Machine learning on SARS-CoV-2 Omicron substitutions shows conservation of TCR-epitope recognition potential

Pieter Meysman[1,2], Benson Ogunjimi[1,2,3], Kris Laukens[1,2], and Sander Wuyts[1]

[1]ImmuneWatch BV, Kessel-Lo, Belgium
[2]University of Antwerp, Antwerp, Belgium
[3]Antwerp University Hospital, Antwerp, Belgium

16 December 2021 (v1.0)

1 Abstract

Omicron, the latest SARS-CoV-2 variant of concern, is characterized by a large number of mutations. Here we evaluate the effect of these mutations on T-cell immunity. A T-cell receptor (TCR)/epitope interaction model was implemented that predicts the interaction between a TCR sequence and a putative target epitope. In general, our data support previous findings, that it appears unlikely that the mutations observed in Omicron variant epitopes will lead to escape of T-cell recognition. While we acknowledge the limitations of this predictive approach, we believe that it can help researchers better understand immune responses in the context of vaccinations and infections by providing fine-grained insights into the dynamics of T-cell receptor/epitope binding affinity.

2 Introduction

Since the start of the SARS-CoV-2 pandemic, several variants of concern (VoCs) have emerged that require special attention. They are associated with an increased transmissibility, increased virulence or a decrease in effectiveness of public health or social measures [1]. Predicting and evaluating the effect of the mutations accumulated in these variants is important, as this can help in defining strategic decisions for controlling the pandemic.

The latest VoC, omicron (SARS-CoV-2 B.1.1.529), emerged in November 2021 and is characterized by several deletions and a large number of mutations (>30 on amino acid level), which is much larger than any previous variant classified as VoC [2, 3]. Several of these mutations have been characterized

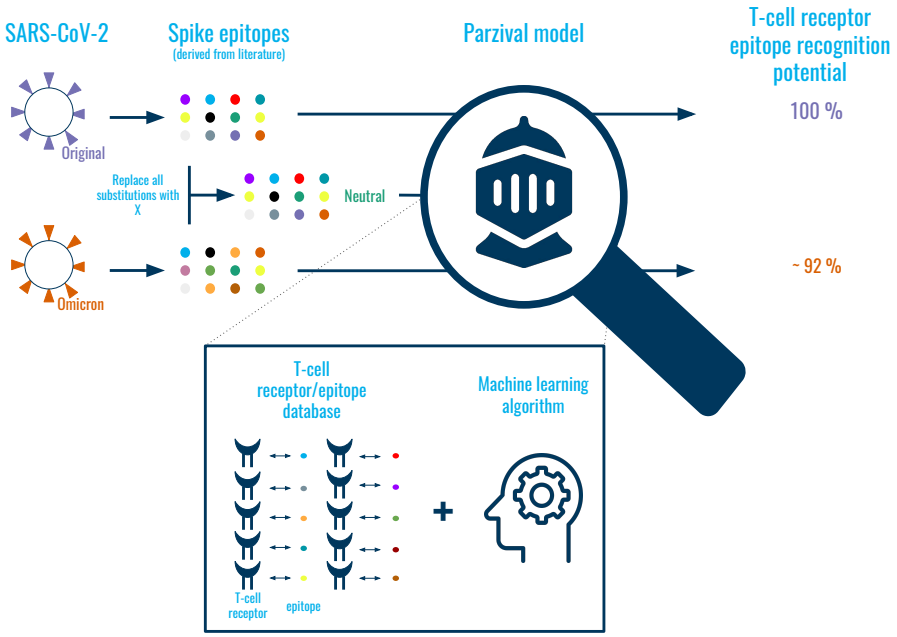

Figure 1: Graphical abstract.

previously and are known to lead to increased transmissibility and higher neutralizing antibody escape [2, 4, 5]. Especially the latter has caused a rise in concerns about omicron's capability to escape the immune system, questioning the effectiveness of naturally acquired or vaccine-induced immunity against the COVID-19 disease caused by this VoC. Given that immunity against COVID-19 is based on both an antibody and a T-cell response [6], and that the antibody response seems to be affected, it is of importance to evaluate how the mutations accumulated in Omicron impact T-cell immunity.

In general, it is unlikely that SARS-CoV-2 variants will escape T-cell recognition as, among other reasons, the T-cell response is characterised by a high number of different epitopes [6]. Nevertheless, a reduction in efficacy can be expected as the acquired mutations could affect MHC presentation or the binding affinity of T-cell receptors (TCRs) to the mutated epitopes. Two studies have already evaluated the impact of Omicron's mutations by analysing which amino acid substitutions were detected within any of SARS-CoV-2's original epitopes [7, 8]. These studies, however, were performed in a conservative way and assumed that every mutation found in an epitope will result in a failure of TCR recognition.

However, an amino acid substitution in a mutated T-cell epitope may still result in recognition by the same T-cell receptor [9]. Validating this novel interaction experimentally in a confident manner is however time-consuming. On the other hand, while being a computationally hard problem, it has been shown

that unseen epitope-TCR interactions can be predicted for epitopes with a small number of substitutions [10]. Thus in this instance, such approach allows to judge if the substitutions will still broadly allow recognition by the same T-cells.

In this study we test this hypothesis and present a fine-grained computational evaluation of the impact of mutations in the SARS-CoV-2 Omicron variant on T-cell receptor recognition. Our approach is based on a novel, in-house developed TCR-epitope interaction model (Parzival) that is able to predict TCR-epitope interactions for mutated epitopes.

3 Results

3.1 Spike protein mutations in CD8+ T-cell epitopes

We found 97 MHC class I Spike protein epitopes from the original SARS-CoV-2 pandemic strain with known T-cell receptor sequences, as reported in the literature. The amino acid mutations on the SARS-CoV-2 Omicron variant [3] were compared against this set of epitopes. Only 16 MHC class I Spike epitopes were affected by a non-synonymous mutation. Three of these epitopes featured an insertion/deletion event in the Omicron variant, which can be assumed to fully disrupt the epitope. The remainder thirteen concerned amino acid substitutions which were the subject of further investigation.

3.2 Omicron substitutions have minor impact on CD8+ TCR specificity

The 13 Omicron Spike proteins substitutions were screened with the Parzival TCR-epitope interaction model to estimate their effect on TCR binding. This model predicts the most likely interaction location based on the TCR and epitope sequence, and the contribution of each amino acid at these positions. An example of two such predictions can be found in Figure 2.

The Omicron mutations were compared to the original epitope as well as an epitope with neutral substitutions at the same position. Overall, as can be seen in Figure 3, the majority of Omicron epitopes retain similar CD8+ T-cell binding scores to those of the original strain. Only five epitopes have a slightly decreased score, indicating that these Omicron epitopes are less likely to be bound by the same T-cells. Of the five with reduced binding potential, three remained stable with a neutral substitution, thus indicating that the Omicron substitutions are worse than a random substitution at the same position. Six epitopes feature binding preferences that are mostly unchanged, and an additional two epitopes have an increased score for the Omicron variant. Notably there was no relation between the number of substitutions and the increase or decrease in score.

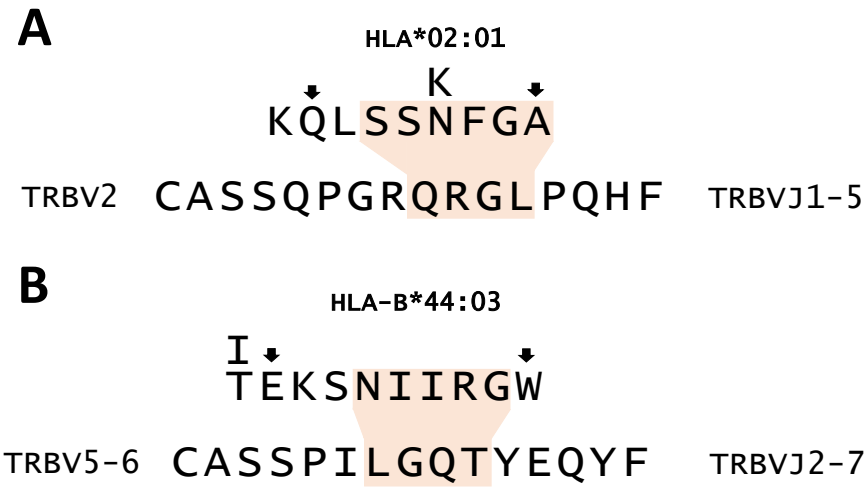

Figure 2: Two examples of Parzival predictions. The beige region denotes the predicted interaction region between the epitope and TCR. The anchor residues with respect to the MHC molecule are indicated by arrows. A) The interaction between the KQL epitope and one of its TCRs, where an Omicron mutation (the letter on top of the epitope depicts the mutation) results in decreased binding potential. (Figure 3) B) the interaction between the TEK epitope and one of its TCRs where the binding is predicted to be unaffected as the mutations is outside of the predicted interaction region. (Figure 3)

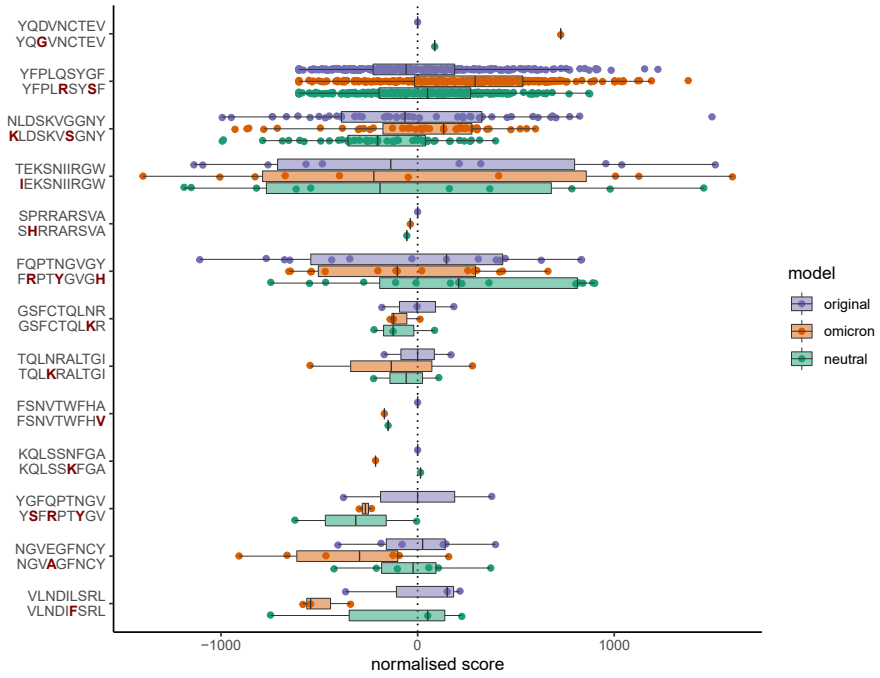

Figure 3: T-cell receptor/epitope binding predictions of the 13 mutated Spike epitopes of the SARS-CoV-2 Omicron variant. Predictions are made for the original SARS-CoV-2 epitopes (original), the mutated Omicron epitopes (omicron) and a neutral amino acid sequence (neutral). The y-axis shows the original epitope (top) and the omicron epitope (bottom) with mutations in the omicron epitope highlighted in dark red (the amino acid sequence of the neutral substitutions are omitted). The x-axis shows the binding prediction score, normalised by subtracting each score from the mean "original" score for each epitope separately. Each dot represents the binding score of one T-cell receptor with its associated epitope (original) and the mutated versions (omicron and neutral).

4 Discussion

Our analysis showed that out of 97 MHC class I Spike protein epitopes under study, only 5 appeared to have a decreased T-cell receptor/epitope binding score due to the mutations gathered in the SARS-CoV-2 Omicron variant. In addition three epitopes are assumed to be fully disrupted by an insertion/deletion event. Taken together, we predicted that 8 out of 97 ($\sim$92%) Spike protein epitopes negatively affect TCR-binding caused by mutations present in the SARS-CoV-2 Omicron variant. As this is only a tiny fraction of epitopes involved in the broad T-cell response, our analysis supports previous findings and predictions that epitopes derived from the SARS-CoV-2 Omicron variant are unlikely to drastically change acquired T-cell immunity [2, 7, 8].

In this manuscript a newly developed algorithm termed Parzival was used that is able to predict TCR-epitope binding interaction strengths. However, this methodology in its current form still suffers from drawbacks. First of all, the area-under-the-ROC-curve, a metric that estimates the quality of the overall prediction quality, is not yet fully meeting expectations. Further algorithmic developments are ongoing to increase the overall quality of the predictive approach [10]. An increasing body of training data will also likely have a positive impact on the algorithmic performance. Secondly, many SARS-CoV-2 epitopes did not have paired alpha-beta chain data. Thus the analysis was currently restricted to β-chain only to allow for larger uniform epitope coverage, with a minor trade-off in loss in performance. Finally, our analysis did not take into account the effect of mutations on MHC presentation, which could also affect T-cell immunity [11]. Nevertheless, we believe that future iterations of Parzival can help researchers better understand immune responses in the context of vaccinations and infections by quickly providing fine-grained insights into the dynamics of T-cell receptor/epitope binding affinity.

5 Conclusions

Using a computational T-cell receptor-epitope interaction model, we predicted that the mutations acquired by the SARS-CoV-2 Omicron variant have limited effect on TCR-binding. While we acknowledge that the algorithmic performance needs further improvement, our data supports previous findings, that it appears unlikely that the mutations observed in Omicron variant epitopes will result in escape of T-cell recognition.

6 Methods

6.1 The Parzival TCR epitope interaction model

The Parzival interaction model is designed to predict how a TCR sequence and a putative target epitope will bind, and the likelihood of binding occurring. The interaction model has been trained against 44,262 unique CD8+ TCR sequences

and 1,006 unique MHC class I target epitopes, which have been collected, curated and re-analysed from scientific literature. On a held-out dataset, Parzival features an area-under-the-ROC-curve of 0.63 on confident non-promiscuous TCR-epitope pairs.

6.2 TCR epitope screening

The known CD8+ TCR sequences against the 16 selected epitopes were filtered to include only those where our algorithm highlighted the TCRβ sequence as critical for the epitope recognition. Parzival then predicted the interaction between the TCRβ and 1) the original epitope, 2) the Omicron variant, 3) a neutral-changed epitope where the differing positions are placed by a 'X' amino acid. Parzival considers any 'X' as neither contributing nor hindering binding in a given position.

7 Acknowledgements

We thank the scientific community for making the data needed to perform this study freely accessible. The knight icon in the graphical abstract was made by Freepik (`https://www.flaticon.com/authors/freepik`)

8 References

References

[1] Classification of omicron (b.1.1.529): Sars-cov-2 variant of concern. `https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern`, Date accessed: 13/12/2021.

[2] Salim S Abdool Karim and Quarraisha Abdool Karim. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *The Lancet*, 398(10317):2126–2128, December 2021.

[3] Gisaid: Tracking of variants. `https://www.gisaid.org/hcov19-variants/`, Date accessed: 13/12/2021.

[4] Allison J Greaney, Tyler N Starr, Pavlo Gilchuk, Seth J Zost, Elad Binshtein, Andrea N Loes, Sarah K Hilton, John Huddleston, Rachel Eguia, Katharine HD Crawford, et al. Complete mapping of mutations to the sars-cov-2 spike receptor-binding domain that escape antibody recognition. *Cell host & microbe*, 29(1):44–57, 2021.

[5] William T Harvey, Alessandro M Carabelli, Ben Jackson, Ravindra K Gupta, Emma C Thomson, Ewan M Harrison, Catherine Ludden, Richard

Reeve, Andrew Rambaut, Sharon J Peacock, et al. Sars-cov-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, 19(7):409–424, 2021.

[6] Alba Grifoni, John Sidney, Randi Vita, Bjoern Peters, Shane Crotty, Daniela Weiskopf, and Alessandro Sette. SARS-CoV-2 human T cell epitopes: Adaptive immune response against COVID-19. *Cell Host & Microbe*, 29(7):1076–1092, July 2021.

[7] Damon H. May, Benjamin Rubin, Sudeb C. Dalai, Krishna Patel, Shahin Shafiani, Rebecca Elyanow, Matthew T. Noakes, Thomas M. Snyder, and Harlan S. Robins. Omicron variant partially escapes the T-cell response induced by SARS-CoV-2 vaccines. `https://www.adaptivebiotech.com/wp-content/uploads/2021/12/omicron_t_cell_impact.pdf`, Date accessed: 13/12/2021.

[8] Andrew D Redd, Alessandra Nardin, Hassen Kared, Evan M Bloch, Brian Abel, Andrew Pekosz, Oliver Laeyendecker, Michael Fehlings, Thomas C Quinn, and Aaron AR Tobian. Minimal cross-over between mutations associated with omicron variant of sars-cov-2 and cd8+ t cell epitopes identified in covid-19 convalescent individuals. *bioRxiv*, 2021.

[9] Alison Tarke, John Sidney, Nils Methot, Esther Dawen Yu, Yun Zhang, Jennifer M. Dan, Benjamin Goodwin, Paul Rubiro, Aaron Sutherland, Eric Wang, April Frazier, Sydney I. Ramirez, Stephen A. Rawlings, Davey M. Smith, Ricardo da Silva Antunes, Bjoern Peters, Richard H. Scheuermann, Daniela Weiskopf, Shane Crotty, Alba Grifoni, and Alessandro Sette. Impact of SARS-CoV-2 variants on the total CD4+ and CD8+ T cell reactivity in infected or vaccinated individuals. *Cell Reports Medicine*, 2(7), July 2021. Publisher: Elsevier.

[10] Pieter Moris, Joey De Pauw, Anna Postovskaya, Sofie Gielis, Nicolas De Neuter, Wout Bittremieux, Benson Ogunjimi, Kris Laukens, and Pieter Meysman. Current challenges for unseen-epitope tcr interaction prediction and a new perspective derived from image classification. *Briefings in Bioinformatics*, 2020.

[11] Benedikt Agerer, Maximilian Koblischke, Venugopal Gudipati, Luis Fernando Montaño-Gutierrez, Mark Smyth, Alexandra Popa, Jakob-Wendelin Genger, Lukas Endler, David M. Florian, Vanessa Mühlgrabner, Marianne Graninger, Stephan W. Aberle, Anna-Maria Husa, Lisa Ellen Shaw, Alexander Lercher, Pia Gattinger, Ricard Torralba-Gombau, Doris Trapin, Thomas Penz, Daniele Barreca, Ingrid Fae, Sabine Wenda, Marianna Traugott, Gernot Walder, Winfried F. Pickl, Volker Thiel, Franz Allerberger, Hannes Stockinger, Elisabeth Puchhammer-Stöckl, Wolfgang Weninger, Gottfried Fischer, Wolfgang Hoepler, Erich Pawelka, Alexander Zoufaly, Rudolf Valenta, Christoph Bock, Wolfgang Paster, René Geyeregger, Matthias Farlik, Florian Halbritter, Johannes B. Huppa, Judith H. Aberle,

and Andreas Bergthaler. Sars-cov-2 mutations in mhc-i-restricted epitopes evade cd8+ t cell responses. *Science Immunology*, 6(57):eabg6461, 2021.