Remove the soil from the Falcon and transfer the biopsies (without liquid) with the aid of a P1000 pipette into a well on a 6-well plate.

BREAKDOWN:

1) In the well add the digestion solution:
   a) 1.3 mL of HBSS 1X.
   b) 350 µL of LIBERASE 1 mg/mL (final concentration: 205 µg/mL = 1.07 Wunsch U/mL).
   c) 40 µL HYALURONIDASE 600U.
   (These Vol are good for 2-3 biopsies/well. Enzymes should be diluted in sterile PBS)
2) Incubate at 37°C under agitation at 750 rpm/min for about 45 minutes (max digestion time-1h: the longer the digestion time, the lower the cell viability).
   Every 15' flick up and down for 1-2' (first with a sterile Pasteur pipette and then with the P1000) to disintegrate the tissue (The cell suspension changes color during disintegration due to a change in pH).

FILTRATION:

1) Once digestion is complete, the suspension should be spiked well and transferred to a 40 µm cell filter placed on a 50 mL Falcon.
2) Wash the well: add 1 mL of HBSS 1X to the well, wash the well, transfer the liquid over the filter; repeat with another 1 mL of HBSS 1X.
3) Raise the filter and retrieve with a P1000 the cells that have remained stuck on the bottom and side of the filter and put them in a 15 mL Falcon (to recover as much as possible). Do not throw away the filter but place it in the well.
4) Transfer the filtered solution to the 15 mL Falcon. Centrifuge the Falcon at 500 g for 7' at 10°C (the 15 mL Falcon gives a better pellet than the 50 mL Falcon).
5) Remove supernatant.
6) Resuspended the pellet in 300 µL of 1x cold pbs.
7) Add 200 µL of PBS 1x cold.
8) Redoing filtering.
9) Choose range 2.5 x 10^5 cells/mL.
10) Count to get the number of cells in 100 µL.

3.3 Single Cell RNAlibrary construction

Single-cell RNA sequencing libraries were prepared using the Singleron GEXSCOPE® Single Cell RNA Library Kit, following the protocol described in the SCOPE-chip SD/HD Manual (2 RXNs/16 RXNs, Version 2). The workflow is detailed below.

3.3.1 Preparation

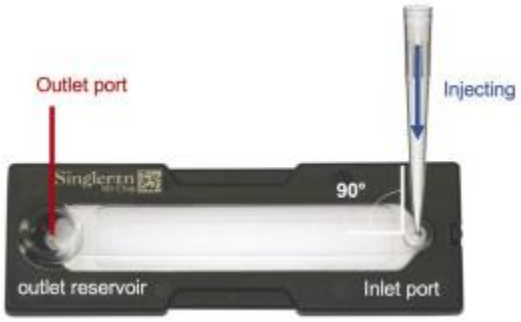
Priming of the SCOPE-chip

1) Prepare PBST (PBS with 0.02% v/v Tween-20) as described in the table below. Vortex and centrifuge briefly.

PBST:

Component	1 RXN (μL)	2 RXNs (μL)
PBS	998	1996
10% Tween-20	2	4
Total	1000	2000

2) Using a 200 μL pipette, inject 200 μL of absolute ethanol into the SCOPE-chip via the inlet port. Dispense slowly over a period of 10 seconds to avoid the introduction of air bubbles.



3) Remove the ethanol from the outlet reservoir without making contact with the outlet port. (Suction applied at the outlet port may lead to the introduction of air in the SCOPE-chip).
4) Repeat the priming with absolute ethanol two additional times.
5) Inject 200 μL of PBST into the SCOPE-chip via the inlet port. Dispense slowly over a period of 10 seconds to avoid the introduction of air bubbles.
6) Remove the liquid from the outlet reservoir without making contact with the outlet port.

7) Repeat the priming with PBST two additional times.
8) Leave 50-100 µL of PBST in the outlet reservoir to avoid the appearance of air in the chip.
9) Cover the SCOPE-chip to protect it from contaminants and let it sit at room temperature.

Barcode Beads Preparation

1) Vortex the Barcode Beads thoroughly for 30 seconds until fully resuspended.
2) For each reaction, pipet 900 µL of the Barcode Beads into a new labeled 1.5 mL tube.
3) Centrifuge briefly, place in a magnetic rack until the liquid is clear.
4) Discard the supernatant without disturbing the beads.
5) Remove the tube from the magnetic rack and resuspend the beads in 1 mL of PBS.
6) Centrifuge briefly, place the tube back to the magnetic rack place until the liquid is clear.
7) Discard the supernatant without disturbing the beads.
8) Wash two additional times with 1 mL of PBS.
9) After the last wash, resuspend the beads in 60 µL of PBS.
10) Store the barcode beads at room temperature for later use.

3.3.2 Single Cell partitioning and mRNAcapture

Lysis Mix Preparation

1) Prepare the Lysis Mix on ice according to the tables below. Vortex, centrifuge briefly, and store on ice.

Lysis Mix (SCOPE-chip HD):

Component	1 RXN（µL） ×1	2 RXNs（µL） ×2.2
Lysis Buffer, Stock	138.7	305
100mM DTT	7.5	16.5
RNase Inhibitor	3.8	8.5
Total	150	330

Cell Suspension Preparation

1) Dilute the cell suspension using pre-chilled PBS to a concentration between 1.5 × 10^5 to 5.0 × 10^5 cells /mL according to the tables below.

SCOPE-chip HD:

Targeted number of cells captured	Recommended cell concentration
9000-12000	（1.5-2.0）×10^5 cells /mL
12000-15000	（2.0-2.5）×10^5 cells /mL
15000-18000	（2.5-3.0）×10^5 cells /mL
18000-24000	（3.0-4.0）×10^5 cells /mL
24000-30000	（4.0-5.0）×10^5 cells /mL

Important: Note that only a subset of the cells loaded on the chip will be captured into the microwells. For example, to capture 10 000 cells on the SCOPE-chip SD (first table, left column), 100 µL of a 3.5 x 105 cells/mL suspension (right column) (35000 cells) is needed to load the SCOPE-chip.

2) Proceed immediately to the loading of the cells into the SCOPE-chip.

Loading of the Cells on the SCOPE-chip

1) Retrieve the previously primed SCOPE-chip.
2) Remove the residual liquid from the outlet reservoir without making contact with the port.
3) Inject 200 µL of pre-chilled PBS into the SCOPE-chip via the inlet port. Dispense slowly over a period of 10 seconds to avoid the introduction of air bubbles.
4) Remove the liquid from the inlet port and outlet reservoir without making contact with the port.
5) Repeat the PBS wash one additional time.
6) Resuspend the diluted cells by pipetting gently up and down.
7) Inject 100 µL (SCOPE-chip SD) or 150 µL (SCOPE-chip HD) of the homogenous cell suspension into the SCOPE-chip via the inlet port. Dispense slowly over a period of 30 seconds to avoid the introduction of air bubbles.

Important: When loading cells into the SCOPE-chip, a drop of liquid can be kept at the inlet port to avoid the introduction of bubbles; The pipette should be kept in a vertical position when loading cells into the SCOPE-chip; Injecting cells in less than 30 seconds could cause the capture of fewer cells into the chip; Injecting cells for over 30 seconds

could increase the cell doublet rate by biasing the distribution of the cell at the entrance of the chip.

8) Remove excess liquid from the outlet reservoir without making contact with the port. (Suction applied at the outlet port may lead to the removal of the cells from the SCOPE-chip and introduction of air in the SCOPE-chip).
9) Let the SCOPE-chip sit for 5 minutes at room temperature to allow the cells to fall into the microwells.
10) Under the microscope, look at the distribution of the cells in the microwells. Ideally, 5 to 15% of the wells should be populated with cells.
11) Inject 200 µL of pre-chilled PBS into the SCOPE-chip via the inlet port. Dispense slowly over a period of 30 seconds to avoid the introduction of air bubbles.
12) Remove the excess of liquid from the outlet reservoir without making contact with the port.
13) Repeat the PBS wash and proceed to the loading of Barcode Beads.

Important: Under the microscope, look for the presence of cells floating above the microwells. If detected, repeat the PBS wash.

Loading of the Barcode Beads

1) Resuspend the previously prepared Barcode Beads by pipetting up and down.
2) Immediately, inject 60 µL of the Barcode Beads resuspension into the SCOPE-chip via the inlet port. Incubate for 1 minute.
3) Inject 100 µL of PBS to the inlet port. Dispense slowly over a period of 30 seconds. The beads will spread inside the SCOPE-chip.
4) Inject 200 µL of PBS into the SCOPE-chip via the inlet port. Dispense slowly over a period of 30 seconds.
5) In a new labeled 1.5 mL tube, collect the excess of beads from the output reservoir without making contact with the outlet port.
6) Repeat the 200 µL PBS wash two additional times to get rid of the excess of Barcode Beads.
7) Under the microscope, look at the distribution of the Barcode Beads. The beads should occupy at least 95% of the total number of microwells.

Important: If less than 95% of the microwells are occupied with beads, use the beads collected in step 5 to load the SCOPE-chip a second time. Place the recovered beads

on a magnetic rack. Remove the supernatant. Resuspend the beads in 30 µL of PBS. Inject the beads in the SCOPE chip following steps 2 to 7 above. If many empty wells are observed at proximity of the outlet reservoir, the recovered Barcode Beads can be injected via the outlet reservoir port.

Cell Lysis and mRNA Capture

1) Inject 100 µL (SCOPE-chip SD) or 150 µL (SCOPE-chip HD) of previously prepared Lysis Mix into the SCOPE-chip via the inlet port. Dispense slowly over a period of 30 seconds to avoid the introduction of air bubbles.
2) Remove the excess liquid from the outlet reservoir without making contact with the port.
3) Incubate the SCOPE-chip at room temperature for 20 minutes. During this time, the cells will be lysed, and the released mRNA molecules will bind to the Barcode Beads.
4) Proceed immediately to the Barcode Beads retrieval.

Barcode Beads Retrieval

1) Place the Singleron Magnetic Rack under the SCOPE-chip and let sit for 1 minute at room temperature.
2) Add 200 µL of pre-chilled Wash Buffer A into the outlet reservoir to remove impurities. Avoid making contact with the port. (***Do not inject the chip)

3) Remove liquid from the outlet reservoir without making contact with the outlet port.
4) Clean the reservoir two additional times, by repeating steps 2 and 3.
5) Place the Singleron Magnetic Rack on top of the SCOPE-chip and incubate for 1 minute at room temperature.

6) While keeping the magnet on top of the SCOPE-chip, add 200 µL of pre-chilled Wash Buffer A to the SCOPE-chip outlet reservoir without making contact with the port. (***Do not inject the chip).
7) Aspirate 200 µL of liquid containing the Barcode Beads with bound mRNA via the inlet port.

8) Transfer the collected Barcode Beads to a pre-cooled labeled 1.5 mL tube.
9) Repeat steps 6 to 8 two additional times to collect most of the Barcode Beads.
10) Proceed to the reverse transcription.

3.3.3 Reverse transcription and cDNA amplification

Reverse Transcription

1) Thaw RT Master Mix and TS Primer reagents at room temperature. Vortex, centrifuge briefly, keep on ice. (Ensure that the reagents are fully thawed).

Important: If a precipitate is observed in the RT Master Mix, mix by vortexing or pipetting up and down until the solution becomes clear.

2) Prepare the RT Mix on ice according to the tables below. Add the reagents in the order listed from top to bottom. Vortex, centrifuge briefly and keep on ice.

RT Mix (SCOPE-chip HD):

Component	1 RXN (µL) × 1	2 RXNs (µL) × 2.2
RT Master Mix	240	528
Cold nuclease-free water	90	198
TS Primer	20	44
RNase Inhibitor	10	22
Reverse Transcriptase	40	88
Total	400	880

3) Briefly centrifuge the tube containing the Barcode Beads previously collected from the SCOPEchip.

4) Place the tube on a magnetic rack. Incubate until the supernatant becomes clear.
5) Carefully discard the supernatant without disturbing the beads.
6) Remove the tube from the magnetic rack. Add 1 mL of pre-chilled Wash Buffer A. Mix well by pipetting up and down.
7) Centrifuge briefly and return the tube to the magnetic rack. Let sit until the supernatant becomes clear. 8. Carefully discard the supernatant without disturbing the beads.
8) Remove the tube from the magnetic rack.
9) Add 500 µL of Wash Buffer B and mix well by pipetting up and down.
10) Centrifuge briefly and return the tube to the magnetic rack. Let sit until the supernatant becomes clear.
11) Carefully discard the supernatant without disturbing the beads.
12) Centrifuge briefly and return the tube to the magnetic rack. Remove all residual buffer using a fine pipet tip.
13) Remove the tube from the magnetic rack and add: 400 µL of the pre-chilled RT Mix for SCOPE-chip HD.
14) Mix by pipetting up and down gently to homogenize the beads.
15) Place the tube in a preheated ThermoMixer at 42°C and shake at 1300 rpm for 90 minutes.

Important: If the cDNA amplification reaction cannot be carried out immediately, incubate the product of the reverse transcription reaction at 70°C for 15 minutes. Store at 4°C overnight.

cDNA Amplification

1) On a thermocycler, set up the cDNA Amplification program according to the following table. The lid temperature of the thermocycler should be set at 105°C.

Lid Temperature 105°C		Reaction Volume 50µL
Step	Temperature	Time
1	95°C	3 min
	98°C	20 sec
2	65°C	45 sec
cycles = 4	72°C	3 min
	98°C	20 sec
3	67°C	20 sec
cycles = 9	72°C	3 min
4	72°C	5 min
5	4°C	Hold

2) Thaw the Amplification Master Mix and A Primer Mix at room temperature. Vortex, centrifuge briefly, keep on ice. (Ensure that the reagents are fully thawed)
3) Prepare the PCR Mix on ice according to the table below. Add the reagents in the order listed, from top to bottom. Vortex, centrifuge briefly, and keep on ice.

PCR Mix (SCOPE-chip HD):

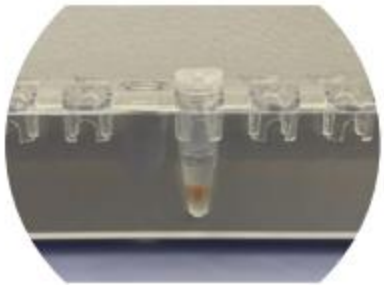
Component	1 RXN (µL) ×1	2 RXNs (µL) ×2.2
Amplification Master Mix	344	756.8
A Primer Mix	64	140.8
Cold nuclease-free water	376	827.2
Amplification Enzyme	16	35.2
Total	800	1760

4) Briefly centrifuge the reverse transcription product and place it in the magnetic rack.
5) Let sit undisturbed until the supernatant becomes clear. Then, carefully discard the supernatant without disrupting the beads.
6) Centrifuge the tube briefly.
7) Place the tube back in the magnetic rack. Using a fine pipet tip to remove all residual liquid.
8) Remove the tube from the magnetic rack and add: 800 µL of the PCR Mix for SCOPE-chip HD 8. Mix well by pipetting up and down.
9) Distribute the mixture evenly into a labeled 8-tube PCR strip by pipetting 50 µL of the PCR Mix + beads into each tube. For SCOPE-chip HD, use two 8-tube PCR strips.
10) Close the caps of the 8-tube strip and place it in the preheated thermocycler.
11) Perform PCR amplification using the cDNA Amplification program.
12) Once the thermocycler reaches 4°C, the cDNA can be stored at 4°C for 48h or at -20°C for up to a week.

cDNA Purification

1) Prepare 2 mL (SCOPE-chip SD) or 4 mL (SCOPE-chip HD) of 80% ethanol per reaction.

2) For SCOPE-chip SD, combine the contents of the 8-tube PCR strip containing the amplified cDNA into a labeled 1.5 mL tube. For SCOPE-chip HD, combine the contents of each 8-tube PCR strip into separate labeled 1.5 mL tubes.
3) Centrifuge briefly and measure the volume with a pipette.
4) Calculate the volume of AMPure magnetic beads equivalent to 0.6× the total volume of the amplified cDNA. (For example, if the volume of the measured product is 400 µL, then use 0.6 × 400 µL = 240 µL of AMPure beads.)
5) Vortex the AMPure beads until homogenized, then add the calculated volume to the amplified cDNA tube(s).
6) Mix well by vortexing for 15 seconds. Incubate at room temperature for 5 minutes.
7) Centrifuge the tube briefly and place in the magnetic rack for 5 minutes or until the liquid is clear.

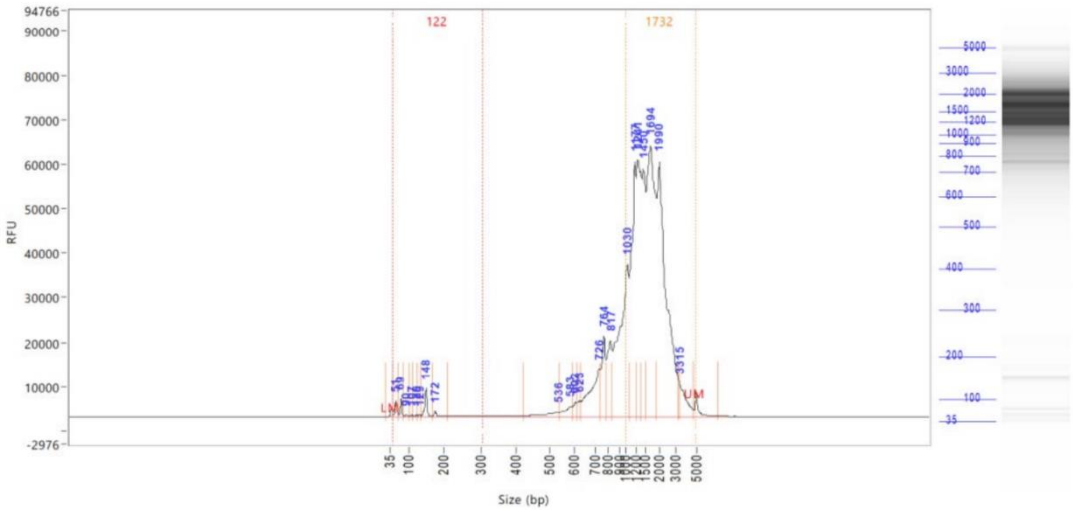


8) Carefully discard the supernatant without disturbing the beads.
9) Keep the tube(s) in the magnetic rack and add 800 µL of freshly prepared 80% ethanol to wash the magnetic beads. Incubate at room temperature for 30 seconds.
10) Carefully discard the supernatant without disrupting the beads.
11) Repeat the 80% ethanol wash one more time.
12) Centrifuge briefly and return the tube in the magnetic rack.
13) Remove all residual ethanol using a fine pipet tip.
14) Keep the lid open to dry the beads for about 2 minutes or until they are no longer shiny (no more than 5 minutes).
15) Remove the tube from the magnetic rack and add 20 µL Elution Buffer (10mM Tris-HCl pH 8.5). Vortex to mix.

16) Incubate at room temperature for 5 minutes. Centrifuge briefly, and place back in the magnetic rack until the liquid is clear.
17) Transfer the supernatant containing the purified cDNA to a new labeled 1.5 mL tube. For SCOPE-chip HD, combine the supernatants from the two 1.5 mL tubes into a single new tube.
18) The cDNA can be stored at 4°C for 72 hours or at -20°C for up to one week.

cDNA QC

1) Take an aliquot (1 µL) of the cDNA and measure its concentration using a Qubit 4 Fluorometer.
2) Take an aliquot of the cDNA (2-5 ng/µL) to measure the fragment size profile using Agilent Fragment Analyzer 5200 or equivalent. (Dilute the cDNA following the fragment analyzer recommendations).
3) The ideal cDNA should meet the following criteria:
   a) Average peak from region between 200 - 5,000 bp: > 1,200 bp;
   b) The fragments between 1,000 – 5,000 bp should represent more than 30% of the total molecules.



4) If the fragments below 300bp represents more than 10% of the total molecules, perform a second AMPure purification according to the table below.

Percentage of fragment between 40bp-300bp	Ratio of AMPure beads added
10% - 20%	0.8 x
21% - 35%	0.7 x
> 35%	0.6 x

It is recommended to dilute the cDNA with nuclease-free water to a volume of 100 µl to perform the second round of AMPure purification.

3.3.4 Library preparation

Fragmentation

1) On a thermocycler, set up the Fragmentation program according to the following table. The lid temperature of the thermocycler should be set at 75°C.

Lid Temperature 75°C		Reaction Volume 35µL
Step	Temperature	Time
1	37°C	10 min
2	65°C	30 min
3	4°C	Hold

2) Thaw the Fragmentation Buffer at room temperature. Vortex 5 to 8 seconds, centrifuge briefly, and keep on ice. If a precipitate is visible, mix by vortexing until the solution is clear.
3) Vortex the Fragmentation Enzyme Mix 5 to 8 seconds before use.
4) Dilute the cDNA with 1xTE according to the following recommendations: a) If the total amount of cDNA is less than 10 ng, use all the cDNA. b) If the total amount of cDNA is between 10 and 50 ng, use 10 ng as input. c) If the total amount of cDNA is above 50 ng, use 50 ng as the input.
5) Prepare the Fragmentation Reaction on ice according to the table below. Add the reagents in the order listed from top to bottom. Vortex for 10 seconds, centrifuge briefly, and keep on ice. Fragmentation Reaction:

Component	Volume（µL）
Diluted cDNA (10ng or 50ng）	26
Fragmentation Buffer	7
Fragmentation Enzyme Mix	2
Total	35

6) Immediately place the PCR tube in the pre-heated thermocycler and run the Fragmentation program.
7) Once the thermocycler reaches 4°C, proceed immediately with the adapter ligation.

Adapter Ligation

1) On a thermocycler, set up the Ligation program according to the table below. The lid temperature of the thermocycler should be off and allowed to cool down to room temperature.

Lid Temperature OFF		Reaction Volume 70µL
Step	Temperature	Time
1	20°C	15 min
2	4°C	Hold

2) Thaw the Adaptor at room temperature, mix well, keep on ice.
3) Gently mix the Ligation Mix and the Ligation booster by pipetting up and down. Centrifuge briefly and keep on ice. Do not vortex.
4) On ice, add the Adapter Ligation reagents directly to the fragmented cDNA according to the table below.
   Adapter Ligation*:

Component	Volume（µL）
Fragmented cDNA	35
Ligation Mix**	30
Ligation booster	1
Adaptor	2.5
Total	68.5

** It is not recommended to prepare a master mix*

***Ligation Mix is viscous. Pipet carefully to ensure an accurate volume*

5) Mix gently by pipetting up and down and centrifuge briefly. Place the reaction tube into the precooled (20°C) thermocycler and proceed with the Ligation PCR program. Do not vortex.
6) Once the thermocycler reaches 4°C, immediately proceed to the purification of the adapter ligated cDNA.

Purification of the Adapter-Ligated cDNA

Important: Vortex the AMPure beads thoroughly before use; Bring the AMPure beads to room temperature by allowing a ~1 mL aliquot to sit for 30 minutes or until they reach room temperature; Pipet the AMPure beads slowly to ensure the accuracy of the retrieved volume.

1) Prepare 0.5 mL of 80% ethanol and 25 µl 0.1x TE per reaction (1:10 dilution of 1×TE with nuclease-free water).
2) Centrifuge the adapter-ligated cDNA from the previous step briefly and measure the volume with a pipette.
3) Calculate the volume of AMPure beads equivalent to 0.5× the total volume of the ligated product. (For example, if the volume of the measured product is 68.5 µL, then 0.5 × 68.5 µL = 34.3 µL of AMPure magnetic beads should be used).
4) Vortex the AMPure beads until homogenized and add the calculated volume to the ligated product.
5) Mix well by vortexing and incubate at room temperature for 5 minutes.
6) Centrifuge the tube briefly and place in the magnetic rack for 5 minutes or until the liquid is clear.
7) Carefully discard the supernatant without disturbing the beads.
8) Keep the tube in the magnetic rack and add 200 µL of freshly prepared 80% ethanol to wash the magnetic beads. Let sit at room temperature for 30 seconds.
9) Carefully discard the supernatant without disrupting the beads.
10) Repeat 80% ethanol wash one more time.
11) Centrifuge briefly and return the tube to the magnetic rack.
12) Remove all residual ethanol using a fine pipet tip.
13) Keep the lid open to dry the beads for about 2 minutes or until the beads are no longer shiny (no more than 5 minutes).
14) Remove the tube from the magnetic rack and add 17 µL of 0.1x TE. (1:10 dilution of 1×TE with nuclease-free water). Vortex to mix.
15) Incubate at room temperature for 5 minutes. Centrifuge briefly, and place back in the magnetic rack until the liquid is clear.
16) Transfer 15 µL the supernatant containing the purified adapter-ligated product to a new labeled PCR tube.
17) The adapter-ligated product can be stored at 4°C for 72 hours or at -20°C for one week.

Library Amplification

1) On a thermocycler, set up the Library Amplification program according to the following table. The lid of the thermocycler should be set at 105°C.

Lid Temperature 105°C		Reaction Volume 50 µL
Step	Temperature	Time
1	98°C	30 sec
2（cycle number varies*）	98°C	10 sec
	65°C	75 sec
3	65°C	5 min
4	4°C	Hold

** The number of cycles should be selected based on the amount of cDNA input as depicted in the table below.*

cDNA input	Number of cycles
50ng	10
10ng	13

2) Thaw the Library Amp Mix v2 and Unique Dual Indexes on ice. Vortex, centrifuge briefly, keep on ice. Ensure that the reagents are fully thawed.

Component	Volume（µL）
Purified adapter-ligated product	15
Library Amp Mix v2	25
Unique Dual Indexes*	10
Total	50

** Choose a different UDI Adapter for each sample and use Illumina pooling recommended strategies. See Appendix for NovaSeq X.*

3) On ice, add the following reagents directly to the purified adapter-ligated product.
4) Mix by pipetting up and down and centrifuge briefly.
5) Place the reaction in the thermocycler and run the Library Amplification program.
6) Once the thermocycler reaches 4°C, immediately proceed to the purification and size selection of the amplified library.

Purification and Dual Size Selection of the Amplified Library

Important: Vortex the AMPure beads thoroughly before use; Bring the AMPure beads to room temperature by allowing a ~1 mL aliquot to sit for 30 minutes or until they reach room temperature; Pipet the AMPure beads slowly to ensure the accuracy of the retrieved volume.

1) Prepare 0.5 mL of 80% ethanol per reaction.

2) Briefly centrifuge the amplified library and measure the volume with a pipette.
3) Calculate the volume of AMPure beads equivalent to 0.5× the total volume of the amplified libraries. (For example, if the volume of the measured product is 50 µL, then use 0.5 × 50 µL = 25 µL of AMPure magnetic beads.)
4) Vortex the AMPure beads until homogenized and add the calculated volume to the amplified libraries. 5. Mix well by vortexing and incubate at room temperature for 5 minutes.
5) Centrifuge the tube briefly and place in the magnetic rack for 5 minutes or until the liquid is clear.
6) Transfer 70 µL of the supernatant containing the amplified library to a new tube.
7) Vortex the AMPure beads stock tube until homogenized and add the 7.5 µL to the amplified libraries.
8) Mix well by vortexing and incubate at room temperature for 5 minutes.
9) Centrifuge the tube briefly and place in the magnetic rack for 5 minutes or until the liquid is clear.
10) Carefully remove and discard the supernatant without disrupting the beads.
11) Keep the tube in the magnetic rack and add 200 µL of freshly prepared 80% ethanol to wash the magnetic beads. Incubate at room temperature for 30 seconds.
12) Carefully discard the supernatant without disrupting the beads.
13) Repeat 80% ethanol wash one more time.
14) Centrifuge briefly and return the tube to the magnetic rack.
15) Remove all residual ethanol using a fine pipet tip.
16) Keep the lid open to dry the beads for about 2 minutes or until the beads are no longer shiny (no more than 5 minutes).
17) Remove the tube from the magnetic rack and add 20 µL of Elution Buffer or 10mM Tris-HCl pH 8.5. Vortex to mix.
18) Incubate at room temperature for 5 minutes. Centrifuge briefly, and place back in the magnetic rack until the liquid is clear.
19) Transfer 18 µL the supernatant containing the purified library to a new labeled tube.
20) The libraries can be stored at -20°C or -80°C for up to three months.

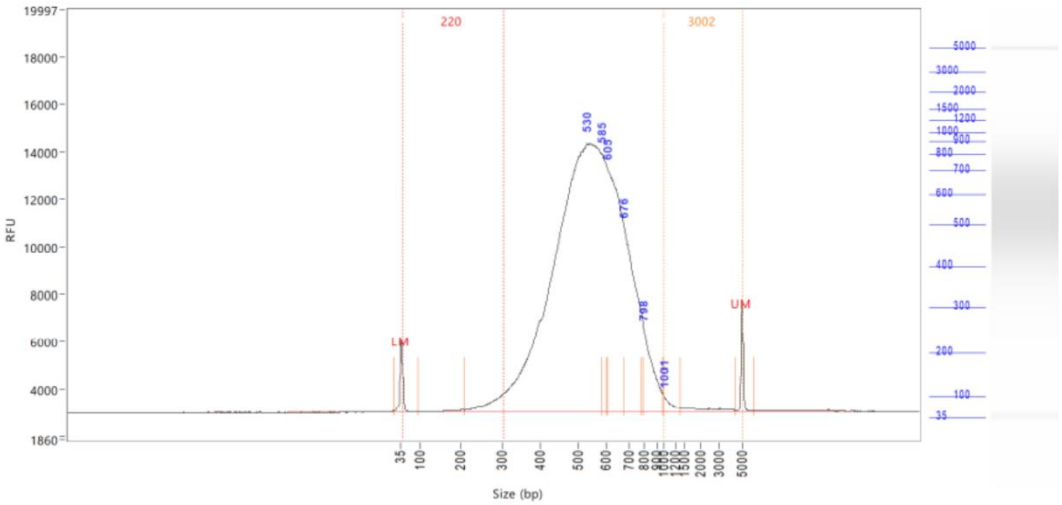
QC of the Library

1) Take an aliquot (1 µL) of the library to measure the concentration using Qubit 4 Fluorometer.
2) Take an aliquot of the library (2 - 5 ng/µL) to measure the fragment size profile using Agilent Fragment Analyzer 5200 or equivalent. (Dilute the library following the fragment analyzer recommendations).
3) The ideal library should meet the following criteria:

   a. The main peak should average between 400 - 700bp when the region analyzed is set between 300 bp and 2000 bp.

   b. The fragments between 900 bp and 1500 bp should account for less than 10% of the molecules.

   c. The fragments below 300 bp should account for less than 10% of the molecules.
4) If a peak is observed around 200 bp, an extra AMPure purification using a bead ratio of 0.8x the volume of your sample should be performed.
5) Qualified libraries can be sent directly for sequencing. Stored at -20°C or -80°C for up to three months.



3.4 Bioinformatic analysis

3.4.1 Celescope

For the analysis of single-cell RNA sequencing data libraries, the Celescope (Singleron Biotechnologies) pipeline was used, a computational framework specifically designed for processing and analyzing scRNA-seq data (https://github.com/singleron-RD/CeleScope/blob/master/doc/assay/multi_rna.md). Celescope includes the preprocessing of raw FASTQ files, the correction of errors in cellular barcodes through edit distance algorithms, sequence demultiplexing, alignment to the reference genome using STAR aligner, and the quantification of gene expression at single cell level. The pipeline automatically generates gene-cell matrices in standard format, applying quality filters to remove cells with low number of genes detected and genes expressed in insufficient number of cells.

3.4.2 Seurat

Single cell RNA sequencing data analysis was conducted using the Seurat v5 package in R environment. Data from different experimental conditions (CTRL1-5 and CD1-5) were combined into a single Seurat object, with unique identifiers assigned to each sample to preserve the cellular origin. The following steps were performed first individually on each sample and then on the final Seurat object, formed by the set of 10 samples. Seurat provides an initial phase of **quality control** using three key factors:

1) **percent.mt** (Mitochondrial gene percentage) Represents the percentage of reads/UMI (Unique Molecular Identifiers) that map to mitochondrial genes in relation to the total number of reads for each cell. It is a crucial indicator of cell quality, as dead or stressed cells tend to have compromised cell membranes that allow the release of cytoplasmic mRNA while mitochondrial mRNA remains trapped. High values (typically >20-25%) suggest dead or low-quality cells.
2) **nCount_RNA** (Total UMI count) Indicates the total number of RNA molecules (UMI) detected per cell. This value indicates the depth of sequencing for each cell. Cells with very low values could be dead cells, empty doublets or cells with poor mRNA capture, while extremely high values could indicate doublets (two or more cells captured together).

3) **nFeature_RNA** (Number of detected genes) Represents the number of unique genes detected (with at least 1 UMI) in each cell. It is an indicator of the transcriptional complexity of the cell. Healthy and viable cells should express thousands of genes, while dead or low-quality cells typically show a reduced number of detected genes. Very low (<500-1000 genes) or very high values may indicate quality problems.

These three parameters are used together to filter low quality cells before downstream analysis. In this work the **cut-off** used are: **percent.mt** ≤ 25, **nFeature_RNA** ≥ 250, **nCount_RNA** ≥ 250. The evaluation of these cut-offs took place through the visualization of the VlnPlot that allows us to identify the anomalous distributions and consequently choose the appropriate values.

Data **normalization** was then performed, allowing the difference in sequencing depth between cells to be corrected. The "LogNormalize" method with a scale factor of 10000 was used for standardization. Following normalization, the selection of the most **variable genes** was made using the "vst" method, essential for subsequent dimensionality reduction analyses. The first 3000 most variable genes have been chosen.

These data were scaled and subjected to **Principal Component Analysis** (PCA) to reduce the dimensionality of the dataset but most of the variance of the samples was preserved. Cell cycle variability (CC.Difference), percentage of mitochondrial genes (percent.mt) and total number of RNA counts (nCount_RNA) were removed by regression analysis. The choice of the number of main components (PC) was determined by analysis of the explained variance (ElbowPlot) and criteria such as a cumulative variance greater than 90% and a percentage variation less than 5% between consecutive PCs. In particular, the first 11 PCs were used.

The identification of distinct cell populations was achieved by clustering algorithms based on k-nearest neighbor graphs. **Uniform Manifold Approximation and Projection** (UMAP) non-linear projections were applied for cell visualization, both before and after data integration via Harmony. **Harmony** corrects batch effects related to the origin of the samples, as it is performed only on the final Seurat object and eliminates the specific variations of the dataset. The **identification of specific marker genes** for each cluster was carried out by means of differential statistical tests using the **FindAllMarkers** package. Genes differentially expressed in the various clusters

have been used as specific markers to identify each cell type. These genes were researched in the literature manually and then assigned to each individual population.
The last step was the removal of duplicates, carried out through the **DoubletFinder** package (v2.0.4.) This process involved:

1) Identifying the optimal pK parameter (representing the density of the nearest k-the most appropriate for differentiating simulated duplicates from single cells) by selecting the maximum bimodality coefficient (BCmetric) from a sweep of parameters.
2) The estimation of the proportion of homotypic duplicates (homotypic.prop) based on the initial entries of the clusters.
3) The calculation of the expected number of duplicates (nExp_poi), assuming a duplicate formation rate of 7.5% and adjusting for non-standard duplicates.

DoubletFinder was then run with these optimized parameters (using the first 20 PCs and a pN of 0.25) to classify and report the likely duplicates within the dataset.
Note that this step has been performed for each individual sample at the end of the analysis, before creating the final unified object in such a way as to have an object without duplicates.

3.4.3 Cell-cell communication analysis

Cell-cell communication analysis was performed using the **CellChat** package (version 1.5.0). This package quantifies the probability of signaling communication between two groups of cells based on mass action, which incorporates the fundamental interaction between ligands and multisubunit receptors along with modulation by cofactors(Wilk et al. 2024). Furthermore, CellChat identifies the signaling pathways and ligand-receptor pairs involved. This identification allows communication circuits between cell populations to be found and visualized in hierarchical networks that highlight the main signal senders and receivers. Default parameters were used for this thesis. Specifically, the analysis was first performed on all samples together without distinction of disease, and then separately on celiac patients and controls.

3.5 Statistical analysis

Statistical analysis to examine the different composition of cell types between the two groups (CD and CTRL) was performed using the Mann-Whitney U test. The Mann-Whitney U test is a widely used nonparametric statistical method for comparing the distributions of two independent samples. Unlike Student's t-test, which requires data to be normally distributed, the Mann-Whitney test does not assume a specific distribution (hence the term “non-parametric”) and instead focuses on the relative position of values. To perform the test, observations from both groups are combined and sorted, and each is assigned a rank. The test then calculates the U statistic by comparing the sum of the ranks of each group, determining the probability that an observation chosen at random from one population will be greater than an observation chosen at random from the other. To interpret the test results, a significance level of 0.05 was adopted. Consequently, a value equal to or less than 0.05 was considered statistically significant, while a value greater than 0.05 was considered statistically insignificant. This makes the Mann-Whitney test particularly effective and robust when analyzing data with significant outliers or when the sample size is small, as is often the case in biological or clinical studies.

4. RESULTS AND DISCUSSION

4.1 Patient characteristics

For this study a total of 12 patients were enrolled:

- Patients with CD.
- Controls.

The 6 patients with CD (5 females and 1 male) showed a mean age of 11 years with a standard deviation 1.5. Five out of 6 were symptomatic patients. They were tested positive for both CD-specific serological markers EMA and anti-tTG (mean = 32.5 U/mL and standard deviation = 10.9) (Table 1).

The 6 controls (5 females and 1 male) showed a mean age of 14 years with a standard deviation of 2.4 and suffered from minor intestinal disorders unrelated to celiac disease. Minor intestinal disorders include: functional abdominal pain, functional dyspepsia and dysphagia somatic. They were all symptomatic subjects. They tested negative for both EMA CD-specific and anti-tTG serological markers (Table 1).

ID sample	***Age***	***Sex***	***Symptomatology***	***Anti-tTG U/mL***	***EMA***	***Histologic classification***	***Diagnosis***
CTRL0	17	F	Epigastric pain	0	Negative	Marsh 0	Functional abdominal pain
CTRL1	14	F	Abdominal pain; Epigastric pain; Severe constipation	0.5	Negative	Marsh 0	Functional Dyspepsia
CTRL2	12	M	Abdominal pain, Diarrhea, Oral aphthosis	0.2	Negative	Marsh 0	Other
CTRL3	16	F	Epigastric pain	0	Negative	Marsh 1	Functional Dyspepsia
CTRL4	11	F	Abdominal pain; Constipation; Epigastric pain; Nausea	0.1	Negative	Marsh 0	Functional Dyspepsia
CTRL5	12	M	Dysphagia	0.2	Negative	Marsh 0	Dysphagia somatic
CD0	12	F	Abdominal pain	35	Positive	Marsh 0	Celiac Disease
CD1	12	M	Abdominal pain	15	Positive	Marsh 3a-3b-3c	Celiac Disease
CD2	12	F	Diarrhea	42	Positive	Marsh 3a-3b-3c	Celiac Disease
CD3	8	F	Dermatitis alopecia areata	44	Positive	Marsh 3a-3b-3c	Celiac Disease
CD4	11	F		34	Positive	Marsh 3	Celiac Disease
CD5	11	F	Abdominal pain; Meteorism	25	Positive	Marsh 3	Celiac Disease

Table 1: Subjects' features. *Clinical, biological and histological features of CD and control subjects. CD, coeliac disease; Ctrl, control; F, female; M, male.*

4.2 Library generation and sequencing

The generation of the scRNA-seq data from intestinal biopsy samples is a fairly long and delicate process with multiple steps starting from tissue dissociation to library generation and sequencing. The raw data are finally used for bioinformatic analysis (Figure 11).

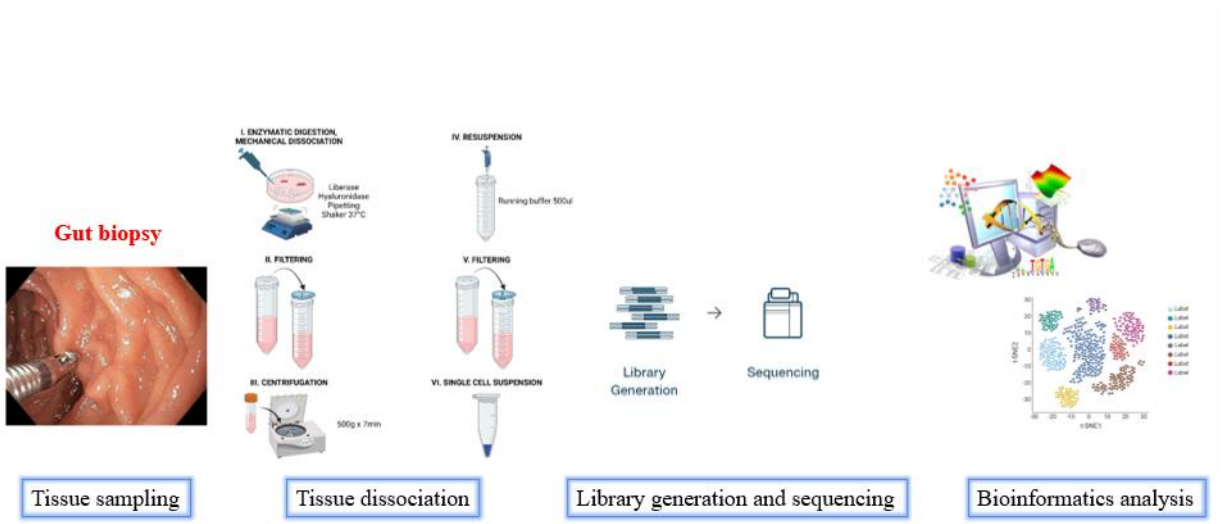


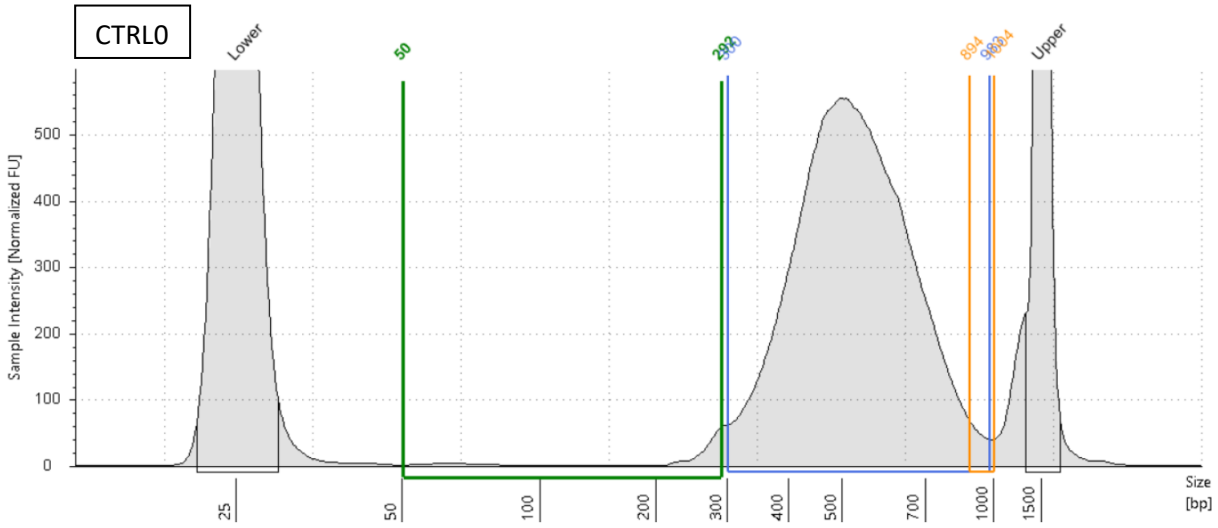
Figure 11: Library generation. *The first step is collecting the sample. Subsequently the sample is dissociated through mechanical and enzymatic breakdown obtaining a suspension of single cells. From this suspension of single cells, the library is generated which will then be sequenced. The raw data obtained are analyzed through bioinformatic analyses.*

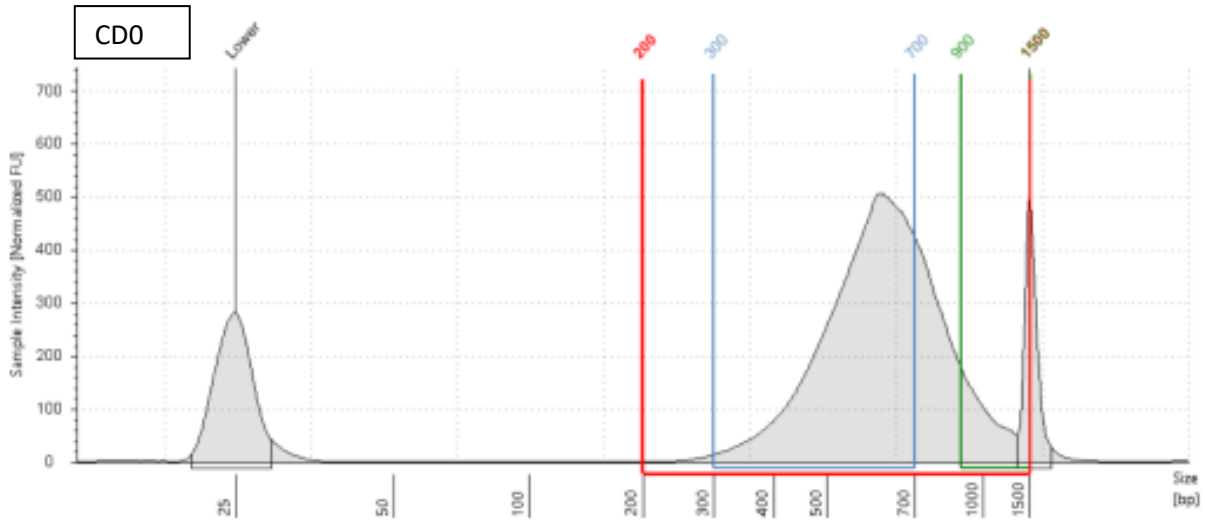
A pilot study was performed to understand whether libraries could be generated from stored or fresh intestinal samples. In this study 4 patients were involved (2 CD patients and 2 controls) and divided in 2 groups:

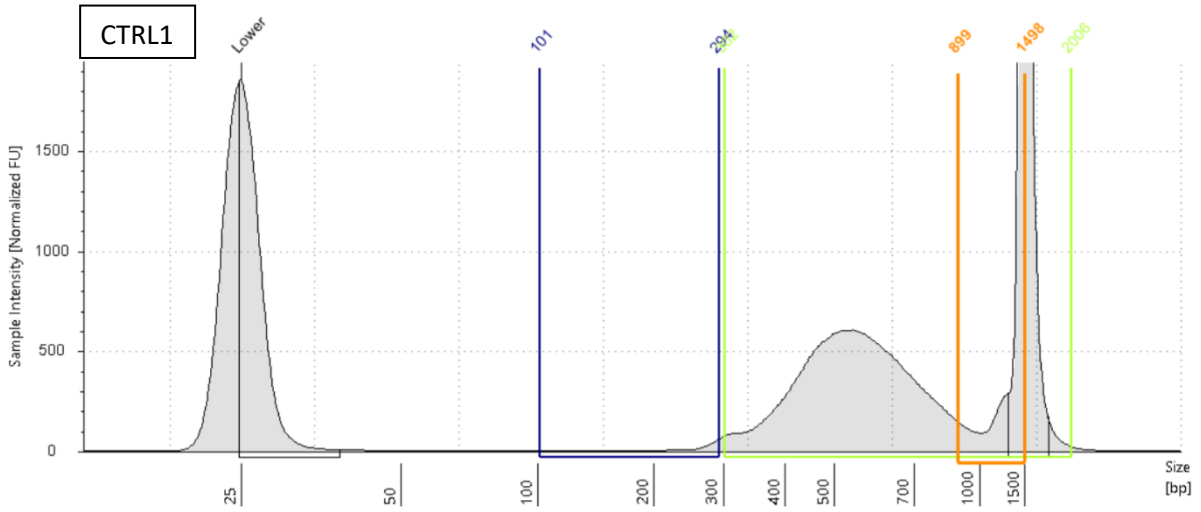
Group A: intestinal biopsies from CD0 and CTRL0 were stored in a tissue preservation buffer (sCelLiVETM) for 3 days before tissue dissociation and scRNA-seq library generation;

Group B: intestinal biopsies from CD1 and CTRL1 were processed after endoscopic collection within 30 minutes.

The sc-RNA seq libraries were successfully generated in both groups by using the mRNA contained in around 10.000 single cells. As we can see from Figure 12, in all samples the average of the main peak was between 400 and 700 bp. Analysing the region between 300 and 2000 bp, fragments between 900 and 1500 bp accounted for less than 10% and fragments less than 300 bp accounted for less than 10%, confirming the success of the generation.


CTRL0
Lower
50
293
834
9834
Upper
Sample Intensity [Normalized FU]
0
100
200
300
400
500
Size
[bp]
25
50
100
200
300
400
500
700
1000
1500


CD0
Lower
209
300
700
900
1500
Sample Intensity [Normalized FU]
0
100
200
300
400
500
600
700
Size
[bp]
25
50
100
200
300
400
500
700
1000
1500


CTRL1
Lower
101
294
899
1498
2006
Sample Intensity [Normalized FU]
0
500
1000
1500
Size
[bp]
25
50
100
200
300
400
500
700
1000
1500

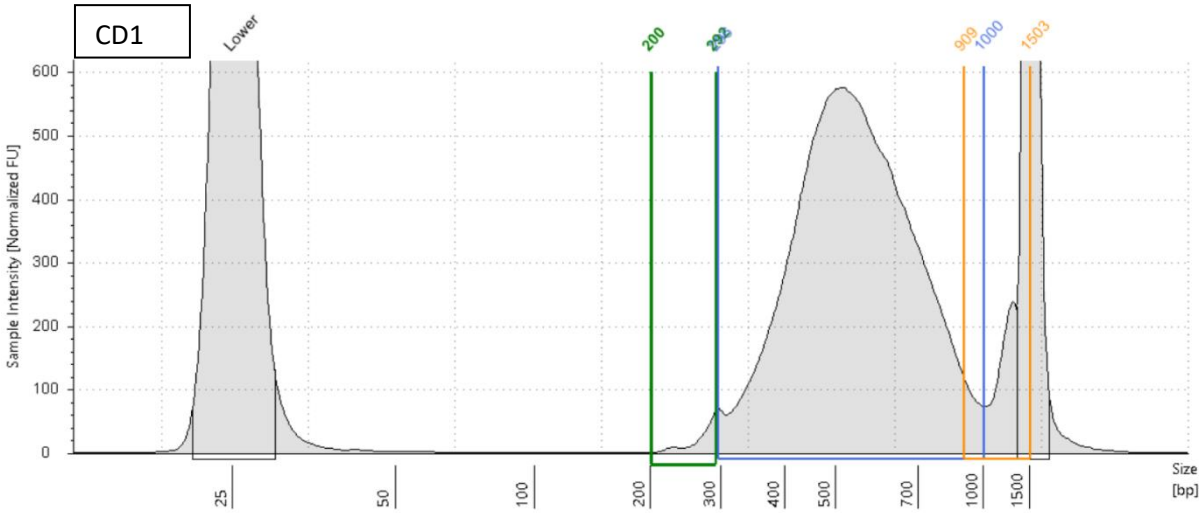


Figure 12: Quality control of scRNA-seq libreries. *The average of the main peak must be between 400 and 700 bp, fragments between 900 and 1500 bp must be less than 10%, and fragments less than 300 bp must represent less than 10%.*

All the 4 sc-RNA libraries were sequenced in service, on an Illumina NovaSeq 6000 sequencing platform. For an in-depth analysis we chose a sequencing configuration set to 150PE (Paired-End at 150 bp), generating a high productivity of 90 Gigabases (Gb), corresponding to 600 million total reads. Finally, the raw sequencing data was delivered via an FTP link for bioinformatics analysis.

The raw data were obtained after library sequencing and analyzed. The bioinformatics analysis starts from a quality control of data and includes nFeature_RNA ≥ 250; nCount_RNA ≥ 250; Percent_mt ≤ 25. In both samples from group A, a high number of cells (about 80%) showed a high percentage of mithocondrial RNA (> 25%) (Figure 13). Following this step, both samples lost approximately 8.000 cells, thus going from approximately 10.000 initial single cells to 2.000 single cells.

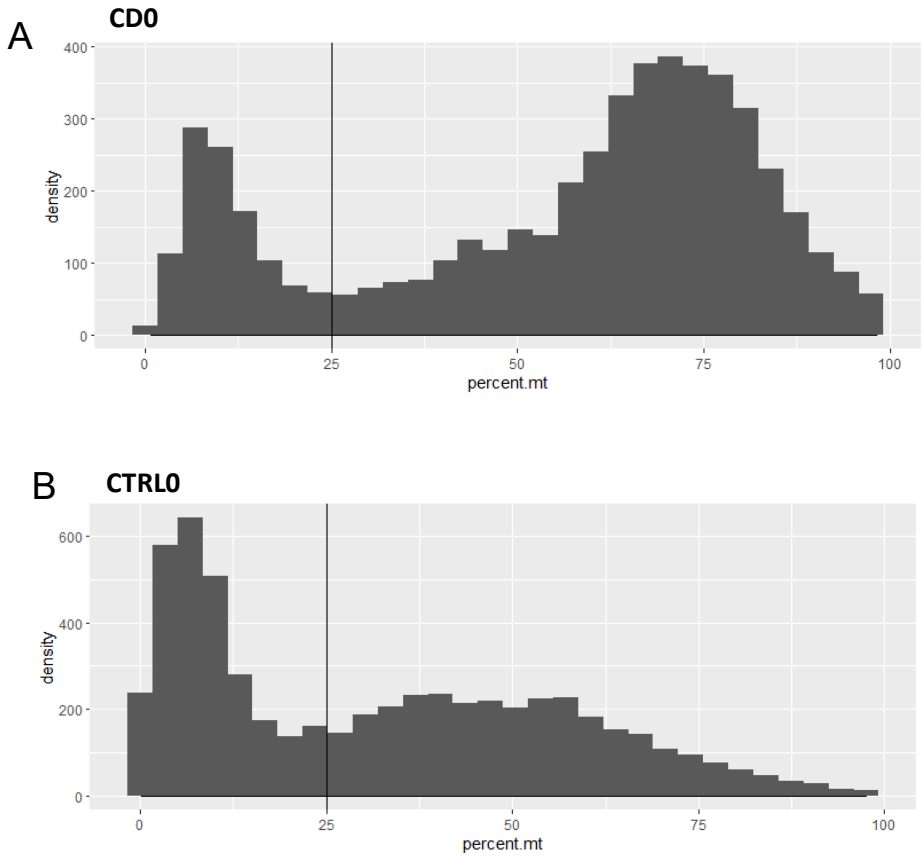


Figure 13: Concentration of mitochondrial RNA. *(A) The density plot shows a bimodal distribution with an initial peak around 10-15% and a second larger peak centered around 75-80% mitochondrial RNA. (B) The density plot shows a high concentration of mitochondrial RNA (peak around 25% and distribution extended up to 100%), indicative of cell death in the biopsies analyzed.*
The black vertical line highlights a reference threshold for evaluating the quality of the sample.

Instead, in group B, the analysis showed that a low number of cells (about 35%) had a high percentage of mitochondrial RNA (<25%). In this case, however, both samples lost approximately 3.500 cells, going from 10.000 initial single cells to 6.500 single cells (Figure 14).

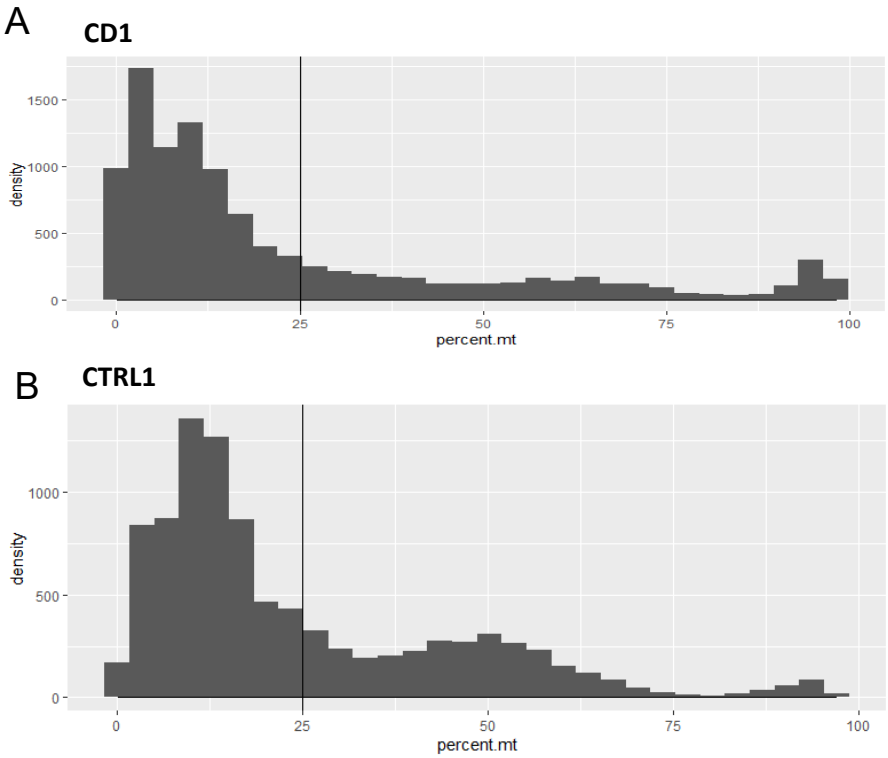


Figure 14: Concentration of mitochondrial RNA. *(A) The density plot shows that the majority of the celiac patient's cells have a mitochondrial RNA concentration of less than 25%. (B) The density plot shows that in controls the concentration of mitochondrial RNA is less than 25%. The black vertical line highlights a reference threshold for evaluating the quality of the sample.*

In this work we decided to use this preservation method and not for example freezing in liquid nitrogen or at -80°C, as it was a process that stressed the cells less than the other two. This pilot study allowed us to understand that to obtain high-quality scRNA-seq data we need to work on fresh samples. We therefore decided to continue the study by generating the scRNA-seq libraries as already carried out for group B. Single cell RNA sequencing libraries were successfully generated from other 8 patients: 4 CD (CD2, CD3, CD4, CD5) and 4 CTRL (CTRL2, CTRL3, CTRL4, CTRL5). A quality control analysis was carried out on 87.710 single cells obtained from sequencing. From the quality control analysis (nFeature_RNA ≥ 250; nCount_RNA ≥ 250; Percent_mt ≤ 25) of these 8 patients, we obtain a dataset of 58.503 single cells. The loss of this 30% of cells is due to the filter linked to the percentage of mitochondrial present in the cells.
The final dataset comprised 71.397 high-quality single cells obtained from 10 patients (CD1-5, CTRL1-5). We therefore decided to process the samples immediately, since storage induces cellular stress, with an increase in mitochondrial RNA (mtRNA) concentration highlighted. Cells exceeding the quality control threshold (Percent_mt ≥25%) were eliminated, and the data obtained from the few remaining cells were of low

quality. Conversely, analysis of the immediately processed sample showed a significantly lower incidence of cellular stress (Percent_mt ≤25%), thereby producing a high-quality final dataset.

4.3 Cluster analysis

The dataset of 71.397 single cells was processed to identify all the intestinal cell types. The first step involved the selection of the most variable genes (HVGs) which represents one of the most critical steps in the analysis of scRNA-seq data. This step is crucial for determining the subsets of genes that have been used for subsequent size reduction and clustering analyses. The genes that have been chosen are those that show biological variability between cells, which is important for then interpreting cellular diversity. In this study we decided to select the first 3.000 genes (Figure 15) as it is a very large and complex dataset.

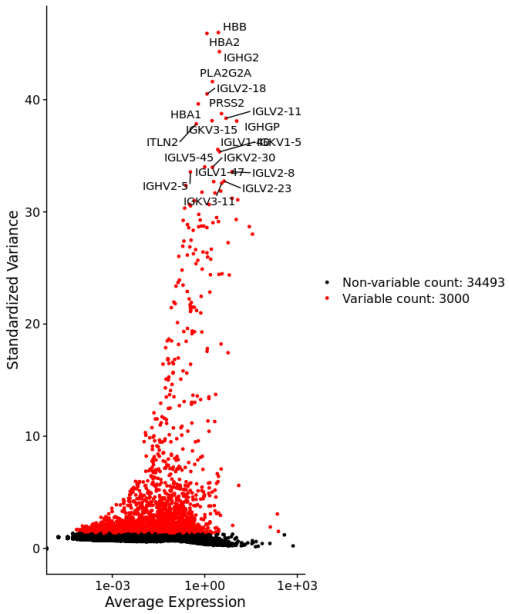


Figure 15: Variable Features. *The scatter plot graph shows the relationship between mean expression (x-axis) and standardized variance (y-axis) for all detected genes. The red dots represent the 3.000 most variable genes selected for the analyses, while the black dots indicate the excluded non-variable genes. Genes with the highest levels of variance are labeled (HBB, HBA2, IGHG2, PLA2G2A, IGLV2-18, PRSS2, IGLV2-11, etc.), many of which encode hemoglobin proteins and immunoglobulins.*

The identification of HVGs was performed with the variance stabilizing transformation (VST) approach which is implemented in the Seurat package used. This method is one of the most robust and effective for selecting biologically informative genes from scRNA-seq data. Subsequently, principal component analysis (PCA) was carried out

on the previously selected first highly variable 3000 genes. The main aim of PCA on scRNA-seq data is the transformation of high-dimensional gene expression matrices into a reduced-dimensional space. This transformation preserves the maximum amount of biological information (variance) while reducing the complexity of the data and facilitating its visualization. The PCA achieves this goal by finding new variables, called principal components (PCs), which are mutually independent and which summarize the information contained in the data. The first main component captures the greatest possible amount of variability present in the dataset, the second captures a smaller part independent of the first, and so on. Each successive component explains an increasingly smaller share of overall variability. Through a systematic analysis of the variance explained by each main component and by the identification of the "inflection point" in the so-called "elbow plot" (Figure 16) the first 11 PCs (in red) appear to be the most significant.

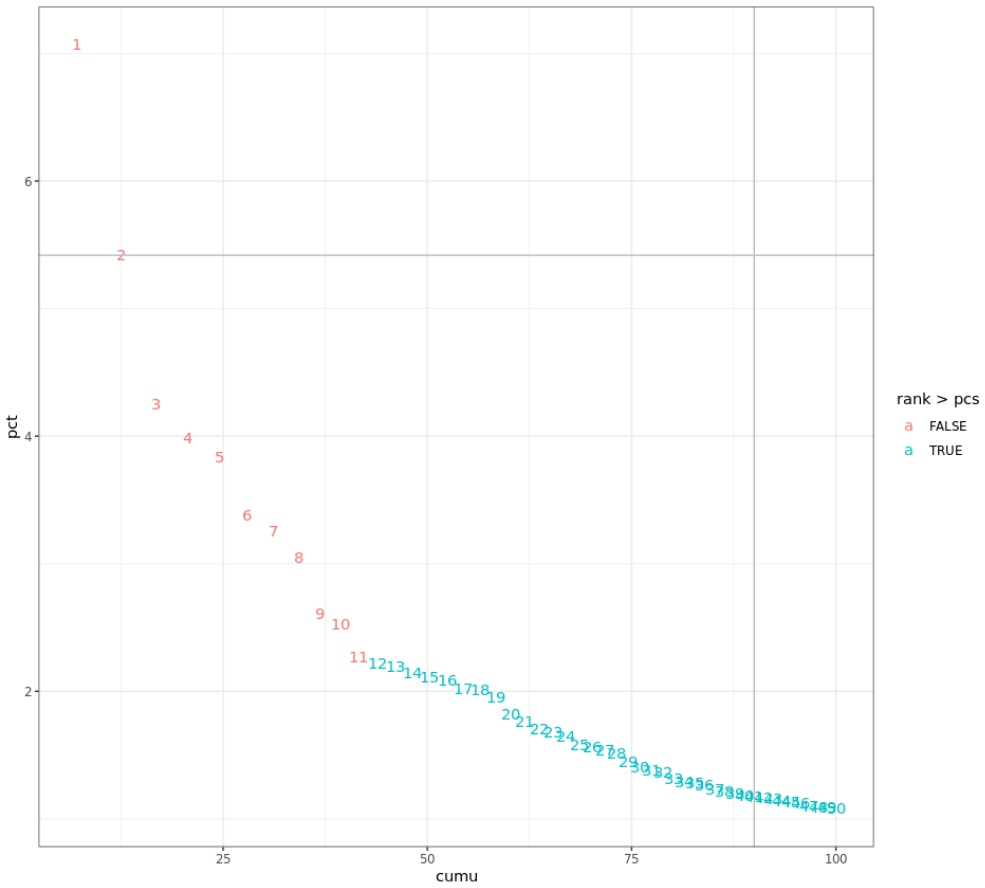


Figure 16: Selecting the optimal number of principal components (PCs) for PC analysis. *The elbow plot graph shows the relationship between the cumulative number of principal components (x-axis) and their standard deviation (y-axis). Red dots (FALSE) indicate PCs with significant variance, while blue dots (TRUE) represent PCs with non-significant variance. The "elbow" of the curve is observed around the first 11.*

We decided to use these 11 PCs for all subsequent analysis. Once we had selected all these parameters we moved on to one of the most crucial steps of this analysis, namely the choice of resolution. Resolution is a numerical parameter that directly influences the number and size of the groups (clusters) of cells that will be identified. This parameter regulates the "sensitivity" of the clustering algorithm (typically the Leiden algorithm, predefined in Seurat) in dividing cells into subgroups. For this analysis we used a resolution range from 0 to 1.5. Low resolution values (for example, between 0.2 and 0.5) tend to produce fewer but larger clusters. This means that the algorithm will be less likely to separate cells that have minimal gene expression differences, grouping them into larger, more general populations. It is useful for identifying the main cell types present. High resolution values (for example, between 0.8 and 1.5, or even higher) tend to generate a greater number of smaller, more specific clusters. With higher resolution, the algorithm becomes more sensitive to small variations in gene expression, allowing the identification of rare cell subtypes or transient cell states within larger populations. However, an excessively high value can lead to "over-clusterization", artificially dividing cell populations that are actually homogeneous. The choice of optimal resolution is often an interactive process and is based on biological understanding of the system under study, visual inspection of the results (Figure 17), and sometimes analysis of the expression of known marker genes. Based on various tests and subsequent clustering evaluations at different resolutions, we opted to use a resolution of 0.4. This value was selected because higher resolutions compromised the stability and internal cohesion of the clusters (or caused cluster fragmentation) and did not lead to a clear separation between cell populations, thereby distorting the biological interpretation of the results. Using a resolution of 0.4, 24 distinct clusters were identified.

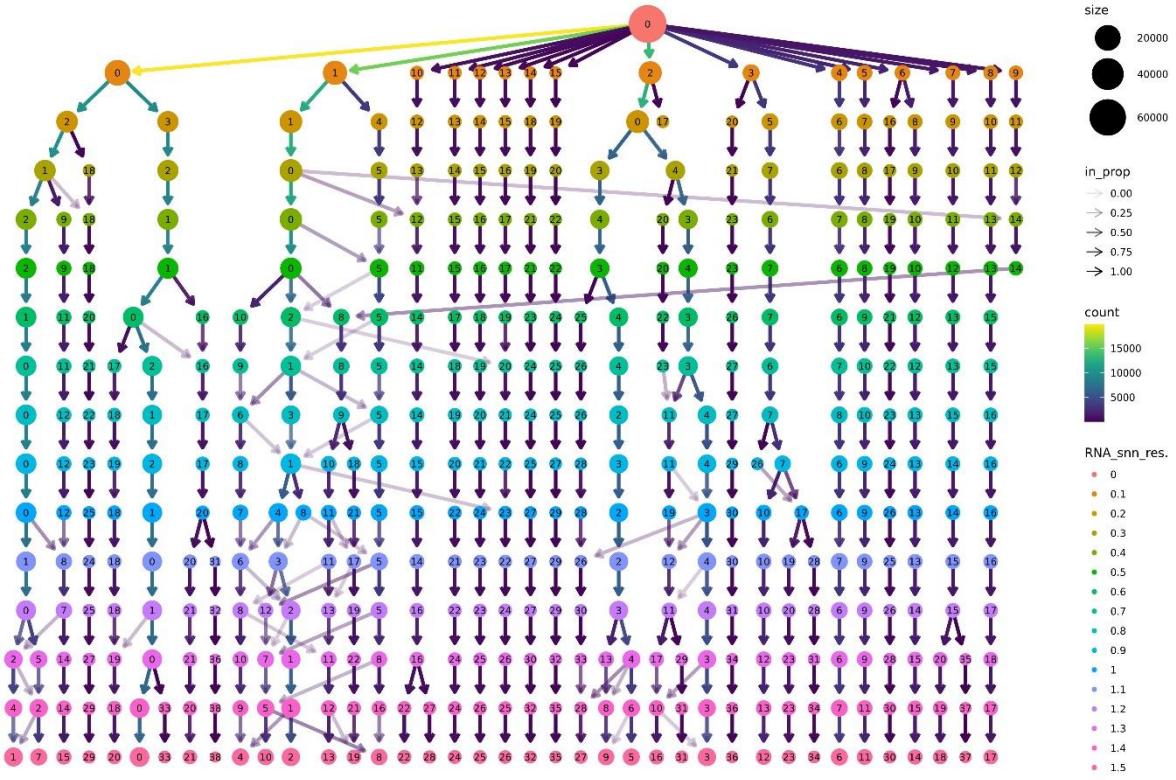


Figure 17: Selection of the optimal resolution for clustering using the Leiden algorithm. *The tree diagram (clustree) shows the evolution of clusters as the resolution parameter (indicated as RNA_snn_res) varies from 0 to 1.5. Each node represents a cluster. The size of each cluster is proportional to the number of cells it contains (size). The main color of each node reflects the number of cells inside it (count). The arrows between layers show how clusters at lower resolutions break down into subclusters as the resolution parameter increases.*

The spatial location of these 24 clusters can be seen via UMAP (Figures 18). The spatial distribution reflects transcriptional similarity between cells: cells with similar gene expression profiles cluster in the same cluster. In this view each cluster is represented by a different color.

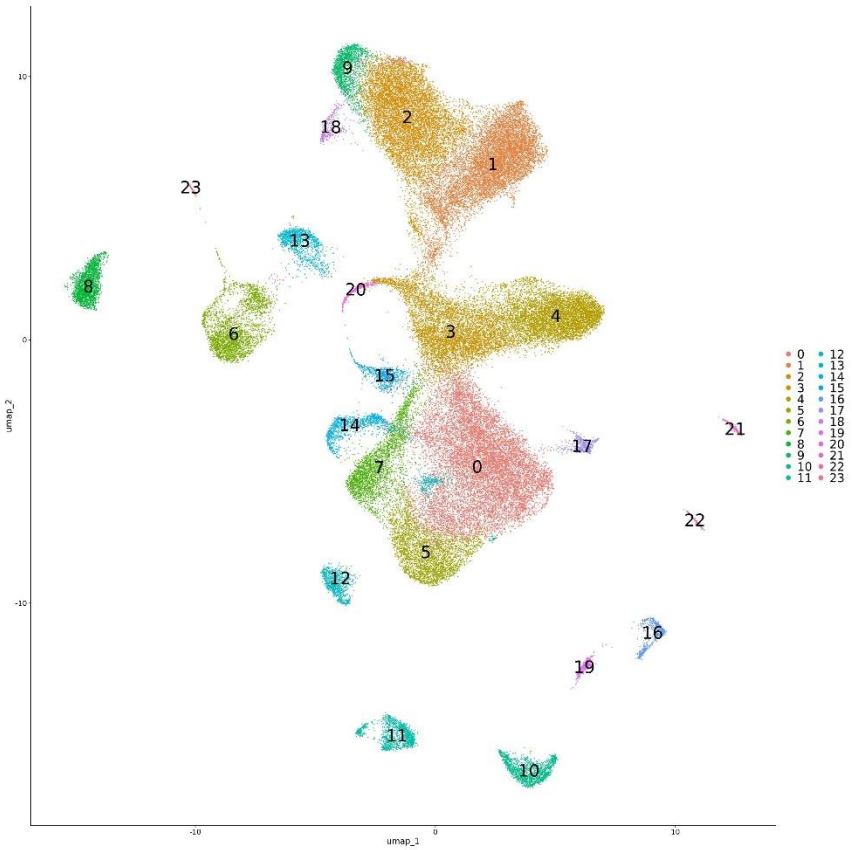


Figure 18: UMAP visualization. *The two-dimensional graph shows the UMAP (Uniform Manifold Approximation and Projection) projection of cells, where each point represents a single cell and the colors distinguish the 24 clusters identified using the Leiden clustering algorithm. The spatial distribution reflects the transcriptional similarity between cells: cells with similar gene expression profiles cluster in the same cluster. The clusters are numbered from 0 to 23 and represent distinct cell populations with specific transcriptional characteristics.*

The next step was to identify and then name each clusterconseidering the differential expressed genes (DEGs) among the 24 clusters.

The DEGs were found comparing the expression of the highly variable 3000 genes in each cluster versus all the others and showed by using heatmap visualization (Figure 19).

The objective of DEGs analysis is to find marker genes exclusively expressed in a cluster and not in the others. If two clusters express the same marker gene and are spatially close, they are grouped together. In particular, clusters 3, 4 and 20 were identified with a single name, going from 24 clusters to 21.

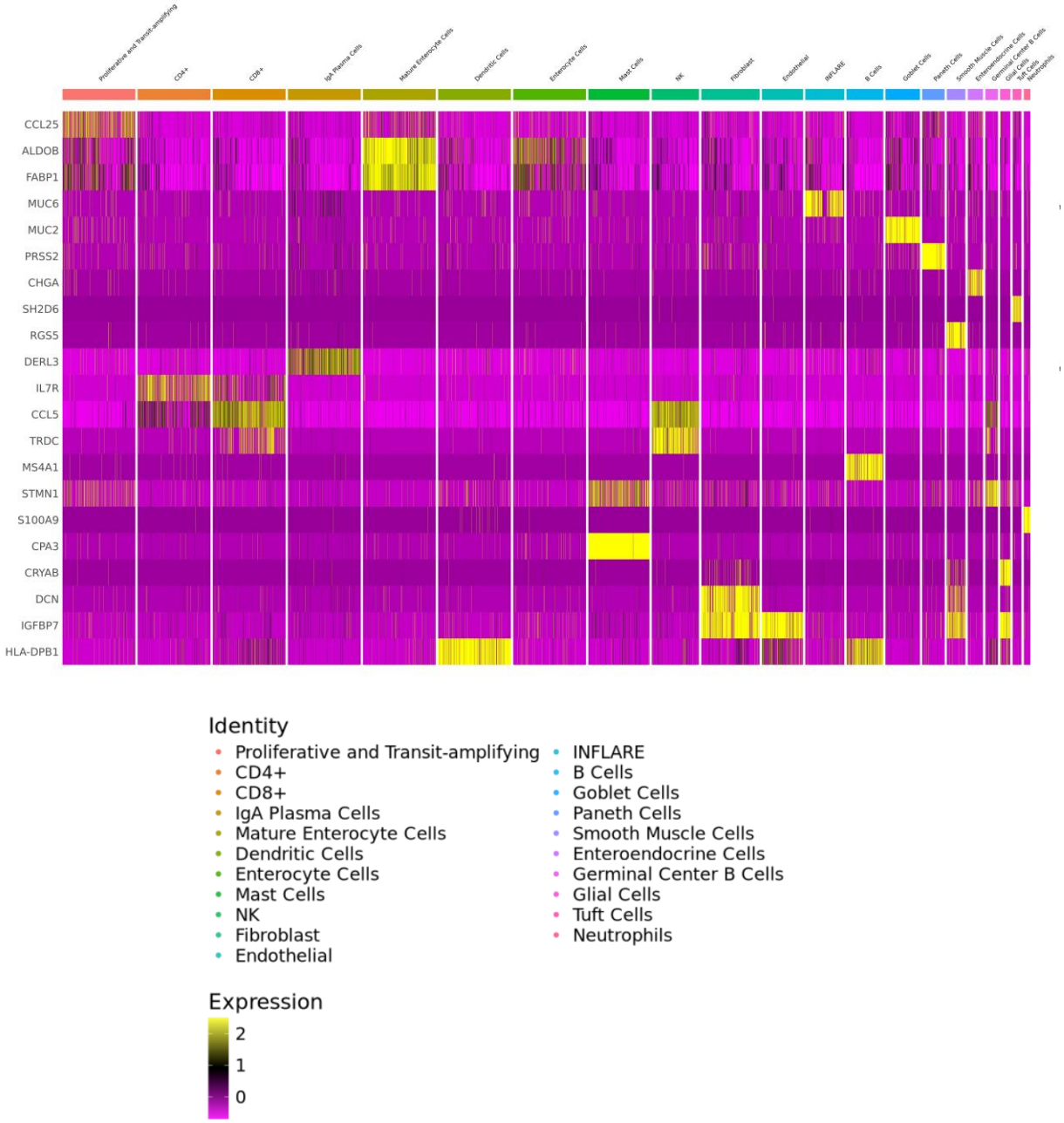


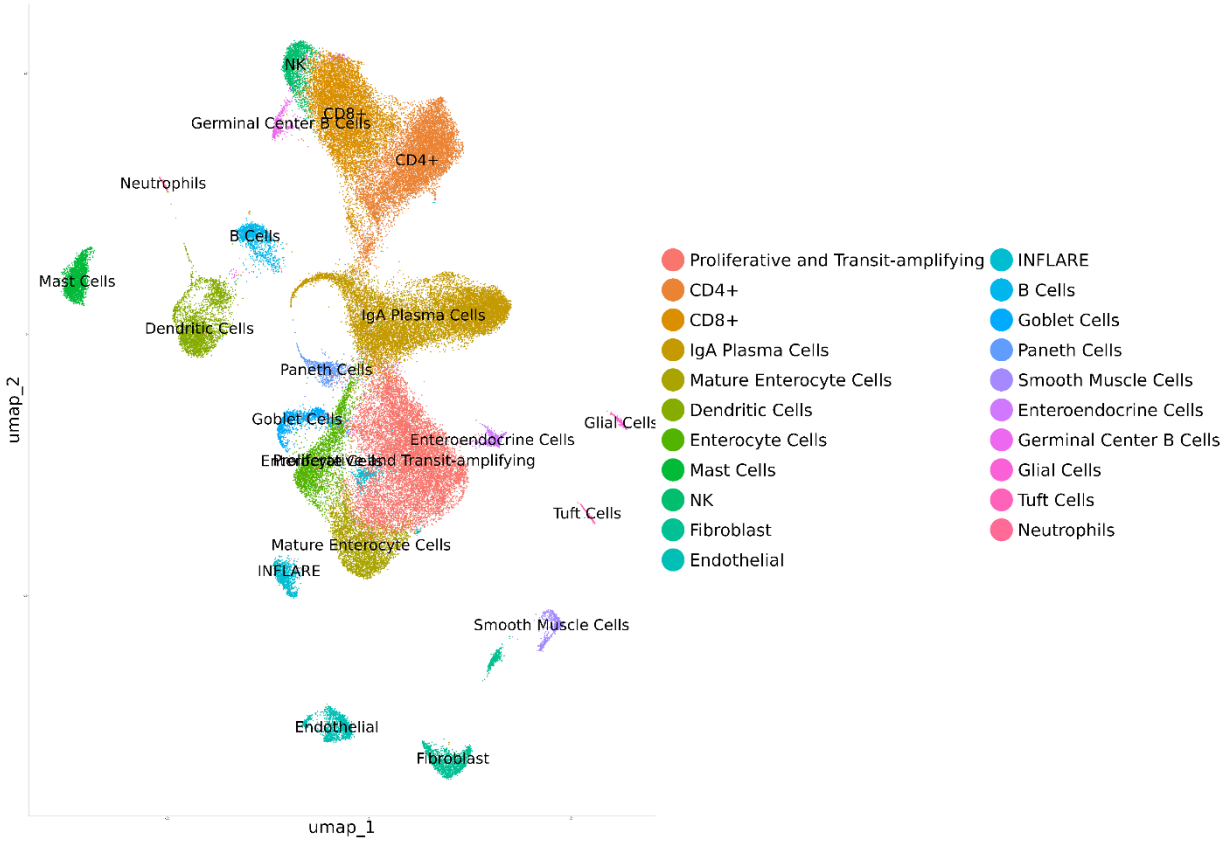
Figure 19: Heatmap of marker gene expression. *The heatmap shows the expression levels of the main marker genes (rows) across different cell clusters (columns). The colors represent the normalized expression intensity: magenta indicates low expression, yellow indicates high expression.*

Each marker gene is associated with a specific cell type and by using the scientific literature all the 21 cell types were identified. In detail, we recognized:

- Epithelial cells: Proliferative and Transit-amplifying, Mature Enterocyte Cells, Enterocyte Cells, INFLARE, Goblet Cells, Paneth Cells, Enteroendocrine Cells, Tuft Cells.

- Immune cells: CD4+, CD8+, IgA Plasma Cells, Dendritic Cells, Mast Cells, NK, B Cells, Germinal Center B Cells, Glial Cells, Neutrophils.
- Mesenchymal and Endothelial cells: Smooth Muscle Cells, Fibroblast, Endothelial Cells.

The 21 clusters were found in all 10 patients, regardless of the condition (Figure 20). However, each cell type is present in different percentages among the enrolled patients (Figure 21).



Figures 20: Annotation of cell types. *The UMAP projection shows the 21 cell clusters identified and annotated according to their specific transcriptional markers. Each cell type is represented by a specific color as indicated in the legend, highlighting the complex cellular architecture of the intestinal tissue analyzed.*

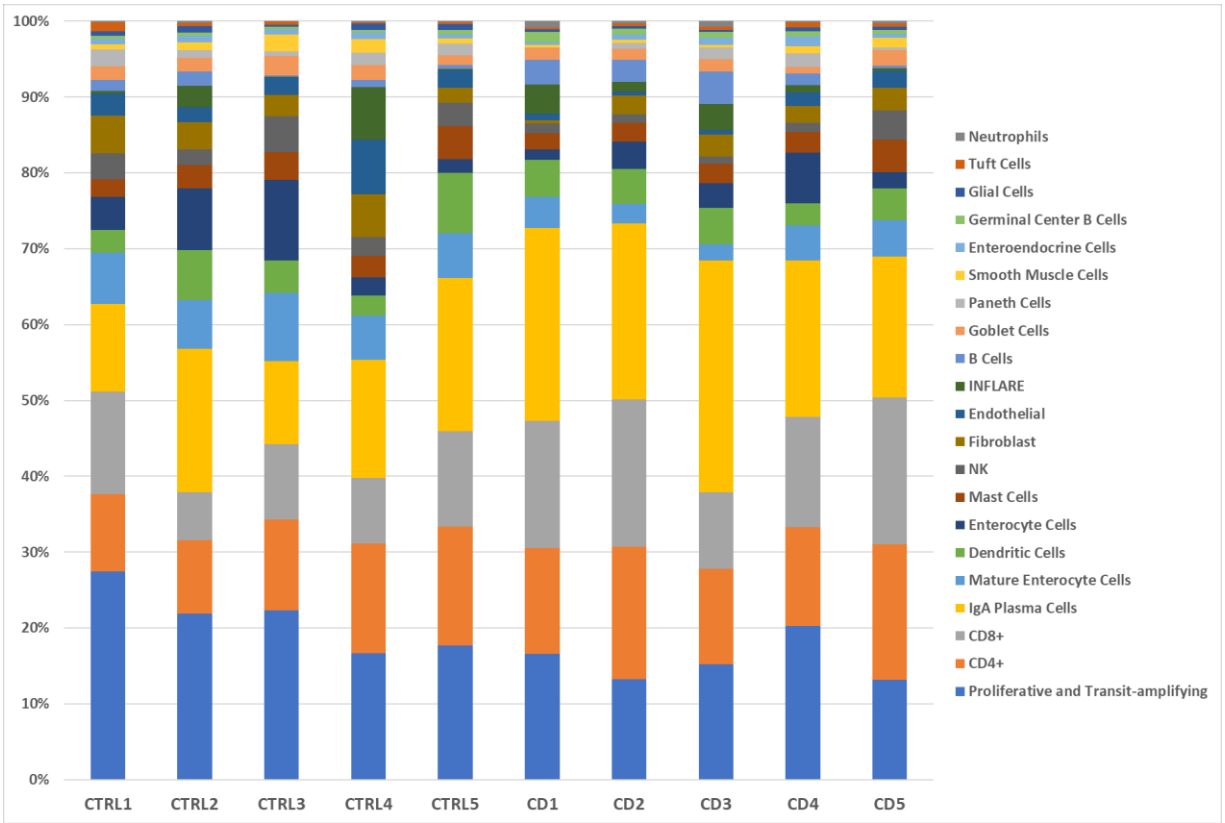


Figure 21: Differential cellular composition between control samples (CTRL) and patients with celiac disease (CD). *The bar plot shows the relative proportion of different cell types in each sample analyzed.*

4.3.1 Epithelial cells compartment

Eight out of 21 clusters belong to the epithelial cell compartment and express the EPCAM marker gene (Figure 22). The EPCAM gene was found in approximately 23.561/71.397 single cells (33%).

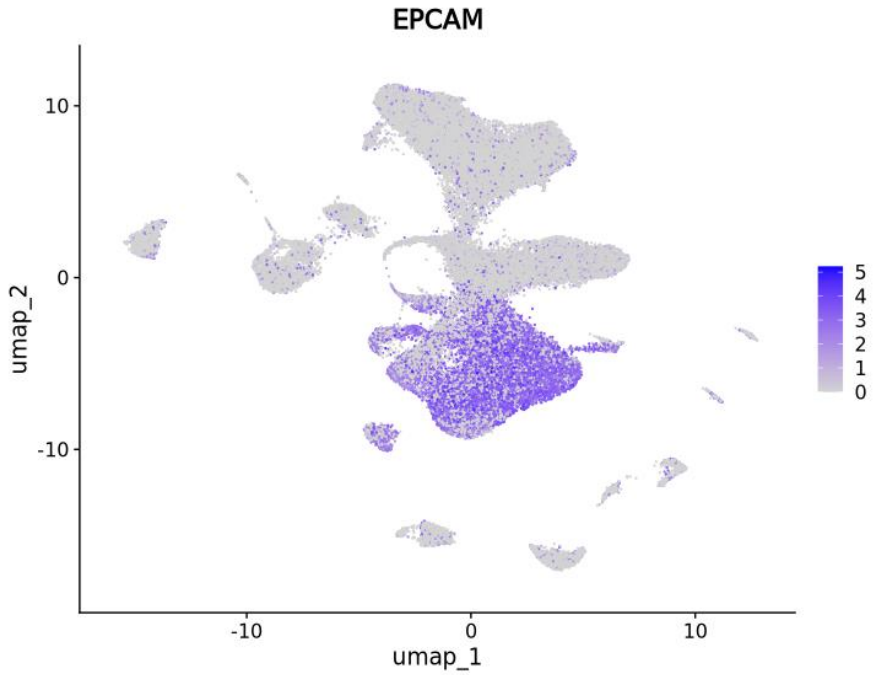


Figure 22: Epithelial cells marker gene. *UMAP (Uniform Manifold Approximation and Projection) shows how epithelial cells express the EPCAM marker genes.*

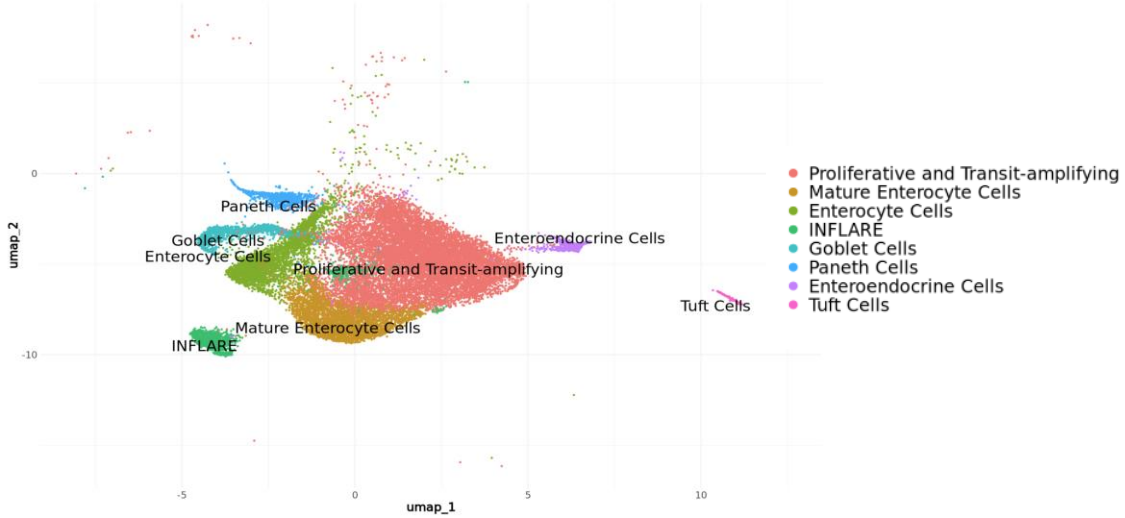
The epithelial cell compartment includes:

- Proliferative and Transit-amplifying are characterized by the expression of two marker genes: MT1G and CCL25. This group represents the largest cell type 55% of the epithelial compartment and includes the stem cell compartment and its differentiating progeny (Figure 23 A, B, C).
- Mature Enterocyte Cells are characterized by the expression of two marker genes: ALDOB, FABP1. This group represents 15% of the epithelial cell compartment. These cells are considered to be in a more advanced state of differentiation, metabolically active, responsible for the absorption of nutrients and located in the upper part of the villus (FitzPatrick et al. 2025) (Figure 23 A, B, C).
- Enterocyte Cells are characterized by the expression of the marker gene: ALDOB and are 12% of the epithelial cell compartment. These cells are in the

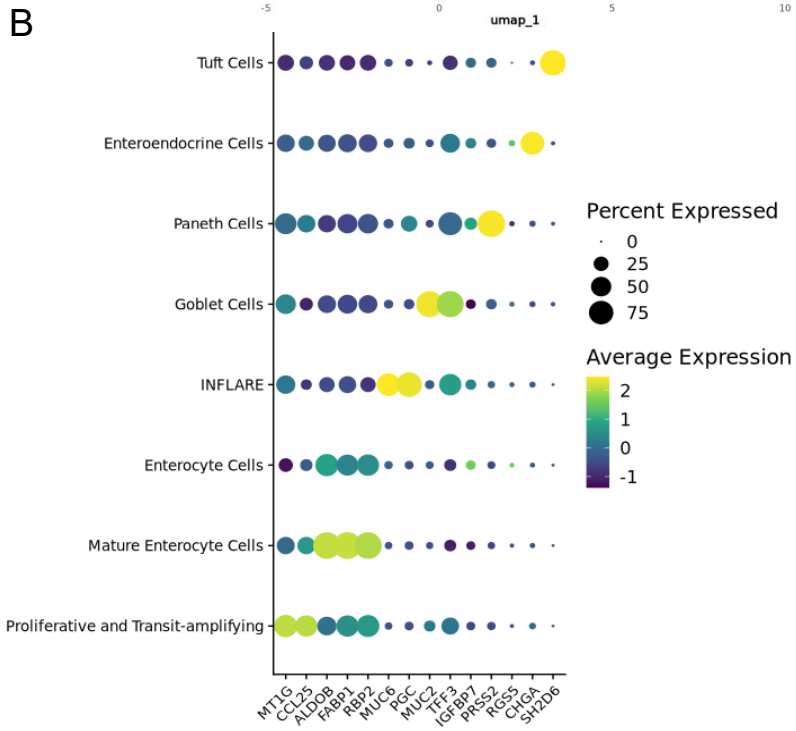
lower part of the villus and are still maturing, they have not fully developed all the structures and metabolic pathways necessary for optimal absorption (Figure 23 A, B, C).

- INFLARE express the marker genes MUC6 and PGC and represent 6% of the epithelial cell population. This cell population has been recently discovered and expressed programs that may contribute to chronic intestinal inflammation (Oliver et al. 2024).. This population is associated with an inflammatory condition (Figure 23 A, B, C).
- Goblet Cells are characterized by the expression of two marker genes: MUC2 and TFF3. It represents 5% of the epithelial cell population. They form the protective layer of mucus that lines the inner surface of the intestine but also are intimately linked to the immune system (Gustafsson and Johansson 2022)(Figure 23 A, B, C).
- Paneth cells express the marker gene PRSS2 and represent 3% of the epithelial cell population. They secrete large quantities of antimicrobial peptides and other signaling molecules, therefore they have a role in immune defense and support for stem cells (Figure 23 A, B, C).
- Enteroendocrine cells express the marker gene CHGA and are 2% of the epithelial cell compartment. They act as chemical sensors so they detect and regulate hormone levels (Figure 23 A, B, C).
- Tuft cells are characterized by the expression of the marker gene SH2D6 and represent 2% of the epithelial cell population. These are chemosensory cells: they detect the presence of parasites or specific luminary signals and, in response, release large amounts of Type 2 cytokines (Figure 23 A, B, C).

A


umap_2
umap_1
Paneth Cells
Goblet Cells
Enterocyte Cells
Proliferative and Transit-amplifying
Enteroendocrine Cells
Tuft Cells
Mature Enterocyte Cells
INFLARE
Proliferative and Transit-amplifying
Mature Enterocyte Cells
Enterocyte Cells
INFLARE
Goblet Cells
Paneth Cells
Enteroendocrine Cells
Tuft Cells

B


Tuft Cells
Enteroendocrine Cells
Paneth Cells
Goblet Cells
INFLARE
Enterocyte Cells
Mature Enterocyte Cells
Proliferative and Transit-amplifying
MT1G
CCL25
ALDOB
FABP1
RBP2
MUC6
PGC
MUC2
TFF3
IGFBP7
PRSS2
RGS5
CHGA
SH2D6
Percent Expressed
0
25
50
75
Average Expression
2
1
0
-1

C

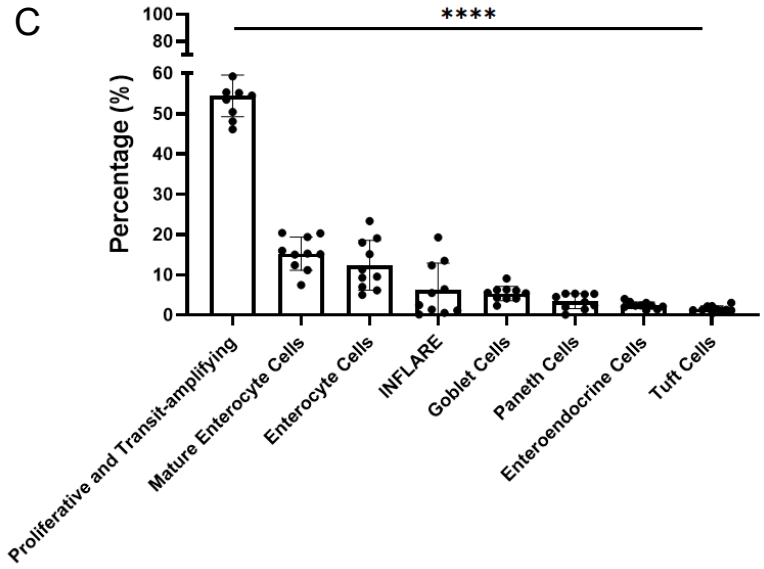
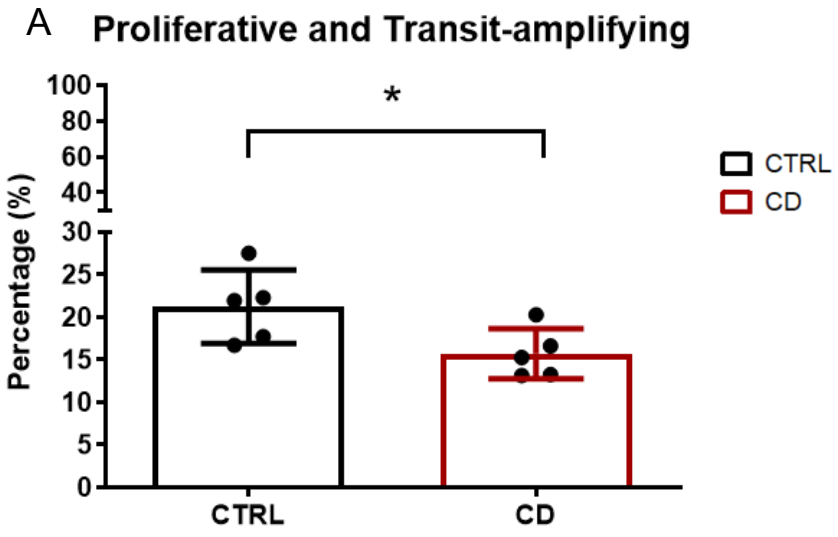

****
Percentage (%)
100
80
60
50
40
30
20
10
0
Proliferative and Transit-amplifying
Mature Enterocyte Cells
Enterocyte Cells
INFLARE
Goblet Cells
Paneth Cells
Enteroendocrine Cells
Tuft Cells

Figure 23: Epithelial cells. *(A) UMAP (Uniform Manifold Approximation and Projection) shows the spatial localization of intestinal epithelial cell types. (B) The dot plot shows the expression of the selected marker genes in the different epithelial cell types. The size of the dots indicates the percentage of cells that express the gene (Percent Expressed), while the color represents the average level of expression (Average Expression) according to the color scale. (C) The bar plot shows the average percentage (± standard deviation) of the epithelial cell types. Statistical analysis using the Mann-Whitney test shows a significant difference in proliferative and transit-amplifying cells number compared to all other cell types (****p-value ≤ 0.0001).*

A comparative analysis (CD vs CTRL) has been performed. We observed that both proliferative and transit-amplifying (p-value = 0.03) and mature enterocytes cells (p−value=0.008) are significantly more abundant in controls than celiac patients (Figure 24 A, B). The reduced number of mature enterocytes in celiac patients mirrors the pathology. Celiac disease is characterized by villous atrophy, which leads to a clear reduction in the number of functional epithelial cells. This atrophy is the direct consequence of the massive destruction of mature cells and the compensatory failure of proliferative cells in the crypts. Although the crypt cells increase in number (crypt hyperplasia), they ultimately fail to complete proper differentiation into mature enterocytes due to the persistent inflammatory microenvironment. Moreover, our data showed also a reduced number of proliferative and transit-amplifying cells revealing a lack of cells able to differentiate in mature enterocytes.



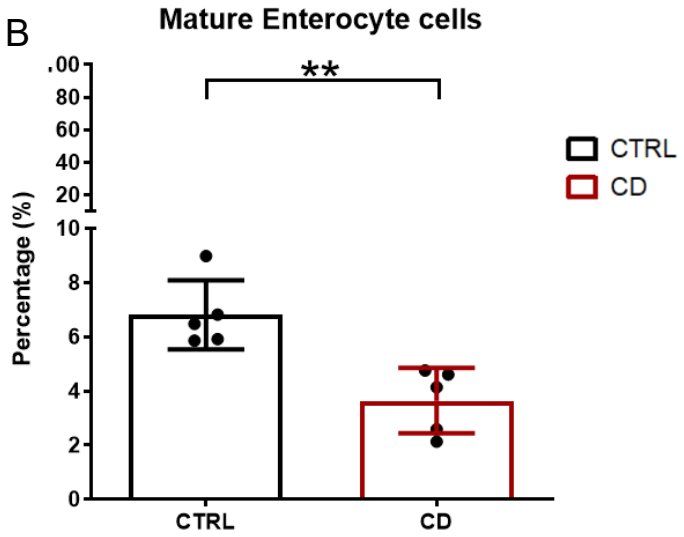


Figure 24: Percentage of proliferative and transit-amplifying cells and mature enterocyte cells. *The bar plot shows the mean percentage (±standard deviation) of the following epithelial cell (A-B) populations by comparing control subjects (CTRL) and celiac patients (CD). Statistical analysis (Mann-Whitney test) found significant proportion in: (A) proliferative and transit-amplifying cells (*p-value= 0.03, (B). Mature Enterocyte Cells (**p-value= 0.008).*

The other cell populations (Enterocyte, INFLARE, Goblet, Paneth, Enteroendocrine and Tuft cells) on the other hand are similarly present in both celiacs and controls without showing significant differences towards either group (Figure 25).

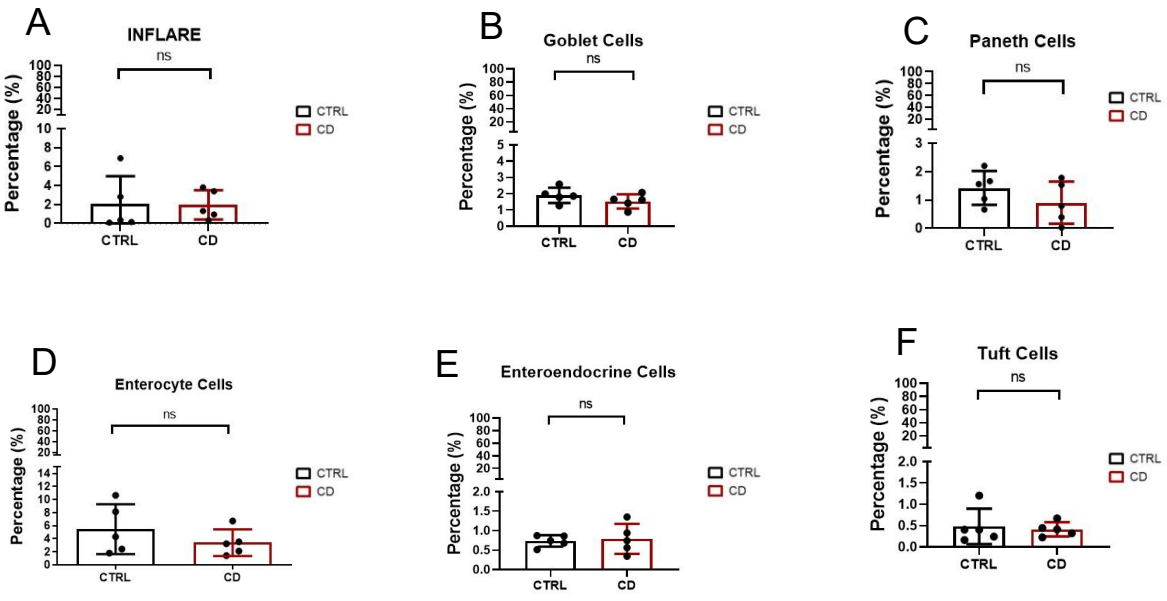


Figure 25: Percentage of INFLARE, Goblet, Paneth, Enterocyte, Enteroendocrine and Tuft cells. *The bar plot displays the mean percentage (± standard deviation) of these cell types(A-F) in control samples and celiac disease patients. Statistical analysis was performed using the Mann-Whitney test, which did not reveal any significant difference between the two groups (ns).*

Finally, an analysis of differentially expressed genes within the epithelial cell compartment was conducted. We obtained 384 genes differentially expressed between celiacs and controls, of which 63 genes up-regulated (red dots), 112 genes down-regulated (blue dots) and the 15 genes most significant are those indicated by the name in the Volcano Plot (Figures 26). The most upregulated genes in celiac patients are associated with immunoglobulins (JCHAIN, IGHA1/2, IGHM) and with Class I MHC molecules (B2M, HLA-B). Conversely, the downregulated genes are involved in lipid metabolism (APOA1/4), carbohydrate metabolism (ALDOB), and Vitamin A transport (RBP2). The results obtained from the differential gene expression analysis are highly consistent with the pathophysiology of celiac disease. Indeed, in the gut of celiac patients there is an increase of immunoglobulins responsible for producing the CD-related autoantibodies (anti-transglutaminase 2). Moreover, the intense persistent activation of the immune system causes structural and functional damage to the intestine, which in turn leads to severe malabsorption of key nutrients (fats and fat-soluble vitamins) (Dotsenko et al. 2021). Finally, the up-regulated Class I MHC molecules in the epithelial cells confirmed the CD-related intestinal inflammation.

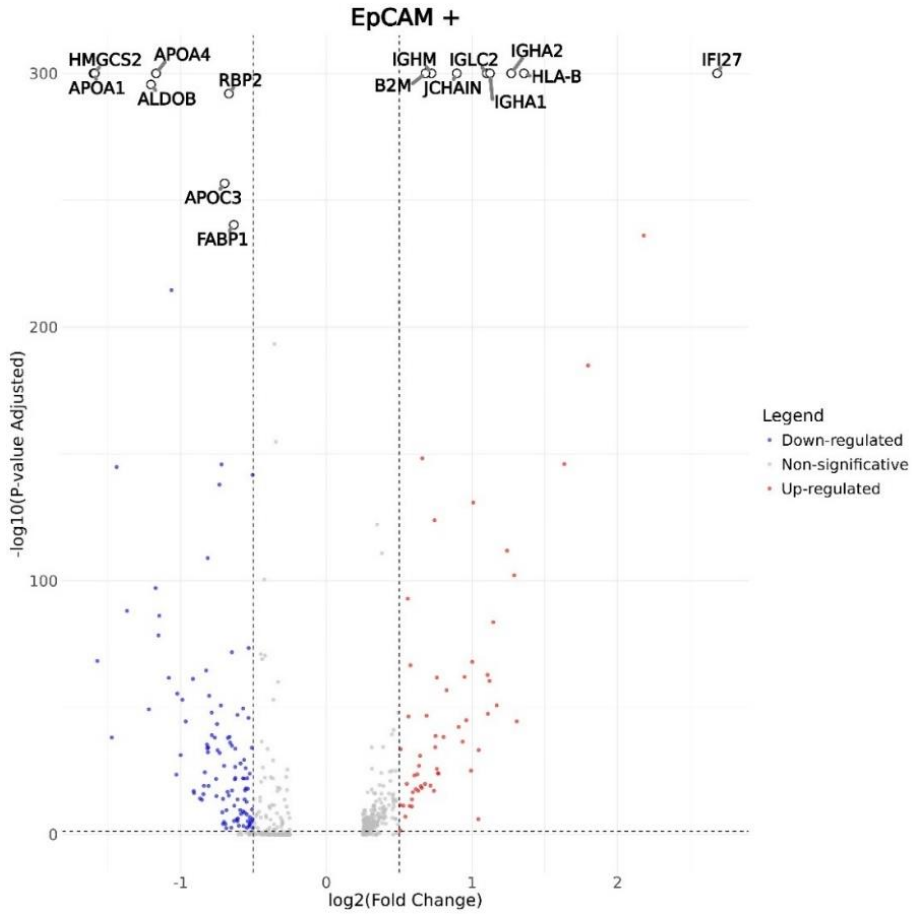


Figure 26: Differential gene expression analysis. *The volcano plot shows genes differentially expressed by comparing epithelial cells from celiac patients versus controls. The x-axis represents log2 (Fold Change) and the y-axis -log10 (adjusted p-value). Red dots indicate significantly up-regulated genes in celiac patients, blue dots indicate down-regulated ones, while gray dots represent genes that are not significantly differentiated. Vertical dotted lines indicate thresholds of biological significance.*

4.3.2 Immune cells compartment

Nine out of 21 clusters are part of the immune cells compartment that include both T and B lymphocytes. In particular the T lymphocytes express the marker gene PTPRC while the B lymphocytes express the marker gene MZB1 (Figure 27). These genes were found in approximately 43.552/71.397 single cells (61%).

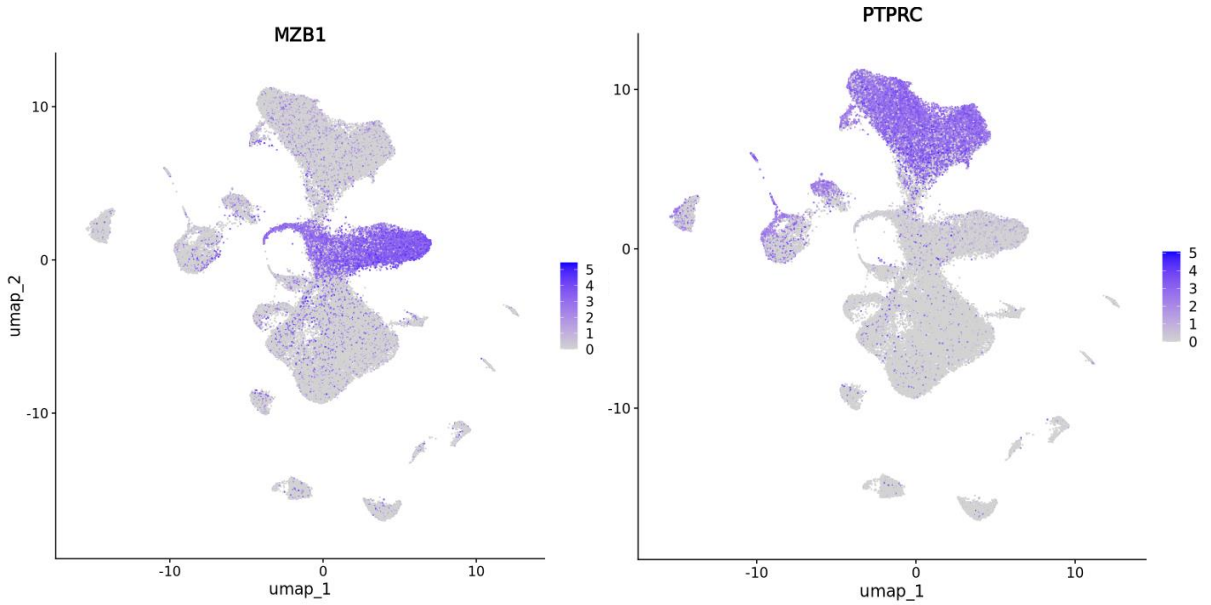


Figure 27: IgA Pasma cells and Immune cells marker gene. *UMAP (Uniform Manifold Approximation and Projection) shows how plasma cell IgA expresses the marker gene MZB1 while immune cells express the marker gene PTPRC.*

The immune cells compartment includes:

- CD4+ cells, characterized by the expression of marker gene IL7R. These cells represent 22% of immune cells compartment. They are also as known T Helpers and are involved in supporting humoral immune responses by promoting B cell proliferation and maturation, germinal center (GC) response, and high-affinity antibody production (Sun et al. 2023) (Figure 28A, B, C).
- CD8+ cells, expressing the marker gene CCL5. This group represents 21% of immune cells population. These are the "killer" cells: play critical roles in fighting against intracellular pathogens (Sun et al. 2023) and contribute to cell turnover and repair (Figure 28A, B, C).
- IgA Plasma Cells, characterized by the expression of the marker genes: DERL3, MZB1 and IGLV2-23. These cells represent 32% of the immune cells

compartment and are the largest immune cell type. They are terminally differentiated B lymphocytes. Their main function is to produce and secrete huge amounts of Immunoglobulin A (SIgA) that are released into the intestinal lumen, where they neutralize toxins and pathogens and prevent commensal bacteria from adhering to the epithelium. They are the first line of defense of the intestine (Mantis, Rol, and Corthésy 2011) (Figure 28A, B, C).

- Dendritic cells, expressing the marker genes HLA-DPB1. This group represents 8% of immune cell population. They act as sentinels, process antigens and present them to T lymphocytes, triggering or tolerating an immune response (Figure 28A, B, C).
- Mast cells, characterized by the expression of the marker genes CPA3 and TPSAB1. This cell type represents 6% of immune cell compartment. They are activated in response to allergens or parasitic infections, rapidly releasing the granules to increase vascular permeability and attract other immune cells (Figure 28A, B, C).
- NK, expressing the marker gene TRDC. These cells are 5% of immune cell population and are part of the innate immune system. They spontaneously detect and destroy infected cells or cancer cells without requiring prior sensitization (they do not require antigen presentation) (Figure 28A, B, C).
- B cells, expressing the marker gene MS4A1. This group are 3% of immune cell population. They recognize antigens and, once activated by CD4+ T lymphocytes, transform into plasma cells or memory B cells (Figure 28A, B, C).
- Germinal center B cells, expressing the marker gene STMN1. This group represent 1% of immune cell compartment. These are B lymphocytes that undergo a process of affinity maturation within the germinal centers (secondary structures of the lymph nodes). This process generates antibodies with an increased ability to bind antigen (Figure 28A, B, C).
- Neutrophils, expressing the marker gene S100A9. These cells represent 1% of immune cell compartment. They are the most abundant leukocytes in the circulation, and have been regarded as first line of defense in the innate arm of the immune system (Rosales 2018) (Figure 28 A, B, C).

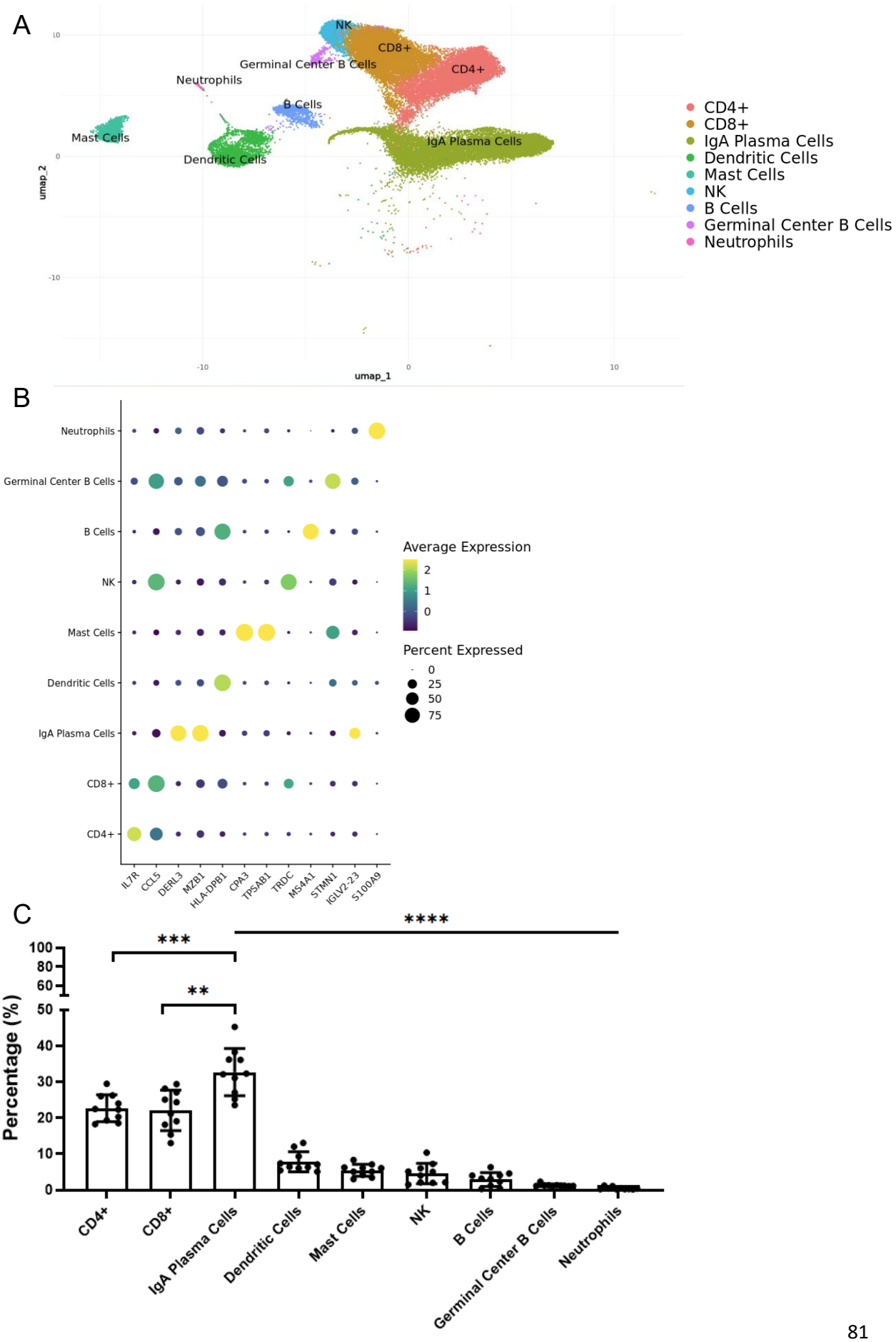

A
NK
CD8+
Germinal Center B Cells
CD4+
Neutrophils
B Cells
Mast Cells
IgA Plasma Cells
Dendritic Cells
umap_2
umap_1
CD4+
CD8+
IgA Plasma Cells
Dendritic Cells
Mast Cells
NK
B Cells
Germinal Center B Cells
Neutrophils
B
Neutrophils
Germinal Center B Cells
B Cells
NK
Mast Cells
Dendritic Cells
IgA Plasma Cells
CD8+
CD4+
IL7R
CCL5
DERL3
MZB1
HLA-DPB1
CPA3
TPSAB1
TRDC
MS4A1
STMN1
IGLV2-23
S100A9
Average Expression
2
1
0
Percent Expressed
0
25
50
75
C
****
***
**
Percentage (%)
100
80
60
50
40
30
20
10
0
CD4+
CD8+
IgA Plasma Cells
Dendritic Cells
Mast Cells
NK
B Cells
Germinal Center B Cells
Neutrophils

Figure 28: Immune cells. *(A) UMAP (Uniform Manifold Approximation and Projection) shows the spatial localization of immune cells in CTRL and CD. (B) The dot plot shows the expression of the selected marker genes in the different epithelial cell types. The size of the dots indicates the percentage of cells that express the gene (Percent Expressed), while the color represents the average level of expression (Average Expression) according to the color scale. (C) The bar plot shows the mean percentage (±standard deviation) of the immune cell compartment in celiac patients and control. Statistical analysis (Mann-Whitney Test) reveals a significant increase in IgA Plasma Cells in celiac patients compared to controls, in contrast to other populations. In particular, the abundance of IgA Plasma Cells was significantly different compared to CD4+ (***p-value= 0.0005) and CD8+ (**p-value= 0.001). Furthermore, IgA Plasma Cells show a statistically very significant difference compared to all remaining immune cells analyzed (****p-value ≤ 0.0001 in all binary comparisons).*

A comparative analysis (CD vs CTRL) has been performed. We observed that the IgA Plasma Cells (p-value= 0.03), CD8+ (p-value= 0.03), Germinal Center B Cells (p-value= 0.01) and Neutrophils (p-value= 0.03) are significantly more abundant in celiac patients vs controls (Figure 29 A-D). The results obtained from these statistical analyzes are in line with what happens during the pathogenesis of celiac disease. The increased presence of IgA plasma cells is crucial for the humoral response, as these cells mass produce IgA antibodies that bind specifically to tissue transglutaminase (tTG), a distinctive event that contributes to chronic inflammation. Upstream of this antibody hyperproduction, an increase in Germinative Center B cells is observed, evidence of a chronically overactive germinative center that matures and selects anti-tTG IgA plasma cells. In parallel, inflammation and tissue damage are triggered and maintained by the cell-mediated response, dominated by CD8+ T lymphocytes, which mediate the direct cellular destruction of enterocytes, leading to inflammation and atrophy of the intestinal villi typical of the disease. The intensity and activity of intestinal inflammation are also reflected by the increase in Neutrophils, cells that are recruited at the site of damage, further amplifying the destruction of the intestinal epithelium.

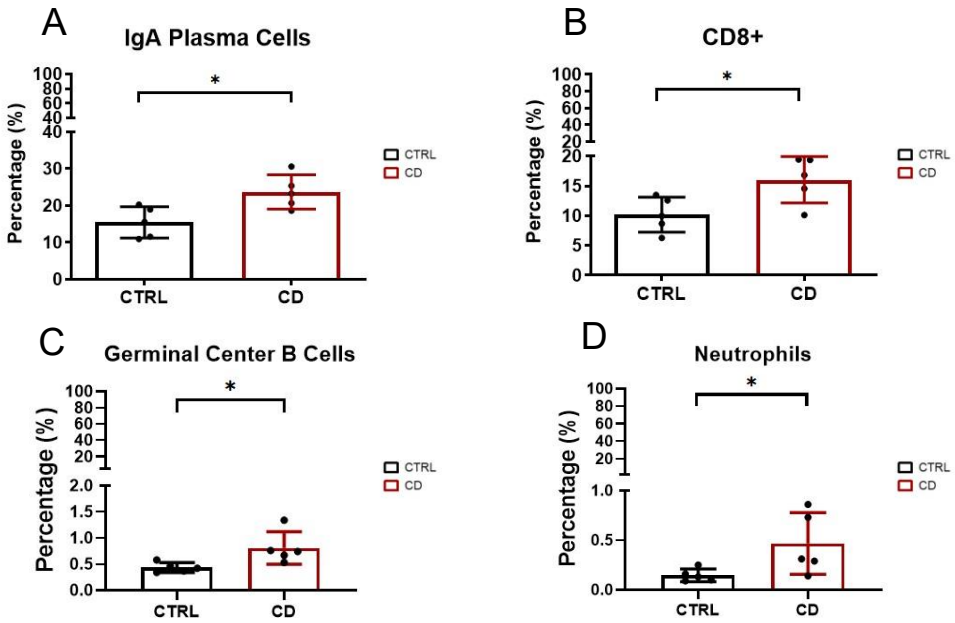


Figure 29: Percentage of IgA Plasma Cells, CD8+, Germinal Center B Cells, Neutrophils. *The bar plot shows the mean percentage (±standard deviation) of the following epithelial cell (A-D) populations by comparing control subjects (CTRL) and celiac patients (CD). Statistical analysis (Mann-Whitney test) found significant proportion in: (A) IgA Plasma Cell (*p-value= 0.03, (B). CD8+ (*p-value= 0.03), (C) Germinal Center B Cells (*p-value= 0.01), (D) Neutrophils (*p-value= 0.03).*

The other cell populations (CD4+, Dendritic cells, Mast cells, NK, B cells, Glial cells) on the other hand are similarly present in both celiacs and controls without showing significant abundance towards either group (Figure 30).

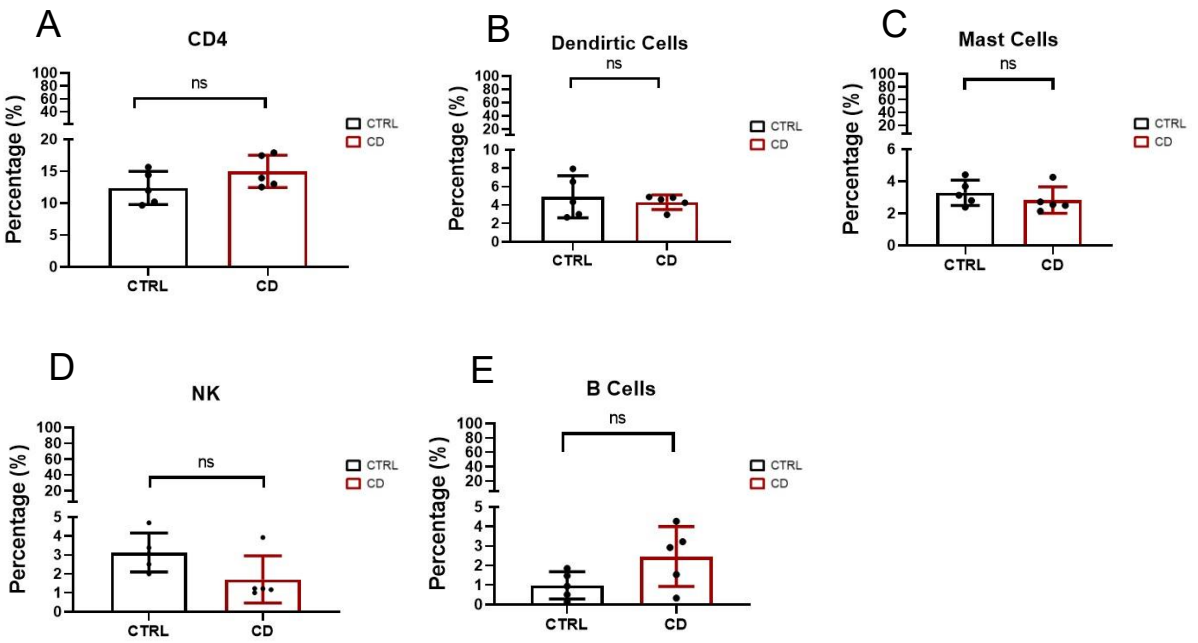


Figure 30: Percentage of CD4+, Dendritic cells, Mast cells, NK, B cells. *The bar plot displays the mean percentage (± standard deviation) of these cell types(A-E) in control samples and celiac disease patients. Statistical analysis was performed using the Mann-Whitney test, which did not reveal any significant difference between the two groups (ns).*

Finally, the analysis of differentially expressed genes between the two groups of patients was carried out. We obtained 387 genes differentially expressed between celiacs and controls, of which 60 genes up-regulated (red), 95 genes down-regulated (blue) and the 15 genes most significant are those indicated by the name. The Volcano Plot (Figure 31) shows that down-regulated genes (blue) in celiacs are associated with nutrient uptake and metabolism (ALDOB, APOC3, APOA4, APOA1) while up-regulated genes (red) are associated with the immune response (STAT1, IGHM, IGHG1). These gene expression changes therefore reflect the complex interaction between intestinal tissue damage, malabsorption, inflammation and autoimmune response that defines celiac disease.

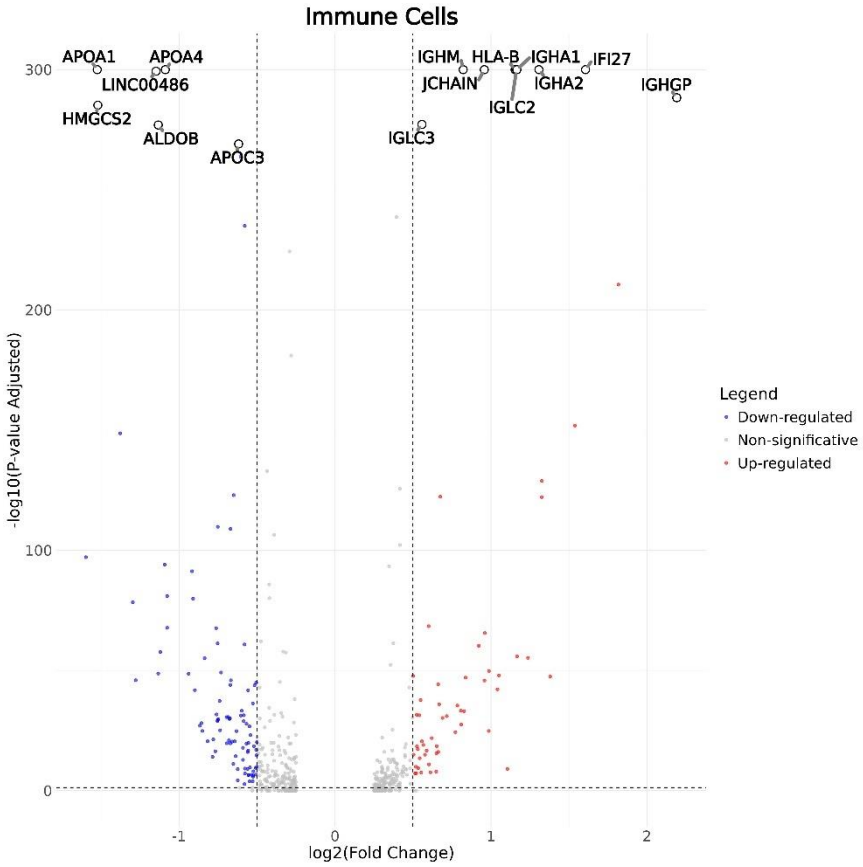


Figure 31: Differential gene expression analysis. *The volcano plot shows genes differentially expressed by comparing immune cells from celiac patients versus controls. The x-axis represents log2 (Fold Change) and the y-axis -log10 (adjusted p-value). Red dots indicate significantly up-regulated genes in celiac patients, blue dots indicate down-regulated ones, while gray dots represent genes that are not significantly differentiated. Vertical dotted lines indicate thresholds of biological significance.*

4.3.3 Mesenchymal, Endothelial and Glial cells

The remaining 6% (4.284/71.397 single cells) of cells comprise:

- Mesenchymal cells are divided into two subfamilies: the fibroblasts and smooth muscle cells. Fibroblasts, express the marker genes DCN and CXCL14. These cells produce and secrete extracellular matrix (ECM) components, such as collagen, which provide structural support and are essential for wound repair and fibrosis. The smooth muscle cells express the marker RGS5 gene. They are the main component of the muscle layers. Their coordinated contraction and relaxation move food along the digestive tract and mix the intestinal contents.
- Endothelial cells express the marker gene IGFBP7. They form the inner wall of blood vessels and are essential for regulating blood flow and permeability.

Glial cells express the marker gene CRYAB. They are part of the Enteric Nervous System (ENS). Similar to astrocytes in the brain, they provide metabolic and structural support to intestinal neurons and are involved in regulating motility, secretion and modulation of local inflammation.

The Mann-Whitney statistical test was carried out which showed a significant increase in endothelial cells in the controls (*p-value=0.01) while there is no significant increase in smooth muscle cells (ns, p-value= 0.1), fibroblasts (ns, p-value= 0.2), glial cells (ns, p-value= 0.1) (Figures 32).

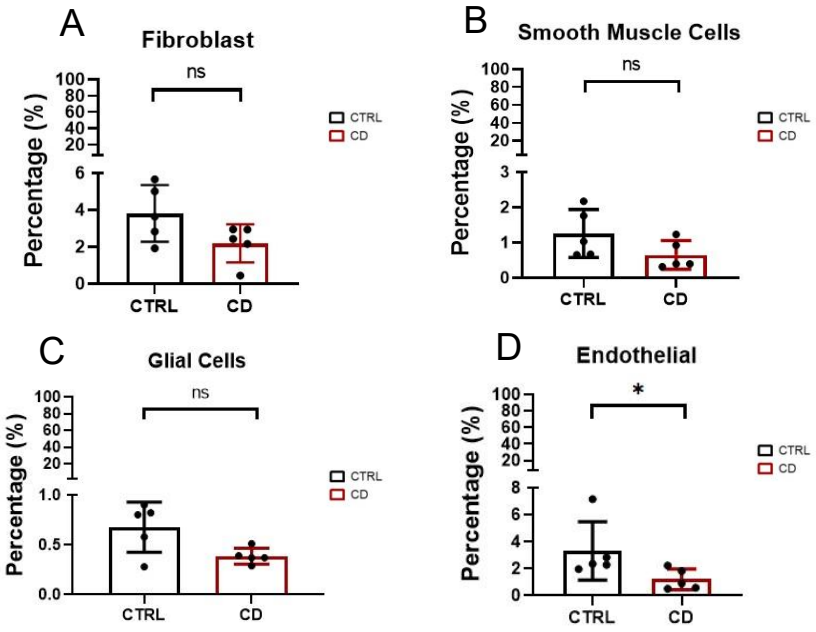


Figure 32: Percentage of Endothelial, Smooth Muscle, Fibroblast and Glial cells. *Bar plot shows the mean percentage (±standard deviation) of the following epithelial cell populations (A-C) by comparing control (CTRL) and celiac patients (CD). Statistical analysis (Mann-Whitney test) found no significant differences (ns) between the two groups for the following populations: (A) Fibroblast (ns, p-value= 0.2), (B) Smooth Muscle Cells (ns, p-value= 0.1) and (C) Glial Cells (ns, p-value= 0.1). The test instead showed that the following are significantly more expressed in controls than celiac: (C) Endothelial (*p-value= 0.01).*

This lower statistical presence of endothelial cells in the celiac, despite the presence of inflammation, is a clear effect of the massive structural destruction of the mucosa, where the reduction in tissue volume and the loss of the capillaries of the villi exceed the compensatory increase in angiogenesis at the level of the crypts.

An analysis of differentially expressed genes was carried out and in both populations, it was up regulated genes associated with immune response (IGHA1, IGHM, IGLC2) while down regulated ones were associated with barrier alteration and metabolism (ALDOB, APOA4, FABP1) in CD vs CTRL.

4.4 Cell-Cell communication analysis

In order to gain deeper insight into the intercellular signaling mechanisms, specifically ligand-receptor interactions, and determine the key cell populations involved, we conducted a cell-cell communication analysis. This analysis shows cell-cell communication evaluated on the basis of the number and the intensity (weight) of interaction.

The analysis was conducted on single cells data from:

- Group 1: all patients (CD and CTRL together).
- Group 2: celiac patients.
- Group 3: controls.

We first performed the analysis on group 1 in order to obtain an overview of the cellular communications that occur in the intestine, excluding the disease. As shown in Figure 33, there is a dense network of communication between all cell types. In particular, we observed the strongest interaction occurs between CD8+ and all the other cell types (Figure 33 B).

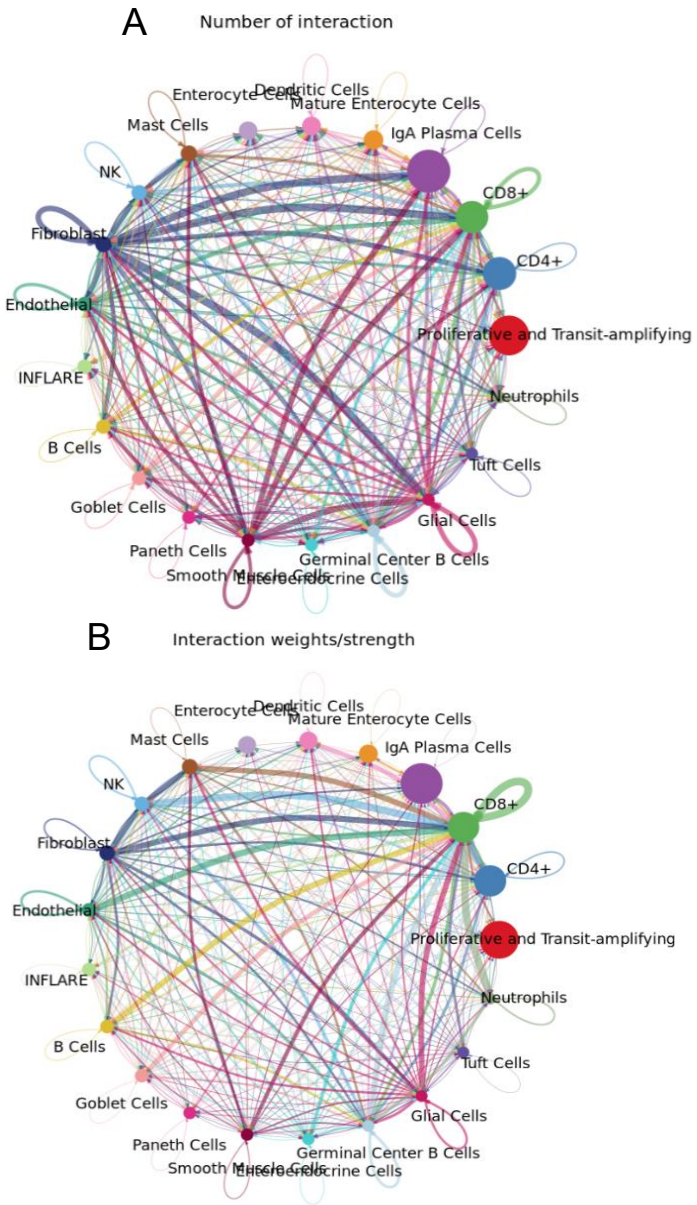


Figure 33: Analysis of cell-cell communications. *Network plots show intercellular interactions based on ligands and receptors. (A) The circle plot shows the number of interactions between the different cell types, where the thickness of the arrows is proportional to the number of interactions between the cell populations. (B) The circle plot shows the intensity (weight) of the interactions.*

Subsequently, the same analysis was performed on both group 2 and 3 in order to highlight differences between celiac patients and controls. High communication between CD8+ and the other cell types was observed in both celiac patients and controls. In celiac patients compared to controls, we found a greater communication strength among immune cells (Figures 34), confirming an overactivation of the immune response.

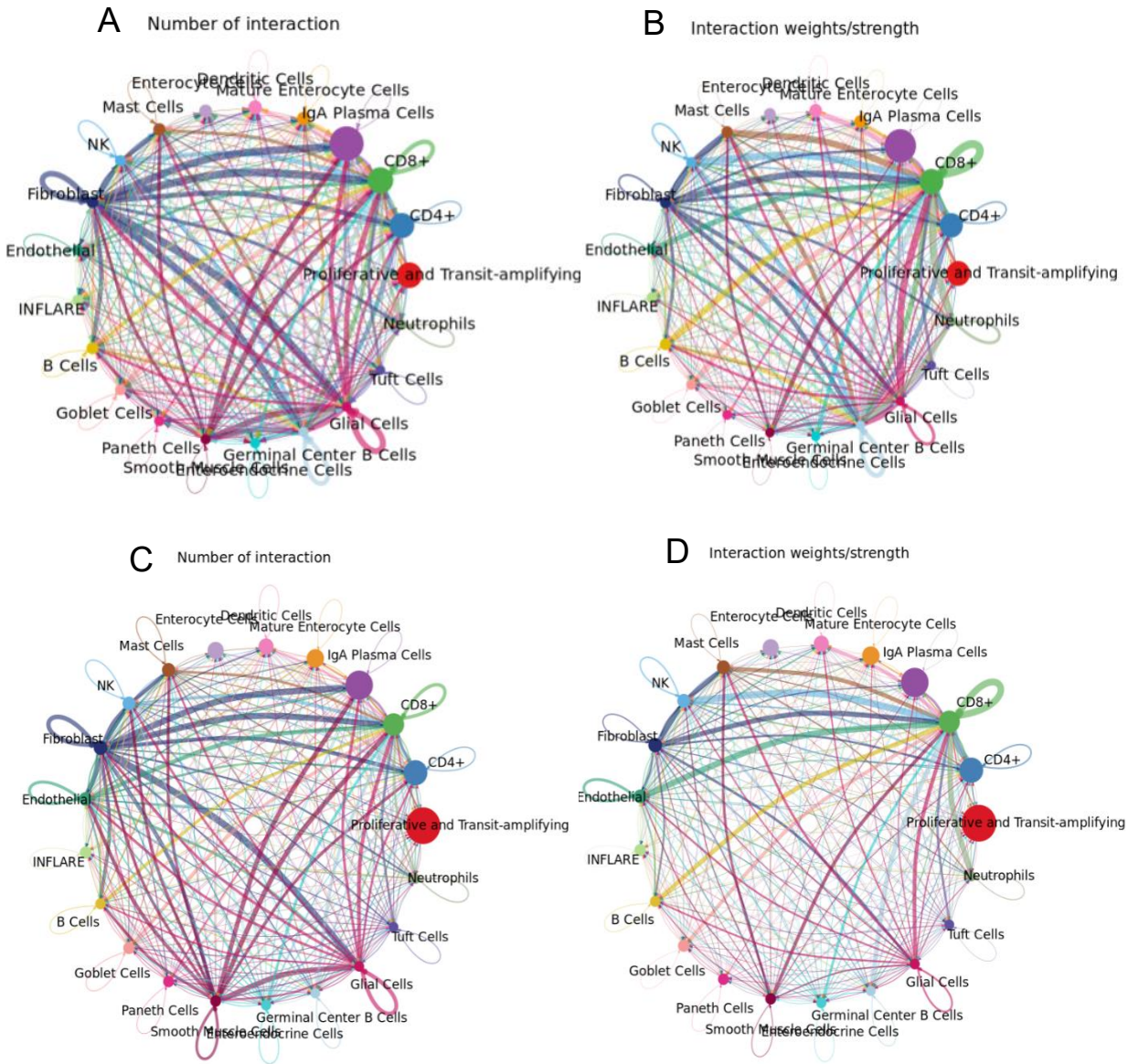


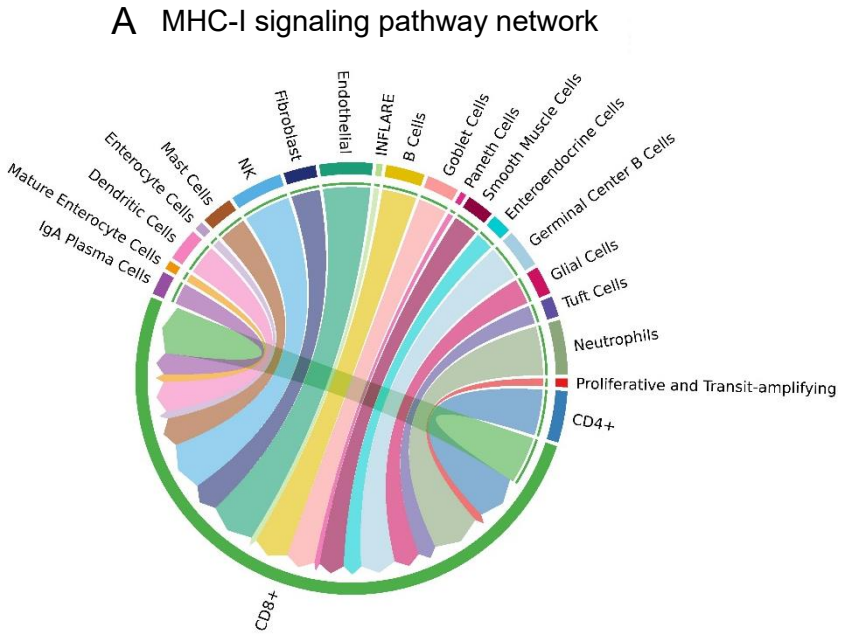
Figure 34: Analysis of cell-cell communications. *Figures A and B the analyzes carried out on celiac patients, while figures C and D on controls. Network plots show intercellular interactions based on ligands and receptors. (A-C) The circle plot shows the number of interactions between the different cell types, where the thickness of the arrows is proportional to the number of interactions between the cell populations. (B-D) The circle plot shows the intensity (weight) of the interactions.*

Furthermore, cell-cell communication analysis identifies the pathways in which each cellular communication is involved. In all groups (1, 2 and 3), fifty pathways were recognized the recognized pathways are ordered on the basis of the weight of ligand-receptor bonds in descending order (data not shown).

The first identified pathway is MHC-I (antigen presentation) and we focused on this to define the pathway network. In particular, this more in depth analysis was performed in celiac patients and controls seprately to indetify CD-related caracteristic.

In both celiac patients and controls we observed a significant involvement of CD8+ cells receptors. In controls the ligand-receptor interaction is unidirectional: all the cell types serve as ligand to CD8+ cells receptors (Figure 35 A). This result is consistent with the function of CD8+ cells which are responsible of immune surveillance always active in the gut (Al Moussawy and Abdelsamed 2022). In CD patients the CD8+-Germinal center B cells interaction is bidirectional (Figure 35 B). CD8+ cells serve as receptor and ligand to germinal center B cells.

Moreover, in CD patients we observed a significant involvement of germinal center B cells receptors. In particular, all the cell types serve as ligand to germinal center B cells. This result is related to the gluten-related activation of the adaptive immune system which leads to the production CD specific autoantibodies detecatable in celiac patients.



B MHC-I signaling pathway network

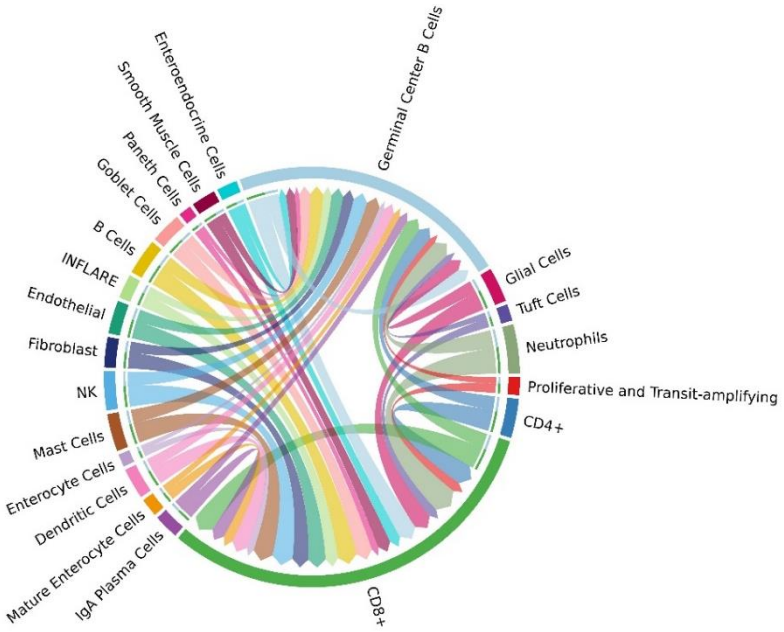


Figure 35: MHC-I signaling pathway. *Figure (A) shows the analysis done on celiac patients, while figure (B) shows the analysis done on controls. The chord diagram shows the molecular interactions within the antigen presentation pathway mediated by MHC-I molecules. The different cell types are represented as colored segments along the circumference, while the curved lines inside indicate the signaling connections between the different cell populations. The thickness of the lines is proportional to the intensity of the interactions.*

5. CONCLUSION

Celiac disease is one of the most common chronic lifelong diseases, with a prevalence of 1% in the general population. A triad of factors (genetic predisposition, gluten and environmental factors) is involved in the pathogenesis of celiac disease that is not yet fully understood. Recent discoveries made with single cell RNA sequencing (scRNA-seq) have significantly improved our understanding of the human intestine, both in normal conditions and also across the spectrum of gut diseases. To study celiac disease with a scRNA-seq approach could provide new insights into the pathogenesis of celiac disease.

This study revealed that the best way to obtained top-notch single cells RNA-seq data is to generate libraries from fresh intestinal biopsy samples, immediately after collection (within 30 minutes). Storage conditions generate high cellular stress level affecting data quality.

Therefore, for this project, single-cell libraries were generated from fresh intestinal biopsy samples from 5 celiac patients e 5 controls. The libraries were sequenced and a large amount of raw data was obtained (approximately 100,000 single cells). After the quality control step (nFeature_RNA ≥ 250; nCount_RNA ≥ 250; Percent_mt ≤ 25) top-notch quality data from 70.000 single cells were included in the final dataset.

Thanks to the cluster analysis all expected intestinal cell populations were identified in 21 clusters. Interestingly inflammatory epithelial cells (INFLARE) were identified. These cells have been recently discovered by Oliver et al., which highlighted their potential mechanisms, including inflammatory signaling pathways, stem and tissue regeneration factors and cell–cell communication pathways. Further studies are needed to explore a potential involvement of INFLAREs in celiac disease. Moreover, a significant increase in some epithelial cell populations (mature enterocytes, proliferative and transit-amplifying cells, and endothelial cells) and a decrease in several immune cell populations (IgA plasma cells, CD8+, germinal center B cells, and neutrophils) were confirmed in CD patients compared to CTRL patients. Subsequently, analysis of cell-cell communication showed bidirectional communication involving CD8+ cells and germinal center B cells only in celiac patients. This is consistent with immuno pathogenesis of celiac disease in which the gluten activate the adaptive immune response leading to the production of CD specific autoantibodies.

The size and robustness of the data obtained from this important project provide the cue for future investigation and insights to improve our knowledge of celiac disease. Therefore, single cells RNA-seq data will be integrated with intestinal spatial transcriptomics to obtain a high-resolution map of cellular populations in gut. To date, intestinal biopsy samples have been already collected and the generated libraries are currently being sequenced. Such integrative approaches will shed light on the mechanisms whereby such populations collaboratively lead to the development of CD.

6. BIBLIOGRAPHY

Abadie, Valérie, Valentina Discepolo, and Bana Jabri. 2012. “Intraepithelial Lymphocytes in Celiac Disease Immunopathology.” *Seminars in Immunopathology* 34(4):551–66. doi:10.1007/s00281-012-0316-x.

Al Moussawy, Mouhamad, and Hossam A. Abdelsamed. 2022. “Non-Cytotoxic Functions of CD8 T Cells: ‘Repentance of a Serial Killer.’” *Frontiers in Immunology* 13. doi:10.3389/fimmu.2022.1001129.

Atlasy, Nader, Anna Bujko, Espen S. Bækkevold, Peter Brazda, Eva Janssen-Megens, Knut E. A. Lundin, Jørgen Jahnsen, Frode L. Jahnsen, and Hendrik G. Stunnenberg. 2022. “Single Cell Transcriptomic Analysis of the Immune Cell Compartment in the Human Small Intestine and in Celiac Disease.” *Nature Communications* 13(1):4920. doi:10.1038/s41467-022-32691-5.

Balaban, Daniel Vasile, Alina Popp, Florentina Ionita Radu, and Mariana Jinga. 2019. “Hematologic Manifestations in Celiac Disease-A Practical Review.” *Medicina (Kaunas, Lithuania)* 55(7):373. doi:10.3390/medicina55070373.

Barker, Jennifer M., and Edwin Liu. 2008. “Celiac Disease: Pathophysiology, Clinical Manifestations, and Associated Autoimmune Conditions.” *Advances in Pediatrics* 55:349–65. doi:10.1016/j.yapd.2008.07.001.

Barker, Nick, Johan H. van Es, Jeroen Kuipers, Pekka Kujala, Maaike van den Born, Miranda Cozijnsen, Andrea Haegebarth, Jeroen Korving, Harry Begthel, Peter J. Peters, and Hans Clevers. 2007. “Identification of Stem Cells in Small Intestine and Colon by Marker Gene Lgr5.” *Nature* 449(7165):1003–7. doi:10.1038/nature06196.

Bengtsson, Martin, Martin Hemberg, Patrik Rorsman, and Anders Ståhlberg. 2008. “Quantification of mRNA in Single Cells and Modelling of RT-qPCR Induced Noise.” *BMC Molecular Biology* 9:63. doi:10.1186/1471-2199-9-63.

Bigaeva, Emilia, Werna T. C. Uniken Venema, Rinse K. Weersma, and Eleonora A. M. Festen. 2020. “Understanding Human Gut Diseases at Single-Cell Resolution.” *Human Molecular Genetics* 29(R1):R51–58. doi:10.1093/hmg/ddaa130.

Blakeley, Paul, Norah M. E. Fogarty, Ignacio Del Valle, Sissy E. Wamaitha, Tim Xiaoming Hu, Kay Elder, Philip Snell, Leila Christie, Paul Robson, and Kathy K. Niakan. 2015. “Defining the Three Cell Lineages of the Human Blastocyst by Single-Cell RNA-Seq.” *Development (Cambridge, England)* 142(20):3613. doi:10.1242/dev.131235.

Burclaff, Joseph, R. Jarrett Bliton, Keith A. Breau, Meryem T. Ok, Ismael Gomez-Martinez, Jolene S. Ranek, Aadra P. Bhatt, Jeremy E. Purvis, John T. Woosley, and Scott T. Magness. 2022. “A Proximal-to-Distal Survey of Healthy Adult Human Small Intestine and Colon Epithelium by Single-Cell Transcriptomics.” *Cellular and Molecular Gastroenterology and Hepatology* 13(5):1554–89. doi:10.1016/j.jcmgh.2022.02.007.

Busslinger, Georg A., Bas L. A. Weusten, Auke Bogte, Harry Begthel, Lodewijk A. A. Brosens, and Hans Clevers. 2021. “Human Gastrointestinal Epithelia of the Esophagus, Stomach, and Duodenum Resolved at Single-Cell Resolution.” *Cell Reports* 34(10):108819. doi:10.1016/j.celrep.2021.108819.

Caio, Giacomo, Umberto Volta, Anna Sapone, Daniel A. Leffler, Roberto De Giorgio, Carlo Catassi, and Alessio Fasano. 2019. “Celiac Disease: A Comprehensive Current Review.” *BMC Medicine* 17(1):142. doi:10.1186/s12916-019-1380-z.

Cardoso-Silva, Danielle, Deborah Delbue, Alice Itzlinger, Renée Moerkens, Sebo Withoff, Federica Branchi, and Michael Schumann. 2019. “Intestinal Barrier Function in Gluten-Related Disorders.” *Nutrients* 11(10):2325. doi:10.3390/nu11102325.

Catassi, Carlo, Elena F. Verdu, Julio Cesar Bai, and Elena Lionetti. 2022. “Coeliac Disease.” *Lancet (London, England)* 399(10344):2413–26. doi:10.1016/S0140-6736(22)00794-2.

Catassi, Giulia N., Alfredo Pulvirenti, Chiara Monachesi, Carlo Catassi, and Elena Lionetti. 2021. “Diagnostic Accuracy of IgA Anti-Transglutaminase and IgG Anti-Deamidated Gliadin for Diagnosis of Celiac Disease in Children under Two Years of Age: A Systematic Review and Meta-Analysis.” *Nutrients* 14(1):7. doi:10.3390/nu14010007.

Cerovic, Vuk, Calum C. Bain, Allan M. Mowat, and Simon W. F. Milling. 2014. “Intestinal Macrophages and Dendritic Cells: What’s the Difference?” *Trends in Immunology* 35(6):270–77. doi:10.1016/j.it.2014.04.003.

Conesa, Ana, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, Daniel J. Gaffney, Laura L. Elo, Xuegong Zhang, and Ali Mortazavi. 2016. “A Survey of Best Practices for RNA-Seq Data Analysis.” *Genome Biology* 17:13. doi:10.1186/s13059-016-0881-8.

Corazza, G. R., and V. Villanacci. 2005. “Coeliac Disease.” *Journal of Clinical Pathology* 58(6):573–74. doi:10.1136/jcp.2004.023978.

Deng, Qiaolin, Daniel Ramsköld, Björn Reinius, and Rickard Sandberg. 2014. “Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells.” *Science (New York, N.Y.)* 343(6167):193–96. doi:10.1126/science.1245316.

Di Sabatino, A., R. Ciccocioppo, F. Cupelli, B. Cinque, D. Millimaggi, M. M. Clarkson, M. Paulli, M. G. Cifone, and G. R. Corazza. 2006. “Epithelium Derived Interleukin 15 Regulates Intraepithelial Lymphocyte Th1 Cytokine Production, Cytotoxicity, and Survival in Coeliac Disease.” *Gut* 55(4):469–77. doi:10.1136/gut.2005.068684.

Dieterich, W., T. Ehnis, M. Bauer, P. Donner, U. Volta, E. O. Riecken, and D. Schuppan. 1997. “Identification of Tissue Transglutaminase as the Autoantigen of Celiac Disease.” *Nature Medicine* 3(7):797–801. doi:10.1038/nm0797-797.

Dotsenko, Valeriia, Mikko Oittinen, Juha Taavela, Alina Popp, Markku Peräaho, Synnöve Staff, Jani Sarin, Francisco Leon, Jorma Isola, Markku Mäki, and Keijo Viiri. 2021. “Genome-Wide Transcriptomic Analysis of Intestinal Mucosa in Celiac Disease Patients on a Gluten-Free Diet and Postgluten Challenge.” *Cellular and Molecular Gastroenterology and Hepatology* 11(1):13–32. doi:10.1016/j.jcmgh.2020.07.010.

Eberwine, J., H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel, and P. Coleman. 1992. “Analysis of Gene Expression in Single Live Neurons.” *Proceedings of the National Academy of Sciences of the United States of America* 89(7):3010–14. doi:10.1073/pnas.89.7.3010.

Escudero-Hernández, Celia. 2021. “Epithelial Cell Dysfunction in Coeliac Disease.” *International Review of Cell and Molecular Biology* 358:133–64. doi:10.1016/bs.ircmb.2020.09.007.

FitzPatrick, Michael E. B., Agne Antanaviciute, Melanie Dunstan, Karolina Künnapuu, Dominik Trzupek, Nicholas M. Provine, Kyla Dooley, Jia-Yuan Zhang, Sophie L. Irwin, Lucy C. Garner, Jane I. Pernes, Ricardo C. Ferreira, Sarah C. Sasson, Dominik Aschenbrenner, Devika Agarwal, Astor Rodrigues, Lucy Howarth, Oliver Brain, Darren Ruane, Elizabeth Soilleux, Sarah A. Teichmann, Calliope A. Dendrou, Alison Simmons, Holm H. Uhlig, John A. Todd, and Paul Klenerman. 2025. “Immune–Epithelial–Stromal Networks Define the Cellular Ecosystem of the Small Intestine in Celiac Disease.” *Nature Immunology* 26(6):947–62. doi:10.1038/s41590-025-02146-2.

Gehart, Helmuth, and Hans Clevers. 2019. “Tales from the Crypt: New Insights into Intestinal Stem Cells.” *Nature Reviews. Gastroenterology & Hepatology* 16(1):19–34. doi:10.1038/s41575-018-0081-y.

Green, Peter H. R., and Christophe Cellier. 2007. “Celiac Disease.” *The New England Journal of Medicine* 357(17):1731–43. doi:10.1056/NEJMra071600.

Gustafsson, Jenny K., and Malin E. V. Johansson. 2022. “The Role of Goblet Cells and Mucus in Intestinal Homeostasis.” *Nature Reviews Gastroenterology & Hepatology* 19(12):785–803. doi:10.1038/s41575-022-00675-x.

Hadjivassiliou, M., M. Mäki, D. S. Sanders, C. A. Williamson, R. A. Grünewald, N. M. Woodroofe, and I. R. Korponay-Szabó. 2006. “Autoantibody Targeting of Brain and Intestinal Transglutaminase in Gluten Ataxia.” *Neurology* 66(3):373–77. doi:10.1212/01.wnl.0000196480.55601.3a.

Hausch, Felix, Lu Shan, Nilda A. Santiago, Gary M. Gray, and Chaitan Khosla. 2002. “Intestinal Digestive Resistance of Immunodominant Gliadin Peptides.” *American Journal of Physiology. Gastrointestinal and Liver Physiology* 283(4):G996–1003. doi:10.1152/ajpgi.00136.2002.

Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir. 1965. “STRUCTURE OF A RIBONUCLEIC ACID.” *Science (New York, N.Y.)* 147(3664):1462–65. doi:10.1126/science.147.3664.1462.

Huang, Huan, Mari Goto, Hiroyuki Tsunoda, Lizhou Sun, Kiyomi Taniguchi, Hiroko Matsunaga, and Hideki Kambara. 2014. “Non-Biased and Efficient Global Amplification of a Single-Cell cDNA Library.” *Nucleic Acids Research* 42(2):e12. doi:10.1093/nar/gkt965.

Husby, S., S. Koletzko, I. R. Korponay-Szabó, M. L. Mearin, A. Phillips, R. Shamir, R. Troncone, K. Giersiepen, D. Branski, C. Catassi, M. Lelgeman, M. Mäki, C. Ribes-Koninckx, A. Ventura, K. P. Zimmer, ESPGHAN Working Group on Coeliac Disease Diagnosis, ESPGHAN Gastroenterology Committee, and European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. 2012. “European Society for Pediatric Gastroenterology, Hepatology, and Nutrition Guidelines for the Diagnosis of Coeliac Disease.” *Journal of Pediatric Gastroenterology and Nutrition* 54(1):136–60. doi:10.1097/MPG.0b013e31821a23d0.

Jaitin, Diego Adhemar, Ephraim Kenigsberg, Hadas Keren-Shaul, Naama Elefant, Franziska Paul, Irina Zaretsky, Alexander Mildner, Nadav Cohen, Steffen Jung, Amos Tanay, and Ido Amit. 2014. “Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types.” *Science (New York, N.Y.)* 343(6172):776–79. doi:10.1126/science.1247651.

Jovic, Dragomirka, Xue Liang, Hua Zeng, Lin Lin, Fengping Xu, and Yonglun Luo. 2022. “Single-cell RNA Sequencing Technologies and Applications: A Brief Overview.” *Clinical and Translational Medicine* 12(3):e694. doi:10.1002/ctm2.694.

Kivelä, Laura, Alberto Caminero, Daniel A. Leffler, Maria Ines Pinto-Sanchez, Jason A. Tye-Din, and Katri Lindfors. 2021. “Current and Emerging Therapies for Coeliac Disease.” *Nature Reviews. Gastroenterology & Hepatology* 18(3):181–95. doi:10.1038/s41575-020-00378-1.

Kivioja, Teemu, Anna Vähärautio, Kasper Karlsson, Martin Bonke, Martin Enge, Sten Linnarsson, and Jussi Taipale. 2011. “Counting Absolute Numbers of Molecules Using Unique Molecular Identifiers.” *Nature Methods* 9(1):72–74. doi:10.1038/nmeth.1778.

Klein, Allon M., Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A. Weitz, and Marc W. Kirschner. 2015. “Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells.” *Cell* 161(5):1187–1201. doi:10.1016/j.cell.2015.04.044.

de Klerk, Eleonora, and Peter A. C. ’t Hoen. 2015. “Alternative mRNA Transcription, Processing, and Translation: Insights from RNA Sequencing.” *Trends in Genetics: TIG* 31(3):128–39. doi:10.1016/j.tig.2015.01.001.

Kolodziejczyk, Aleksandra A., Jong Kyoung Kim, Valentine Svensson, John C. Marioni, and Sarah A. Teichmann. 2015. “The Technology and Biology of Single-Cell RNA Sequencing.” *Molecular Cell* 58(4):610–20. doi:10.1016/j.molcel.2015.04.005.

Korponay-Szabó, I. R., T. Halttunen, Z. Szalai, K. Laurila, R. Király, J. B. Kovács, L. Fésüs, and M. Mäki. 2004. “In Vivo Targeting of Intestinal and Extraintestinal Transglutaminase 2 by Coeliac Autoantibodies.” *Gut* 53(5):641–48. doi:10.1136/gut.2003.024836.

Lebwohl, Benjamin, David S. Sanders, and Peter H. R. Green. 2018. “Coeliac Disease.” *Lancet (London, England)* 391(10115):70–81. doi:10.1016/S0140-6736(17)31796-8.

Li, Xinmin, and Cun-Yu Wang. 2021. “From Bulk, Single-Cell to Spatial RNA Sequencing.” *International Journal of Oral Science* 13(1):36. doi:10.1038/s41368-021-00146-0.

Lindfors, Katri, Carolina Ciacci, Kalle Kurppa, Knut E. A. Lundin, Govind K. Makharia, M. Luisa Mearin, Joseph A. Murray, Elena F. Verdu, and Katri Kaukinen. 2019. “Coeliac Disease.” *Nature Reviews. Disease Primers* 5(1):3. doi:10.1038/s41572-018-0054-z.

Lönnberg, Tapio, Valentine Svensson, Kylie R. James, Daniel Fernandez-Ruiz, Ismail Sebina, Ruddy Montandon, Megan S. F. Soon, Lily G. Fogg, Arya Sheela Nair, Urijah Liligeto, Michael J. T. Stubbington, Lam-Ha Ly, Frederik Otzen Bagger, Max Zwiessele, Neil D. Lawrence, Fernando Souza-Fonseca-Guimaraes, Patrick T. Bunn, Christian R. Engwerda, William R. Heath, Oliver Billker, Oliver Stegle, Ashraful Haque, and Sarah A. Teichmann. 2017. “Single-Cell RNA-Seq and Computational Analysis Using Temporal Mixture Modelling Resolves Th1/Tfh Fate Bifurcation in Malaria.” *Science Immunology* 2(9):eaal2192. doi:10.1126/sciimmunol.aal2192.

Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R. Bialas, Nolan Kamitaki, Emily M. Martersteck, John J. Trombetta, David A. Weitz, Joshua R. Sanes, Alex K. Shalek, Aviv Regev, and Steven A. McCarroll. 2015. “Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets.” *Cell* 161(5):1202–14. doi:10.1016/j.cell.2015.05.002.

Maglio, Mariantonia, Antonella Tosco, Renata Auricchio, Francesco Paparo, Barbara Colicchio, Erasmo Miele, Luciano Rapacciuolo, and Riccardo Troncone. 2011. "Intestinal Deposits of Anti-Tissue Transglutaminase IgA in Childhood Celiac Disease." *Digestive and Liver Disease: Official Journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver* 43(8):604–8. doi:10.1016/j.dld.2011.01.015.

Mantis, N. J., N. Rol, and B. Corthésy. 2011. "Secretory IgA's Complex Roles in Immunity and Mucosal Homeostasis in the Gut." *Mucosal Immunology* 4(6):603–11. doi:10.1038/mi.2011.41.

Mardis, Elaine R. 2008. "Next-Generation DNA Sequencing Methods." *Annual Review of Genomics and Human Genetics* 9:387–402. doi:10.1146/annurev.genom.9.081307.164359.

Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, Michael S. Braverman, Yi-Ju Chen, Zhoutao Chen, Scott B. Dewell, Lei Du, Joseph M. Fierro, Xavier V. Gomes, Brian C. Godwin, Wen He, Scott Helgesen, Chun Heen Ho, Gerard P. Irzyk, Szilveszter C. Jando, Maria L. I. Alenquer, Thomas P. Jarvie, Kshama B. Jirage, Jong-Bum Kim, James R. Knight, Janna R. Lanza, John H. Leamon, Steven M. Lefkowitz, Ming Lei, Jing Li, Kenton L. Lohman, Hong Lu, Vinod B. Makhijani, Keith E. McDade, Michael P. McKenna, Eugene W. Myers, Elizabeth Nickerson, John R. Nobile, Ramona Plant, Bernard P. Puc, Michael T. Ronan, George T. Roth, Gary J. Sarkis, Jan Fredrik Simons, John W. Simpson, Maithreyan Srinivasan, Karrie R. Tartaro, Alexander Tomasz, Kari A. Vogt, Greg A. Volkmer, Shally H. Wang, Yong Wang, Michael P. Weiner, Pengguang Yu, Richard F. Begley, and Jonathan M. Rothberg. 2005. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." *Nature* 437(7057):376–80. doi:10.1038/nature03959.

McCarty, Thomas R., Corey R. O'Brien, Anas Gremida, Christina Ling, and Tarun Rustagi. 2018. "Efficacy of Duodenal Bulb Biopsy for Diagnosis of Celiac Disease: A Systematic Review and Meta-Analysis." *Endoscopy International Open* 6(11):E1369–78. doi:10.1055/a-0732-5060.

Metzker, Michael L. 2010. "Sequencing Technologies - the next Generation." *Nature Reviews. Genetics* 11(1):31–46. doi:10.1038/nrg2626.

Meyerson, Matthew, Stacey Gabriel, and Gad Getz. 2010. "Advances in Understanding Cancer Genomes through Second-Generation Sequencing." *Nature Reviews. Genetics* 11(10):685–96. doi:10.1038/nrg2841.

Nisticò, L., C. Fagnani, I. Coto, S. Percopo, R. Cotichini, M. G. Limongelli, F. Paparo, S. D'Alfonso, M. Giordano, C. Sferlazzas, G. Magazzù, P. Momigliano-Richiardi, L. Greco, and M. A. Stazi. 2006. "Concordance, Disease Progression, and Heritability of Coeliac Disease in Italian Twins." *Gut* 55(6):803–8. doi:10.1136/gut.2005.083964.

Oliver, Amanda J., Ni Huang, Raquel Bartolome-Casado, Ruoyan Li, Simon Koplev, Hogne R. Nilsen, Madelyn Moy, Batuhan Cakir, Krzysztof Polanski, Victoria Gudiño, Elisa Melón-Ardanaz, Dinithi Sumanaweera, Daniel Dimitrov, Lisa Marie Milchsack, Michael E. B. FitzPatrick, Nicholas M. Provine, Jacqueline M. Boccacino, Emma Dann, Alexander V. Predeus, Ken To, Martin Prete, Jonathan A. Chapman, Andrea C. Masi, Emily Stephenson, Justin Engelbert, Sebastian Lobentanzer, Shani Perera, Laura Richardson, Rakeshlal Kapuge, Anna Wilbrey-Clark, Claudia I. Semprich, Sophie Ellams, Catherine Tudor, Philomeena Joseph, Alba Garrido-Trigo, Ana M. Corraliza, Thomas R. W. Oliver, C. Elizabeth Hook, Kylie R. James, Krishnaa T. Mahbubani, Kourosh Saeb-Parsy, Matthias Zilbauer, Julio Saez-Rodriguez, Marte Lie Høivik, Espen S. Bækkevold, Christopher J. Stewart, Janet E. Berrington, Kerstin B. Meyer, Paul Klenerman, Azucena Salas, Muzlifah Haniffa, Frode L. Jahnsen, Rasa Elmentaite, and Sarah A. Teichmann.

2024. “Single-Cell Integration Reveals Metaplasia in Inflammatory Gut Diseases.” *Nature* 635(8039):699–707. doi:10.1038/s41586-024-07571-1.

Papalexi, Efthymia, and Rahul Satija. 2018. “Single-Cell RNA Sequencing to Explore Immune Cell Heterogeneity.” *Nature Reviews. Immunology* 18(1):35–45. doi:10.1038/nri.2017.76.

Petropoulos, Sophie, Daniel Edsgärd, Björn Reinius, Qiaolin Deng, Sarita Pauliina Panula, Simone Codeluppi, Alvaro Plaza Reyes, Sten Linnarsson, Rickard Sandberg, and Fredrik Lanner. 2016. “Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos.” *Cell* 167(1):285. doi:10.1016/j.cell.2016.08.009.

Rosales, Carlos. 2018. “Neutrophil: A Cell with Many Roles in Inflammation or Several Cell Types?” *Frontiers in Physiology* 9:113. doi:10.3389/fphys.2018.00113.

Salazar, Carolina, Jennyfer M. García-Cárdenas, and César Paz-y-Miño. 2017. “Understanding Celiac Disease From Genetics to the Future Diagnostic Strategies.” *Clinical Medicine Insights. Gastroenterology* 10:1179552217712249. doi:10.1177/1179552217712249.

Salmi, T. T., P. Collin, I. R. Korponay-Szabó, K. Laurila, J. Partanen, H. Huhtala, R. Király, L. Lorand, T. Reunala, M. Mäki, and K. Kaukinen. 2006. “Endomysial Antibody-Negative Coeliac Disease: Clinical Characteristics and Intestinal Autoantibody Deposits.” *Gut* 55(12):1746–53. doi:10.1136/gut.2005.071514.

Salmi, Teea T., Kaisa Hervonen, Kaija Laurila, Pekka Collin, Markku Mäki, Outi Koskinen, Heini Huhtala, Katri Kaukinen, and Timo Reunala. 2014. “Small Bowel Transglutaminase 2-Specific IgA Deposits in Dermatitis Herpetiformis.” *Acta Dermato-Venereologica* 94(4):393–97. doi:10.2340/00015555-1764.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. “DNA Sequencing with Chain-Terminating Inhibitors.” *Proceedings of the National Academy of Sciences of the United States of America* 74(12):5463–67. doi:10.1073/pnas.74.12.5463.

Shoaran, Maryam, Hani Sabaie, Mehrnaz Mostafavi, and Maryam Rezazadeh. 2024. “A Comprehensive Review of the Applications of RNA Sequencing in Celiac Disease Research.” *Gene* 927:148681. doi:10.1016/j.gene.2024.148681.

Sollid, Ludvig M., Jason A. Tye-Din, Shuo-Wang Qiao, Robert P. Anderson, Carmen Gianfrani, and Frits Koning. 2020. “Update 2020: Nomenclature and Listing of Celiac Disease-Relevant Gluten Epitopes Recognized by CD4+ T Cells.” *Immunogenetics* 72(1–2):85–88. doi:10.1007/s00251-019-01141-w.

Stuart, Tim, and Rahul Satija. 2019. “Integrative Single-Cell Analysis.” *Nature Reviews. Genetics* 20(5):257–72. doi:10.1038/s41576-019-0093-7.

Sulic, Ana-Marija, Kalle Kurppa, Tiina Rauhavirta, Katri Kaukinen, and Katri Lindfors. 2015. “Transglutaminase as a Therapeutic Target for Celiac Disease.” *Expert Opinion on Therapeutic Targets* 19(3):335–48. doi:10.1517/14728222.2014.985207.

Sun, Lina, Yanhong Su, Anjun Jiao, Xin Wang, and Baojun Zhang. 2023. “T Cells in Health and Disease.” *Signal Transduction and Targeted Therapy* 8:235. doi:10.1038/s41392-023-01471-y.

Svensson, Valentine, Kedar Nath Natarajan, Lam-Ha Ly, Ricardo J. Miragaia, Charlotte Labalette, Iain C. Macaulay, Ana Cvejic, and Sarah A. Teichmann. 2017. “Power Analysis of Single-Cell RNA-Sequencing Experiments.” *Nature Methods* 14(4):381–87. doi:10.1038/nmeth.4220.

Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B. Tuch, Asim Siddiqui, Kaiqin Lao, and M. Azim Surani. 2009a. “mRNA-Seq Whole-Transcriptome Analysis of a Single Cell.” *Nature Methods* 6(5):377–82. doi:10.1038/nmeth.1315.

Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B. Tuch, Asim Siddiqui, Kaiqin Lao, and M. Azim Surani. 2009b. “mRNA-Seq Whole-Transcriptome Analysis of a Single Cell.” *Nature Methods* 6(5):377–82. doi:10.1038/nmeth.1315.

Taniguchi, Kiyomi, Tomoharu Kajiyama, and Hideki Kambara. 2009. “Quantitative Analysis of Gene Expression in a Single Cell by qPCR.” *Nature Methods* 6(7):503–6. doi:10.1038/nmeth.1338.

Trapnell, Cole. 2015. “Defining Cell Types and States with Single-Cell Genomics.” *Genome Research* 25(10):1491–98. doi:10.1101/gr.190595.115.

Treutlein, Barbara, Doug G. Brownfield, Angela R. Wu, Norma F. Neff, Gary L. Mantalas, F. Hernan Espinoza, Tushar J. Desai, Mark A. Krasnow, and Stephen R. Quake. 2014. “Reconstructing Lineage Hierarchies of the Distal Lung Epithelium Using Single-Cell RNA-Seq.” *Nature* 509(7500):371–75. doi:10.1038/nature13173.

Trynka, Gosia, Karen A. Hunt, Nicholas A. Bockett, Jihane Romanos, Vanisha Mistry, Agata Szperl, Sjoerd F. Bakker, Maria Teresa Bardella, Leena Bhaw-Rosun, Gemma Castillejo, Emilio G. de la Concha, Rodrigo Coutinho de Almeida, Kerith-Rae M. Dias, Cleo C. van Diemen, Patrick C. A. Dubois, Richard H. Duerr, Sarah Edkins, Lude Franke, Karin Fransen, Javier Gutierrez, Graham A. R. Heap, Barbara Hrdlickova, Sarah Hunt, Leticia Plaza Izurieta, Valentina Izzo, Leo A. B. Joosten, Cordelia Langford, Maria Cristina Mazzilli, Charles A. Mein, Vandana Midah, Mitja Mitrovic, Barbara Mora, Marinita Morelli, Sarah Nutland, Concepcion Nunez, Suna Onengut-Gumuscu, Kerra Pearce, Mathieu Platteel, Isabel Polanco, Simon Potter, Carmen Ribes-Koninckx, Isis Ricano-Ponce, Stephen S. Rich, Anna Rybak, Jose Luis Santiago, Sabyasachi Senapati, Ajit Sood, Hania Szajewska, Alexandra Zhernakova, Cisca Wijmenga, Spanish Consortium Genetics Coelia, PreventCD Study Grp, and WTCCC. 2011. “Dense Genotyping Identifies and Localizes Multiple Common and Rare Variant Association Signals in Celiac Disease.” *Nature Genetics* 43(12):1193-U45. doi:10.1038/ng.998.

Tye-Din, Jason A., Heather J. Galipeau, and Daniel Agardh. 2018. “Celiac Disease: A Review of Current Concepts in Pathogenesis, Prevention, and Novel Therapies.” *Frontiers in Pediatrics* 6:350. doi:10.3389/fped.2018.00350.

Venteicher, Andrew S., Itay Tirosh, Christine Hebert, Keren Yizhak, Cyril Neftel, Mariella G. Filbin, Volker Hovestadt, Leah E. Escalante, McKenzie L. Shaw, Christopher Rodman, Shawn M. Gillespie, Danielle Dionne, Christina C. Luo, Hiranmayi Ravichandran, Ravindra Mylvaganam, Christopher Mount, Maristela L. Onozato, Brian V. Nahed, Hiroaki Wakimoto, William T. Curry, A. John Iafrate, Miguel N. Rivera, Matthew P. Frosch, Todd R. Golub, Priscilla K. Brastianos, Gad Getz, Anoop P. Patel, Michelle Monje, Daniel P. Cahill, Orit Rozenblatt-Rosen, David N. Louis, Bradley E. Bernstein, Aviv Regev, and Mario L. Suvà. 2017. “Decoupling Genetics, Lineages, and Microenvironment in IDH-Mutant Gliomas by Single-Cell RNA-Seq.” *Science (New York, N.Y.)* 355(6332):eaai8478. doi:10.1126/science.aai8478.

Wang, Yalong, Wanlu Song, Jilian Wang, Ting Wang, Xiaochen Xiong, Zhen Qi, Wei Fu, Xuerui Yang, and Ye-Guang Chen. 2020. “Single-Cell Transcriptome Analysis Reveals Differential Nutrient Absorption Functions in Human Intestine.” *The Journal of Experimental Medicine* 217(2):e20191130. doi:10.1084/jem.20191130.

Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. “RNA-Seq: A Revolutionary Tool for Transcriptomics.” *Nature Reviews. Genetics* 10(1):57–63. doi:10.1038/nrg2484.

Warren, Luigi, David Bryder, Irving L. Weissman, and Stephen R. Quake. 2006. “Transcription Factor Profiling in Individual Hematopoietic Progenitors by Digital RT-PCR.” *Proceedings of the National Academy of Sciences of the United States of America* 103(47):17807–12. doi:10.1073/pnas.0608512103.

Watson, J. D., and F. H. C. Crick. 1953. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.” *Nature* 171(4356):737–38. doi:10.1038/171737a0.

Wilk, Aaron J., Alex K. Shalek, Susan Holmes, and Catherine A. Blish. 2024. “Comparative Analysis of Cell–Cell Communication at Single-Cell Resolution.” *Nature Biotechnology* 42(3):470–83. doi:10.1038/s41587-023-01782-z.

Yan, Liying, Mingyu Yang, Hongshan Guo, Lu Yang, Jun Wu, Rong Li, Ping Liu, Ying Lian, Xiaoying Zheng, Jie Yan, Jin Huang, Ming Li, Xinglong Wu, Lu Wen, Kaiqin Lao, Ruiqiang Li, Jie Qiao, and Fuchou Tang. 2013. “Single-Cell RNA-Seq Profiling of Human Preimplantation Embryos and Embryonic Stem Cells.” *Nature Structural & Molecular Biology* 20(9):1131–39. doi:10.1038/nsmb.2660.

Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, Tobias D. Wheeler, Geoff P. McDermott, Junjie Zhu, Mark T. Gregory, Joe Shuga, Luz Montesclaros, Jason G. Underwood, Donald A. Masquelier, Stefanie Y. Nishimura, Michael Schnall-Levin, Paul W. Wyatt, Christopher M. Hindson, Rajiv Bharadwaj, Alexander Wong, Kevin D. Ness, Lan W. Beppu, H. Joachim Deeg, Christopher McFarland, Keith R. Loeb, William J. Valente, Nolan G. Ericson, Emily A. Stevens, Jerald P. Radich, Tarjei S. Mikkelsen, Benjamin J. Hindson, and Jason H. Bielas. 2017. “Massively Parallel Digital Transcriptional Profiling of Single Cells.” *Nature Communications* 8:14049. doi:10.1038/ncomms14049.

Ziegenhain, Christoph, Beate Vieth, Swati Parekh, Björn Reinius, Amy Guillaumet-Adkins, Martha Smets, Heinrich Leonhardt, Holger Heyn, Ines Hellmann, and Wolfgang Enard. 2017. “Comparative Analysis of Single-Cell RNA Sequencing Methods.” *Molecular Cell* 65(4):631-643.e4. doi:10.1016/j.molcel.2017.01.023.