Paper Title

Lack of diversity in T cell receptor sequencing studies

Authors and Affiliations

Yu-Ning Huang⁺

Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, CA, USA

yuninghu@usc.edu

ORCID: <https://orcid.org/0000-0003-1697-4267>

Kerui Peng⁺

Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, CA, USA

keruipen@usc.edu

ORCID: <https://orcid.org/0000-0002-9362-4618>

Alice B. Popejoy

Department of Biomedical Data Science (DBDS), Stanford University, Stanford, CA, USA

apopejoy@stanford.edu

ORCID: <https://orcid.org/0000-0003-4976-0628>

Jieting Hu

Department of Translational Genomics, Keck School of Medicine, University of Southern California, CA, USA

jjietingh@usc.edu

Theodore Scott Nowicki

Department of Pediatrics, Division of Pediatric Hematology/Oncology, University of California, Los Angeles, CA, USA

TNowicki@mednet.ucla.edu

ORCID: <https://orcid.org/0000-0002-3815-6137>

Stefan M. Gold

Institut für Neuroimmunologie und Multiple Sklerose (INIMS), Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany

Charité – Universitätsmedizin Berlin, Freie Universität Berlin, Humboldt Universität zu Berlin, Klinik für Psychiatrie und Psychotherapie, Campus Benjamin Franklin, Berlin, Germany

Charité – Universitätsmedizin Berlin, Freie Universität Berlin, Humboldt Universität zu Berlin, Medizinische Klinik m.S. Psychosomatik, Campus Benjamin Franklin, Berlin, Germany

stefan.gold@charite.de

ORCID: <https://orcid.org/0000-0001-5188-4799>

Twitter: @GoldLabCharite

Lluís Quintana-Murci

Human Evolutionary Genetics Unit, CNRS UMR2000, Institut Pasteur, 25-28 rue du Dr Roux, Paris, France

Human Genomics and Evolution, Collège de France, 11 Place Marcelin Berthelot, Paris, France

quintana@pasteur.fr

ORCID : <https://orcid.org/0000-0003-2429-6320>

Twitter: @QuintanaMurci

Macarena Fuentes-Guajardo

Departamento de Tecnología Médica, Facultad de Ciencias de la Salud, Universidad de Tarapacá, Arica 1000009, Chile

mafuentesg@academicos.uta.cl

ORCID: <https://orcid.org/0000-0002-2486-1628>

Mikhail Shugay

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

Pirogov Russian National Research Medical University, Moscow, Russia

mikhail.shugay@gmail.com

ORCID: <https://orcid.org/0000-0001-7826-7942>

Victor Greiff

Department of Immunology, University of Oslo, Norway

victor.greiff@medisin.uio.no

ORCID: <https://orcid.org/0000-0003-2622-5032>

Twitter: @victorgreiff

Amanda M. Burkhardt

Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, CA, USA

aburkhar@usc.edu

ORCID: <https://orcid.org/0000-0002-7326-474X>

Houda Alachkar

Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, CA, USA

alachkar@usc.edu

Serghei Mangul

Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, CA, USA

serghei.mangul@gmail.com

ORCID: <https://orcid.org/0000-0003-4770-3443>

Twitter: @smangul1

+ These authors contributed equally to this work.

Correspondence: serghei.mangul@gmail.com

Introduction

The emergence of high-throughput sequencing techniques and the development of bioinformatics tools provide efficient ways to profile human adaptive immune receptor repertoire (AIRR) including repertoires of T cell receptors (TCRs) and immunoglobulins^{1,2}. Despite these advancements, the representation of populations in the AIRR sequencing (AIRR-Seq) studies remains unknown. The knowledge of AIRR in diverse populations will potentially elucidate genomic and phenotypic similarities or differences in the human immune system and immune-related diseases across populations^{3,4}.

It was previously demonstrated that the majority of the participants in genome-wide association studies (GWAS) were of European ancestry^{5,6}. Efforts to investigate the diversity in AIRR-Seq stand in stark contrast⁴. To address this emerging issue and to further develop a strategic roadmap to engage diverse populations in AIRR-seq studies, we investigated the race/ethnicity/ancestry (REA) diversity in TCR sequencing (TCR-Seq) studies. We also discussed the challenges and opportunities of including underrepresented populations in TCR-Seq research and how findings from this study may guide future efforts to increase diversity in the field of immunogenomics.

Results

We performed a survey of 114 TCR-Seq studies including 3263 study participants (Supplemental Table 1) on their REA information. We examined the current stage of REA information availability across TCR-Seq studies, we also presented the proportion of population groups from various aspects based on the available information. Information was extracted from the text of TCR-Seq publications or directly acquired from the corresponding authors. Details on how studies were selected and methods used to obtain information about study participants are described in the Supplemental Materials.

The majority of TCR-Seq studies lack study participants' race, ethnicity, ancestry (REA) information

We first examined the availability of study participants' REA information in TCR-Seq studies from publications. Less than 20% of the surveyed studies included such information in the text of the paper (Figure S1a). We then obtained study participants' REA information from additional 21 studies by sending inquiries to corresponding authors. In total, we collected the study participants' REA information from 42 TCR-Seq studies across 3263 individuals for the subsequent analysis (Figure S1a).

We further explored the availability of population information through PubMed Medical Subject Headings (MeSH) terms search. Less than 4% among 2,480 immunogenetics studies included both "immunogenetics" and "population" MeSH terms. Surprisingly, only 0.3% among 37,824 TCR-Seq studies (publications tagged with "Receptors, Antigen, T-Cell" MeSH term) included "population" MeSH term. Observations were similar between the manual text check on publications and the MeSH term search, that most of the TCR-Seq studies lack study population information.

Finally, we investigated the reasons for the unavailability of study participants' REA information. Among studies without available study participants' REA information, 6% of the authors did not share data due to privacy concerns and limitations in study designs. First, the authors had concerns about violating the U.S. Health Insurance Portability and Accountability Act (HIPAA) rules. Second, the authors were not able to share participants' information due to the limitations of the study protocols that were already approved by the Institutional Review Board (IRB). For example, REA information was not included during the recruitment phase, or the study participants' information was not approved to be shared with researchers that were from other institutions. Last, some studies utilized de-identified samples, which prevented authors from obtaining the study participants' information.

Individuals of non-European descent are underrepresented in TCR-Seq studies

Firstly, we analyzed the REA diversity among the examined TCR-Seq studies. Nearly 60% of the studies included European or European descent participants, while only 10% of the studies included African or African descent participants (Figure S1b). From the perspective of the study participants, more than 80% of study participants were reported to be European REA groups, followed by 9% of participants from Asian REA groups, 4% of participants of African REA groups (Figure S1b).

Secondly, we examined the proportion of TCR-Seq studies that were conducted in a single REA group and multiple REA groups. In total 33 out of 42 studies were conducted in a single self-reported ancestry or race/ethnicity group (Figure S1c). While European-based studies were dominant and followed by Asian-based studies, notably, no studies were conducted in African populations alone. Nine of the examined studies included study participants from multiple REA groups, all of which predominantly consisted of European populations. (Supplemental Table 1).

Thirdly, we focused on the ethnic information among 12 U.S.-based TCR studies. Only 4% of study participants were self-reported as Hispanic, which indicated that Hispanics were highly underrepresented in TCR-Seq studies, relative to their proportion of the U.S. population (Figure S1d). Furthermore, we investigated the temporal dynamics of REA diversity in TCR-Seq studies. From 2009 to 2021, there was a 10% increase in the proportion non-European individuals (Figure S1e). Despite the increase, the distribution of study participants based on their ethnicity, race, ancestry is still highly disproportionate in TCR-Seq studies.

Additionally, we examined the relationship between sample sizes and the population component in TCR-Seq studies. The results showed that the average number of study participants of European ancestry was much greater than the average number of study participants of Asian and African ancestries (Figure S1f). Lastly, in light of the COVID-19 pandemic, the pattern of skewed diversity in TCR-Seq studies on COVID patients has also been observed. Substantial bias towards participants from European or of European ancestry was observed across all five COVID TCR-Seq studies that we evaluated (Figure S1g) and four studies mainly included European populations (Figure S1h).

Discussion

Our study is the first to report the REA information availability and REA diversity in TCR-Seq studies. Similar to GWAS, our findings reveal that individuals of non-European ancestry were severely underrepresented in TCR-Seq studies. The disproportionate distribution of study participants' self-reported ancestry or race/ethnicity information in the TCR-Seq studies might restrict our understanding of disease pathology in diverse populations, confine the discovery of immunogenomics variants across populations⁶, and hinder the advancement in precision medicine^{7,8}. Furthermore, the severe lack of available REA information in most of the TCR-Seq studies was reported. This might potentially limit the reuse of TCR-Seq data for secondary analysis to infer novel population-specific TCR alleles for improving the representation of the diverse populations in current reference databases.

According to a previously published work, over 50% of 448 researchers and clinical genetics professionals surveyed considered REA important in clinical settings.⁹ Immunogenomics studies could potentially be beneficial to a broad range of populations if they are conducted across diverse populations. The comprehensive understanding of AIRR in diverse populations will accelerate its translational and clinical applications, and eventually promote health equity⁷.

There are a few limitations in our study. First, the unavailability of genetic ancestry information may prevent us from accurately categorizing the study participants. Ultimately, it is preferable to

examine ancestry by genetics-based methodologies. The genetic ancestry depicts the single-nucleotide variants across geographic origin groups and depicts the extent of single-nucleotide variants among individuals of different ancestries¹⁰. However, none of the studies we examined performed genotyping or other computational methods to infer the study participants' genetic ancestry information. The use of the self-reported ancestry or race/ethnicity in TCR-seq studies may be considered to be reasonable at the current stage¹¹. Second, the unified protocol of reporting study participants' REA information may introduce bias in our analysis. The lack of standardized terminologies of the REA information might cause inconsistency among researchers, and make it challenging to conduct secondary analysis. Developing and adopting standardized experimental protocols and computational methods to report or infer genetic ancestry in the field of immunogenomics were urgently needed¹².

In conclusion, based on the results from this preliminary study, we recognize the opportunities and challenges in addressing the diversity in AIRR-Seq studies. The inclusion of diverse populations in TCR-Seq studies is crucial in future work.

Code Availability

All the code used to produce the results of this study, including figures, is available at GitHub: <https://github.com/Mangul-Lab-USC/Diversity-in-TCR-Seq-studies>

Acknowledgment

S.M. is partially supported by National Science Foundation grants 2041984. M.S. is supported by the Ministry of Science and Higher Education of the Russian Federation grant no. 075-15-2020-807. We thank all the authors (Dr. Anastasia A. Minervina, Dr. Mikhail V. Pogorelyy, Dr. Grigory A Efimov, Dr. Zheming Lu, Dr. Yang Ke, Dr. Xiao Liu, Dr. Mascha Binder, Dr. Dmitriy M. Chudakov, Dr. Nathalie Bedard, Dr. Jun S. Liu, Dr. Michelle Miron, Dr. Maura Rossetti, Dr. Satu Mustjoki, Dr. Atsunari Kawashima, Dr. Matthew A. Brown, Dr. Jorge Correale, and Dr. Tae Jin Kim) who shared the information of the study participants with us.

Conflict of interest statement

None declared.

References

1. Brown, A. J. *et al.* Augmenting adaptive immunity: progress and challenges in the quantitative engineering and analysis of adaptive immune receptor repertoires. *Mol. Syst. Des. Eng.* **4**, 701–736 (2019).
2. Bradley, P. & Thomas, P. G. Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. *Annu. Rev. Immunol.* **37**, 547–570 (2019).

3. Goulielmos, G. N. *et al.* The genetics and molecular pathogenesis of systemic lupus erythematosus (SLE) in populations of different ancestry. *Gene* **668**, 59–72 (2018).
4. Peng, K. *et al.* Diversity in immunogenomics: the value and the challenge. *Nat. Methods* (2021).
5. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
6. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
7. Precision medicine needs an equity agenda. *Nat. Med.* **27**, 737 (2021).
8. Greiff, V., Yaari, G. & Cowell, L. G. Mining adaptive immune receptor repertoires for biological and clinical information using machine learning. *Curr. Opin. Syst. Biol.* **24**, 109–119 (2020).
9. Popejoy, A. B. *et al.* Clinical Genetics Lacks Standard Definitions and Protocols for the Collection and Use of Diversity Measures. *Am. J. Hum. Genet.* **107**, 72–82 (2020).
10. Jorde, L. B. & Bamshad, M. J. Genetic Ancestry Testing: What Is It and Why Is It Important? *JAMA* **323**, 1089–1090 (2020).
11. Oni-Orisan, A., Mavura, Y., Banda, Y., Thornton, T. A. & Sebro, R. Embracing Genetic Diversity to Improve Black Health. *N. Engl. J. Med.* **0**, null (2021).
12. Morales, J. *et al.* A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* **19**, 21 (2018).
13. Corrie, B. D. *et al.* iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.* **284**, 24–41 (2018).

Supplementary Materials

Figure

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Supplemental Figure 1. Survey of 114 TCR-Seq studies of the study participants' Race, ethnicity, and ancestry (REA) information. (a) Pie chart for the availability of the study participants' REA information of the TCR-Seq studies. (b) Bar chart for the comparisons of the proportion of reported REA information among total study participants and total studies. (c) Pie chart for the proportion of the studies that reported with a single REA group (Mono-ancestry) and multiple REA groups (Trans-ancestry). (d) Pie chart for the proportion of Hispanic and non-Hispanic individuals in the U.S.-based TCR-Seq studies. (e) Bar plot for the changes in the proportion of study participants' REA information over years. (f) Boxplot for the number of study participants in each REA group among the TCR-Seq studies. (g) Pie chart for the proportion of COVID patients' REA information. (h) Bar plot for the proportion of the COVID patients' REA information among the five covid cohorts based on the number of the COVID patients.

Method

Examine the ancestry availability and information across TCR-Seq studies and categorize the study participants for analysis

Among current AIRR studies, we chose to focus on the T cell receptor sequencing (TCR-Seq) studies. We checked for the study participants' ethnicity, race, ancestry information among the published TCR studies that have available TCR-Seq data through PubMed and online public repositories, including iReceptor¹³ and immuneACCESS® (Adaptive Biotechnologies), till January 2021. We collected either the study participants' self-reported ancestry or race/ethnicity. We included the studies that included ancestry or race/ethnicity information in the study design. Conversely, email inquiries were made to the authors to acquire relevant information. Two follow-up email inquiries were made. After the email inquiries, we acquired a total of 42 studies with available study participants' self-reported ancestry or race/ethnicity. Then we categorized the study participants into European, Asian, African, and other population groups. Due to the great heterogeneity in the way others replied to our inquiry, we made inferences of the most suitable category for each study participant for the analysis.

Use the Medical Subject heading (MeSH) term to estimate the proportion of studies that have included the population factor in the MeSH

We used the Medical Subject Heading (MeSH) term to estimate the availability of population information on PubMed. We used “(“immunogenetics”[Mesh] NOT ”Models, Animal”[Mesh]) AND “population groups”[Mesh]” to evaluate the proportion of immunogenetics studies that included population factors in the studies based on MeSH term search. Similarly, we used “(“Receptors, Antigen, T-Cell”[Mesh] NOT ”Models, Animal”[Mesh]) AND “Population Groups”[Mesh],” to evaluate the proportion of the TCR studies that include population factors in the studies based on MeSH terms.

Table

Supplemental Table 1. The summary table of the overall 42 TCR-Seq studies that were included in the study

Study Title	Number of study participants	Category	Type of information	Population labels in the study
Freeman, J. D., Warren, R. L., Webb, J. R., Nelson, B. H. & Holt, R. A. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. <i>Genome Res.</i> 19, 1817–1824 (2009).	335	European	Self-report ancestry	Caucasian
Wang, C. et al. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 107, 1518–1523 (2010).	1	Asian	Self-report ancestry	East Asian
Warren, R. L. et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. <i>Genome Res.</i> 21, 790–797 (2011).	3	European	Self-report ancestry	Caucasian
Putintseva, E. V. et al. Mother and child T cell receptor repertoires: deep profiling study. <i>Front. Immunol.</i> 4, 463 (2013).	9	European	Self-report ancestry	Caucasian
Britanova, O. V. et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. <i>J. Immunol. Baltim. Md</i> 192, 2689–2698 (2014).	39	European	Self-report ancestry	Caucasian
Zvyagin, I. V. et al. Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 111, 5980–5985 (2014).	6	European	Self-report ancestry	Russian
Robert, L. et al. CTLA4 blockade broadens the peripheral T-cell receptor repertoire. <i>Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.</i> 20, 2424–2432 (2014).	21	European, Other	Ethnicity, race	Caucasian, Hispanic
Gaide, O. et al. Common clonal origin of central and resident memory T cells following skin immunization. <i>Nat. Med.</i> 21, 647–653 (2015).	11	European	Self-report ancestry	Caucasian
Wu, J., Liu, D., Tu, W., Song, W. & Zhao, X. T-cell receptor diversity is selectively	26	Asian	Self-report ancestry	Chinese

skewed in T-cell populations of patients with Wiskott-Aldrich syndrome. J. Allergy Clin. Immunol. 135, 209–216 (2015).				
Harden, J. L., Hamm, D., Gulati, N., Lowes, M. A. & Krueger, J. G. Deep Sequencing of the T-cell Receptor Repertoire Demonstrates Polyclonal T-cell Infiltrates in Psoriasis. F1000Research 4, 460 (2015).	19	European, Asian, African	Ethnicity	Black, Hispanic, White, Asian
Britanova, O. V. et al. Dynamics of Individual T Cell Repertoires: From Cord Blood to Centenarians. J. Immunol. Baltim. Md 1950 196, 5005–5013 (2016).	65	European	Self-report ancestry	Caucasian
Chen, Z. et al. T cell receptor β -chain repertoire analysis reveals intratumour heterogeneity of tumour-infiltrating lymphocytes in oesophageal squamous cell carcinoma. J. Pathol. 239, 450–458 (2016).	7	Asian	Self-report ancestry	Han
Rossetti, M. et al. TCR repertoire sequencing identifies synovial Treg cell clonotypes in the bloodstream during active inflammation in human arthritis. Ann. Rheum. Dis. 76, 435–441 (2017).	30	European	Ethnicity	Caucasian
Seay, H. R. et al. Tissue distribution and clonal diversity of the T and B cell repertoire in type 1 diabetes. JCI Insight 1, e88242 (2016).	33	European, African, Other	Ethnicity, race	Caucasian, African American, Hispanic
Li, B. et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. Nat. Genet. 48, 725–732 (2016).	3	European	Ethnicity	White
Zvyagin, I. V. et al. Tracking T-cell immune reconstitution after TCR $\alpha\beta$ /CD19-depleted hematopoietic cells transplantation in children. Leukemia 31, 1145–1153 (2017).	24	European	Self-report ancestry	Caucasian
Ramesh, M., Hamm, D., Simchoni, N. & Cunningham-Rundles, C. Clonal and constricted T cell repertoire in Common Variable Immune Deficiency. Clin. Immunol. Orlando Fla 178, 1–9 (2017).	66	European	Self-report ancestry	Caucasian
Savola, P. et al. Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. Nat.	82	European	Self-report ancestry	Finnish

Commun. 8, 15869 (2017).				
Wang, T. et al. The Different T-cell Receptor Repertoires in Breast Cancer Tumors, Draining Lymph Nodes, and Adjacent Tissues. Cancer Immunol. Res. 5, 148–156 (2017).	16	Asian	Self-report ancestry	Han
Abdel-Hakeem, M. S., Boisvert, M., Bruneau, J., Soudeyns, H. & Shoukry, N. H. Selective expansion of high functional avidity memory CD8 T cell clonotypes during hepatitis C virus reinfection and clearance. PLoS Pathog. 13, e1006191 (2017).	5	European	Self-report ancestry	French Caucasian, Swiss Caucasian
Pogorelyy, M. V. et al. Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. PLoS Comput. Biol. 13, e1005572 (2017).	10	European	Self-report ancestry	Caucasian
Kargl, J. et al. Neutrophils dominate the immune cell composition in non-small cell lung cancer. Nat. Commun. 8, 14381 (2017).	140	European, Asian	Ethnicity	White, Asian
Herati, R. S. et al. Successive annual influenza vaccination induces a recurrent oligoclonotypic memory response in circulating T follicular helper cells. Sci. Immunol. 2, (2017).	43	European, African	Ethnicity	White, African American
Emerson, R. O. et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat. Genet. 49, 659–665 (2017).	560	European, Asian, African	Ethnicity, race	Non-Hispanic/Latino o White, Hispanic/Latino White, African American, Asian non-Hispanic/Latino , Asian-Hispanic/Latino, American Indian or Alaska Native, Native Hawaiian or Pacific Islanders, Unknown Hispanic/Latino
Jia, Q. et al. Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. Nat. Commun. 9, 5361 (2018).	15	Asian	Self-report ancestry	Chinese
Komech, E. A. et al. CD8+ T cells with characteristic T cell receptor beta motif	25	European	Self-report ancestry	Caucasian

are detected in blood and expanded in synovial fluid of ankylosing spondylitis patients. <i>Rheumatol. Oxf. Engl.</i> 57, 1097–1104 (2018).				
Pogorelyy, M. V. et al. Precise tracking of vaccine-responding T cell clones reveals convergent and personalized response in identical twins. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 115, 12704–12709 (2018).	6	European	Self-report ancestry	Russian
Patas, K. et al. T Cell Phenotype and T Cell Receptor Repertoire in Patients with Major Depressive Disorder. <i>Front. Immunol.</i> 9, 291 (2018).	10	European	Self-report ancestry	European
Miron, M. et al. Human Lymph Nodes Maintain TCF-1hi Memory T Cells with High Functional Potential and Clonal Diversity throughout Life. <i>J. Immunol. Baltim. Md</i> 1950 201, 2132–2140 (2018).	5	European, Other	Ethnicity, race	Hispanic, White
Carnero Contentti, E., Farez, M. F. & Correale, J. Mucosal-Associated Invariant T Cell Features and TCR Repertoire Characteristics During the Course of Multiple Sclerosis. <i>Front. Immunol.</i> 10, 2690 (2019).	110	European	Ethnicity	Caucasian
Lee, M. et al. Preferential Infiltration of Unique V γ 9J γ 2-V δ 2 T Cells Into Glioblastoma Multiforme. <i>Front. Immunol.</i> 10, 555 (2019).	4	Asian	Ethnicity	Korean
Simnica, D. et al. High-Throughput Immunogenetics Reveals a Lack of Physiological T Cell Clusters in Patients With Autoimmune Cytopenias. <i>Front. Immunol.</i> 10, 1897 (2019).	47	European	Self-report ancestry	Caucasian
Ramien, C. et al. T Cell Repertoire Dynamics during Pregnancy in Multiple Sclerosis. <i>Cell Rep.</i> 29, 810-815.e4 (2019).	23	European	Self-report ancestry	Caucasian/ White
Minervina, A. A. et al. Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. <i>eLife</i> 10, (2021).	2	European	Ethnicity	Russian (Central European part of Russia)
Lin, Y.-H. et al. Dissecting efficiency of a 5' rapid amplification of cDNA ends (5'-RACE) approach for profiling T-cell receptor beta repertoire. <i>PLoS One</i> 15,	2	Asian	Self-report ancestry	Asian

e0236366 (2020).				
Hanson, A. L. et al. Altered Repertoire Diversity and Disease-Associated Clonal Expansions Revealed by T Cell Receptor Immunosequencing in Ankylosing Spondylitis Patients. <i>Arthritis Rheumatol.</i> Hoboken NJ 72, 1289–1302 (2020).	85	European	Ethnicity	Australians of white European
Schultheiß, C. et al. Next-Generation Sequencing of T and B Cell Receptor Repertoires from COVID-19 Patients Showed Signatures Associated with Severity of Disease. <i>Immunity</i> 53, 442-455.e4 (2020).	36	European	Self-report ancestry	Caucasian
Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. <i>Nat. Med.</i> 26, 842–844 (2020).	16	Asian	Self-report ancestry	Han
Nolan, S. 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. <i>Res. Sq.</i> (2020) doi:10.21203/rs.3.rs-51964/v1.	1146	European, Asian, African, Other	Self-report ancestry, race	Asian, Caucasian, Hispanic, African American, Native American
Shomuradova, A. S. et al. SARS-CoV-2 Epitopes Are Recognized by a Public and Diverse Repertoire of Human T Cell Receptors. <i>Immunity</i> 53, 1245-1257.e5 (2020).	34	European	Ethnicity	Jewish, Ukrainian, Russian, German, Armenian, Greek, Tajik, Tatars, Georgian, Osetian, Mordvin, Chuvash
Nishida, K. et al. Clinical importance of the expression of CD4+CD8+ T cells in renal cell carcinoma. <i>Int. Immunol.</i> 32, 347–357 (2020).	104	Asian	Self-report ancestry	Japanese
Kusnadi, A. et al. Severely ill COVID-19 patients display impaired exhaustion features in SARS-CoV-2-reactive CD8+ T cells. <i>Sci. Immunol.</i> 6, (2021).	39	European, Asian, African	Ethnicity	Arab, Black British, Chinese, Indian, White British, White Other