UNIVERSIDADE NOVA DE LISBOA

MScCBBi

MASTER IN
COMPUTATIONAL BIOLOGY & BIOINFORMATICS

DIOGO MARTINS FONSECA
BSc in Cell and Molecular Biology

Single-cell transcriptional landscape of circulating T_{fh} and T_{reg} cells in COVID-19 patients

Sep, 2023

NOVA SCHOOL OF SCIENCE & TECHNOLOGY
NOVA MEDICAL SCHOOL
NOVA IMS Information Management School
DESDE 1902 INSTITUTO DE HIGIENE E MEDICINA TROPICAL UNIVERSIDADE NOVA DE LISBOA

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF LIFE SCIENCES

SINGLE-CELL TRANSCRIPTIONAL LANDSCAPE OF CIRCULATING TFH AND TREG CELLS IN COVID-19 PATIENTS

DIOGO MARTINS FONSECA

BSc in Cell and Molecular Biology

Adviser: Luís Ricardo Silva Graça
Full Professor, University of Lisbon

Co-advisers:

Saumya Kumar,
Postdoctoral Researcher, IMM

Ana Rita Fialho Grosso
Assistant Professor, FCT-NOVA

Examination Committee:

Chair:	Paula Maria Theriaga Mendes Bernardo Gonçalves, Associate Professor, FCT-NOVA
Adviser:	Luís Ricardo Silva Graça, Full Professor, University of Lisbon
Examiner:	Alexandre Augusto da Silva Figueiredo e Raposo, Senior Postdoctoral Researcher, IMM

MASTER IN COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

NOVA University Lisbon
September, 2023

ACKNOWLEDGMENTS

Firstly, I would like to express a special acknowledgment to my supervisor, Professor Doctor Luís Graça for graciously welcoming me into his project and laboratory. Thank you for all the knowledge you have imparted to me, and for all the experiences I have been able to live through.

I would also like to express a special acknowledgment to Saumya Kumar and Rui Vieira. Thank you for everything you have taught me, for your availability, and motivation.

My gratitude extends to the Instituto de Medicina Molecular for offering the resources and space essential for the progress of this project.

I would like to thank all the members of Luís Graça's lab for fostering a supportive and constructive work environment. Your collective guidance and assistance have been invaluable, and I am grateful for the knowledge and support I have received from each one of you.

Finally, a very special thanks to my family for everything they have taught me and continue to teach me, every day. To Maria, my best friend and counselor, who supports me unconditionally.

ABSTRACT

The emergence of COVID-19 has posed an unprecedented challenge to global healthcare systems and socioeconomic progress. Understanding the immune response against SARS-CoV-2 is crucial for developing effective therapeutic and prophylactic strategies and enhancing preparedness for future airborne virus threats. A key aspect of this response is the regulation of antibody production, which not only drives protective immunity following infection but also plays a pivotal role in vaccine efficacy. Equally important is comprehending the immunosuppressive mechanisms in severely ill patients to prevent excessive immune responses, mitigate immunopathology, and promote controlled antiviral immunity. Tfh cells are essential for guiding B cells to generate high-affinity neutralizing antibodies, while Treg cells act as immune suppressors to prevent immune overactivity. In our study, we employed advanced computational methods, including single-cell transcriptomics analysis, to unravel the intricate dynamics of both Tfh and Treg populations elicited in response to SARS-CoV-2 infection. We conducted single-cell resolution analyses of circulating Tfh cells (cTfh) and Treg cells in severely ill COVID-19 patients. For this, we sorted Tfh and Treg cells from peripheral blood samples and performed single-cell RNA sequencing. In COVID-19 patients, cTfh cell clusters exhibited segregation based on two critical factors: activation status and cTfh subtype, a phenomenon not observed in cTfh cells from healthy donors. Furthermore, our investigation into Treg cell heterogeneity in both healthy and COVID-19 conditions unveiled distinct cell populations varying in activation states and Th signatures. In conclusion, our research offers a comprehensive resource on the transcriptomics of various Tfh and Treg populations and their responses to viral infections.

Keywords: Viral Infection; Immune response; Tfh cells; Treg cells; Single-cell RNA-Seq.

RESUMO

O aparecimento da COVID-19 constituiu um desafio sem precedentes para os sistemas de saúde mundiais e para o progresso socioeconómico. A compreensão da resposta imunitária contra o SARS-CoV-2 é crucial para o desenvolvimento de estratégias terapêuticas e profiláticas eficazes e para melhorar a preparação para futuras ameaças de vírus. Um aspeto fundamental desta resposta é a regulação da produção de anticorpos, que não só impulsiona a imunidade protetora após a infeção, mas também desempenha um papel fundamental na eficácia das vacinas. Igualmente importante é compreender os mecanismos imunossupressores em doentes graves para evitar respostas imunitárias excessivas e promover uma imunidade antiviral controlada. As células Tfh são essenciais para guiar as células B na geração de anticorpos neutralizantes de alta afinidade, enquanto as células Treg atuam como supressores imunitários para evitar a sobre atividade imunitária. Neste estudo, utilizámos métodos computacionais avançados, incluindo a análise transcriptómica de células individuais, para desvendar a complexa dinâmica das populações de Tfh e Treg em resposta à infeção por SARS-CoV-2. Realizámos análises transcriptómicas a células Tfh circulantes (cTfh) e células Treg em doentes graves com COVID-19. Para isso, separámos as células cTfh e Treg de amostras de sangue periférico e realizámos a sequenciação do RNA de cada célula. Nos doentes com COVID-19, as células cTfh apresentaram segregação com base em dois importantes fatores: estado de ativação e o subtipo, um fenómeno não observado nas células cTfh de dadores saudáveis. Além disso, a nossa investigação sobre a heterogeneidade das células Treg, tanto em dadores saudáveis como em pacientes com COVID-19, revelou populações celulares distintas que variam em termos de estados de ativação e subtipo. Em conclusão, a nossa investigação oferece um recurso detalhado sobre o transcriptoma de várias populações de Tfh e Treg e as suas respostas a infeções virais.

Palavras chave: Infeção viral; Resposta imunitária; Células Tfh; Células Treg; single-cell RNA-seq.

CONTENTS

LIST OF FIGURES

1. Introduction

1.1 Infectious diseases in modern times

In the modern era, a series of highly contagious diseases, with viral pandemics (such as influenza or coronaviruses) being particularly prominent, have caused profound devastation to live and livelihoods worldwide. The risks of spillover of infectious diseases – triggered by pathogenic microorganisms such as viruses, bacteria, fungi, and apicomplexans – have increased due to recent societal changes. Despite significant advances across the globe in terms of improved sanitation and healthcare access, these risks persist [1], [2].

In December 2019, a viral disease named COVID-19, caused by the novel coronavirus SARS-CoV-2, was detected in Wuhan (China), serving as a significant example of how such risks can swiftly become a global concern. The swift and worldwide spread of COVID-19 underscored the vulnerability of societies to unforeseen health crises and emphasized the ongoing need for vigilant preparedness and robust healthcare systems to mitigate the impact of future disease outbreaks.

In this context, enhancing our comprehension of the immune response to pathogenic microorganisms holds the potential to advance therapeutic and prophylactic strategies and improve preparedness for future outbreaks. In this respect, a better understanding of mechanisms involved in the regulation of antibody production is critical, not only regarding the protective immune response following infection but also as a key mechanism in effective vaccines.

1.2 The Immune Response - Overview

Immunity, or the capacity to protect our body from infection and cancer, consists of three defensive mechanisms. The first defense involves anatomical and chemical barriers, i.e., skin, mucosal surfaces, and antimicrobial peptides produced within these tissues, which in combination form an effective protective barrier to prevent the invasion of microbes into the human body. Nevertheless, pathogens evolved methods to circumvent this initial defense barrier, and other protective immune responses come into play [3].

Once a pathogen manages to penetrate the epithelial barriers of the host, it will then confront the innate immune mechanisms. The innate immune system consists of several cellular and molecular components that work synergistically to detect and eradicate pathogens providing immediate, non-specific defense. The initiation of this response relies on pattern-recognition receptors (PRRs), expressed on the surface of innate immune cells, or soluble molecules like complement system, that can recognize pathogen-associated molecular patterns

(PAMPs). The interaction between PAMPs and PRRs triggers a signaling cascade, activating innate immune cells, such as macrophages, and natural killer cells, acting as an essential early resistance to pathogen replication [4].

Finally, the adaptive immune responses are tailored to the specific pathogen, since they rely on the activation of antigen-specific lymphocytes, specifically B and T cells, which possess highly variable antigen receptors capable of recognizing and binding to antigens exhibited by professional antigen-presenting cells (APCs), in the case of T cells, or directly on the target pathogen for B cells. The lymphocyte activation, which is followed by clonal expansion, and differentiation, gives rise to antigen-specific effector cells that precisely eliminate the pathogen. A key property of this highly specialized system is the production of immunological memory, characterized by the maintenance of long-lived immune cells after the infection is cleared, resulting in quicker and more efficient responses upon a second encounter with the same antigen [5].

Regarding adaptive immunity, this highly specific response encompasses two general classes: cell-mediated immunity and humoral immunity. Cell-mediated immunity relies on T cells, which, upon recognizing their specific antigen through binding between MHC proteins and exogenous antigens, undergo differentiation into various functional subclasses of effector T cells with distinct specializations for diverse activities. These subsets carry out distinct functions - CD8 T cells recognize and neutralize infected or abnormal cells by cell cytotoxicity [6]. The CD4 effector subsets stimulate or suppress the action of their target immune cells [7].

On the other hand, humoral immunity, mediated by B cells, entails the production and distribution of antibodies. These molecules are produced to detect and neutralize detrimental agents, such as infectious viruses, bacteria, and fungi, through specific binding to their respective target antigens. While certain microbial antigens can directly trigger B cell activation without any additional stimulus, the process of affinity maturation in B cells and their differentiation into memory cells and high-affinity antibody-secreting B cells typically require interactions with helper T cells located in the germinal center [5].

1.3 T follicular helper and T regulatory cells in the immune system

After antigen exposure through the B cell receptor (BCR), antigen-activated naïve B cells undergo proliferation, somatic hypermutation, and affinity-based selection leading to the generation of antibody-secreting cells (ASCs) – plasma cells – and memory B cells [8]. This process predominantly takes place within secondary lymphoid organs (SLOs) and depends on T follicular helper (Tfh) cells costimulation – ICOSL:ICOS and CD40:CD40L signaling – and cytokine production (IL-21). Therefore, Tfh cells are a $CD4^+$ T cell subpopulation that provides help to B cells, stimulating a long-lasting antibody production with an enhanced affinity towards their targets.

Tfh cells originate from naïve $CD4^+$ T cells. Once stimulated, some $CD4^+$ T cells go through a complex, multistage differentiation process in which BCL6 expression plays a vital role, directing their transformation into Tfh cells. This process promotes the development of

Tfh cell fate and their specialized functions, resulting in increased CXCR5 chemokine receptor expression and the suppression of CCR7. Consequently, these changes facilitate the migration of Tfh cells towards B cell follicles, which is essential for germinal center formation during an active immune response [9]. Tfh cell deficiency limits the development of germinal centers (GCs), which results in detectable impairments in the generation of antibodies [10], [11]. The fate decision driving the initial process of differentiation of a naïve $CD4^+$ T cell towards a Tfh is not fully established. However, it has been shown that the presence of IL-2 can regulate this fate choice, with low levels of this cytokine favoring Tfh differentiation [12], [13].

Over the past decade, extensive research has focused on unraveling the intricacies of Tfh cell biology within secondary lymphoid organs. This research has yielded substantial progress in our comprehension of the origin and function of these cells. Nonetheless, the occurrence of $CXCR5^+$ $CD4^+$ Tfh cells extends beyond lymphoid organs and can also be found in the human bloodstream. Within the blood, a distinct population of long-lived memory cells shares functional characteristics with Tfh cells, acting as a circulating counterpart and referred to as circulating Tfh cells (cTfh). These cTfh cells are derived from germinal centers and have the potential to function as circulating memory cells that can be quickly mobilized for immune reactions, effectively supporting naive or memory B cells in generating antibodies. Human cTfh cells are a minor subset of $CD4^+$ T cells found in the peripheral blood. They are composed of distinct subsets characterized by distinct phenotypes and functions, determined by the presence or absence of specific markers such as chemokine receptors CXCR3 and CCR6, PD-1, and ICOS [14]. Unlike Tfh cells in SLOs, cTfh cells lack Bcl-6 protein expression. Instead, the presence of CXCR3 and CCR6 allows for the categorization of three primary subsets within circulating Tfh cells: $CXCR3^+CCR6^-$ cells, which resemble Th1 cells and are thus referred to as cTfh1 cells; $CXCR3^-CCR6^-$ cells, resembling Th2 cells and termed cTfh2 cells; and $CXCR3^-CCR6^+$ cells, resembling Th17 cells and referred to as cTfh17 cells. Within the cTfh1, cTfh2, and cTfh17 subsets, distinct subpopulations can be identified based on the expression of PD-1 and ICOS. These subpopulations include an activated subset characterized by high expression of ICOS and PD-1 ($ICOS^+PD\text{-}1^{++}$), as well as two quiescent subsets with varying levels of ICOS and PD-1 expression ($ICOS^-PD\text{-}1^+$ and $ICOS^-PD\text{-}1^-$). The functions of each distinct cTfh subtype in various states exhibit notable disparities between mice and humans, and further investigation is needed to determine their origins and contributions to antibody responses [14], [15].

Investigating the cTfh response following SARS-CoV-2 infection has emerged as an essential contributor in advancing our comprehension of the adaptive immune response against viruses. Due to the virus's capacity to impact both the upper and lower respiratory systems, effective cell-mediated immunity is crucial for clearing the virus initially and establishing immunological memory. The functions of cTfh play a central role in these crucial aspects.

Studies have demonstrated a substantial expansion of cTfh cells in COVID-19 patients, which aligns with observations made in other viral infections and post-vaccination scenarios [16], [17]. This increase in cTfh cell populations has been found to strongly correlate with heightened levels of neutralizing antibodies, suggesting their pivotal role in orchestrating a robust immune response against the virus [18]. Furthermore, the study of these $CD4^+$ T cell subtypes has allowed the identification of distinct subsets of cTfh cells associated with SARS-

CoV-2 infection, providing additional insights into how these cells contribute to both protective and pathogenic immune responses against SARS-CoV-2 infection [19].

Continuing research on the cTfh response holds the key to comprehending its dynamics and regulation, which could have a major impact on the development of vaccines and therapeutic strategies against viral infections.

On the other hand, immunological homeostasis stems from the intricate interplay between protective immune responses and regulatory mechanisms. Any disruption in achieving the proper counterbalance to adaptive immune responses can result in the development of autoimmunity and malignancies. Thus, Treg cells represent a subpopulation of $CD4^+$ T cells with suppressive capabilities, working to counteract immune responses, crucial to maintaining immune homeostasis. An insufficient number of Treg cells can instigate life-threatening autoimmune reactions, while an excessive abundance can induce immune suppression. They are generally characterized by the expression of the transcription factor FOXP3, which is essential for their development and suppressive functions [20]. Various mechanisms have been suggested to explain the suppressive actions of Treg cells, encompassing both direct cell-to-cell contact and the involvement of soluble factors. These mechanisms entail a diverse range of accessory molecules, including cell surface molecules like CTLA4, CD25, TIGIT, CD39, and CD73, as well as cytokines such as IL-10, TGF- β . Additionally, secreted, or intracellular molecules like granzymes and cyclic AMP also contribute to the Treg-mediated suppression [21], [22].

Treg cells exhibit heterogeneity and are characterized by their dynamic nature. Emerging research indicates that Treg cells possess the ability to adapt and specialize within distinct environmental contexts, enabling them to customize their functions and maintain homeostasis across diverse tissues and immune dynamics [23]. One example of this specialization is the regulatory counterpart of Tfh cells, named T follicular regulatory (Tfr). Tfr cells can be found in lymphoid tissue and in circulation, and are specialized in the regulation of GC reaction and prevention of autoimmunity [10].

Sustained FOXP3 expression is vital for preserving the Treg cell program and its suppressive properties outside the thymus. This requires the synergetic action of a network of transcription factors with FOXP3 to stabilize the expression of the Treg cell-specific genes, ensuring the maintenance of a stable and differentiated Treg cell population essential for immune homeostasis [24].

The significance of Treg cells extends to various aspects of viral infections, including COVID-19 instigated by the SARS-CoV-2 virus. This specific subset of $CD4^+$ T cells assumes a pivotal role in controlling the humoral and cellular immune responses triggered by infection, preventing hyperreactive immune responses that could potentially result in tissue damage and the exacerbation of pathological conditions. Emerging research underscores the pivotal role of Treg cells in the complex battle against COVID-19, particularly among patients with mild SARS-CoV-2 infection, where they appear to wield substantial influence over the orchestration of inflammatory responses [25]. However, in severe cases, the immunosuppressive potency of Treg cells appears compromised, raising compelling questions about their intricate involvement in disease progression [19].

1.4 Computational biology and bioinformatics in modern biological research

Over the past decade, significant strides have been made in the modern biological and biomedical domains, with computational biology and bioinformatics emerging as driving forces behind these advancements. This progress has been greatly facilitated by remarkable technological breakthroughs in the field, empowering researchers to push the boundaries of scientific knowledge and accelerate discoveries.

The 14.8-billion DNA bp sequenced in the Human Genome Project [26], as well as all the impacts that this milestone has brought to research, are unquestionably the greatest proof of the essential role that computational biology and bioinformatics have in the field [27]. Without the ability to effectively manage and analyze the massive amount of data generated by the Human Genome Project, a power conferred by computational biology and bioinformatics, the current development in the areas of life sciences would not have been possible.

Furthermore, the emergence of other high-throughput technologies has enabled the qualitative and quantitative analysis of the target molecules of different "omics" studies at a much larger scale than was previously possible. Consequently, the multi-omics field – the genome, epigenome, transcriptome, and proteome – has enhanced the ability to tackle complex biological questions generating big data, which necessitates sophisticated bioinformatic analysis to derive meaningful insights [28], [29].

The transcriptome, in particular, is the complete set of RNA transcripts produced by the genome within a cell or a population of cells at a given time. The analysis of RNA molecules – transcriptomics – enables the capture of a snapshot of actively transcribed genes undergoing translation into proteins – an intermediary between the genotype and the phenotype. Besides this, the transcriptome exhibits remarkable dynamics, due to its high responsiveness to cellular processes and environmental stimuli. Thus, by analyzing transcriptome, we can obtain an early and accurate measurement of cells' response to stimuli [30].

The emergence of Next Generation Sequencing (NGS) technologies has revolutionized the study of the transcriptome by enabling high-throughput and cost-effective sequencing of RNA molecules (RNA-seq). In the past decade, the widespread usage of bulk RNA-seq methods has enabled the analysis of gene expression patterns by quantifying transcripts across a population of cells. However, this approach is constrained to determining the average expression levels of transcripts within the cell population, thereby unaware of cell population heterogeneity [31].

With around 30 to 40 trillion cells comprising the human body, each one harboring a massive amount of biological information, diverse cell types coexist within tissues and organs. Despite sharing morphological similarities, cells exhibit heterogeneity in their particular molecular profile, and conversely, cells with similar molecular profiles may manifest morphological differences. Comprehending these cells' heterogeneity and intricate interactions becomes crucial in uncovering the human body's specialized and diverse physiological states, regulatory mechanisms, and diseases. To obtain a fine-grained understanding of the transcriptomic

landscape within heterogeneous cell populations, the utilization of single-cell RNA sequencing (scRNA-seq) techniques becomes essential [32].

Nowadays, scRNA-seq is a well-established and widely used technique for analyzing the transcriptome of individual cells. It serves as a valuable tool for investigating genotype-phenotype relationships, while also enabling a detailed understanding of dynamic gene expression at a single-cell resolution. Over recent years, advancements in sequencing technologies have led to the development of various scRNA-seq protocols. Each of these methods has distinct characteristics, offering its advantages and disadvantages, leading to different outcomes. The selection of a particular scRNA-seq technology should consider the trade-off between the research objectives and the cost of sequencing [33].

Many infectious diseases have eluded conventional research and treatment methods and persist as significant health challenges. Through single-cell analysis, we gain a deeper understanding of the cellular diversity that determines the immune response to pathogens. scRNA-seq can provide significant insights into the molecular and cellular responses of individual host cells during infection. Moreover, scRNA-seq contributes to the discovery of novel antibodies by providing a thorough comprehension of antigen specificity and the diverse repertoire of B cell clones produced during immune responses. This critical knowledge facilitates the identification and development of antibodies capable of effectively neutralizing pathogens. Additionally, scRNA-seq can empower the monitoring of immune responses to vaccination, providing valuable insights into the cellular dynamics and molecular changes that occur following immunization. By harnessing the potential of scRNA-seq, researchers can gain unprecedented depth into the intricate mechanisms governing host-pathogen interactions and advance the development of innovative strategies for combating infectious diseases [34].

1.5 Aim of the study

The aim of the study was the elucidation of the transcriptional signature of two T cell subsets - cTfh and Treg cells - in a viral infection context and at single-cell resolution.

The cTfh cells play a pivotal role in guiding B cell responses and antibody production, crucial for effective immune defense against infections, including viruses. Conversely, Treg cells act as immune suppressors, preventing excessive immune responses, while their balance is critical to prevent immunopathology and promote controlled antiviral immunity.

The study of cTfh and Treg populations that are present in circulation at a low frequency among the leukocytes, rather than studying the entire pool of $CD4^+$ T cells or PBMCs using a single-cell transcriptomics approach, enables a comprehensive understanding of their heterogeneities and their distinct roles in orchestrating immune responses against viral infections, revealing aspects that have thus far remained incompletely elucidated.

In the first part, we aimed to identify the important cell subsets in cTfh and Treg cells, and their assignment, on both healthy donors and COVID-19 patients, based on the expression of genes characteristic of known cell populations. Following this, we aimed to further understand the transcriptional landscape of our target T cell populations, by pinpointing transcription factors closely associated with each cellular phenotype, bestowing the study with a granular understanding of the regulatory underpinnings intrinsic to individual subpopulations.

This approach is especially important as transcription factors have a low level of expression, being hard to document within single-cell transcriptomes.

Disentangling the mechanisms through which Tfh and Treg cells modulate immune responses and contribute to either immunopathology or protection forms the foundational premise for advancing subsequent translational applications. We, therefore, aim to provide insights that can steer the development of targeted interventions to enhance vaccine efficacy and therapeutic strategies that modulate immune responses in the case of cTfh response or predict treatment responses in the case of Treg cells.

2. Materials and Methods

2.1 Human samples

Fresh peripheral blood samples were obtained from four severely ill patients with COVID-19, admitted to the Centro Hospitalar Universitário Lisboa Norte, in May and June of 2020, and from four adult healthy volunteers. The ages of the COVID-19 patients ranged from 51 to 67, with a median age of 59.5, while for the healthy donors, the ages ranged from 51 to 61, with a median age of 56.

This study was approved by the Lisbon Academic Medical Center Ethics Committee. Everyone involved in the study gave written informed consent for the specimen collection and data analysis.

2.2 Sample Preparation and cell sorting

Peripheral blood mononuclear cells (PBMCs) were isolated from ex-vivo blood samples [35]. Prior to cell sorting, cell suspensions were obtained, and cells were stained with FITC-anti-human CD127 (BioLegend, 1/100), PeCy-7-anti-human CD25 (BioScience, 1/25), Bv711-anti-human CD4 (Biolengend, 1/100), PE-anti-human CXCR5 (BioLegend, 1/100). The Live/Dead Fixable Aqua Dead Stain Kit (Life Technologies, 1/1000) was employed in cell viability assays or live/dead cell discrimination. After double exclusion of duplets by forward and side scatter, single cells of different immune cells populations were sorted on a BD FACSAria Fusion cell sorter as $CD4^{+}CD127^{-}CD25^{+}$ Treg and $CD4^{+}CD127^{+}CD25^{-}CXCR5^{+}$ Tfh cells at the Flow Cytometry Facility of Instituto Medicina Molecular.

2.3 RNA processing for single-cell sequencing and RNA expression quantification

To generate 3′ single-cell transcriptomics datasets from single cells, droplet-based mRNA libraries were produced using a Chromium system (10x Genomics). cDNA synthesis, amplification, and 10x barcoded sequencing libraries were constructed according to the manufacturer's protocol for the 10x single cell 3′ kit (Chromium Next GEM Single Cell 3′ v3.1). All libraries were sequenced on a HiSeq 4000 to achieve an average depth of 60 thousand reads per cell. All experimental procedures up to this stage were carried out by expert members of the Luis Graça Lab.

FASTQ files containing raw reads data from the single-cell sequencing were aligned to the human reference genome GRCh38 (Genome Reference Consortium Human Build 38) using the CellRanger pipeline (version 7.1.0) from the 10x Genomics workflow.

2.4 Downstream analysis

The realm of single-cell analysis is gradually evolving from a challenging specialized field to an established and cutting-edge discipline. When analyzing, it's crucial to consider the significant attributes associated with single-cell RNA-seq datasets.

The first important characteristic of scRNA-seq data is a prevalent challenge in single-cell analysis: dropouts. These arise when the detection of low-abundance transcripts is hindered by technical constraints, leading to inaccuracies [36]. Consequently, scRNA-seq datasets often exhibit a pronounced sparsity, characterized by an abundance of zero values within the data matrix. However, the scope for correcting the data, including addressing high dimensionality, and measurement noise, is constrained. There is a risk of overcorrection that might inadvertently eliminate essential biological insights encoded within the data. Although there is no universally effective algorithmic solution for all the data, it is possible to establish a series of broad steps for pre-processing single-cell data and address the aforementioned challenges. The pipeline comprises six essential stages: filtering data, normalization, feature selection, dimensionality reduction, clustering, and cell type annotation.

The single-cell datasets were analyzed individually in R (version 4.1.3) using the Seurat package (version 4.3.0)[37]. The downstream analysis was performed using the provided workflow on the Seurat website (https://satijalab.org/seurat/).

2.4.1 Quality control

Single-cell RNA sequencing analysis assumes that each occurrence in the dataset represents measurements from a single cell. Low-quality cells, such as empty droplets or doublets, may violate this assumption and can be detected as containing very few genes or aberrantly high gene counts, respectively. In addition, assessing the percentage of reads that map to the mitochondrial genome, and creating a cutoff based on the percentage of counts from mitochondrial genes per barcode is typically performed since dying cells often exhibit extensive mitochondrial contamination. Therefore, filtering non-informative cells and genes holds significant importance not only for accurately deriving biological insights but also for streamlining the computational workload throughout the analysis.

Considering that there were few datasets, quality control was executed through manual assessment involving an examination of the distribution of different quality control covariates (nFeature_RNA, nCount_RNA, and percent.mt). Subsequently, outliers were identified and then excluded from each dataset.

2.4.2 Normalization

The presence of technical biases associated with cell-to-cell variation, such as cell-specific capture efficiency, sequencing depth, or amplification variability, results in considerable disparities in raw scRNA-seq counts that can mask genuine biological disparities.

The core objective behind normalizing scRNA-seq data is to mitigate these biases, ensuring the accuracy and comparability of gene expression measurements across individual

cells, thereby facilitating the extraction of biologically meaningful insights from scRNA-seq experiments [38].

In this study, the *NormalizeData* function was employed. This function utilizes the *LogNormalize* method, a global-scaling normalization technique, to normalize feature expression measurements for individual cells. This involves dividing the feature expression by the total counts for that cell, followed by multiplication with a scale factor (10,000), and ultimately applying a log transformation to the outcome.

2.4.3 Feature selection and scaling of the data

Single-cell RNA-seq datasets often yield tens of thousands of genes, many of which may be not informative or confounding. Therefore, a standard preprocessing workflow encompasses a feature selection step. This stage is designed to identify and select a subset of relevant genes which could encompass those that are most informative or exhibit the highest variability. Through prioritizing relevant features, such as genes associated with cell identity or functional changes, feature selection amplifies result interpretability, reduces noise, and contributes to a deeper understanding of cellular dynamics. This renders it a pivotal step in extracting meaningful insights from intricate, high-dimensional single-cell transcriptomic data.

The identification and selection of highly variable features (2000) was done using the *FindVariableFeatures* function.

Prior to dimensional reduction techniques such as Principal Component Analysis (PCA), it's vital to apply a linear transformation known as scaling. This process adjusts the magnitude of individual genes (mean = 0 and standard deviation = 1), ensuring that the contribution of each gene to variance corresponds to its biological relevance. This adjustment empowers PCA, a variance-based method, to effectively capture the most informative patterns of cellular variability. The absence of scaling could lead to higher-expressed genes overpowering the analysis, potentially overshadowing genes with lower expression yet carrying crucial biological insights. The data was scaled using the *ScaleData* command, accompanied by regressing out the sequencing depth, mitochondrial, and ribosomal expression percentages for each cell, in order to remove unwanted variation in scRNA-seq data, helping to focus the analysis on the biologically meaningful variations in gene expression.

2.4.4 Dimensionality reduction

As mentioned above, the high-dimensional nature of the scRNA-seq data leads to challenges in visualization, interpretation, and computational efficiency. In this pipeline, the first effort to trim down data dimensionality was feature selection. To proceed, dimensionality reduction algorithms were employed to further curtail the dimensions of scRNA-seq data, which play a fundamental role in facilitating data visualization. Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) are among the most commonly used techniques in the realm of scRNA-seq data.

PCA is a method that identifies the axes along which a set of points exhibits the highest variance, essentially representing a greater degree of dispersion. Subsequently, the data is

projected onto these axes, referred to as the principal components (PC). This approach enables the compression of complex expression profiles into a reduced set of uncorrelated variables (principal components), capturing the most informative sources of cellular heterogeneity.

UMAP, on the other hand, serves as a non-linear technique that facilitates the visualization of this reduced-dimensional data in a way that retains the local relationships among cells. This capability facilitates the identification of cell clusters and trajectories, providing a comprehensive perspective on the underlying biological dynamics.

For all the analysis, the *RunPCA* was used. To determine the 'dimensionality' of the dataset, the number of PCs to be used was deduced by considering the significant PCs identified through JackStraw analysis, together with the insights from the "Elbow Plot". Additionally, an evaluation of the genes that contributed most to each of the PCs was performed using the *DimHeatmap* command. The number of PCs retained for downstream analyses were the following: for cTfh cells, 15 on healthy donors data, 16 on COVID-19 patients data; for Treg cells, 17 on healthy donors data, and 17 on COVID-19 patients data.

2.4.5 Clustering

In the context of scRNA-seq data analysis, clustering emerges as a crucial step, holding remarkable importance in uncovering the complexities of cellular diversity. Its role lies in defining distinct clusters of cells, culminating in an organized and more interpretable partitioning of the vast and unstructured initial dataset.

Community detection is widely used to understand the arrangement of complex networks, an example of which is the entire set of cells and their distinct expression profiles. In this study, we employed the Louvain algorithm, a heuristic approach designed for detecting communities. The algorithm functions through an iterative process of relocating nodes between communities to optimize a modularity measure. Modularity, in turn, assesses the effectiveness of grouping network nodes into distinct communities. By employing the Louvain algorithm on scRNA-seq data, it becomes possible to unveil biologically significant cell subpopulations that are delineated by common gene expression patterns.

The clustering step was executed using Seurat's *FindNeighbors* and *FindClusters* functions. The *FindNeighbors* function requires as input the dimensionality of the dataset previously defined. The *FindClusters* command involves a resolution parameter, which establishes the level of detail in the downstream analysis. Elevating this parameter value results in the generation of a higher number of communities or clusters. Depending on the analysis, different resolutions were used for cluster assessment. To prevent the selection of a resolution that deviated from accurately representing the actual number of cell populations and their heterogeneity, each resolution choice was consistently accompanied by an assessment of gene expression patterns. Furthermore, a cluster tree was created using the *clustree* function [39] as an additional validation step. Clustering used the following resolutions: for cTfh cells, 0.2 on healthy donors data, 0.4 on COVID-19 patients data; for Treg cells, 0.4 on healthy donors data and 0.3 on COVID-19 patients data.

2.4.6 Cell Type Annotation

Following the completion of dataset pre-processing and the determination of the number of clusters, the subsequent task entailed unraveling the distinct cellular phenotypes associated with each cluster within the dataset. And to do so, differential gene expression was computed using the *FindAllMarkers* function with the default settings: logFC threshold above 0.25; P-value of less than 0.05 using the non-parametric Wilcoxon Rank Sum test. Using the *FindAllMarkers* command, we conduct comparisons between each cluster and all other clusters in order to pinpoint differential expressed genes.

2.4.7 Data integration

Recent technological progress has significantly enhanced our capacity to produce high-throughput single-cell gene expression data across various methodologies that capture diverse cellular attributes. One prominent analytical hurdle involves the integration of these datasets from multiple sources or experimental conditions, mitigating batch effects, and enhancing the accuracy of downstream analyses.

Within Seurat's toolkit, commands such as *FindIntegrationAnchors* and *IntegrateData* are present. These tools enable the identification of common sources of variation among different datasets and the subsequent integration of these datasets into a unified framework. The first step involves therefore the identification of cells with shared biological states between multiple scRNA-seq datasets using the *FindIntegrationAnchors* function. These anchors serve as reference points for aligning cells from different datasets into one harmonize dataset, using the *IntegrateData* function.

In this study, we employed integration to jointly analyze three datasets of cTfh cells sourced from COVID-19 patients and four datasets of Treg cells, all derived from COVID-19 patients. Throughout the analysis, default parameters were consistently applied, except for employing 3000 variable features instead of the default 2000.

2.5 Gene Set Variation Analysis (GSVA)

Gene Set Variation Analysis (GSVA) stands as a computational technique that estimates the enrichment of predefined pathways or the functional annotations of specific samples. In contrast to the majority of Gene Set Enrichment (GSE) methods, GSVA adopts a comprehensive perspective on gene set behavior, thereby fostering robust and versatile analyses centered around pathways. This characteristic renders it a potent resource for decoding high-throughput transcriptomic data.

For the purpose of inferring the enrichment of pathways within our cell subsets, a GSVA was conducted, using the Bioconductor's *GSVA* package [40]. In this procedure, Normalized Enrichment Scores (NES) were computed for distinct gene sets (listed in the annexes) across the various clusters present in each dataset.

2.6 Regulatory Network Analysis

To uncover distinct regulatory networks within each cluster, a Regulatory Network Analysis was conducted using the well-established *pySCENIC* pipeline [41]. This pipeline entails three key steps. Firstly, genes that are coexpressed with transcription factors (TFs) are identified through *GRNBoost*. Given the exclusive dependency of GRNBoost modules on co-expression patterns, these modules include both direct and indirect targets, along with positively and negatively correlated ones, as well as potential false positives. In the subsequent stages, only the positively correlated modules are considered, and to determine the modules founded on direct-binding regulatory interactions, each coexpression module undergoes cis-regulatory motif analysis using the *RcisTarget* algorithm. The final step involves the assessment of regulon activity in individual cells.

This workflow was applied to our Tfh dataset from COVID-19 patients, performed separately for each patient - resulting in three distinct analyses. To integrate the outcomes, a rank-product analysis was employed, employing the *RankProduct* function [42]. The objective was to capture regulons consistently enriched within each cluster.

3. RESULTS

3.1 Sex-specific cTfh cells in healthy donors

By analyzing the transcriptome of ~21,000 cTfh cells, we assessed the heterogeneity within cTfh cells from healthy donors. A Uniform Manifold Approximation and Projection (UMAP) visualization of these cells revealed their partition into three main clusters (Figure 1A). We observed that the predominant factor contributing to the segregation of the three resultant clusters and the primary source of variability was attributed to donor origin, with cluster 1 notably stemming from a specific donor (Donor 4), as depicted in Figure 1B.

The assessment of differentially expressed genes in each cluster revealed sex-linked genes in both cluster 0 and cluster 1 (Figure 1C). Cluster 0 represents cells expressing higher levels of genes located on the Y chromosome, such as UTY and PRKY, whereas cluster 1 exhibited high expression of X-linked genes (Figure 1C and D). Finally, cluster 2, predominantly consisting of cells sourced from donors 1, 2, and 3, exhibited strong expression of granzymes and other cytotoxic-related elements.

Consequently, with an approach designed to investigate heterogeneity among cells of the same type (cTfh cells). The donor characteristics, namely sex, can be an important source of variation, not necessarily related to cell biology.

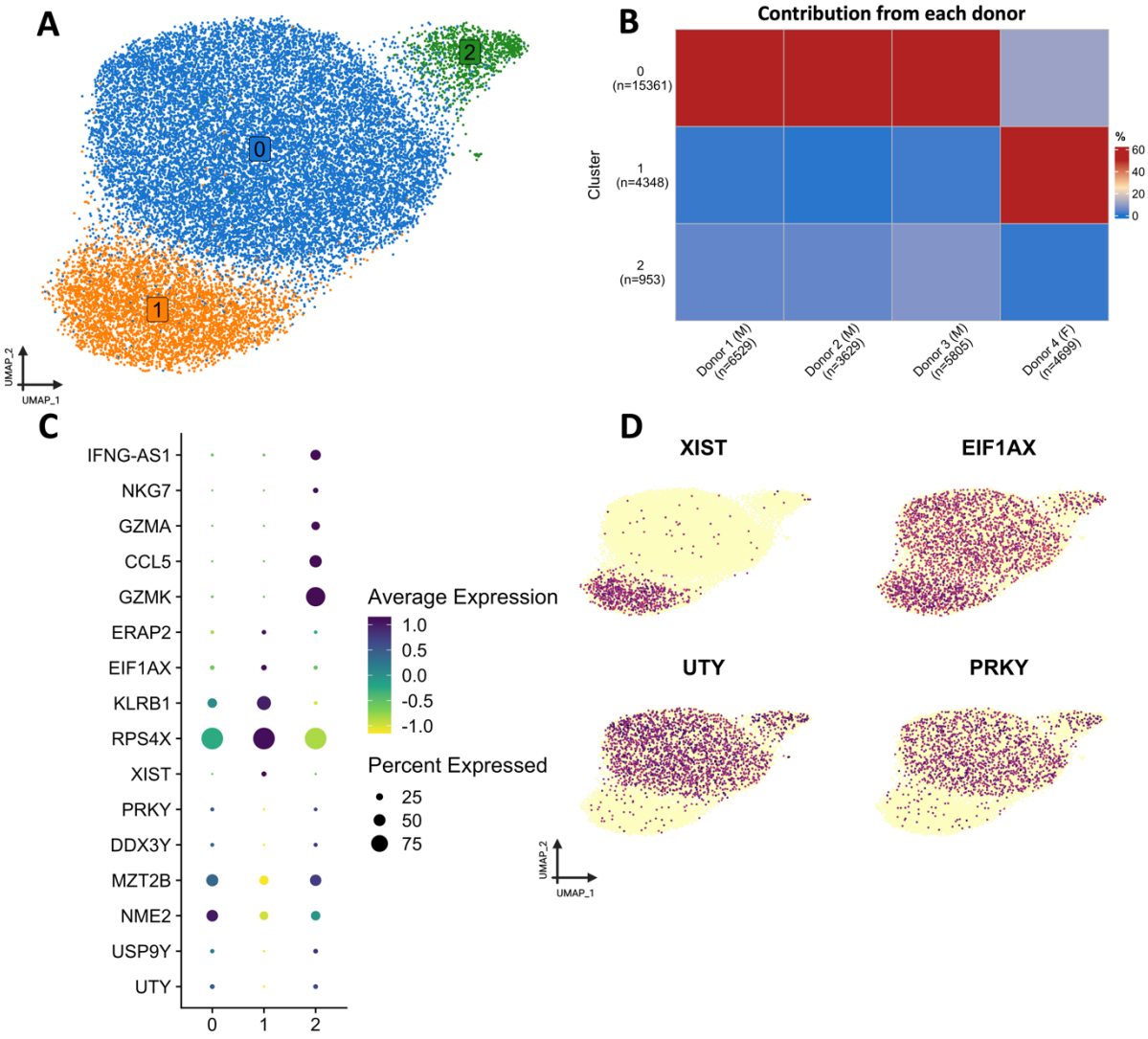

Figure 1—**The single-cell transcriptome of cTfh cells displays sex-specific characteristics. (A)** The UMAP representing all cTfh cells from four healthy donors revealed the emergence of three distinct clusters. **(B)** The heatmap illustrates the distribution of the frequency of cTfh cell subpopulations within each cluster across the different donors, including three males (donors 1, 2, and 3) and one female donor (donor 4). Cluster 1 consists entirely of cells from donor 4, reflecting the impact of donor effect on the dataset. **(C)** Dotplot showing the top differentially expressed genes for each individual cluster. **(D)** Projection of two X chromosome genes (XIST and EIFAX) and two Y chromosome genes (UTY and PRKY) expression onto the UMAP.

3.2 GZMK expression as the primary source of segregation among cTfh cells from healthy donors

To ensure the accuracy and reliability of our downstream analyses and interpretations, it became imperative to mitigate the donor effect from the dataset, as it was merely acting as a confounding factor that may overshadow genuine biological variations that could exist among cTfh cells.

After removing the donor effect, our subsequent analysis of cTfh cells unveiled the presence of three main clusters (Figure 2A). As illustrated in Figure 2B, following batch correction, the donor-specificity had no impact on the clustering of these cells. Upon conducting a more

detailed analysis of these cells, which included examining the loadings of the PCs and identifying differentially expressed genes, we identified GZMK expression as the predominant source of segregation among healthy donor cTfh cells. Cluster 1 represents cells highly enriched for GZMK and cytotoxicity-associated transcripts (NKG7, GZMA), as well as the chemokine CCL5. Cluster 1 also exhibited expression of CXCR3. Cluster 0, in contrast, was characterized by low GZMK expression (GZMK-), while exhibiting CCR6 expression together with some other genes associated with a Th17 signature (Figure 2C and D). We investigated different clustering resolutions. Even with higher clustering resolutions, the algorithm failed to effectively segregate the cells based on CCR6 and CXCR3 expression (Figure S1). Cluster 2 emerged as a distinct cluster, consisting of a mere 22 cells. Notably, this cluster exhibited a robust and statistically significant expression of HLA-DR genes, typically associated with activated T cells.

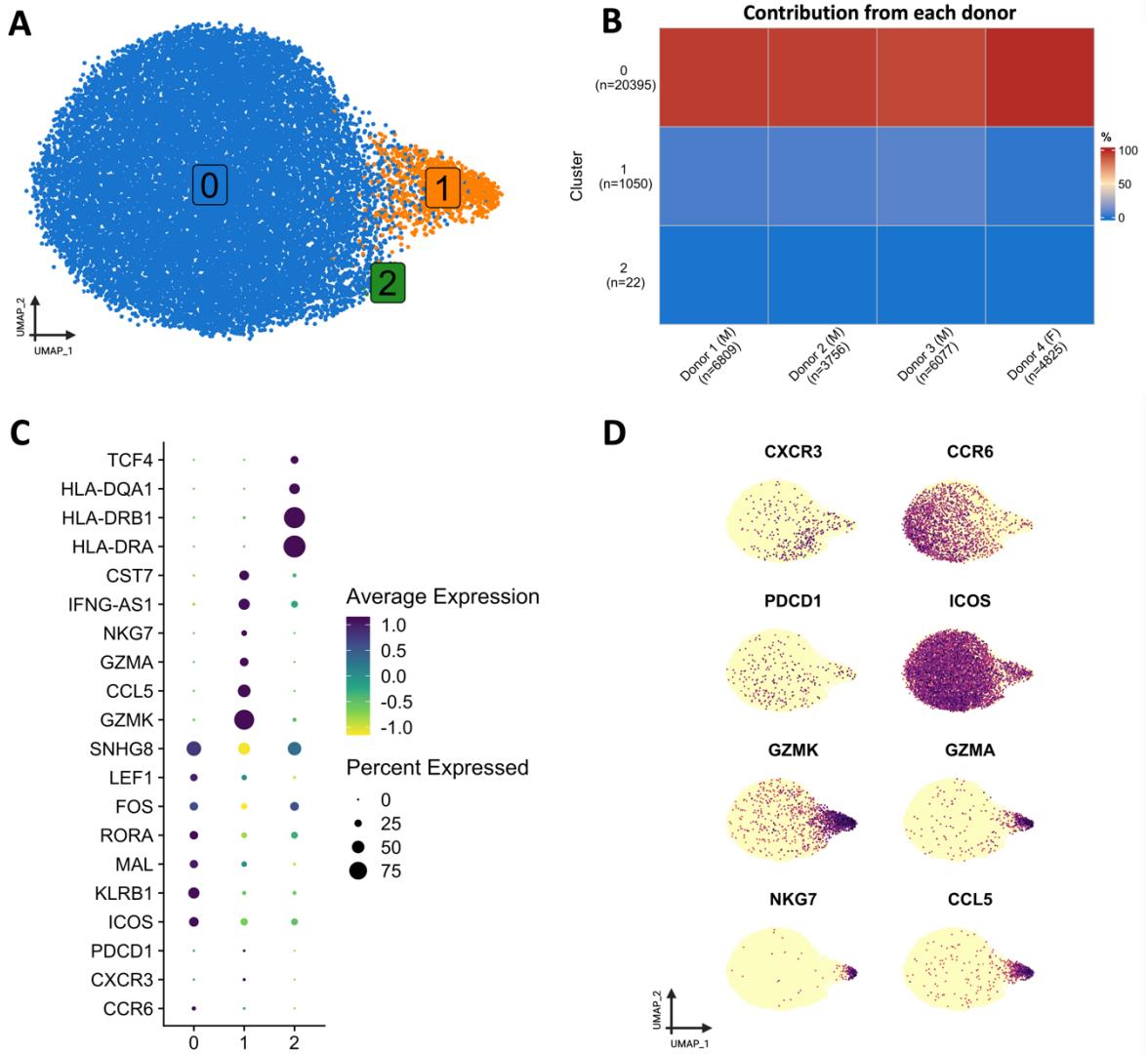

Figure 2— **Analysis of the scRNA-seq dataset of cTfh cells from healthy donors after correcting the donor effect.** **(A)** UMAP projection of all cTfh cells from all four healthy donors. Three distinct cTfh clusters emerged. **(B)** The heatmap illustrates the contribution of each donor to each cluster. After removing the sex effect, the donor specificity did not significantly affect one particular cluster. **(C)** Dotplot displaying the top differential expressed genes for individual clusters, with ribosomal and mitochondrial genes excluded. **(D)** Projection of the expression of genes

associated with the signature of distinct cTfh subtypes (CXCR3 and CCR6) and cTfh maturation stage (ICOS and PDCD1), as well as markers from cluster 1.

3.3 Distinct clustering patterns and phenotypic shifts in cTfh cells of COVID-19 patients

Following the integration of the single-cell transcriptomes of all cTfh cells collected from three COVID-19 patients, our analysis unveiled five distinct cTfh cell subsets, each characterized by a particular transcriptional program (Figure 3.1A). The donor-specificity had no impact on the clustering of cTfh cells from COVID-19 patients (Figure 3.1B). The segregation of the clusters was essentially defined along two axes: the well-defined cTfh subtypes (cTfh1, cTfh2, and cTfh17 subsets), characterized by the expression of CXCR3 and CCR6, and cell-state specific phenotypes reflecting their activity (i.e., active, or quiescent), as determined by the expression of PDCD1 and ICOS (Figure 3.1C and D). Cluster 0 was identified as cTfh17 cells, displaying a quiescent state and a robust CCR6 expression. Cluster 1 was assigned as cTfh1 cells due to pronounced CXCR3 expression. Cluster 2 was annotated as representing cTfh2 cells, characterized by the absence of CXCR3 and CCR6 expression, in addition to exhibiting a phenotype consistent with quiescent T cells. Cluster 3 consisted of $CXCR3^{+}$ cells exhibiting a strong expression of GZMK and cytotoxicity-associated genes. Finally, Cluster 4 was identified as activated cTfh1 cells, as it expressed CXCR3 and displayed a markedly activated state, with strong expression of PDCD1 and HLA-DR genes (Figure 3.1D and E).

Transcription factors tend to be expressed at low levels, being difficult to study by scRNA-seq. Consequently, to gain further insights into the regulatory networks governing the distinct cTfh cell clusters identified in our study, we conducted a comprehensive Regulatory Network Analysis using the well-established *pySCENIC* pipeline (Figure 3.2). Our analysis of the cTfh17 and cTfh2 clusters, characterized by a more resting cellular profile, unveiled a common regulatory signature. Within the top fifteen regulons the cTfh17 and cTfh2 clusters shared ten regulons, indicating a convergent regulatory profile (Figure 3.2). cTfh17 and cTfh2 clusters were found to be under the control of transcription factors associated with resting T cells, including FOXP1 and ETS1. Despite the heterogeneity observed within the cTfh1 compartment, all cell populations within this group exhibited a shared up-regulation of ZNF580 and CEBPB transcription factors associated with IL-8 production and immune responses, particularly inflammation (Figure 3.2). GZMK-expressing cTfh cells were notably regulated by TBX21, the master Th1 transcription factor, underscoring their Th1 signature. Additionally, our analysis revealed the upregulation of STAT1 in activated cTfh1 cells, suggesting a pivotal role for STAT1 in this activated subset (Figure 3.2).

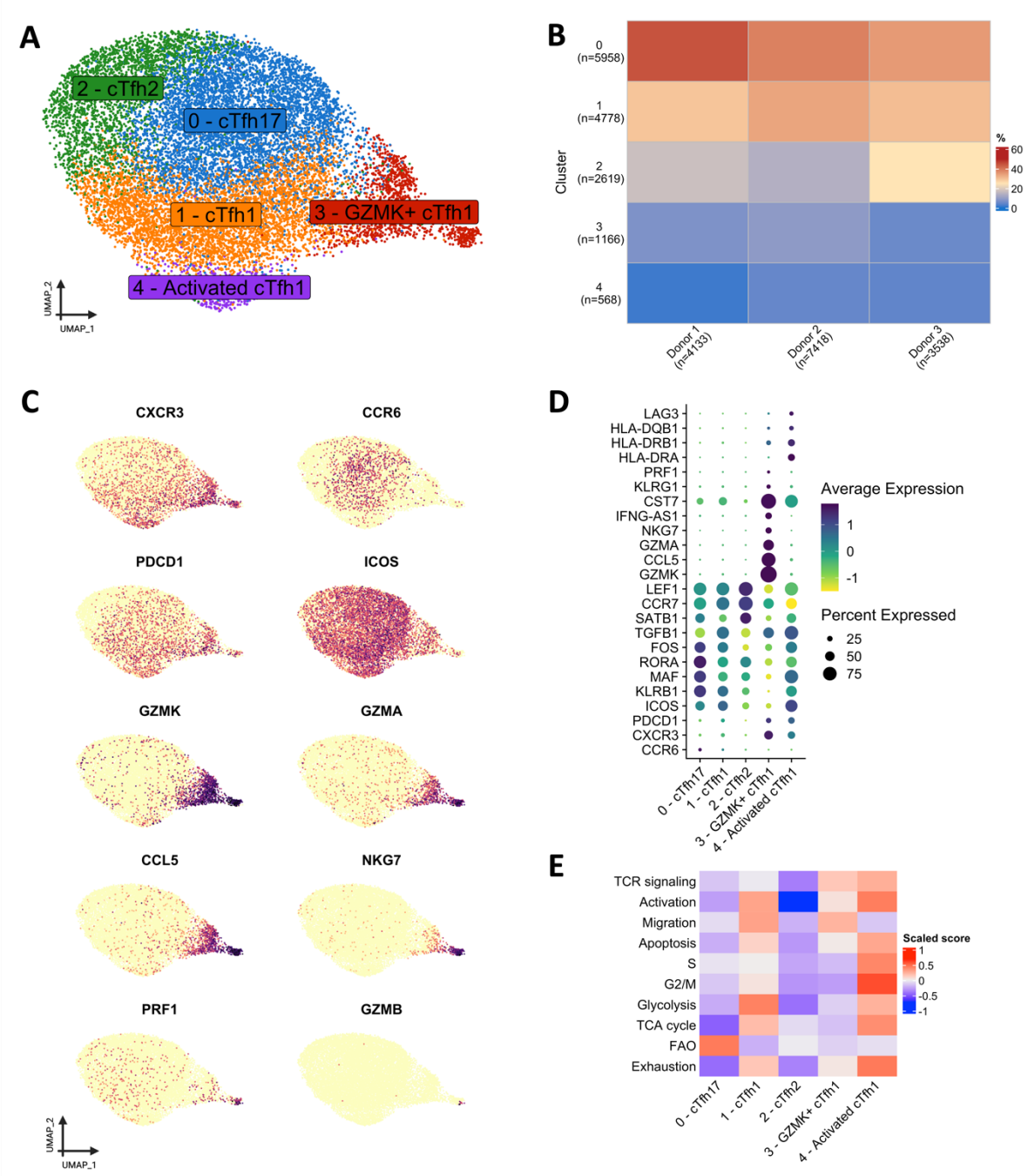

Figure 3.1— **scRNA-seq of cTfh cells from blood uncovers distinct subsets within the cTfh cell population. (A)** UMAP visualization of the cell populations after integration with Seurat. We obtained five different clusters of cTfh cells. **(B)** Heatmap showing the distribution of each cell population across the different donors. **(C)** UMAP feature plots showing the expression of well-defined markers of cTfh cell activation (ICOS and PDCD1), genes that define cTfh1, cTfh2, and cTfh17 subsets (CXCR3 and CCR6), and the expression of GZMK and other co-expressed genes.

(D) Dotplot displaying the top differential expressed genes for individual clusters, with ribosomal and mitochondrial genes excluded. **(E)** Heatmaps illustrating the GSVA enrichment scores of feature pathways for each cluster.

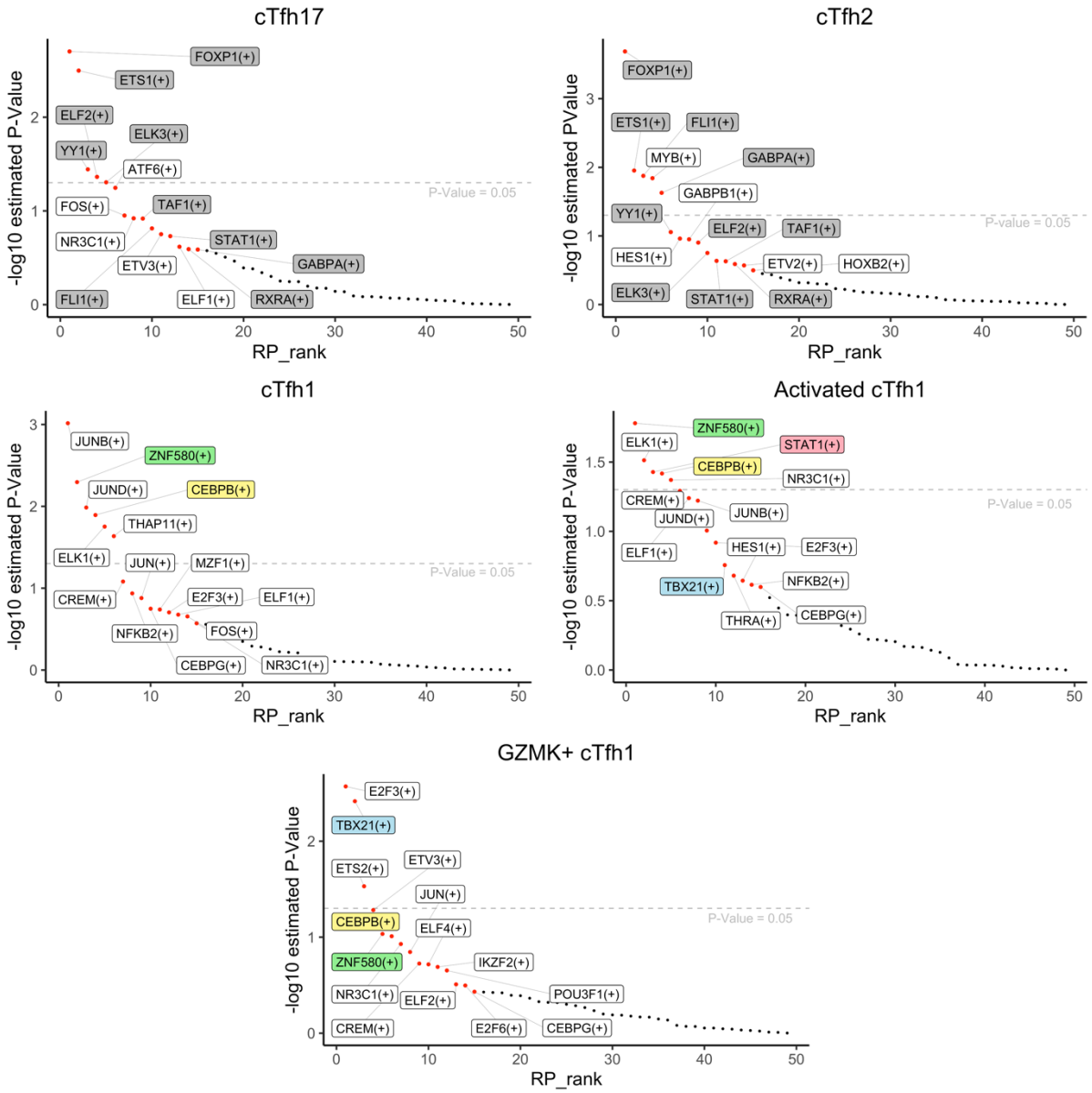

Figure 3.2— **Regulon specificity score for each cTfh cluster from the COVID-19 dataset.** The top fifteen regulons in each donor are highlighted in red and labeled on the plots. From these lists of regulons, ten of them were shared between the resting non-cTh1 clusters (grey label). The cTfh1 clusters showed a common upregulation of ZNF580 and CEBPB transcription factors. TBX21, the master Th1 transcription factor, was regulating GZMK-expressing cTfh cells, while STAT1 was enriched in the Activated cTfh1 cluster.

3.4 Unveiling unique subsets within Treg cells of healthy donors through scRNA-seq

Upon analyzing the scRNA-seq datasets obtained from Treg cells sourced from the blood of four healthy donors, our investigation unveiled the presence of four distinct clusters, as illustrated in Figure 4A. Importantly, these clusters were uniformly distributed among the different donors, as depicted in Figure 4B.

To gain insights into the nature of each cluster, we conducted an analysis of differentially expressed genes within each cluster, enabling their characterization. Cluster 0 exhibited a strong expression profile of genes typically associated with resting T cells, including CCR7 and TCF7 (Figure 4C), which was further confirmed by GSVA analysis (Figure 4D). The remaining three clusters (Clusters 1-3) displayed a heightened state of activation, as evidenced in Figure 4D. Cluster 1, emerged as the most activated, characterized by a robust expression of HLA-DR genes. Cluster 2 displayed a pronounced Th17 signature, marked by the expression of CCR6, KLRB1, and RORC, while cluster 3 was identified as GZMK^{+} Treg cells due to the striking expression of GZMK. Evaluation of the predicted suppressive capacity, conducted through GSVA analysis, unveiled that Clusters 1, 2, and 3 exhibited a transcriptome with genes characteristic of a robust suppressive capacity (Figure 4D).

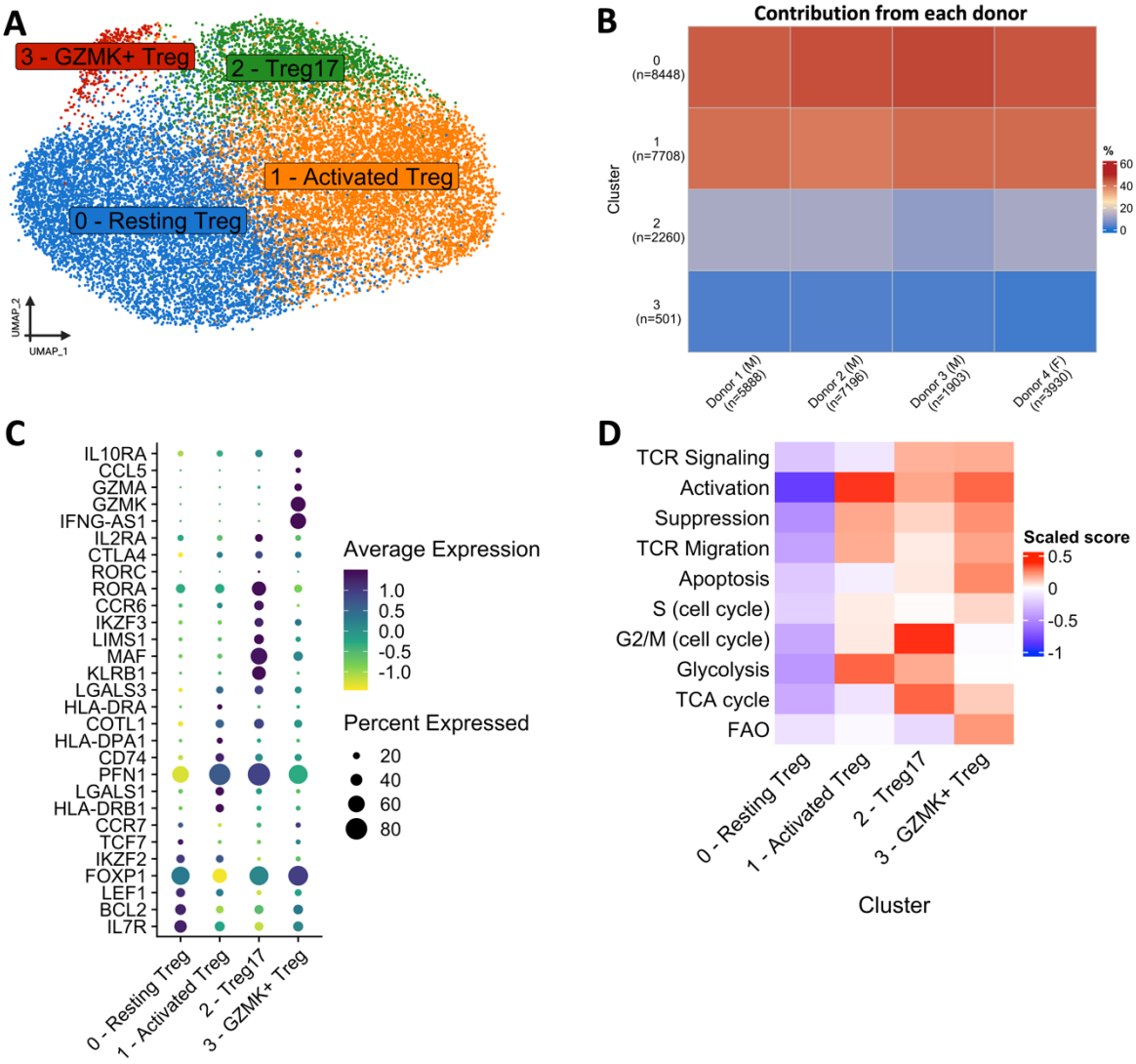

Figure 4—**scRNA-seq of Treg cells from the blood of healthy donors shows four different cell populations**. **(A)** UMAP visualization representing the four distinct clusters assigned to the different cell populations. **(B)** Heatmap showing the distribution of each cell population across the different donors. **(C)** Dotplot displaying the top

differentially expressed genes for individual clusters, with ribosomal and mitochondrial genes excluded. **(D)** Heatmaps illustrating the GSVA enrichment scores of feature pathways specific to Treg cells for each cluster.

3.5 Six distinctive subsets revealed by scRNA-seq among Treg cells from the blood COVID-19 patients

Next, we studied a dataset containing the integration of more than 23,000 cells obtained from the blood of four COVID-19 patients with severe disease (characterized as requiring hospital admission). The analysis of this dataset led to the identification of six distinct clusters (Figure 5A). The contribution of each patient to the final dataset is visually represented in Figure 5B. The segregation of cells in this context was primarily dictated by two critical factors: the activation status of the cells and their distinctive signature concerning polarization towards distinct effector $CD4^+$ T cells (i.e., Th1, Th17). The two biggest clusters, Cluster 0, and Cluster 1, exhibited an expression profile consistent with quiescent Treg cells, characterized by the presence of CCR7 and TCF7 expression (Figure 5C and D). Cluster 2 perfectly exemplified the active state of a Treg cell, with a $CCR7^{low}TCF7^{low}$HLA-DR^{high} profile. Cluster 5 mirrored the signature of cluster 2 cells, albeit with a slightly less active expression pattern. Clusters 3 and 4 exhibited a high expression of genes associated with the Th17 and Th1 signatures, respectively, revealing their distinct CD4 effector lineage characteristics. Cluster 3 cells were enriched for $CCR6^+KLRB1^+RORC^+$ cells, while cluster 4 cells were enriched for $CXCR3^+$ cells, a chemokine associated with type-1 immune responses. Thus, the cells in these clusters were annotated as Treg17 and Treg1, respectively.

GSVA analysis unveiled a predicted progressive increase in suppressive function, as indicated by spanning from resting to activated subsets. Remarkably, the Treg17 and Treg1 clusters demonstrated some of the highest signature scores for TCR signaling and suppression activity, among all the identified clusters. The HLA-DR^{high} clusters exhibited the most pronounced score of T cell activation. However, it is noteworthy that, based on the GSVA results, what distinguishes clusters 2 and 5 is their remarkable association with a suppression signature, coupled with the expression of cell cycle gene sets.

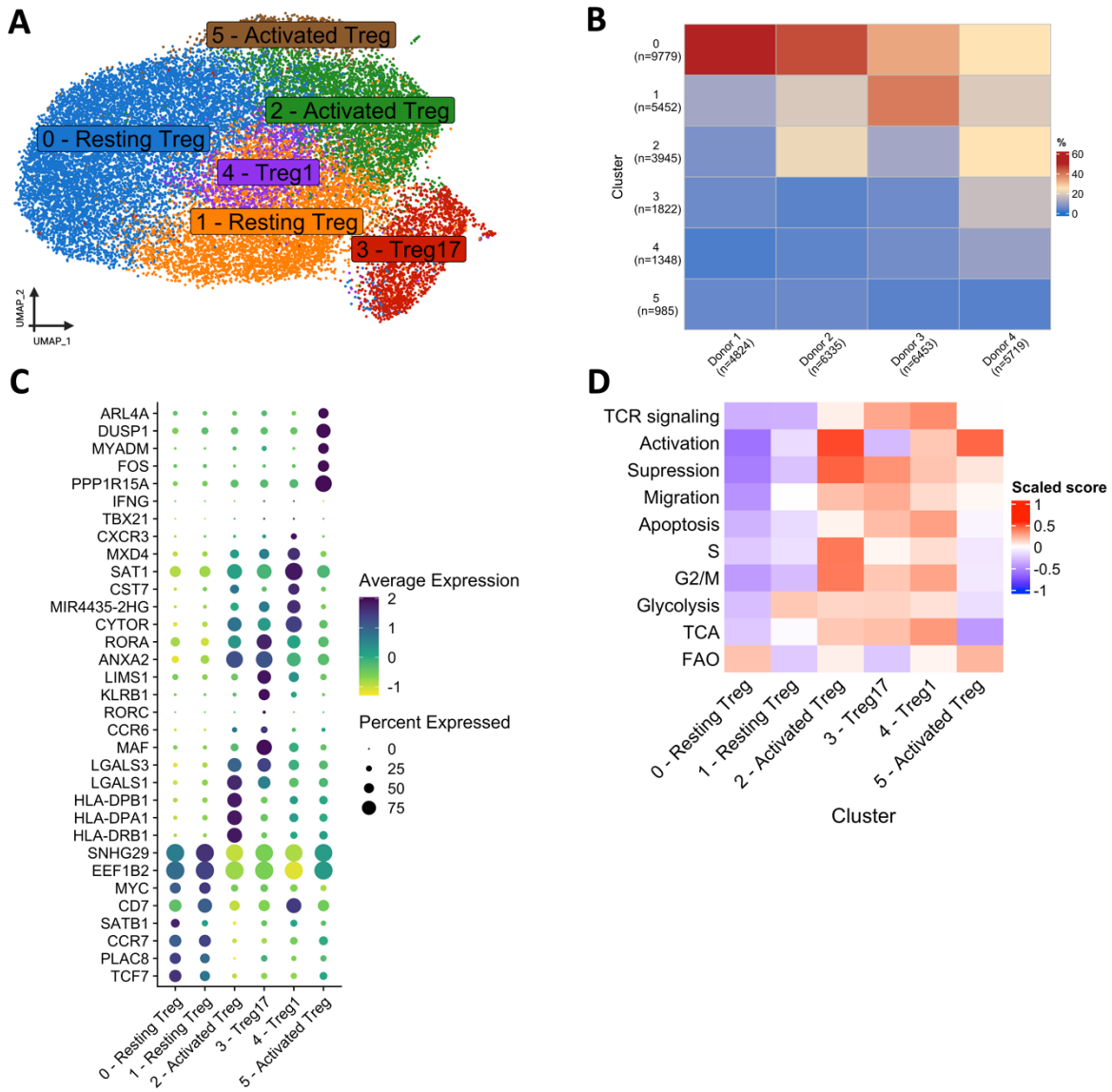

Figure 5—**scRNA-seq of Treg cells from the blood of severe COVID-19 patients. (A)** UMAP visualization of the integrated dataset (n=23,331). We found six subpopulations of Treg cells in the blood of COVID-19 patients. **(B)** Heatmap showing the distribution of each cell population across the different donors. **(C)** Dotplot displaying the top differentially expressed genes for individual clusters, with ribosomal and mitochondrial genes excluded. **(D)** Heatmap illustrating the GSVA enrichment scores of feature pathways specific to Treg cells for each cluster.

4. Discussion

Protection against external agents, such as viruses, hinges upon the effectiveness of the B cell response and the production of high-affinity antibodies. This protection is enhanced by the pivotal activity of Tfh cells. Conversely, Treg cells play the role of immune suppressors, wielding the power to control excessive or misdirected immunity. Consequently, comprehending and characterizing the intricate workings of these two $CD4^+$ T cell subsets serves as a foundational goal for both the design of prophylactic and therapeutic vaccines. Yet, the inherent heterogeneities and distinct function of these two distinct cell types, led us to examine the cell populations through a single-cell transcriptomics approach. Indeed, the heterogeneity of Tfh and Tfr cells remain poorly understood in the context of immune responses against viral infections. Our study elucidated the transcriptome and diversity of cTfh and Treg cell subpopulations that have manifested in response to a viral infection, shedding light on the intricacies of their emergence and dynamics.

To investigate the transcriptome of the two subsets of T cells - cTfh and Treg cells - in the context of viral infection and with a single-cell resolution, it was essential to first analyze the transcriptional profile of these cell populations in healthy donors. We generated one scRNA-seq dataset from cTfh cells obtained from the blood of four healthy donors, which were initially combined (before sequencing), and subsequently demultiplexed using the *Souporcell* algorithm to distinguish individual donor contributions. The approach of pooling cTfh cells from four healthy donors into a scRNA-seq dataset, followed by subsequent demultiplexing using specialized tools to differentiate individual donor contributions, not only demonstrates cost-effectiveness but also effectively mitigates technical variability in the dataset.

However, the initial analysis of the scRNA-seq dataset derived from cTfh cells of healthy donors uncovered an unexpected influence of donor effects on the clustering of the cells (Figure 1B). Upon closer examination of the expression profiles within the distinct clusters, it became evident that the differential gene expression was primarily associated with genes located on the sex chromosomes (Figure 1C and D). Notably, the pronounced expression of the XIST gene in cluster 1, compared to those in the other clusters, emerged as a particularly significant finding. The XIST gene plays a crucial role in the phenomenon of lyonization, a developmental process in which one of the two X chromosomes in each cell is randomly chosen for early inactivation. The XIST gene encodes for an RNA that coats one of the X chromosomes, resulting in a transcriptional silencing [43]. In the context of scRNA-seq analysis of cTfh cells, this observation, coupled with the absence of gene expression associated with the Y chromosome in cells from donor 4, in contrast to cells from the remaining donors, shows that among a dataset that is very homogeneous (all cells are FACS-sorted Tfh cells) the sex-specific transcripts are readily detected. The putative impact of sex variations in Tfh biology remains to be addressed. However, sex differences in the immune system, including in Tfh cells, have been noted, including differences in immune cell composition, cytokine and chemokine production, and susceptibility to certain diseases. The humoral response, in response to infection and vaccination, has been noted to be slightly but significantly higher in females. These

differences can be influenced by a variety of factors, including sex hormones, genetics, and environmental factors [44]–[46]. To further understand the significance of sex-specific cTfh cell populations, future studies should encompass a broader spectrum of approaches and analyses. Analyzing publicly available datasets of cTfh cells or conducting supplementary validation experiments, such as flow cytometry, to confirm the existence of sex-specific cTfh cells at the protein level may be a way to better understand this observation.

The pronounced donor-related segregation observed within the scRNA-seq dataset had noteworthy implications for our analysis. This donor effect introduced a confounding factor that may overshadow biologically relevant differences. Therefore, it became crucial to disentangle this donor-related variability from any genuine biological variations that could exist among cTfh cells. By using the *ScaleData* function, this factor was removed from the possible segregation drivers in the dataset.

After comprehensively analyzing the single-cell RNA-seq datasets derived from cTfh cells, we were able to determine the distinctive states adopted by these cells isolated from healthy donors and COVID-19 patients. In the context of cTfh cells from healthy donors, a simple clustering strategy unveiled that the driving variance lay in the degree of GZMK expression. Specifically, Cluster 0, characterized by the absence of this granzyme subtype, exhibited robust CCR6 expression. Notably, within Cluster 0, the algorithm failed to effectively segregate cells expressing CCR6, even at higher resolutions (Figure 2 and S1). Conversely, the analysis of cTfh cells from severe COVID-19 patients revealed a dominant partition based on two key factors: activation status and cTfh subtypes (Figure 3.1 C and D). In the context of this dataset, we can readily delineate distinct clusters based on the nomenclature established in the comprehensive review by Schmitt *et al.* [15]. Specifically, Cluster 0 aligns with $CCR6^+$ cells, characterizing them as cTfh17 cells; Clusters 1, 3, and 4 are $CXCR3^+$ cells, representing cTfh1 cells; and, lastly, Cluster 2 corresponds to cTfh2 cells, distinguished by the absence of both CCR6 and CXCR3 expression. A pivotal study led by Kiner and his team [47], employing scRNA-seq, hypothesized that the phenotypes of Teff cells are primarily distinguished by the infecting agents rather than conventional Th cell types. Furthermore, their findings suggested that the transcriptional signatures of $CD4^+$ T cells resist clear categorization into discrete Th cell subtypes, even in scenarios where focused differentiation might be expected. However, within the scope of our research, which centers on the analysis of cTfh cells in the context of a specific viral infection, we observed a notable divergence from this paradigm. Here, we found that the clustering of cTfh cells indeed aligns with Th subtypes. It is essential to clarify that this interpretation of cell heterogeneity via clustering does not negate the presence of plasticity between these states, particularly given that these cells coexist within a continuum in our dataset, signifying substantial transcriptional homogeneity [48], [49]. So, it suggests that, under the influence of an acute viral infection, the infection triggers the emergence of a polarized continuum, which can be effectively separated using clustering algorithms.

We also found that the COVID-19 cTfh cells in the active state ($ICOS^+$ and $PD1^+$) are predominantly cTfh1 ($CXCR3^+$ cells), cells reported to be associated with immune responses against intracellular pathogens, such as viral infections (Figure 3.1C). This finding aligns with previous observations regarding the behavior of cTfh cell populations following viral infections, underscoring the pivotal role of $CXCR3^+$ cells in enhancing the immune response against

viruses, particularly in stimulating antibody production [50]. On the other hand, in the context of healthy individuals, our results revealed a noteworthy contrast: there is a scarcity of active cTfh cells (22 cells) (Figure 2). This observation underlines the substantial disparities in the cTfh cell landscape between healthy individuals (with most cells in a quiescent state) and individuals affected by COVID-19 (with a large proportion of activated cTfh cells), emphasizing the substantial changes in cTfh cell dynamics caused by viral infection.

In both our study groups, consisting of healthy donors and COVID-19 patients, an intriguing observation emerged regarding the presence of $GZMK^+$ cells, which clearly exhibited a type-1 associated phenotype, as indicated by their co-expression of CXCR3. Remarkably, when scrutinizing the $GZMK^+$ cell population from COVID-19 patients, a thorough examination of the regulatory network employing the well-established pySCENIC algorithm corroborated their type-1 identity by unveiling the presence of the master transcriptional factor that governs the Th1 responses within this cluster. This discovery provided strong evidence of their distinctive differentiation pathway. Further, this $GZMK^+CXCR3^+CCL5^+NKG7^+GZMA^+$ cell population is not exclusive to active immune responses or inflamed conditions. Its presence in the cTfh cell repertoire of healthy donors raises intriguing possibilities. One plausible interpretation is that these cells may originate from past viral encounters, which resulted in the induction of a type-1 response. This finding can underscore the potential persistence and imprint of prior immune challenges on the cTfh cell landscape, already described in other viral infections [51].

The study conducted by Meckiff et al. [19], provided the first characterization of cytotoxic cTfh cells in humans. However, their findings primarily revealed the expression of granzyme B, a well-studied granzyme. In contrast, GZMK, found in our study, is not associated with cytotoxic responses, and thought to be considered as a different population. The function of GZMK, produced by different cell types associated with type-1 responses remains relatively obscure [52]. Further, $GZMK^+$ $CD4^+$ and $CD8^+$ T cells have been observed in various inflammatory diseases, consistently associated with the absence of GZMB expression, but their function has not been explored [53]. For all these reasons, investigating the functions and significance of GZMK-expressing cells, and cTfh cells, within inflammatory contexts holds considerable relevance.

Our subsequent objective was to pinpoint the specific transcription factors responsible for governing the expression of these distinctive genes across various cell subsets within cTfh subsets from COVID-19 patients (Figure 3.2). The clusters in a resting state (cTfh17 and cTfh2 clusters) showed a common regulatory profile sharing 10 of the top 15 highest scored regulon. Consistently on the top of both clusters are FOXP1 and ETS1, transcription factors reported to be preferentially regulating T cells in the quiescent state [54], [55]. In the context of cTfh1 cells from COVID-19 patients, the prominent transcription factors orchestrating gene regulation were consistently identified as ZNF580 and CEBPB. ZNF580 plays a critical role in augmenting the release of the pro-inflammatory cytokine IL-8. Elevated IL-8 levels are often associated with an enhanced inflammatory response [56]. This finding may suggest that cTfh1 cells with increased ZNF580 expression are contributing to a more robust inflammatory environment, which can be beneficial in responding to infections. Notably, the action of CEBPB plays a crucial role in modulating positively the expression of numerous genes associated with

inflammatory processes, including the proinflammatory cytokine IL-12 [57]. IL-12 is known to be a key mediator in promoting robust immune responses, particularly in the induction of IFNG gene production and the facilitation of the Th1 cell differentiation [58]. This observation underscores the pivotal position of the CEBPB transcription factor within these specific cTfh1 cell populations, as it actively influences the immune response by driving the expression of IL-12. Finally, we found that the activated cTfh1 cells are significantly regulated by the transcription factor STAT1. This finding suggests a regulatory mechanism similar to that previously described in $CD8^+$ T cells, where STAT1 plays a pivotal role. In $CD8^+$ T cells, STAT1 is recognized for its involvement in direct type I interferon (IFN) signaling, a process essential for the efficient expansion of $CD8^+$ T cells in response to viral infections in vivo. Moreover, the function of STAT1 extends to enhancing the survival of activated $CD8^+$ T cells, as evidenced by its role in promoting the persistence of these cells upon viral infection in vivo [59]. The parallels between the regulatory roles of STAT1 in cTfh1 cells and $CD8^+$ T cells underscore the potential significance of STAT1 in orchestrating antiviral responses.

We also profiled with single-cell resolution the Treg cell heterogeneity in the context of a COVID-19 infection. For that, our work elucidated the phenotype of Treg cells in healthy donors and in viral infection (Figures 4 and 5). We resolved Treg cells from healthy donors and COVID-19 patients into four and six subsets, respectively, continuously encompassing the resting and activated stages, as well as lineage-committed subsets of Treg cells. Among them, activated Tregs, characterized by high expression of HLA-DR genes in both datasets [60], are the ones assigned with the highest suppressive capacity.

As $CD4^+$ T cells undergo functional specialization into distinct subsets, they acquire specific effector functions that orchestrate diverse immune responses. Similarly, various Treg cell subtypes can utilize specialized migratory, functional, and homeostatic strategies to modulate diverse inflammatory reactions, including Th1, Th17, Th1/17, Th2, and Tfh subpopulations [61]–[63]. In addition to their ability to adjust to distinct inflammatory contexts, a prior investigation also indicated that Treg cells may exhibit genetic signatures tailored to specific tissues [23]. Our results revealed the existence of a population of Treg17 cells in both conditions. Nevertheless, despite the typical association of Treg17 cells with the suppression of immune responses driven by Th17 cells (primarily involved in controlling infections by extracellular bacteria in mucosal tissues, such as the gut), a prior study suggested the presence of Treg17 cells in the peripheral blood of healthy individuals. The role of Treg17 cells beyond the gut and their involvement in regulating immune responses orchestrated by peripheral T helper cells remains unclear [64].

In COVID-19 patients, our analysis also identified a cluster of $CXCR3^+$ Treg cells. This cluster represents a distinct population of cells with specialized immunoregulatory mechanisms to restrain Th1 cell-mediated inflammation, as described in previous studies [65]. In the context of a COVID-19 infection, where an exaggerated and imbalanced immune response can result in severe complications, these $CXCR3^+$ Treg cells can have a role in maintaining the immune homeostatic state and fine-tuning the immune reaction. Their presence suggests a mechanism to prevent an overactive Th1 response, which can contribute to cytokine storm and tissue damage [66].

Collectively, using a single-cell transcriptomics approach, our data elucidate the specific cellular states that cTfh and Treg populations adopt in response to an acute viral infection. Our data showed that upon encountering a stimulus, such as a viral infection, cTfh cells undergo polarization, acquiring distinct effector CD4 T cell signatures. We observed the emergence in both conditions of cTfh1 cells, with the acquisition by some of those cells of GZMK expression. The significance of a $GZMK^+$ cell population of cTfh1 cells requires further dissection. We were also able to identify cell subsets in the human Treg compartment, on healthy donors and COVID-19 patients. Our results revealed an unexpected presence of Treg17 cells in both conditions that, in the future, should be further investigated. Thus, by unraveling the heterogeneity of cTfh and Treg populations in response to viral infections and delineating their predicted function at the single-cell level, we believe that this work contributes significantly to enhancing our comprehension of human Tfh and Treg subsets, but also holds great potential for advancing the development of more efficacious vaccines and therapeutic interventions against viral diseases.

5. REFERENCES

[1] R. E. Baker *et al.*, "Infectious disease in an era of global change," *Nat. Rev. Microbiol.*, vol. 20, no. 4, pp. 193–205, 2022, doi: 10.1038/s41579-021-00639-z.

[2] A. S. Fauci, "Infectious diseases: Considerations for the 21st century," *Clin. Infect. Dis.*, vol. 32, no. 5, pp. 675–685, 2001, doi: 10.1086/319235.

[3] S. Şenel, "An overview of physical, microbiological and immune barriers of oral mucosa," *Int. J. Mol. Sci.*, vol. 22, no. 15, 2021, doi: 10.3390/ijms22157821.

[4] J. Parkin and B. Cohen, "Overview of the Immune System," *Lancet*, vol. 357, pp. 1777–89, 2011, doi: 10.1016/B978-0-323-39981-4.00004-X.

[5] A. S. Clem, "Fundamentals of vaccine immunology," *J. Glob. Infect. Dis.*, vol. 3, no. 1, pp. 73–78, 2011, doi: 10.4103/0974-777X.77299.

[6] N. S. Joshi and S. M. Kaech, "Effector CD8 T Cell Development: A Balancing Act between Memory Cell Potential and Terminal Differentiation," *J. Immunol.*, vol. 180, no. 3, pp. 1309–1315, 2008, doi: 10.4049/jimmunol.180.3.1309.

[7] A. J. Zajac and L. E. Harrington, *Immune Response to Viruses: Cell-Mediated Immunity ☆*, no. July. Elsevier Inc., 2014.

[8] M. Akkaya, K. Kwak, and S. K. Pierce, "B cell memory: building two walls of protection against pathogens," *Nat. Rev. Immunol.*, vol. 20, no. 4, pp. 229–238, 2020, doi: 10.1038/s41577-019-0244-2.

[9] C. G. Vinuesa, M. A. Linterman, D. Yu, and I. C. M. MacLennan, "Follicular Helper T Cells," *Annu. Rev. Immunol.*, vol. 34, no. 1, pp. 335–368, May 2016, doi: 10.1146/annurev-immunol-041015-055605.

[10] J. Deng, Y. Wei, V. R. Fonseca, L. Graca, and D. Yu, "T follicular helper cells and T follicular regulatory cells in rheumatic diseases," *Nat. Rev. Rheumatol.*, vol. 15, no. 8, pp. 475–490, 2019, doi: 10.1038/s41584-019-0254-2.

[11] S. Crotty, "T Follicular Helper Cell Biology: A Decade of Discovery and Diseases," *Immunity*, vol. 50, no. 5, pp. 1132–1148, 2019, doi: 10.1016/j.immuni.2019.04.011.

[12] H. Qi, "T follicular helper cells in space-time," *Nat. Rev. Immunol.*, vol. 16, no. 10, pp. 612–625, 2016, doi: 10.1038/nri.2016.94.

[13] D. DiToro *et al.*, "Differential IL-2 expression defines developmental fates of follicular versus nonfollicular helper T cells," *Science (80-.).*, vol. 361, no. 6407, p. eaao2933, Sep. 2018, doi: 10.1126/science.aao2933.

[14] H. Ueno, "Human Circulating T Follicular Helper Cell Subsets in Health and Disease," *J. Clin. Immunol.*, vol. 36, pp. 34–39, 2016, doi: 10.1007/s10875-016-0268-3.

[15] N. Schmitt, S. E. Bentebibel, and H. Ueno, "Phenotype and functions of memory Tfh cells in human blood," *Trends Immunol.*, vol. 35, no. 9, pp. 436–442, 2014, doi: 10.1016/j.it.2014.06.002.

[16] M. Koutsakos *et al.*, "Circulating TFH cells, serological memory, and tissue compartmentalization shape human influenza-specific B cell immunity," *Sci. Transl. Med.*, vol. 10, no. 428, pp. 1–16, 2018, doi: 10.1126/scitranslmed.aan8405.

[17] C. Schultheiß *et al.*, "Next-Generation Sequencing of T and B Cell Receptor Repertoires from COVID-19 Patients Showed Signatures Associated with Severity of Disease," *Immunity*, vol. 53, no. 2, pp. 442-455.e4, 2020, doi: 10.1016/j.immuni.2020.06.024.

[18] S. E. Bentebibel *et al.*, "Induction of ICOS+CXCR3+CXCR5+ T H cells correlates with antibody responses to influenza vaccination," *Sci. Transl. Med.*, vol. 5, no. 176, pp. 1–11, 2013, doi: 10.1126/scitranslmed.3005191.

[19] B. J. Meckiff *et al.*, "Imbalance of Regulatory and Cytotoxic SARS-CoV-2-Reactive CD4+ T Cells in COVID-19," *Cell*, vol. 183, no. 5, pp. 1340-1353.e16, 2020, doi: 10.1016/j.cell.2020.10.001.

[20] S. Sakaguchi, N. Mikami, J. B. Wing, A. Tanaka, K. Ichiyama, and N. Ohkura, "Regulatory T Cells and Human Disease," *Annu. Rev. Immunol.*, vol. 38, pp. 541–566, 2020, doi: 10.1146/annurev-immunol-042718-041717.

[21] S. Sakaguchi, K. Wing, Y. Onishi, P. Prieto-Martin, and T. Yamaguchi, "Regulatory T

cells: How do they suppress immune responses?," *Int. Immunol.*, vol. 21, no. 10, pp. 1105–1111, 2009, doi: 10.1093/intimm/dxp095.
[22] A. Schmidt, N. Oberle, and P. H. Krammer, "Molecular mechanisms oftreg-mediatedt cell suppression," *Front. Immunol.*, vol. 3, no. MAR, pp. 1–20, 2012, doi: 10.3389/fimmu.2012.00051.
[23] R. J. Miragaia *et al.*, "Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation," *Immunity*, vol. 50, no. 2, pp. 493-504.e7, 2019, doi: 10.1016/j.immuni.2019.01.001.
[24] A. Liston and D. H. D. Gray, "Homeostatic control of regulatory T cell diversity," *Nat. Rev. Immunol.*, vol. 14, no. 3, pp. 154–165, 2014, doi: 10.1038/nri3605.
[25] P. Seepathomnarong *et al.*, "Regulatory T Cells Decreased during Recovery from Mild COVID-19," *Viruses*, vol. 14, no. 8, pp. 1–11, 2022, doi: 10.3390/v14081688.
[26] J. Craig Venter *et al.*, "The sequence of the human genome," *Science (80-.).*, vol. 291, no. 5507, pp. 1304–1351, 2001, doi: 10.1126/science.1058040.
[27] E. S. Lander *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001, doi: 10.1038/35057062.
[28] A. Baysoy, Z. Bai, R. Satija, and R. Fan, "The technological landscape and applications of single-cell multi-omics," *Nat. Rev. Mol. Cell Biol.*, 2023, doi: 10.1038/s41580-023-00615-w.
[29] M. Krassowski, V. Das, S. K. Sahu, and B. B. Misra, "State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing," *Front. Genet.*, vol. 11, no. December, pp. 1–17, 2020, doi: 10.3389/fgene.2020.610798.
[30] F. Horns *et al.*, "Engineering RNA export for measurement and manipulation of living cells," *Cell*, 2023, doi: https://doi.org/10.1016/j.cell.2023.06.013.
[31] X. Li and C. Y. Wang, "From bulk, single-cell to spatial RNA sequencing," *Int. J. Oral Sci.*, vol. 13, no. 1, pp. 1–6, 2021, doi: 10.1038/s41368-021-00146-0.
[32] V. Svensson *et al.*, "Power analysis of single-cell rnA-sequencing experiments," *Nat. Methods*, vol. 14, no. 4, pp. 381–387, 2017, doi: 10.1038/nmeth.4220.
[33] G. Chen, B. Ning, and T. Shi, "Single-cell RNA-seq technologies and related computational data analysis," *Front. Genet.*, vol. 10, no. APR, pp. 1–13, 2019, doi: 10.3389/fgene.2019.00317.
[34] F. Ginhoux, A. Yalin, C. A. Dutertre, and I. Amit, "Single-cell immunology: Past, present, and future," *Immunity*, vol. 55, no. 3, pp. 393–404, 2022, doi: 10.1016/j.immuni.2022.02.006.
[35] L. Graca, *T-Follicular Helper Cells - Methods and Protocols*, 1st ed. New York: Humana New York, NY.
[36] P. Qiu, "Embracing the dropouts in single-cell RNA-seq analysis," *Nat. Commun.*, vol. 11, no. 1, p. 1169, 2020, doi: 10.1038/s41467-020-14976-9.
[37] T. Stuart *et al.*, "Comprehensive Integration of Single-Cell Data," *Cell*, vol. 177, no. 7, pp. 1888-1902.e21, 2019, doi: 10.1016/j.cell.2019.05.031.
[38] C. Hafemeister and R. Satija, "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression," *Genome Biol.*, vol. 20, no. 1, p. 296, 2019, doi: 10.1186/s13059-019-1874-1.
[39] L. Zappia and A. Oshlack, "Clustering trees: a visualization for evaluating clusterings at multiple resolutions," *Gigascience*, vol. 7, no. 7, pp. 1–9, 2018, doi: 10.1093/gigascience/giy083.
[40] S. Hänzelmann, R. Castelo, and J. Guinney, "GSVA: Gene set variation analysis for microarray and RNA-Seq data," *BMC Bioinformatics*, vol. 14, 2013, doi: 10.1186/1471-2105-14-7.
[41] S. Aibar *et al.*, "SCENIC: Single-cell regulatory network inference and clustering," *Nat. Methods*, vol. 14, no. 11, pp. 1083–1086, 2017, doi: 10.1038/nmeth.4463.
[42] F. Del Carratore, A. Jankevics, R. Eisinga, T. Heskes, F. Hong, and R. Breitling, "RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets," *Bioinformatics*, vol. 33, no. 17, pp. 2774–2775, 2017, doi: 10.1093/bioinformatics/btx292.
[43] K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, and B. Panning, "Xist RNA and the mechanism of X chromosome inactivation," *Annu. Rev. Genet.*, vol. 36, pp. 233–278, 2002,

doi: 10.1146/annurev.genet.36.042902.092433.
[44] K. D. Cook *et al.*, "T Follicular Helper Cell-Dependent Clearance of a Persistent Virus Infection Requires T Cell Expression of the Histone Demethylase UTX," *Immunity*, vol. 43, no. 4, pp. 703–714, 2015, doi: 10.1016/j.immuni.2015.09.002.
[45] S. L. Klein and K. L. Flanagan, "Sex differences in immune responses," *Nat. Rev. Immunol.*, vol. 16, no. 10, pp. 626–638, 2016, doi: 10.1038/nri.2016.90.
[46] N. M. Wilkinson, H.-C. Chen, M. G. Lechner, and M. A. Su, "Sex Differences in Immunity," *Annu. Rev. Immunol.*, vol. 40, no. 1, pp. 75–94, Apr. 2022, doi: 10.1146/annurev-immunol-101320-125133.
[47] E. Kiner *et al.*, "Gut CD4+ T cell phenotypes are a continuum molded by microbes, not by TH archetypes," *Nat. Immunol.*, vol. 22, no. 2, pp. 216–228, 2021, doi: 10.1038/s41590-020-00836-7.
[48] J. J. O'Shea and W. E. Paul, "Mechanisms Underlying Lineage Commitment and Plasticity of Helper CD4+ T Cells," *Science (80-.).*, vol. 327, no. 5969, pp. 1098–1102, Feb. 2010, doi: 10.1126/science.1178334.
[49] K. M. Murphy and B. Stockinger, "Effector T cell plasticity: flexibility in the face of changing circumstances," *Nat. Immunol.*, vol. 11, no. 8, pp. 674–680, 2010, doi: 10.1038/ni.1899.
[50] J. Zhang *et al.*, "Circulating CXCR3+ Tfh cells positively correlate with neutralizing antibody responses in HCV-infected patients," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.1038/s41598-019-46533-w.
[51] J. Niessl *et al.*, "Persistent expansion and Th1-like skewing of HIV-specific circulating T follicular helper cells during antiretroviral therapy.," *EBioMedicine*, vol. 54, p. 102727, Apr. 2020, doi: 10.1016/j.ebiom.2020.102727.
[52] A. C. Bouwman, K. R. van Daalen, S. Crnko, T. ten Broeke, and N. Bovenschen, "Intracellular and Extracellular Roles of Granzyme K ," *Frontiers in Immunology* , vol. 12. 2021, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fimmu.2021.677707.
[53] Y. Gao, G. Dunlap, M. Elahee, and D. A. Rao, "Patterns of T-Cell Phenotypes in Rheumatic Diseases From Single-Cell Studies of Tissue.," *ACR open Rheumatol.*, vol. 3, no. 9, pp. 601–613, Sep. 2021, doi: 10.1002/acr2.11296.
[54] X. Feng *et al.*, "Foxp1 is an essential transcriptional regulator for the generation of quiescent naive T cells during thymocyte development," *Blood*, vol. 115, no. 3, pp. 510–518, Jan. 2010, doi: 10.1182/blood-2009-07-232694.
[55] N. K. Bhat *et al.*, "Reciprocal expression of human ETS1 and ETS2 genes during T-cell activation: regulatory role for the protooncogene ETS1.," *Proc. Natl. Acad. Sci.*, vol. 87, no. 10, pp. 3723–3727, May 1990, doi: 10.1073/pnas.87.10.3723.
[56] R. DangLi, W. HeKong, L. JiQin, Z. MingHua, and Z. WenCheng, "ROS-induced ZNF580 expression: a key role for H2O2/NF-κB signaling pathway in vascular endothelial inflammation," *Mol. Cell. Biochem.*, vol. 359, no. 1, pp. 183–191, 2012, doi: 10.1007/s11010-011-1013-0.
[57] S. Uematsu *et al.*, "The C/EBPβ Isoform 34-kDa LAP Is Responsible for NF-IL-6-Mediated Gene Induction in Activated Macrophages, but Is Not Essential for Intracellular Bacteria Killing1," *J. Immunol.*, vol. 179, no. 8, pp. 5378–5386, Oct. 2007, doi: 10.4049/jimmunol.179.8.5378.
[58] G. Trinchieri, S. Pflanz, and R. A. Kastelein, "The IL-12 Family of Heterodimeric Cytokines: New Players in the Regulation of T Cell Responses," *Immunity*, vol. 19, no. 5, pp. 641–644, Nov. 2003, doi: 10.1016/S1074-7613(03)00296-6.
[59] M. Quigley, X. Huang, and Y. Yang, "STAT1 Signaling in CD8 T Cells Is Required for Their Clonal Expansion and Memory Formation Following Viral Infection In Vivo1," *J. Immunol.*, vol. 180, no. 4, pp. 2158–2164, Feb. 2008, doi: 10.4049/jimmunol.180.4.2158.
[60] C. Baecher-Allan, E. Wolf, and D. A. Hafler, "MHC Class II Expression Identifies Functionally Distinct Human Regulatory T Cells1," *J. Immunol.*, vol. 176, no. 8, pp. 4622–4631, Apr. 2006, doi: 10.4049/jimmunol.176.8.4622.
[61] D. J. Campbell and M. A. Koch, "Phenotypical and functional specialization of FOXP3+ regulatory T cells," *Nat. Rev. Immunol.*, vol. 11, no. 2, pp. 119–130, 2011, doi: 10.1038/nri2916.

[62] B. Höllbacher, T. Duhen, S. Motley, M. M. Klicznik, I. K. Gratz, and D. J. Campbell, "Transcriptomic Profiling of Human Effector and Regulatory T Cell Subsets Identifies Predictive Population Signatures," *ImmunoHorizons*, vol. 4, no. 10, pp. 585–596, Oct. 2020, doi: 10.4049/immunohorizons.2000037.

[63] S. Sakaguchi, D. A. A. Vignali, A. Y. Rudensky, R. E. Niec, and H. Waldmann, "The plasticity and stability of regulatory T cells," *Nat. Rev. Immunol.*, vol. 13, no. 6, pp. 461–467, 2013, doi: 10.1038/nri3464.

[64] D. C. Kim BS, Lu H, Ichiyama K, Chen X, Zhang YB, Mistry NA, Tanaka K, Lee YH, Nurieva R, Zhang L, Yang X, Chung Y, Jin W, Chang SH, "Generation of RORγt+ Antigen-Specific T Regulatory 17 Cells from Foxp3+ Precursors in Autoimmunity.," *Cell Rep*, vol. 21, no. 1, pp. 195–207, 2017, [Online]. Available: doi: 10.1016/j.celrep.2017.09.021.

[65] M. A. Koch, G. Tucker-Heard, N. R. Perdue, J. R. Killebrew, K. B. Urdahl, and D. J. Campbell, "The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation," *Nat. Immunol.*, vol. 10, no. 6, pp. 595–602, 2009, doi: 10.1038/ni.1731.

[66] L. Yang, X. Xie, Z. Tu, J. Fu, D. Xu, and Y. Zhou, "The signal pathways and treatment of cytokine storm in COVID-19," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, p. 255, 2021, doi: 10.1038/s41392-021-00679-0.

6. Supplementary Figures

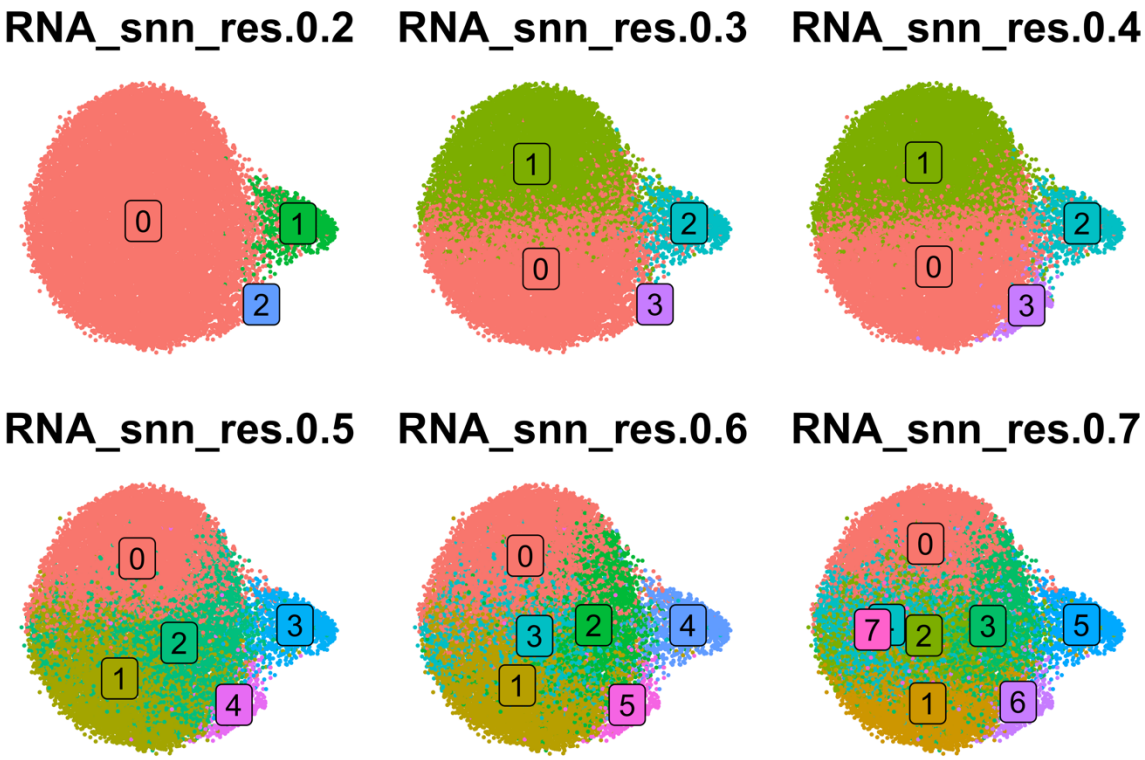

Figure S1— **Clustering resolutions of healthy cTfh cells.** Panel of UMAP for the different resolutions doesn't show clustering based on CCR6 expression.

UNIVERSIDADE NOVA
DE LISBOA