Identification of SARS-CoV-2 T-cell Epitopes for Assessing T-cell Immunity

Eunok Lee[a,b], Kerrie Sandgren[a,b], Gabriel Duette[a,b], Vicki V. Stylianou[a,b], Rajiv Khanna[c,d], John-Sebastian Eden[a,b,e,f], Emily Blyth[b,g,h], David Gottlieb[b,g,h], Anthony L. Cunningham[a,b] and Sarah Palmer[a,b]

[a] Centre for Virus Research, The Westmead Institute for Medical Research, Westmead, New South Wales, Australia
[b] School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, New South Wales, Australia
[c] Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia
[d] QIMR Berghofer Centre for Immunotherapy and Vaccine Development, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
[e] Centre for Infectious Diseases and Microbiology, The Westmead Institute for Medical Research, Westmead, New South Wales, Australia
[f] Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Westmead, New South Wales, Australia
[g] Centre for Cancer Research, The Westmead Institute for Medical Research, Westmead, New South Wales, Australia
[h] BMT and Cell Therapies Program, Westmead Hospital, Westmead, New South Wales, Australia

Acknowledgements

- All authors who kindly shared SARS-CoV-2 genomic data on GISAID.
- BEAT COVID-19 grant sponsored by Snow Medical.
- NSW Health COVID-19 Research Grants.
- We acknowledge with gratitude the participants of this study.

1 Challenges for selecting T-cell epitopes against SARS-CoV-2

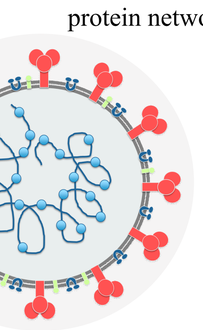
- Recombination between coronaviruses is common.
- The number of mutations within SARS-CoV-2 genome is increasing.
- Human Leukocyte Antigen (HLA) alleles are polymorphic.
- Cellular immunity against SARS-CoV-2 is cross-reactive to seasonal human coronaviruses that cause the common cold.

2 Aims

- Design a novel immunoinformatics analysis pipeline that selects for immunogenic peptides from non-variable and topologically important (i.e. highly networked) protein regions.
- Monitor genetic mutations within Spike and Nucleocapsid protein sequences and determine if the immunogenic peptides selected by our immunoinformatics analysis pipeline avoid these mutations.

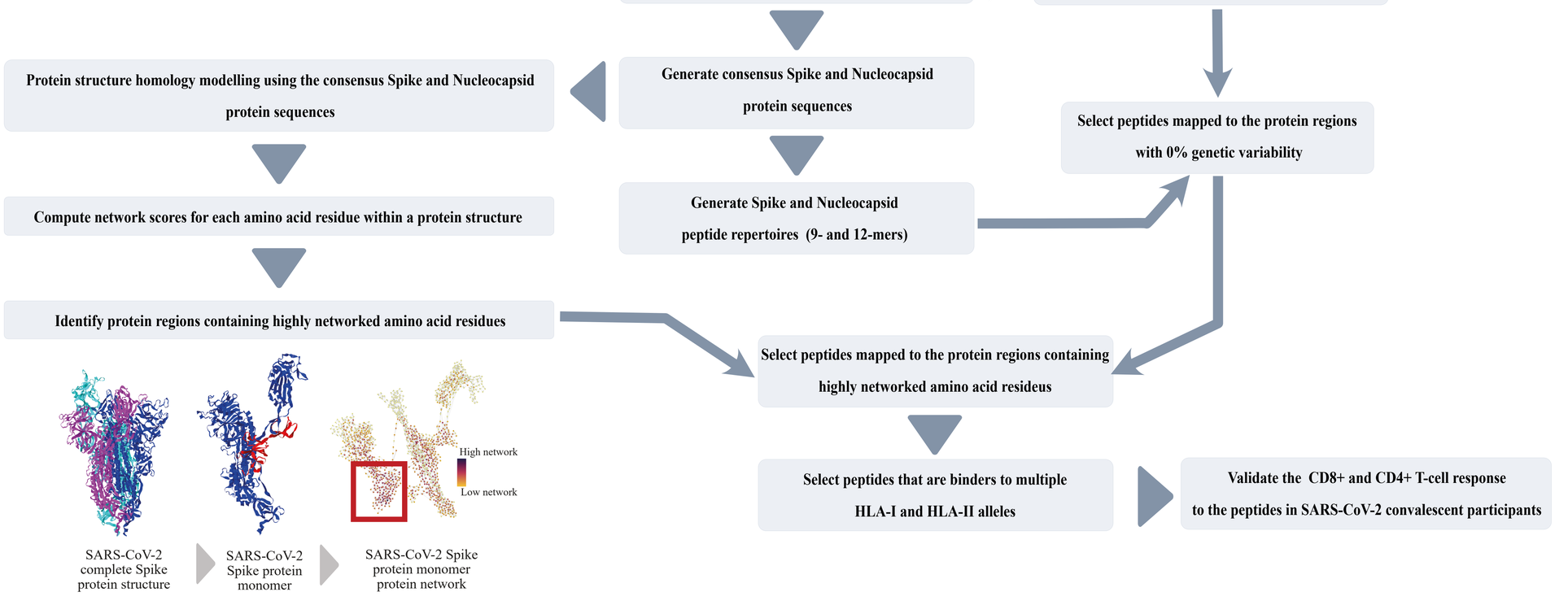
3 New immunoinformatics analysis pipeline for T-cell epitope selection

- The regions comprised of "highly networked" amino acid residues are topologically important for the maintenance of tertiary and quaternary viral protein structures.
- In human immunodeficiency virus (HIV)-infected individuals with diverse HLA class I alleles, targeting epitopes from these topologically important (i.e. highly networked) regions with cytotoxic T-cells provided virologic control (Gaiha et al., 2019).
- Therefore, out immunoinformatics analysis pipeline integrated open-access databases/tools with protein network analysis.



- Our target SARS-CoV-2 proteins for selecting highly networked T-cell epitope derived peptides are Spike and Nucleocapsid.

Spike
Nucleocapsid



4 Highly networked T-cell epitope derived peptides for SARS-CoV-2 specific T-cell immunity assays

- We have identified a total of 57 highly networked T-cell epitope derived peptides from Spike and Nucleocapsid proteins.
- The peptides identified from Spike protein avoid mutations that enhance viral infectivity (Lee et al., 2021).
- Of these, 18 peptides have limited genetic homology to seasonal human coronaviruses making them promising candidates for SARS-CoV-2 specific T-cell immunity assays.

Peptides derived from highly networked protein regions

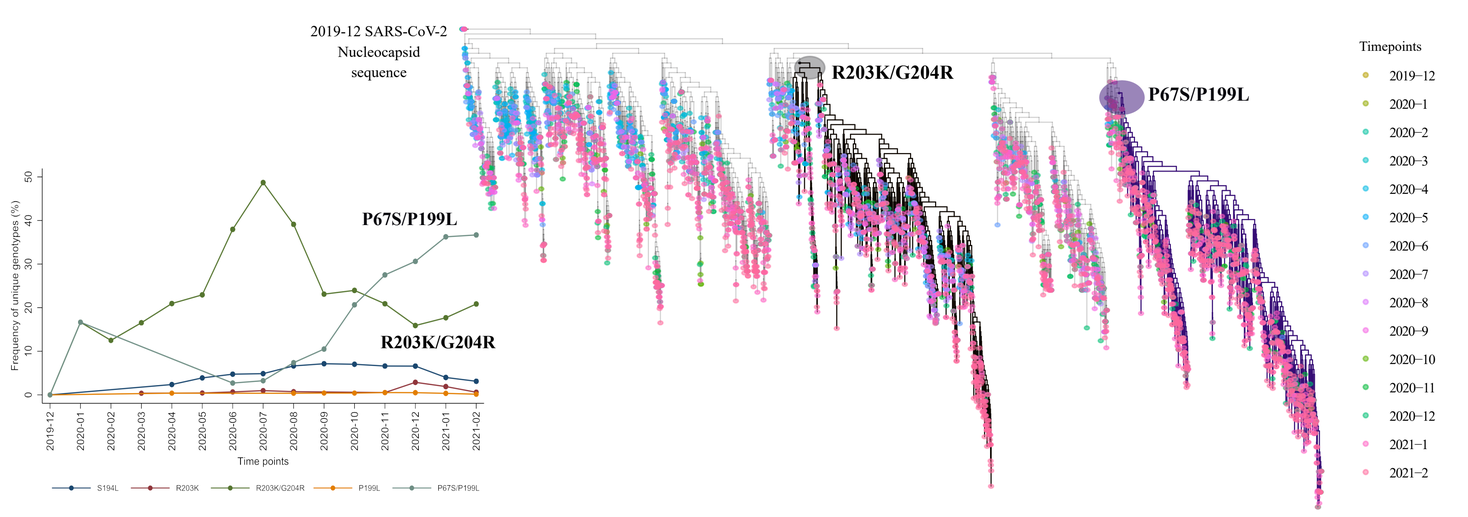
9-mers				12-mers			
NC N-terminal	NTASWFTAL			NC N-terminal	QIGYYRRATRRI	■	
NC N-terminal	TASWFTALT			NC N-terminal	IGYYRRATRRIR	■	
NC N-terminal	IIWVATEGA	■		NC N-terminal	GYYRRATRRIRG	■	
NC N-terminal	TTLPKGFYA			NC N-terminal	IIWVATEGALNT	■	
NC C-terminal	RTATKAYNV	■		NC N-terminal	GIIWVATEGALN	■	
NC C-terminal	ILLNKHIDA	■	■	S protein	LFLPFFSNVTWF	■	■
S protein	FLPFFSNVT			S protein	FLPFFSNVTWFH	■	■
S protein	VTWFHAIHV	■	■	S protein	LPFFSNVTWFHA	■	■
S protein	TLDSKTQSL			S protein	PFFSNVTWFHAI	■	■
S protein	FQFCNDPFL	■		S protein	FFSNVTWFHAIH	■	■
S protein	FCNDPFLGV			S protein	FSNVTWFHAIHV	■	■
S protein	PLVDLPIGI			S protein	SNVTWFHAIHVS	■	■
S protein	YLQPRTFLL	■		S protein	NVTWFHAIHVSG	■	■
S protein	AVDCALDPL	■		S protein	VTWFHAIHVSGT	■	■
S protein	FSTFKCYGV	■		S protein	CTFEYVSQPFLM	■	■
S protein	NVYADSFVI			S protein	YVGYLQPRTFLL	■	
S protein	RVVVLSFEL	■	■	S protein	EKGIYQTSNFRV	■	
S protein	SIIAYTMSL	■	■	S protein	KGIYQTSNFRVQ	■	
S protein	SVTTEILPV	■		S protein	FNFNGLTGTGVL	■	
S protein	LLQYGSFCT	■		S protein	TWRVYSTGSNVF	■	
S protein	QLNRALTGI	■		S protein	WRVYSTGSNVFQ	■	
S protein	KQIYKTPPI			S protein	SNVFQTRAGCLI	■	
S protein	LLFNKVTLA	■		S protein	NVFQTRAGCLIG	■	
S protein	GLTVLPPLL	■		S protein	QSIIAYTMSLGA	■	■
S protein	MIAQYTSAL	■		S protein	SIIAYTMSLGAE	■	■
S protein	ALLAGTITS			S protein	IIAYTMSLGAEN	■	■
S protein	ITSGWTFGA	■		S protein	IAYTMSLGAENS	■	■
S protein	WTFGAGAAL	■		S protein	EMIAQYTSALLA	■	
S protein	LQIPFAMQM	■		S protein	QIPFAMQMAYRF	■	
S protein	QMAYRFNGI	■		S protein	IPFAMQMAYRFN	■	
S protein	VLYENQKLI	■		S protein	PFAMQMAYRFNG	■	
S protein	ALNTLVKQL	■		S protein	FAMQMAYRFNGI	■	
S protein	KQLSSNFGA	■					
S protein	VLNDILSRL	■					
S protein	RLDKVEAEV	■					

■ T-cell epitope derived peptides with top 5% network scores, binding levels to HLA-I and HLA class I mediated antigen processing and immunogenicity scores

■ T-cell epitope derived peptides with top 5% network scores and binding levels to HLA-II

■ T-cell epitope derived peptides with <20% genetic homology to four seasonal human coronaviruses (229E, HKU1, NL63, OC43)

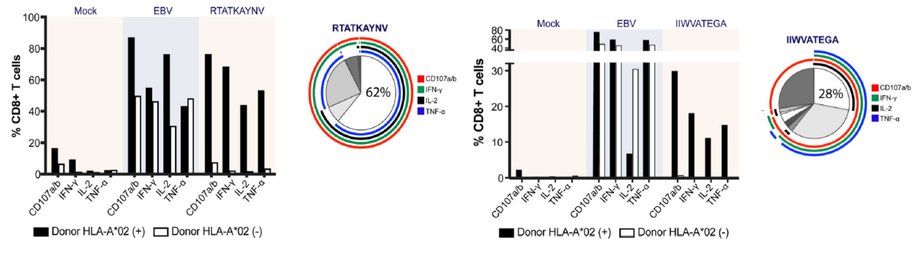
5 A predictive fitness model for Nucleocapsid protein identified two linked mutations which contribute to an increase in the frequency of SARS-CoV-2 variants



- Within Nucleocapsid protein sequences, the predicted fitness model identified P67S/P199L mutation within global SARS-CoV-2 variants sampled from January 2020 to February 2021. The number of Nucleocapsid sequences containing P67S/P199L mutation increased during this period, suggesting that this mutation can enhance viral fitness.
- We also found another mutation, R203K/G204R, within global SARS-CoV-2 variants sampled from January 2020 to February 2021. However, we observed a rapid increase followed by a fast decline in the number of Nucleocapsid sequences containing R203K/G204R mutation from May 2020 to September 2020. This fluctuation in the number of viral variants is possibly due to: 1) immunological control; and/or 2) cross-immunity from previous exposure to other seasonal human coronaviruses that contribute to the persistence of viral variants that contain R203K/G204R mutation.
- None of the highly networked T-cell epitope derived peptides identified from Nucleocapsid contained P67S/P199L and/or R203K/G204R mutations.

6 T-cell epitope derived peptides identified from Nucleocapsid elicit CD8+ T-cell polyfunctional/effector response in SARS-CoV-2 convalescent participants

- When peripheral blood mononuclear cells obtained from participants 1-2 months postrecovery were stimulated by using two Nucleocapsid-derived peptides, we observed robust production of interleukin-2 (IL-2), interferon gamma (IFN-g), tumour necrosis factor alpha (TNF-alpha) and a marker for degranulation of CD8+ T-cells (CD107a/b).
- Importantly, 28-62% of the responding CD8+ T-cells were polyfunctional exhibiting four effector functions simultaneously.



7 Conclusions

- Our immunoinformatics analysis pipeline defined 57 SARS-CoV-2 immunogenic peptides within highly networked (i.e. topologically important regions of Nucleocapsid and Spike proteins that avoid genetic mutations that enhance viral fitness or infectivity.
- Of these, 18 had limited homology to seasonal human coronaviruses and therefore are promising candidates for distinguishing SARS-CoV-2-specific immune response from pre-existing coronavirus immunity.
- Importantly, CD8+ T-cells obtained from COVID-19 survivors exhibited polyfunctional/effector responses to highly networked T-cell epitope derived peptides identified from Nucleocapsid protein, providing a proof of concept that our immunoinformatics analysis pipeline selects novel immunogens which can elicit polyfunctional SARS-CoV-2-specific T-cell response.

Gaiha GD, Rossin EJ, Urbach J, Landeros C, Collins DR, Nwonu C, Muzhingi I, Anahtar MN, Waring OM, Piechocka-Trocha A, Waring M, Worrall DP, Ghebremichael MS, Newman RM, Power KA, Allen TM, Chodosh J, Walker BD. 2019. Structural topology defines protective CD8+ T cell epitopes in the HIV proteome. Science 364:480–484. https://doi.org/10.1126/science.aav5095.
Lee E, Sandgren K, Duette G, Stylianou VV, Khanna R, Eden J-S, Blyth E, Gottlieb D, Cunningham AL, Palmer S. 2021. Identification of SARS-CoV-2 nucleocapsid and spike T-cell epitopes for assessing T-cell immunity. J Virol 95:e02002-20. https://doi.org/10.1128/JVI.02002-20.
Tarke A, Sidney J, Kidd CK, Dan JM, Ramirez SI, Yu ED, Mateus J, da Silva Antunes R, Moore E, Rubiro P, Methot N, Phillips E, Mallal S, Frazier A, Rawlings SA, Greenbaum JA, Peters B, Smith DM, Crotty S, Weiskopf D, Grifoni A, Sette A. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. Cell Rep Med. 2021 Feb 16;2(2):100204. doi: 10.1016/j.xcrm.2021.100204. Epub 2021 Jan 26. PMID: 33521695; PMCID: PMC7837622.