

RESEARCH

Open Access



The T-cell repertoire of Spanish patients with COVID-19 as a strategy to link T-cell characteristics to the severity of the disease

Fernando Marín-Benesiu^{1,2†}, Lucia Chica-Redecillas^{1,2†}, Verónica Arenas-Rodríguez², Esperanza de Santiago², Silvia Martínez-Diz³, Ginesa López-Torres⁴, Ana Isabel Cortés-Valverde⁴, Catalina Romero-Cachinero⁵, Carmen Entrala-Bernal⁶, Francisco Javier Fernandez-Rosado⁶, Luis Javier Martínez-González^{1,2*†} and Maria Jesus Alvarez-Cubero^{1,2,7†}

Abstract

Background The architecture and dynamics of T cell populations are critical in orchestrating the immune response to SARS-CoV-2. In our study, we used T Cell Receptor sequencing (TCRseq) to investigate TCR repertoires in 173 post-infection COVID-19 patients.

Methods The cohort included 98 mild and 75 severe cases with a median age of 53. We amplified and sequenced the TCR β chain Complementary Determining Region 3 (CDR3b) and performed bioinformatic analyses to assess repertoire diversity, clonality, and V/J allelic usage between age, sex and severity groups. CDR3b amino acid sequence inference was performed by clustering structural motifs and filtering validated reactive CDR3b to COVID-19.

Results Our results revealed a pronounced decrease in diversity and an increase in clonal expansion in the TCR repertoires of severe COVID-19 patients younger than 55 years old. These results reflect the observed trends in patients older than 55 years old (both mild and severe). In addition, we identified a significant reduction in the usage of key V alleles (TRBV14, TRBV19, TRBV15 and TRBV6-4) associated with disease severity. Notably, severe patients under 55 years old had allelic patterns that resemble those over 55 years old, accompanied by a skewed frequency of COVID-19-related motifs.

Conclusions Present results suggest that severe patients younger than 55 may have a compromised TCR repertoire contributing to a worse disease outcome.

Keywords T cells, SARS-Cov2, Immunoinformatics, COVID-19, Adaptative immunology, TCR repertoire

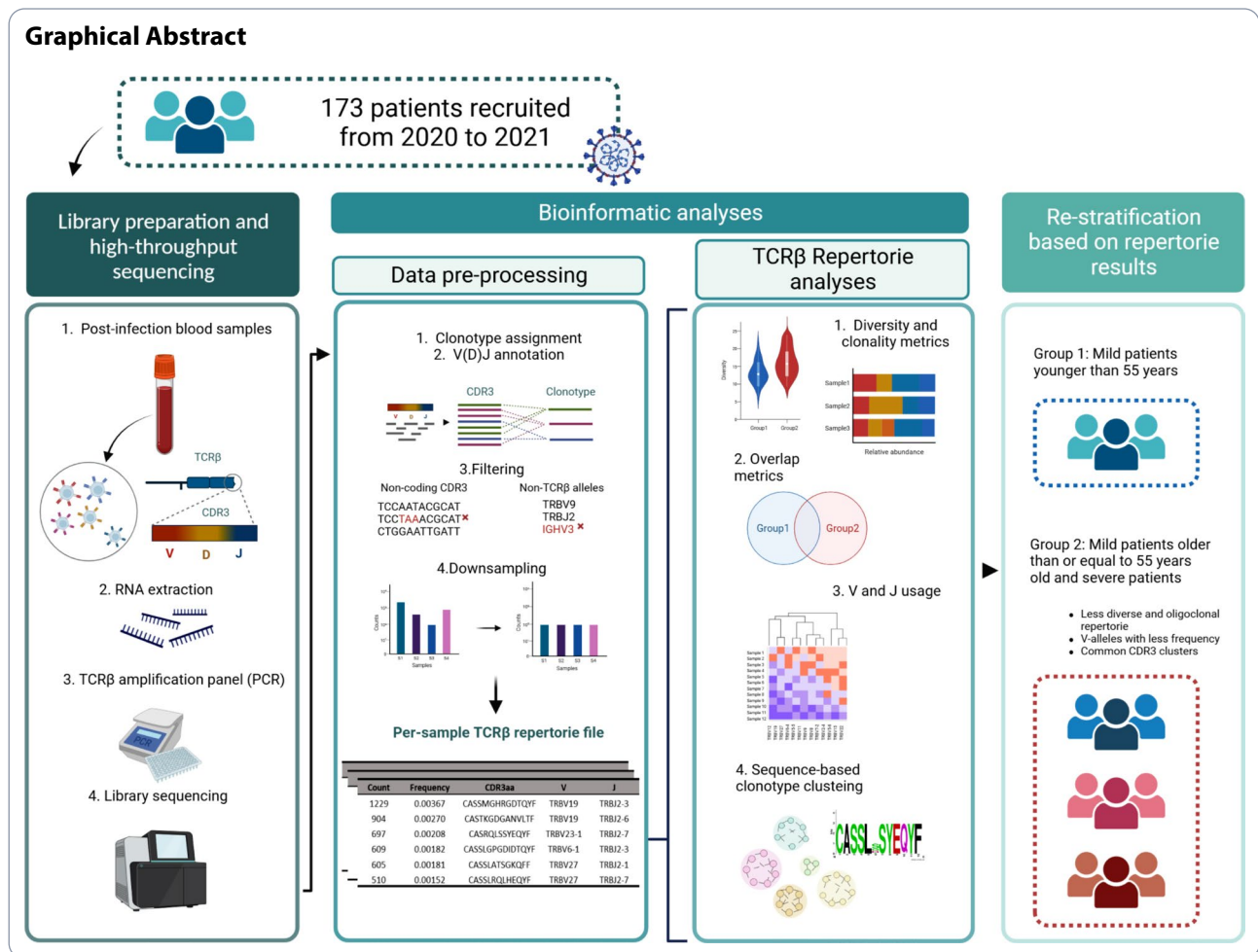
[†]Luis Javier Martínez-González and Maria Jesus Alvarez-Cubero contributed equally to this work.

*Correspondence:
Luis Javier Martínez-González
luisjavier.martinez@genyo.es

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



Background

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection exhibits a wide range of symptoms, leading to diverse outcomes in the severity of COVID-19 disease. This viral infection triggers adaptive antigen-specific responses, facilitates viral clearance, and initiates the formation of immune memory [1]. Single-cell multi-omics analyses have unveiled a lack of coordination between innate and adaptive responses in progressive COVID-19 [2]. T lymphocytes, particularly CD4 and CD8 T cells, recognize multiple antigens, and play a crucial role in influencing disease resolution [3]. These T cell responses persist even after asymptomatic infection [4], suggesting their potential as a correlate of COVID-19 immunological responses and supporting the feasibility of T cell-based diagnostic assays for SARS-CoV-2 [5]. Significant correlations have been reported between circulating and tissue-resident memory T and B cells, supporting the validity of blood-based measures as a convenient, albeit imperfect, approach to assessing overall T-cell responses [6].

A crucial aspect of T-cell populations is their extensive diversity, originating from the sequences of their T-cell receptor (TCR). This diversity primarily arises from variable (V), diversity (D, only present in TCR-beta chain), and joining (J) gene segments, collectively forming the Complementarity Determining Region 3 (CDR3) of the TCR, generated through random and imprecise somatic recombination [7]. Bulk TCR-beta (TCR β) repertoire analysis (bulk TCRseq) involves studying the diversity and characteristics of TCR sequences such as V(D)J usage, repertoire overlap and TCR motif-based clustering, crucial for immune responses [8].

It has been demonstrated the potential to predict SARS-CoV-2 immunity from TCR profiles. In mild cases, the TCR repertoire remains diverse, with specific clonotypes, including SARS-CoV-2-specific ones, compared to severe cases [9]. While few studies have identified strong associations between specific V-, D-, and J-segment usage and prior SARS-CoV-2 exposure, severe cases show notable patterns in specific TCR V β gene segments [10]. We report this TCR β diversity and bioinformatic analysis in a large and new Spanish cohort unvaccinated COVID-19

patients. We suggest changes in TCR diversity, clonality, V usage, and motif-usage.

Methods

The main aspects of the applied methodology are described below, from the study population to sequencing and TCR bioinformatics analysis. For further details regarding inclusion criteria, parameter descriptions and bioinformatic pipelines, refer to Supplementary Material 1: Supplementary Methods.

Study population

A cohort of 173 patients ($n=98$ mild and $n=75$ severe) was recruited between 2020 and 2021 from primary care centres and hospital facilities in Granada. “Tempus™ Blood RNA Tubes” were collected from each patient for total RNA extraction. Clinical symptoms were recorded, including dermatological involvement, anosmia, ageusia, myalgia, headache, fever, dyspnoea, asthenia and cough. All of them follow-up inclusion criteria based on WHO classification. Inclusion criteria were revised periodically to update database trying to have balanced samples according to age, gender and severity. Those in mild disease were characterized by fever, malaise, cough, upper respiratory symptoms, and/or less common features of COVID-19 (headache, loss of taste or smell etc.). Moreover, patients in severe disease group fulfil hospital admission and the following features: (i) hypoxia: $SPO_2 \leq 93\%$ on atmospheric air or $PaO_2:FiO_2 < 300$ mmHg (SF ratio < 315); tachypnoea: in respiratory distress or RR (respiratory rate) > 30 breaths/minutes; or more than 50% involvement seen on chest imaging [11, 12]. SARS-Cov2 infection was confirmed by positive Reverse Transcription Polymerase Chain Reaction (RT-PCR) or by positive Immunoglobulin M (IgM) antibody test; and at the same timeline after COVID-19 recovery.

To investigate changes in the composition of T lymphocyte repertoires with age, we categorized the subjects into two groups: individuals aged 55 years or older, and individuals younger than 55 years. We chose this age threshold based on previous scientific publications that extensively characterized the genetic, clinical, and cellular attributes of this specific cohort [11, 12]. This age ensured a balanced distribution between the groups, facilitating a more robust and reliable comparison.

Differences symptoms were studied by severity, gender, and age group. The study protocol was approved by the Granada Research Ethics Committee (CEI Granada) with internal code 1329-N-21. Written informed consent was obtained from all participants in accordance with the principles of the Helsinki Declaration.

Preparation of libraries and high-throughput sequencing

RNA was extracted from whole blood using the “Tempus™ Spin RNA Isolation Reagent” kit (Applied Biosystems), checked its quality, and normalized it to 1 μ g. For analyzing TCR repertoires, we employed the “AmpliSeq for Illumina Immune Repertoire Plus, TCR beta Panel” (Illumina). Our library preparation included converting RNA into cDNA, amplifying specific cDNA regions through PCR, and partial amplicon digestion. We then attached Index adapters for hybridization, followed by PCR enrichment, library normalization, pooling, and sequencing on the NextSeq 500 platform (Illumina). This platform utilized paired-end sequencing (150 bp x 2) and achieved a final library depth of around 1 million reads per sample.

Repertoire data analysis

FASTQ files were processed using MiXCR v4.0.0 [13] on the Illumina platform. Default settings included read filtering and alignment to TCR β locus segments (V, D, J, C) to assemble clonotypes. Repertoire files were created for each sample, containing annotated V(D)J alleles, clonotype abundance, frequency, and CDR3 sequences. VDJtools v1.2.1 [14] was used to filter non-functional clonotypes and removed non TRC β alleles. To address library size differences, we downsample normalized repertoires. Minimum acceptable depth was determined using Tukey criterion, excluding one sample for being excessively small. Downsampling enables direct metric comparisons, which would be biased when comparing repertoires of differing sequencing depths. While low frequency clonotypes may be affected by this process, their loss is less crucial when analysing diversity and clonality metrics dominated by expanded clonotypes [15]. Repertoires were analyzed using immunarch R package. Common metrics calculated included Chao1 estimator for clonotype richness, normalized Shannon-Wiener and Gini-Simpson indices for α -diversity, and DE50 index and Gini inequality coefficient for clonality assessment. For TCR repertoire overlap, two approaches were applied: based on exclusively CDR3b amino acid sequence or combined with V and J allele. Parameters calculated included public clonotypes, Jaccard index, and Morisita's overlap index. Repertoire overlap also assessed for top N expanded clonotypes. V and J allele usage were calculated by computing relative frequency. For hierarchical clustering, weighted counts for each allele were computed. Finally, each V and J allele detected were subjected for correlation analysis against the serum biomarker variables C-Reactive Protein (CRP), D-dimer, ferritin and Lactate Dehydrogenase (LDH), collected in previous studies [12]. In addition, V and J alleles were subjected for another correlation analysis against key immune cellular subpopulation characterized by Mass Cytometry

(CyTOF) in previous studies from our group such as lymphocytes (CD3+) and monocyte-macrophage subpopulations [11].

Clustering and motif analysis

Top 100 expanded clones for each repertoire were clustered with Grouping of Lymphocyte Interactions by Paratope Hotspots 2 (GLIPH2) algorithm [16]. COVID-positive (COVID+) clonotypes were selected based on SARS-CoV-2 antigen binding validation present in Multiplex Identification of Antigen-Specific T-Cell Receptors Assay (MIRA) dataset [17]. Frequency distributions of motif-filtered clusters were compared among patient groups. Structural signatures visualized using WebLogo v2.8.2 [18]. Motif-filtered data were used to construct logistic regression models for distinguishing patient groups based on severity and age divisions. A Monte Carlo simulation with 1000 iterations was performed, splitting data into 75% training and 25% testing divisions. Models were subsequently processed for stepwise feature selection based on the Akaike Information Criterion (AIC). Predictions were made with the selected model for each iteration, and ROC curves with AUC were calculated. The average AUC was determined after 1000-fold Monte Carlo cross-validation, recording the number of motif-filtered variables considered relevant in logistic regression models.

Statistics

Chi-square test was applied to symptomatology variables with Cramer's V for effect size. Lilliefors Kolmogorov-Smirnov test checked normality of continuous variables. Mann-Whitney U-test was employed for comparing non-normal distributions and Kruskal-Wallis test for three or more groups, with Dunn test for multiple comparisons. T-test and ANOVA plus Tukey test used for normal distributions. For multiple comparisons tests, p-values were

corrected with Bonferroni-Holm adjustment. Effect size coefficients computed for each test. Two-tailed P-value less than 0.05 was considered significant. Spearman's correlation value was calculated for each V or J comparison. P-values of correlation analyses were adjusted using Bonferroni correction, multiplying the raw p-value by the total number of serum or CyTOF variables analyzed. A significance level of 0.05 was set. Logistic regression binomial models were built using the glm function. The stepAIC function from MASS v.7.3 package was specified with the direction argument set to "both", allowing for feature selection through both removals and additions of variables to the model. Analyses involving R packages were performed with R v 4.1.3.

Results

Association of symptomatology between groups of patients

A summary cohort description can be found in Table 1. Supplemental figures and tables can be found in Supplementary Material 2: Supplementary figures and tables. Prior to our analysis of the T-lymphocyte repertoire, we conducted studies to investigate changes in symptom distribution based on patient severity, sex, and age group. The results can be found in Supplementary Table 1. The chi-square analysis did not reveal significant associations between severity and the variables of sex, age group, or the combined group of both variables (Supplementary Table 2).

A significant association between patient severity and the presence of various symptoms was found. These symptoms include dermatological conditions ($p=0.004$), agnosia ($p=0.04$), myalgias ($p<0.001$), fever ($p<0.001$), dyspnea ($p<0.001$), asthenia ($p<0.001$), and cough ($p<0.001$). Fever (Cramer's $V=0.66$), shortness of breath (Cramer's $V=0.59$), and asthenia (Cramer's $V=0.508$) were the most strongly associated symptoms. When categorized according to age groups (<55 and ≥ 55), no significant associations were found. However, there was a weak association between fever and gender, with fever being more frequent in women than men ($p=0.011$, Cramer's $V=0.197$).

Changes in TCR repertoire diversity, clonality and overlap

For the comparative analysis of T lymphocyte repertoires, we performed sample normalization through downsampling. After downsampling, our TCRseq libraries yielded a total of 334,599 clones (Supplementary Fig. 1).

When comparing richness estimators, diversity indices, and clonality indices, we found significant differences between patients under 55 years old and those who were ≥ 55 years old (details in Supplementary Fig. 2). Specifically, for patients aged ≥ 55 years old, we observed a repertoire with reduced clone richness (Chao1: $p=0.034$,

Table 1 Main characteristics of the study population and patient groups

	Mild (n=98)	Severe (n=75)	Total (n=173)
Severity subgroup, n (%)			
Asymptomatic	37 (37.8%)		37 (21.4%)
Mild	61 (62.2%)		61 (35.3%)
Severe		65 (86.7%)	65 (37.6%)
Severe/Critical		10 (13.30%)	10 (5.80%)
Sex, n (%)			
Male	45 (45.9%)	40 (53.3%)	85 (49.1%)
Female	53 (54.1%)	35 (46.7%)	88 (50.9%)
Age (yr), median [Q1, Q3]	52 [46, 62]	55 [48.5, 63]	53 [48, 64.5]
More than 55 years, n (%)			
< 55	58 (59.2%)	35 (46.7%)	93 (53.8%)
≥ 55	40 (40.8%)	40 (53.3%)	80 (46.2%)

$r_{tb}=0.188$), as well as decreased overall diversity (normalized Shannon-Wiener: $p=0.034$, $r_{tb}=0.188$) and diminished diversity targeted towards more expanded clonotypes (Gini-Simpson: $p=0.006$, $r_{tb}=0.258$). This reduction in clonotype richness and diversity led to a poorer repertoire as evidenced by a decrease in Gini index (Gini: $p=0.004$, $d=0.450$) and higher clonal dominance (DE50: $p=0.006$, $r_{tb}=0.244$).

No significant differences were observed when dividing patients into mild and severe categories. However, when severity and age groups are combined, severe patients under 55 years old have a less diverse and more oligoclonal repertoire than mild patients of the same age group. Additionally, the results indicate that the diversity and clonality of this group of patients is similar to that of older patients, either mild or severe. (Supplementary Table 3, Supplementary Tables 4, and Fig. 1A). These differences exhibited a moderate effect size for all metrics (Chao1: $p<0.001$, $E^2_R=0.071$, normalized Shannon-Wiener: $p<0.001$, $E^2_R=0.096$, Gini-Simpson: $p<0.001$, $E^2_R=0.100$, Gini: $p<0.001$, $E^2_R=0.107$, DE50: $p<0.001$, $E^2_R=0.095$). No significant differences were found in diversity metrics based on patient sex, nor in the combined group of sex and severity (see Supplementary Table 3). However, significant differences were observed in the combined group of sex and age, indicating a diverse repertoire in female patients under 55 years old and a moderate increase in

male patients under 55, suggesting a substantial influence of age when comparing these metrics.

The proportion of hyperexpanded clones is significantly higher in the ≥ 55 group ($p=0.003$, $r_{tb}=0.261$) as illustrated in Supplementary Fig. 2B. A non-significant increase in the proportion of hyperexpanded clones was also observed in severe patients as compared to mild patients ($p=0.052$, $r_{tb}=0.172$) (details in Supplementary Fig. 2B and Supplementary Fig. 3). As with previous metrics, the combined classification of patients by severity and age group indicates that severe patients under the age of 55 have a distribution of hyperexpanded and rare clones that is more similar to those over the age of 55 than mild patients in the same age group. (rare: $p=0.003$, $E^2_R=0.081$, hyperexpanded: $p<0.001$, $E^2_R=0.116$, Fig. 1B and Supplementary Table 4). According to sex comparisons, changes were only observed in the group that combine with age (Supplementary Tables 5 and 6). A significant decrease in the proportion of hyper-expanded clones was observed in female patients under 55, consistent with diversity results.

Regarding overlap metrics (public clonotypes, Jaccard index and Morisita's index), we have found a high level of uniqueness, considering a clonotype the combination of its CDR3 amino acid sequence and its V and J allele (Supplementary Fig. 5A). This uniqueness is also reported when we compare the overlap at different repertoire

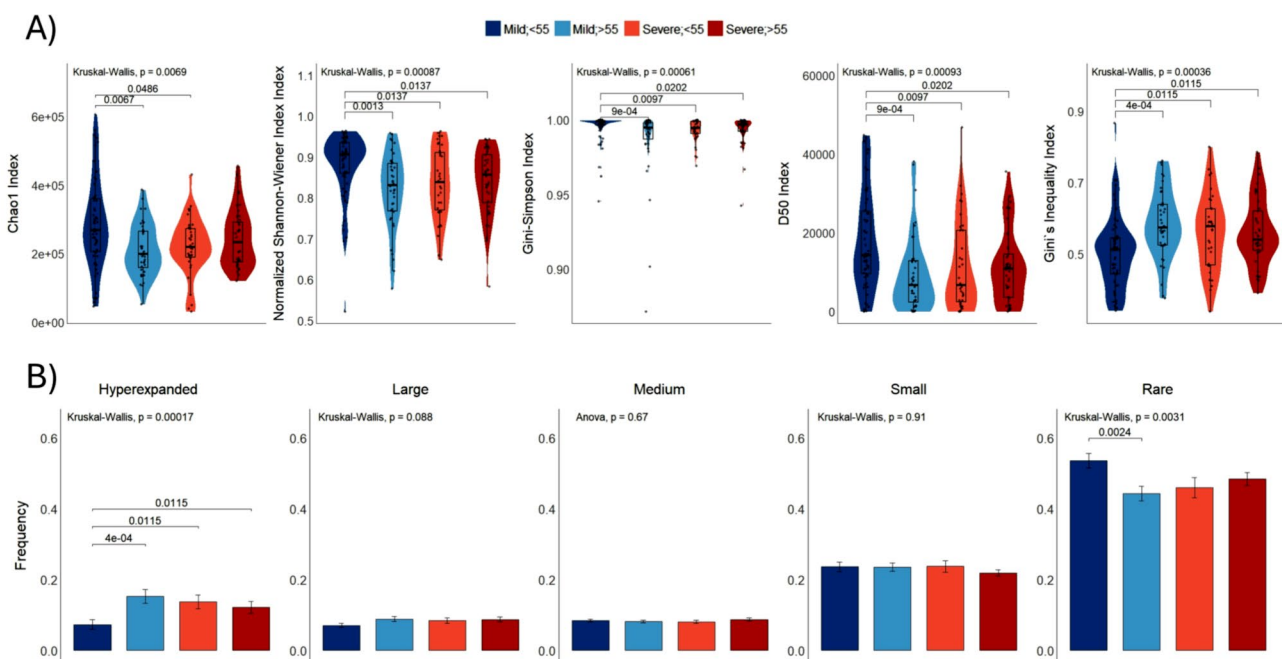


Fig. 1 Main results in α -diversity and clonality analyses for the combined age and severity group. **(A)** CHAO1 richness estimator, normalized Shannon-Wiener α -diversity indices and Gini-Simpson index, DE50 clonality index and Gini coefficient of inequality for patient groups. **(B)** Differences in the proportion of hyperexpanded, large, medium, small and rare clones. Classification of clones from their clonal expansion is based on the frequency intervals of a clonotype in a repertoire: Rare ($0-1e^{-5}$), Small ($1e^{-5}-1e^{-4}$), Medium ($1e^{-4}-0.001$), Large ($0.001-0.01$), Hyperexpanded ($0.01-1$). For comparisons of more than two groups, one-way Kruskal-Wallis test with Dunn's test as posteriori test. For multiple post-hoc comparisons the p-value was corrected with the Bonferroni-Holm method. Significance level of 0.05

depths (first 50,000, 10,000, 5000, 1000, 500 and 100 expanded clonotypes) and also the hyperexpanded ones (Supplementary Fig. 5B, Supplementary Fig. 6). The more clonally expanded, the less degree of overlap. When comparing the overlap between patients within the same group, we found that mild and severe patients maintain a certain degree of overlap, particularly with the Morisita's index. These findings suggest that the degree of overlap is inversely proportional to the level of clonal expansion. When comparing the overlap between patients within the same group, we found that mild and severe patients maintained a certain degree of overlap, particularly in Morisita's index. It is interesting to note that mild patients younger than 55 years showed greater similarity in their repertoires, in contrast to severe patients older than 55 years. Additionally, we observed similar trends when considering only the CDR3 amino acid sequence as clonotype. (Supplementary Fig. 7).

V and J allele usage in severity and age groups

The main differences in TCR repertoire architecture are related to severity and age. This section studies severity, age and their combination. The analysis of V allele usage revealed significant changes in allele frequencies when examining severity, age groups, and combined groups (see Fig. 2A-C). The results indicated a decrease in TRBV19 allele frequency in severe patients ($p=0.008$, $r_{tb}=0.236$). Additionally, the frequencies of TRBV12-3 ($p=0.002$, $rrb=0.271$) and TRBV10-3 ($p=0.045$, $rrb=0.177$) were found to be reduced in the ≥ 55 group. TRBV14 frequencies exhibited a significant decrease among severe patients ($p=0.029$, $rrb=0.194$) and in those aged ≥ 55 ($p=0.041$, $rrb=0.181$), with even more pronounced changes observed in TRBV6-4 ($p<0.001$, $rrb=0.314$ and $p<0.001$). When comparing the combined age and severity groups, the results are consistent with the previously reported changes in diversity and clonality in the frequencies of V alleles such as TRBV14 ($p=0.001$, $E^2_R=0.09$), TRBV6-4 ($p<0.001$, $E^2_R=0.106$), and TRBV15 ($p=0.0017$, $E^2_R=0.09$). Hierarchical clustering using overall V and J allele counts does not clearly differentiate patient clusters by severity, age group, or sex (see Fig. 2D-E). Nevertheless, some V and J alleles exhibit a higher usage across all patients. TRBV5-1, TRBV20-1, TRBV29-1, TRBV19, TRBV7-2, and TRBV28 have the greatest representation, whereas TRBV9 and TRBV10-3 are more frequent in certain individuals. TRBJ2-7, TRBJ2-1, TRBJ2-5, and TRBJ2-3 were the most prevalent J alleles, while another group consisting of TRBJ1-1, TRBJ1-2, and TRBJ2-2 displayed a more diverse distribution. The Jensen-Shannon divergence index detected subgroups with comparable allelic usage distributions, specifically for V alleles (Supplementary Fig. 8). Hierarchical clustering dendrograms provided clearer classifications of clades,

although heterogeneity in severity, age, and sex persisted within those clades.

The correlation analyses show the significance of a small group of V and J alleles with serum levels of CRP, ferritin, D-dimer, and LDH in different patient groups and subgroups (Supplementary Material 3). We highlight the previously mentioned TRBV15, which has a negative correlation with ferritin (Spearman's $\rho = -0.302$, p value=0.007) and CRP (Spearman's $\rho = -0.268$, p value=0.032) in the total patient group, as well as with D-dimer in mild patients (Spearman's $\rho = -0.421$, p value=0.040). Additionally, a significant negative correlation was detected for TRBV6-4 with CRP levels (Spearman's $\rho = -0.267$, p value=0.033) in the total patient group and for TRBV19 with D-dimer levels in the same group (Spearman's $\rho = -0.332$, p value=0.004).

Regarding the correlations with CyTOF populations, we highlight the one detected in mild patients between TRBV12-5 and NCM-CD11b+ (Spearman's $\rho=0.577$, p value=0.004), and between TRBV19 and the same cell group (Spearman's $\rho=0.497$, p value=0.033). We also highlight the negative correlation between TRBV6-4 and the T lymphocyte population (Spearman's $\rho=0.610$, p value=0.015).

CDR3b sequence inference by GLIPH2 and MIRA

We applied GLIPH2 to the top 100 most expanded clonotypes per repertoire, identifying 16 local- 20 global-alignment clusters (Additional file 4). Each cluster has on average 5–6 CDR3b sequences, with some global clusters having 11–13 sequences. Network analysis of these 36 clusters revealed some composed exclusively of expanded CDR3b in mild (cluster LSPLA_4_22), severe (cluster S%GSYE_EGQST), or the severe subgroup under 55 years old combined with ≥ 55 patients (cluster YSSGE_4_22) (Fig. 3A, Additional file 3), aligning with diversity, clonality, and gene usage results. However, most clusters contained CDR3b sequences from all four patient subgroups or unrelated subgroups.

To enhance the functional significance, we filtered CDR3b sequences from the 36 global and local clusters, prioritizing those experimentally validated as COVID reactive in the MIRA assay (labelled as "COVID+"). We then extracted and summed the frequencies of these COVID+ sequences by repertoire, grouping them by their GLIPH2 cluster. After this, the number of clusters was reduced reducing from 36 to 24 by eliminating clusters without CDR3b COVID+ sequences.

Our results showed that four clusters with CDR3b "COVID+" sequences (S%GYE_AEFGHLRSVWY, SS%YE_AGST, SP%YE_GHNRST, SL%SYE_DGNST)

exhibited significant frequency differences between combined severity and age groups, despite not being exclusive in the GLIPH2 network (Fig. 3B). Clusters

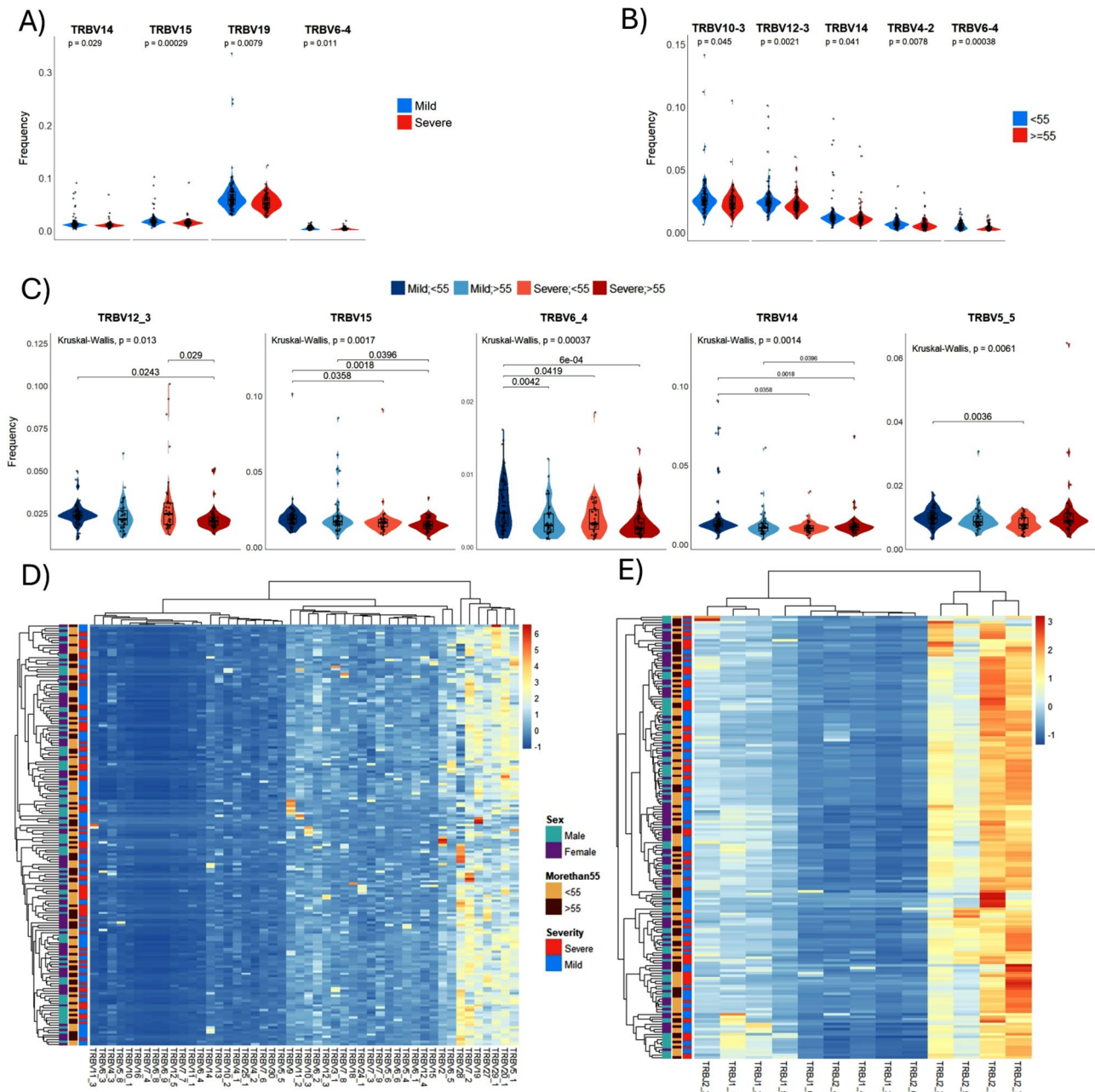


Fig. 2 Main results of the analyses using genes. For allele selection we excluded those clonotypes with ambiguous TRBV and TRBJ annotation. **(A)** V alleles with significant differences in frequency between mild and severe patients and **(B)** patients younger and older than 55 years. **(C)**. Significant TRBV frequencies between the combined age and severity group. The allele frequency is calculated by weighting the degree of clonal expansion of the clonotype in which it is found. **(D)** Heatmap with hierarchical clustering of the total number of patients with the total number of V alleles and **(E)** J alleles detected. The expression levels in both heatmaps have been represented by z-score standardization of the total counts for each allele. Mann-Whitney U test was performed for two-group comparisons and one-way Kruskal-Wallis test for 4-group comparisons, plus post-hoc comparisons by Dunn's test and corrected by the Bonferroni-Holm method (significance level of 0.05)

S%GYE_AEFGHLRSVWY and SL%SYE_DGNST were of particular interest as they showed a tendency for severe patients under 55 to have a profile similar to that of patients over 55, although not as pronounced as in the previous sections. This less pronounced tendency is reflected in the logistic regression models, with variable

number of times each cluster is selected as relevant. However, another GLIPH2 cluster (SPT%AGE-ASV) is highlighted as a relevant motif in both age groups and in the combined age-severity groups (Supplementary Tables 7 and 8).

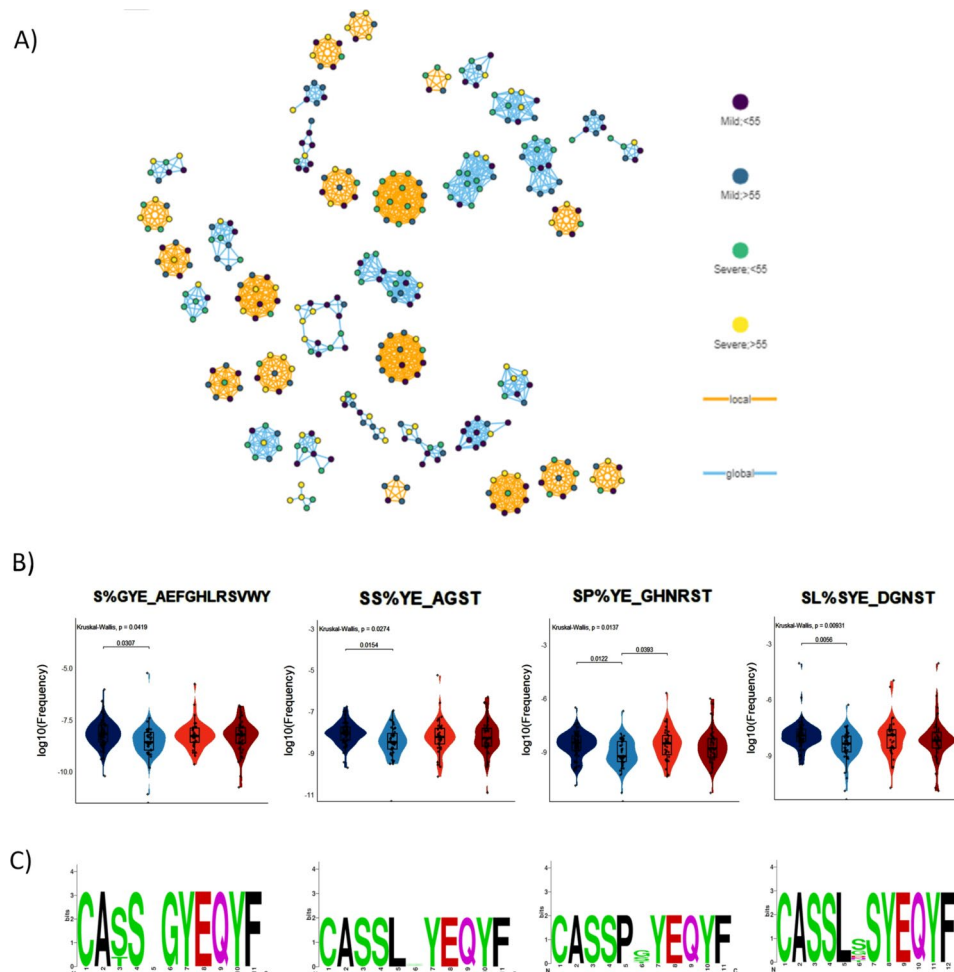


Fig. 3 Main results of the GLIPH2 analysis. **(A)** Network with the 36 global and local clusters initially identified after selection of the 100 most expanded clonotypes of each repertoire. **(B)** Box-violin plot of the 4 clusters identified by GLIPH2 after selection and filtering of CDR3b COVID+ sequences and with significant changes between age and sex groups. **(C)** Sequence logos of the CDR3b COVID+ sequences of each GLIPH2 cluster shown in (B). Mann-Whitney U test test was performed for two-group comparisons and one-way Kruskal-Wallis test for 4-group comparisons plus Dunn's test and corrected by the Bonferroni-Holm method (significance level of 0.05)

Differences between repertoires were notable upon visualizing the frequencies of CDR3b 'COVID+' sequences in each group (see Supplementary Fig. 9). These differences remained significant after removing outliers. WebLogo analysis identified signature variations at intermediate positions in the CDR3b sequence (Fig. 3C), influencing antigen affinity based on molecule type(s) or binding strength. Further examination of CDR3b COVID+ sequences revealed several previously mentioned V alleles. Notably, TRBV12-3 was present in CDR3b chains of all four selected groups, with minor presence of TRBV5-5 and TRBV14/TRBV04-02, indicating ambiguous V allele coding (Supplementary Material 4).

Discussion

Here, we reported a non-significant decrease in diversity, as well as an increase in clonality in severe patients compared to mild patients, consistent with other studies [19, 20]. Present results are consistent with findings supporting that COVID-19 infection and its severity are associated with T-cell lymphopenia [20–23]. Additionally, a reduction in diversity and the proportion of hyper-expanded clones during convalescent phase compared to discharge phase has been reported in recovered patients [24]. Due to a more sustainable T-lymphocyte response and a greater clonal expansion in severe patients [19], the lack of significant diversity and clonality changes can be attributed to the transition of mild patients from a discharge to a convalescent state.

Differences in diversity and clonality are more pronounced when patients are divided into <55 and ≥55 age groups, with significant differences observed. Previous

studies have shown that diversity of TCR repertoires of COVID-19 patients decreased with age [7, 25, 26]. However, our study has revealed that severe patients younger than 55 have significant changes in diversity and clonality, with values very similar to those patients older than 55, regardless of severity. Given the consistent production of naive T lymphocytes in younger patients, a plausible explanation for this phenomenon could be the dysfunction of interferon 1 (INF 1) in disease severity [27].

Regarding repertoire overlap, our findings indicate significant disparity in the similarity of repertoires between mild and severe patients, particularly with hyperexpanded clones. The level of repertoire overlap is low, and it decreases as clonal expansion increases. This result follows the findings of Chang et al., where a minimal sharing of GLIPH2 clustered clonotypes between mild and severe patients was detected. A possible explanation for the limited overlap in repertoire between groups is the variation in antigen selection, leading to the use of different sequence motifs targeting the same antigen [19, 28]. Previous studies have demonstrated a greater degree of overlap between patients within the same severity group.

In the present study, we identified a predominant use of V alleles TRBV7-2, TRBV20-1, and TRBV27 and J alleles such as TRBJ2-1, TRBJ2-3, and TRBJ2-7 in almost all COVID-19 patients. These alleles are involved in high skewness of V/J usage between COVID-19 patients and healthy donors [21–23, 28]. Our analyses show a significant decrease in the frequency of TRBV19, TRBV15, TRBV6-4, and TRBV14 among severe patients compared to mild cases. These findings suggest a potential association between the frequency of these alleles and disease severity. As the differential usage of V/J genes between different severity groups is limited [19], present finding is of relevance for the differentiation of mild and severe cases based on V alleles frequency.

Interestingly severe patients <55 have a frequency of TRBV15, TRBV14 and TRBV6-4 similar to ≥55 patients, following the same line as diversity and clonality results. Several studies have reported noteworthy changes in mentioned alleles regarding COVID-19 severity [10, 19]. Both, our study, and data from Chen et al., found lower frequencies of TRBV19, TRBV15, and TRBV6-4 alleles. The exception was that Chen et al. found a pronounced increase in TRBV14 allele frequency among severe patients. This event is mainly due to TCR affinity for superantigenic regions of SARS-Cov2 [10]. On the other hand, Li et al. reported decreased TRBV15 and other V alleles in convalescent patients [20]. Based on their findings, we suggest that these discrepancies may be due to significant changes in TRBV14 frequencies during the later phases of infection, which may have a greater impact. The decreased occurrence of TRBV6-4 allele in severe patients compared to mild ones is linked to lower

presence of Tc17/MAIT cells; in both acute illness and post-infection phases as these cells have a preferential TRBV6-4 usage [27]. Moreover, TRBV19 and TRBV15 vary T lymphocyte subpopulations at lower levels and changes in their frequencies are associated with an intricate dynamic in T cell repertoires [27].

Finally, we have identified clusters with distinct motifs in central regions of CDR3, which is consistent with previous findings [19, 21, 29, 30]. We observed that certain clusters of CDR3b MIRA-positive sequences have a frequency distribution in severe patients younger than 55; with a tendency that follows the same line as diversity, clonality and gene usage results. However, caution should be taken since the presence of sequence motifs could be influenced by cohort-specific characteristics. This phenomenon was detected using deep learning analysis with variational autoencoder for finding CDR3b signatures linked to COVID-19 severity between two large TCR repertoire datasets [31]. In addition, the frequencies of well differentiated GLIPH2 clusters can vary notably between samples belonging to the same group [32].

Present study has some limitations; TCR repertoire architecture changes could be influenced by Human Leukocyte Antigen (HLA) genotype and TCR capture methodology biases [7, 33]. In addition, bulk TCRb chain analysis cannot capture the absolute clonal identity which comprises paired alpha and beta chains [34]. Another limitation to consider is that the collected data were without fractionating the PBMCs (Peripheral Blood Mononuclear Cells), making it difficult to distinguish specific responses from CD8+ or CD4+ T lymphocytes. In addition, TCR repertoires are strongly influenced by other variables intrinsic to the study cohort or at an individual level [32]. Finally, filtering CDR3b sequences in GLIPH2 clusters may have excluded unannotated sequences with unknown specificity to SARS-CoV-2 antigens. Future studies, such as a more detailed inference of CDR3b sequences and HLA data, will address these limitations using present reported data.

Conclusions

In summary, TCR repertoire analysis is a powerful tool to accurately study changes in T-cell populations, considering the complex dynamics to which they are subjected during COVID-19 infection. Present analysis shows that severe patients below 55 have a similar repertoire of characteristics in terms of diversity, clonality and gene usage to patients above 55 regardless of their prognosis.

Abbreviations

AU	Area Under the Curve
AIC	Akaike Information Criterion
ROC	Receiving Operating Characteristic Curve
CDR3	Complementarity Determining Region 3
CEI	Ethics Committee
CI	Confidence Interval

cDNA	Complementary DNA
COVID-19	Coronavirus Disease-2019
CRP	C reactive protein
GLIPH2	Grouping of Lymphocyte Interactions by Paratope Hotspots 2
HLA	Human Leukocyte Antigen
ICU	Intensive care unit
IgM	Immunoglobuline M
ILC2	Innate Lymphoid Cells 2
IL-6	(interleukin-6)
MERS	Middle East Respiratory Syndrome
MIRA	Multiplex Identification of Antigen-Specific T-Cell Receptors Assay
SARS	Severe Acute Respiratory Syndrome
TCR	T-Cell Receptor
TRBJ	T Bell Receptor Beta-chain J segment
TRBV	T Bell Receptor Beta-chain V segment
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-024-00654-0>.

Supplementary Material 1: Supplementary Methods. This document goes into detail on the methodology used, in particular on the statistics of diversity, clonality and overlap indices, as well as a more detailed description of the data processing in each [results](#) section.

Supplementary Material 2: Supplementary figures and tables. This file contains the supplemental figures and tables generated for this study and mentioned within the manuscript.

Supplementary Material 3: Results of correlation analysis between TRBV and TRBJ alleles against CDRP, LDH, ferritin and D-dimer. This excel book contains the results of applying Spearman's correlation analysis in the whole set of samples and dividing them by Severity, Age and combined groups.

Supplementary Material 4: Main GLIPH2 sequence clustering results (xlsx). This excel book contains several excel sheets which summarizes the most important characteristics of all clusters found when applying the GLIPH2 algorithm, such as V/J allele annotations, CDR3b belonging each cluster and antigen sequence affinities.

Acknowledgements

We want to thank all donors, nursery and clinicians to make this study possible. We also want to thank Biorender for providing the resources necessary to create the Graphical Abstract. Present published results are a part of the PhD thesis of the candidate Fernando Marín Benesiu in the Biochemistry and Molecular Biology Doctoral Program of the University of Granada. We would also like to thank the University of Granada for providing us with the necessary software tools to carry out our work, such as the sequencing platforms and the BioRender academic account to perform the graphical abstract.

Author contributions

LJMG, MJAC, CEB, FJFR and SMD designed the study. SMD, GLT, CEB, FJFR, AICV and CRC, collected patient's samples. SMD, GLT, AICV, CRC, LCR, VAR and FMB managed clinical variables. ES an VAR provided experimental and high-throughput sequencing processes. FMB an LCR performed data and computational analyses. FMB, LJMG and MJAC wrote and reviewed the manuscript. SMD y MJAC obtained the research funding.

Funding

This research received contributions from "Development and Innovation in Biomedicine and Health Sciences in Andalusia (FEDER)", funded by Conserjería de Salud y Familias de la Junta de Andalucía [grant number PECOVID-0006-2020] and from project "Use of NGS and CYTOF methodology to characterize COVID-19 patients" funded by Secretaría General de Universidades Investigación y Tecnología de la Junta de Andalucía [grant number CV20-36740].

Data availability

The code scripts used in this study are available at GitHub (<https://github.com/fmarinb/bulk-TCR-beta-chain-repertoire-analysis-COVID-19>). The TCR β -sequencing data sets used for this project are deposited at the European Genome-Phenome Archive with accession number EGAD50000000477.

Declarations

Ethical approval

The study protocol was approved by the Granada Research Ethics Committee (CEI Granada) with internal code 1329-N-21. Written informed consent was obtained from all participants in accordance with the principles of the Helsinki Declaration.

Competing interests

Authors Carmen Entrala-Bernal and Francisco Javier Fernandez-Rosado are employed by LORGEN G.P. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

¹Department of Biochemistry, Molecular Biology III and Immunology, Faculty of Medicine, University of Granada, Parque Tecnológico de la Salud, Avd. de la Investigación nº 11, Tower C. 11th floor, Granada 18071, Spain

²Centre for Genomics and Oncological Research: Pfizer, Andalusian Regional Government, GENYO, University of Granada, Parque Tecnológico de la Salud, Granada, Spain

³Preventive Medicine and Public Health Service, Hospital Universitario Clínico San Cecilio, Granada, Spain

⁴Casería de Montijo Health Center, Granada Health District, Granada, Spain

⁵Nursery Department, DUE 15 Sanitary Center Almanjajar, Granada, Spain

⁶LORGEN G.P, Ciencias de la Salud - Business Innovation Centre (BIC), Granada, PT, Spain

⁷Ibs Granada, Biosanitary Research Institute of Granada, Granada, Spain

Received: 22 April 2024 / Accepted: 6 August 2024

Published online: 04 September 2024

References

- Sette A, Sidney J, Crotty S. T cell responses to SARS-CoV-2. *Annu Rev Immunol*. 2023; 41:343–73.
- Unterman A, Sumida TS, Nouri N, Yan X, Zhao AY, Gasque V et al. Single-cell multi-omics reveals dysynchrony of the innate and adaptive immune system in progressive COVID-19. *Nat Commun*. 2022;13.
- Vardhana S, Baldo L, Morice WG 2nd, Wherry EJ. Understanding T cell responses to COVID-19 is essential for informing public health strategies. *Sci Immunol*. 2022;7(71):eabo1303.
- Reynolds CJ, Swadling L, Gibbons JM, Pade C, Jensen MP, Diniz MO et al. Discordant neutralizing antibody and T cell responses in asymptomatic and mild SARS-CoV-2 infection. *Sci Immunol*. 2020;5.
- Yu ED, Wang E, Garrigan E, Goodwin B, Sutherland A, Tarke A, et al. Development of a T cell-based immunodiagnostic system to effectively distinguish SARS-CoV-2 infection and COVID-19 vaccination status. *Cell Host Microbe*. 2022;30:388–e3993.
- P. Poon MM, Rybkina K, Kato Y, Kubota M, Matsumoto R, Bloom NI, et al. SARS-CoV-2 infection generates tissue-localized immunological memory in humans. *Sci Immunol*. 2021;6(65):eabl9105
- Gutierrez L, Beckford J, Alachkar H. Deciphering the TCR Repertoire to Solve the COVID-19 Mystery. *Trends Pharmacol Sci*. Elsevier Ltd; 2020. pp. 518–30.
- Kockelbergh H, Evans S, Deng T, Clyne E, Kyriakidou A, Economou A, et al. Utility of Bulk T-Cell Receptor Repertoire Sequencing Analysis in understanding Immune responses to COVID-19. *Diagnostics*. Multidisciplinary Digital Publishing Institute (MDPI); 2022.
- Zornikova KV, Sheetikov SA, Rusinov AY, Iskhakov RN, Bogolyubova AV. Architecture of the SARS-CoV-2-specific T cell repertoire. *Front Immunol*. Frontiers Media S.A.; 2023.

10. Hongying Cheng M, Zhang S, Porritt RA, Noval Rivas M, Paschold L, Willscher E et al. Superantigenic character of an insert unique to SARS-CoV-2 spike supported by skewed TCR repertoire in patients with hyperinflammation. *Proc Natl Acad Sci U S A*. 2020;117(41):25254–62.
11. Martínez-Diz S, Marin-Benesiu F, López-Torres G, Santiago O, Díaz-Cuellar JF, Martín-Esteban S et al. Relevance of TMPRSS2, CD163/CD206, and CD33 in clinical severity stratification of COVID-19. *Front Immunol*. 2022;13.
12. Martínez-Diz S, Morales-Álvarez CM, García-Iglesias Y, Guerrero-González JM, Romero-Cachinero C, González-Cabezuelo JM et al. Analyzing the role of ACE2, AR, MX1 and TMPRSS2 genetic markers for COVID-19 severity. *Hum Genomics*. 2023;17.
13. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12:380–1.
14. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput Biol*. 2015;11.
15. Pagliuca S, Gurnari C, Hong S, Zhao R, Kongkiatkamon S, Terkawi L et al. Clinical and basic implications of dynamic T cell receptor clonotyping in hematopoietic cell transplantation. *JCI Insight*. 2021;6.
16. Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol*. 2020;38:1194–202.
17. Nolan S, Vignali M, Klinger M, Dines JN, Kaplan IM, Svejnova E et al. A large-scale database of T-cell receptor beta (TCR β) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. *Res Sq*. 2020;4:rs.3.rs-51964
18. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. *Genome Res*. 2004;14:1188–90.
19. Chang CM, Feng PH, Wu TH, Alachkar H, Lee KY, Chang WC. Profiling of T cell repertoire in SARS-CoV-2-infected COVID-19 patients between mild Disease and Pneumonia. *J Clin Immunol*. 2021;41:1131–45.
20. Li Y, Hu J, Wang Y, Liu D, Shi Y, Zhang J et al. T-Cell repertoire characteristics of asymptomatic and re-detectable positive COVID-19 patients. *Front Immunol*. 2022;12.
21. Schultheiß C, Paschold L, Simnica D, Mohme M, Willscher E, von Wenserski L, et al. Next-generation sequencing of T and B cell receptor repertoires from COVID-19 patients showed signatures Associated with Severity of Disease. *Immunity*. 2020;53:442–e4554.
22. Niu X, Li S, Li P, Pan W, Wang Q, Feng Y et al. Longitudinal analysis of T and B cell receptor repertoire transcripts reveal dynamic Immune response in COVID-19 patients. *Front Immunol*. 2020;11.
23. Park JJ, Lee KAV, Lam SZ, Moon KS, Fang Z, Chen S. Machine learning identifies T cell receptor repertoire signatures associated with COVID-19 severity. *Commun Biol*. 2023;6.
24. Luo L, Liang W, Pang J, Xu G, Chen Y, Guo X et al. Dynamics of TCR repertoire and T cell function in COVID-19 convalescent individuals. *Cell Discov*. 2021;7.
25. Bartleson JM, Radenkovic D, Covarrubias AJ, Furman D, Winer DA, Verdin E. SARS-CoV-2, COVID-19 and the aging immune system. *Nat Aging*. Springer; 2021. pp. 769–82.
26. Rydyznski Moderbacher C, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, et al. Antigen-Specific Adaptive immunity to SARS-CoV-2 in Acute COVID-19 and associations with Age and Disease Severity. *Cell*. 2020;183:996–e101219.
27. Notarbartolo S, Ranzani V, Bandera A, Guarini P, Bevilacqua V, Rita Putignano A et al. Integrated longitudinal immunophenotypic, transcriptional, and repertoire analyses delineate immune responses in patients with COVID-19. *Sci. Immunol*. 2021;6(62):eabg5021
28. Hou X, Wang G, Fan W, Chen X, Mo C, Wang Y, et al. T-cell receptor repertoires as potential diagnostic markers for patients with COVID-19. *Int J Infect Dis*. 2021;113:308–17.
29. Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, et al. SARS-CoV-2 epitopes are recognized by a public and Diverse Repertoire of Human T Cell Receptors. *Immunity*. 2020;53:1245–e12575.
30. Simnica D, Schultheiß C, Mohme M, Paschold L, Willscher E, Fitzek A et al. Landscape of T-cell repertoires with public COVID-19-associated T-cell receptors in pre-pandemic risk cohorts. *Clin Transl Immunol*. 2021;10.
31. Sidhom JW, Baras AS. Deep learning identifies antigenic determinants of severe SARS-CoV-2 infection within T-cell repertoires. *Sci Rep*. 2021;11.
32. Dannebaum R, Suwalski P, Asgharian H, Du Zhipei G, Lin H, Weiner J et al. Highly multiplexed immune repertoire sequencing links multiple lymphocyte classes with severity of response to COVID-19. *EClinicalMedicine*. 2022;48:101438.
33. Pai JA, Satpathy AT. High-throughput and single-cell T cell receptor sequencing technologies. *Nat Methods Nat Res*; 2021. p. 881–92.
34. Valkiers S, de Vrij N, Gielis S, Verbandt S, Ogunjimi B, Laukens K, et al. Recent advances in T-cell receptor repertoire analysis: bridging the gap with multi-modal single-cell RNA sequencing. *Immunoinformatics*. 2022;5:100009.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.