Defining background shared antibody sequences between unrelated healthy individuals (public clonotypes) to support future studies on specific infectious disease related conditions

Arthur Chang, MD[1] and Mark Hicar MD, PhD[2]

[1]Division of Infectious Diseases, Department of Pediatrics, University of Nebraska Medical Center [2]Division of Infectious Diseases, Department of Pediatrics, Jacobs School of Medicine.

Introduction

Public clonotypes, antibodies against specific antigens in unrelated individuals that have genetic similarities, have been shown in a variety of infections, including SARS-CoV-2 and HIV. Likely, there are shared antibody responses between individuals for many infections. To explore antibody responses that would coincide with specific infectious diseases that may set off chronic illnesses, such as Multiple Sclerosis or Alzheimer's disease, defining the background shared clonotypes is needed to differentiate disease from normal background public clonotype responses.

Methods

Methods: Heavy chain variable sequences were retrieved from public biorepositories (Bioproject PRJNA486667)[1] composed of 43 healthy, 114 HIV+ with broadly neutralizing antibodies (BNAB), and 91 HIV+ with non-broadly neutralizing antibodies (NNAB) subjects. We utilized the Immcantation[3,4] package of software run on our SUNY Buffalo computational cluster. Constant region sequences were annotated with using PRESTo. Duplicate sequences were collapsed to a single row with sequence count retained and sequences with 1 sequence count were removed. Subject ID and group ID were annotated into the sequences before being aligned and further annotated using IgBLAST with ChangeO. Clonal groups were determined using ChangeO requiring IGHV, IGHJ, and CDR3 amino acid sequence to be perfectly matched. Figures and statistics were generated with immcantation, excel, and graphpad prism 8.

Results

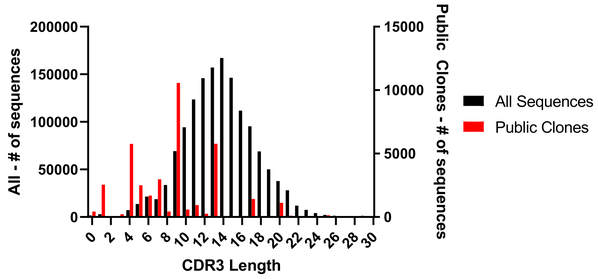
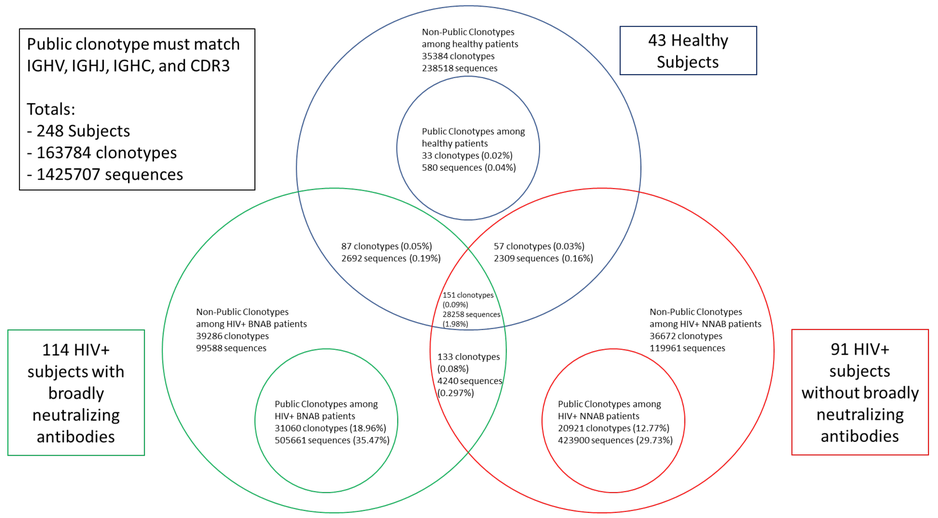


Figure 1: CDR3 length of all sequences and Public clones. Public clones CDR3 length smaller than that of all sequences. 240 sequences are now shown for CDR3 lengths beyond 30, none of which were public clones.

Results

Figure 2: Overlaps of clones and sequences. Clonotypes are defined by perfect matches of IGHV, IGHJ, IGHC assignment by IgBLAST. They are also perfectly matched by CDR3 amino acid sequence.

Public clonotypes are clonotypes found between two or more subjects.



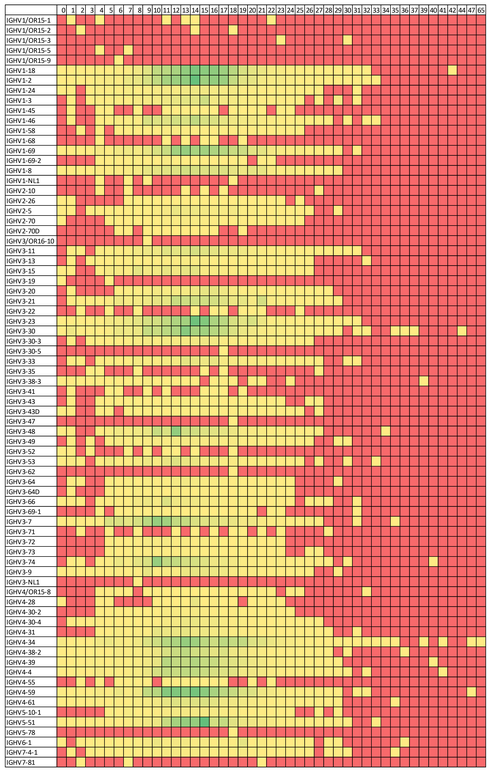


Figure 3: IGHV and CDR3 length heat map for all sequences in dataset

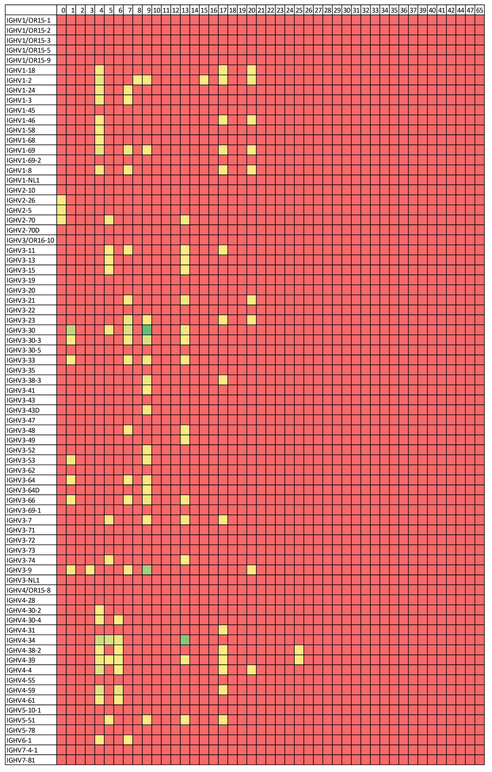


Figure 4: IGHV and CDR3 length heat map for public clonotypes found in healthy, HIV with broadly neutralizing antibodies and HIV subjects without broadly neutralizing antibodies.

Clone ID	IGHV	IGHJ	IGHC	CDR3 AA sequence
12478	IGHV3-30	IGHJ6	IGHG	FPLAPCSRSTS
12479	IGHV3-30	IGHJ6	IGHG	FPLAPSSKSTS
64250	IGHV4-34	IGHJ6	IGHG	GPSVFPLAPSSKSTS
64251	IGHV4-34	IGHJ6	IGHG	GPSVFPLAPCSRSTS
111783	IGHV3-9	IGHJ6	IGHG	FPLAPCSRSTS
111784	IGHV3-9	IGHJ6	IGHG	FPLAPSSKSTS

Tale 1: Common public clones in all 3 groups. Targets of these clones are not clear although have been identified in multiple studies such as in Hepatitis B vaccination[5], Ebola[6], and healthy controls[7]. May represent common infections, vaccination, or regulatory/autoimmune functions[2].

Conclusion

This early work has identified several public clonotypes that are shared among subjects who are HIV positive and otherwise healthy people. Defining the sequences commonly seen between individuals can assist in specifying antibody responses specific to disease states from larger sequence databases.

- May represent vaccination or common infections
- May represent common regulatory and / or autoimmune antibodies.
- Growing dataset can be used as a negative control to screen out common clonotypes compared to interested population.
- Identifying these common public clones can further help screen out these sequences from future analysis

References

1.Roskin KM, et al. Aberrant B cell repertoire selection associated with HIV neutralizing antibody breadth. Nat Immunol. 2020 Feb;21(2):199-209. doi: 10.1038/s41590-019-0581-0.
2.Chang AJ, et al. Clonal expansion and markers of directed mutation of IGHV4-34 B cells in plasmablasts during Kawasaki disease. Mol Immunol. 2022 May;145:67-77. doi: 10.1016/j.molimm.2022.03.011.
3.Vander Heiden JA, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 30, 1930-2 (2014). doi:10.1093/bioinformatics/btu138
4.Gupta NT, et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics 31, 3356-8 (2015). doi:10.1093/bioinformatics/btv359
5.Galson JD, et al. B-cell repertoire dynamics after sequential hepatitis B vaccination and evidence for cross-reactive B-cell activation. Genome Med 8, 68 (2016). https://doi.org/10.1186/s13073-016-0322-z
6.Davis CW, et al. Longitudinal Analysis of the Human B Cell Response to Ebola Virus Infection. Cell. 2019 May 30;177(6):1566-1582.e17. doi: 10.1016/j.cell.2019.04.036.
7.Thörnqvist L, Ohlin M. Data on the nucleotide composition of the first codons encoding the complementary determining region 3 (CDR3) in immunoglobulin heavy chains. Data Brief. 2018 May 4;19:337-352. doi: 10.1016/j.dib.2018.04.125.

University at Buffalo The State University of New York