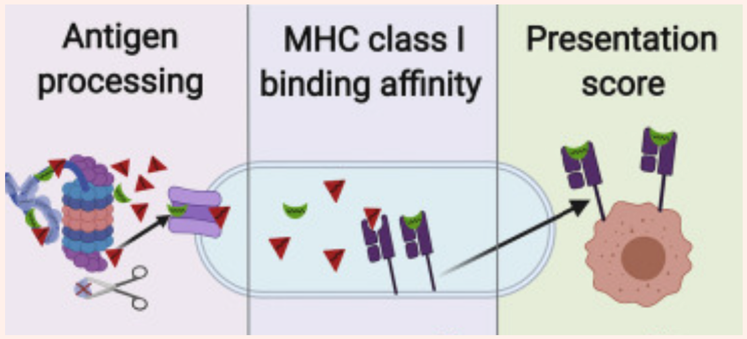
How do mutations in viruses affect their interactions with HLA?



FARKHUNDA AYOBI
DATA SCIENCE APPRENTICESHIP
UNIVERSITY COLLEGE ROOSEVELT

PURPOSE

- When an antigen enters the body, it's broken down into peptides.
- Peptides are recognized by HLA class I and presented to CD8 T cells, which mount an immune response.
- The stronger the binding affinity between HLA and peptide, the quicker and stronger the immune response.
- ***Purpose 1.*** Visualize binding/interactions after specific mutations.
- ***Purpose 2.*** Visualize the correlation between societal restrictions and COVID-19 cases across different countries.

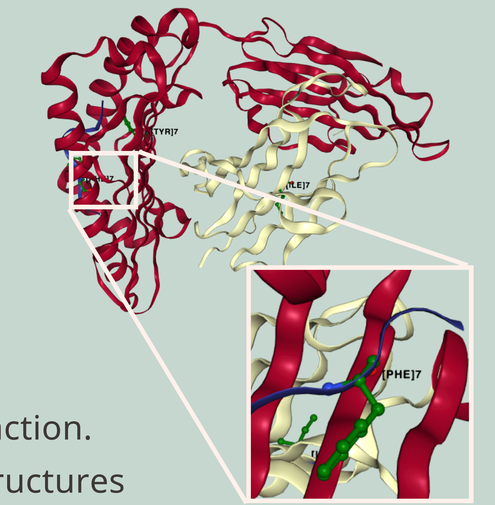
RELEVANCE

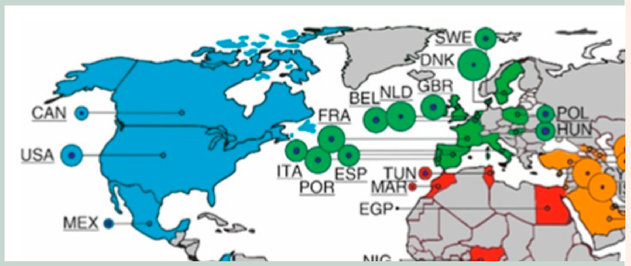
- It helps in understanding immunological responses
  - What is the most dangerous variant?
  - It can be used for developing vaccines and personalized medicine.
- Knowing correlations will influence countries' pandemic response policies, leading to better preparedness for the future.

ANALYSIS

- Data: **Protein Data Bank (PBD)**
  - Each protein has a specific code ie.7RTD
- Package: **NGLViwer**
  - R can read the protein code and create the 3D structure of proteins
  - This package can highlight and label amino acids
- Package: **Bio3d**
  - Load pdb files
- Mutated certain amino acids to visualize change in the interaction. However, no changes were observed because creating 3D structures requires other xyz coordinates.
- ***Plan B:***
  - Data: Our World in Data
  - Variables: COVID-19 cases, stringency index

```
NGLVieweR("7RTD") |>
  addRepresentation("cartoon")
```





FUTURE PLANS

- Obtain the data and tidy it.
- Work on making a bivariate choropleth map.
- Investigate the relationship between COVID-19 cases and SI.

CHALLENGES

- Tried many different packages and sources for 3D structures.
- Will need to do an entirely new project by Feb 20th.

SOURCES

```r
NGLVieweR("6xcn") %>%
  addRepresentation("cartoon")


row_501 <- df[df$resno == 501, ] |> print()



spike_structure$atom$resid[spike_structure$atom$resno == 501 & spike_structure$atom$resid == "ASN"] <- "TYR"

write.pdb(spike_structure, "mutated_spike_protein.pdb")

NGLVieweR("mutated_spike_protein.pdb") |>
  addRepresentation("cartoon") |>
  addRepresentation("ball+stick", param = list(
    colorScheme = "element",
    colorValue = "green",
    sele = "501")) |>
  stageParameters(backgroundColor = "#C6D7CF") |>
  addRepresentation("label",
                    param = list(
                      sele = "501",
                      labelType = "format",
                      labelFormat = "[%(resname)s]%(resno)s", # or enter custom text
                      labelGrouping = "residue", # or "atom" (eg. sele = "20:A.CB")
                      color = "black",
                      fontFamiliy = "sans-serif",
                      xOffset = 1,
                      yOffset = 0,
                      zOffset = 0,
                      fixedSize = TRUE,
                      radiusType = 1,
                      radiusSize = 1.5, # Label size
                      showBackground = FALSE
                      # backgroundColor="black",
                      # backgroundOpacity=0.5
                    )
  )
```

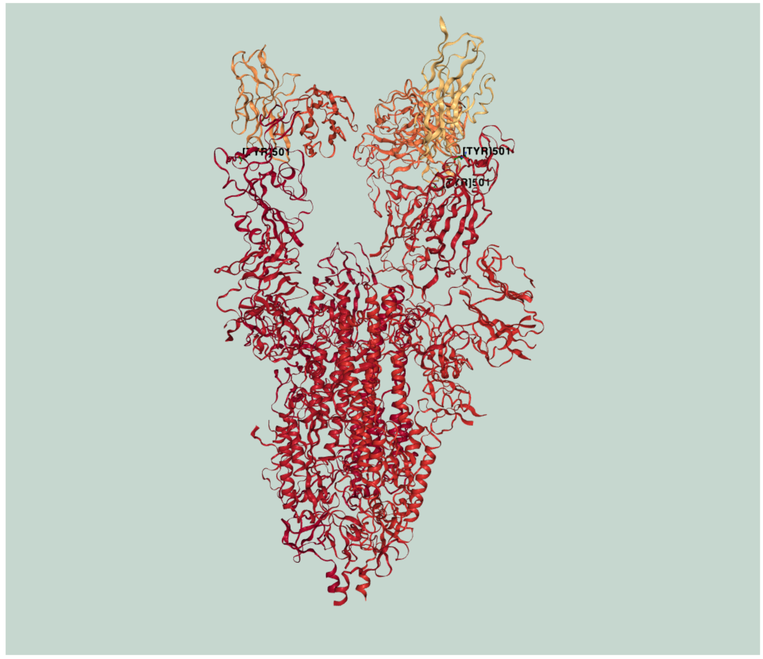
311:1 clean version R Script

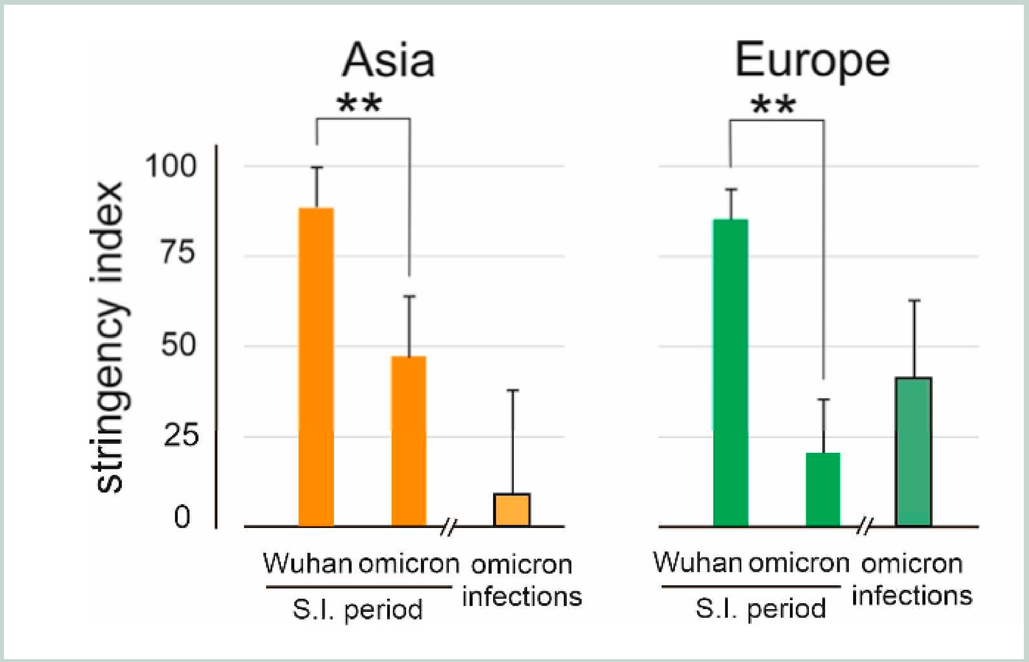
Console

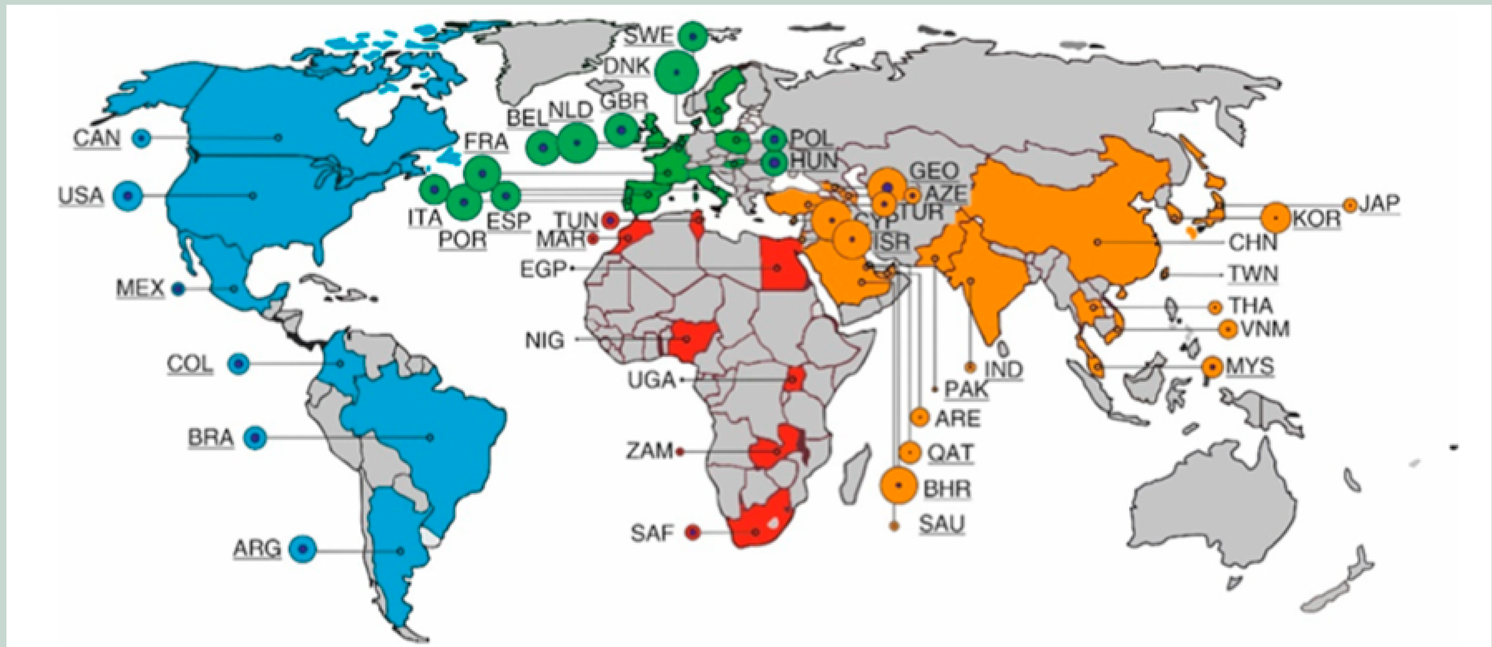
spike_structure	List of 8
view	Large NGLVieweR (8 elements, 2.3 MB)
Values	
pdb_id	"6xcn"
residue_to_mutate	10

Files Packages Help Viewer Presentation

Zoom Export Publish







	HLA_variant	AA_code
HLA:HLA00001 A*01:01:01:01 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA02169 A*01:01:01:02N 200 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA14798 A*01:01:01:03 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA15760 A*01:01:01:04 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA16415 A*01:01:01:05 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA16417 A*01:01:01:06 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA16436 A*01:01:01:07 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA16651 A*01:01:01:08 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA16652 A*01:01:01:09 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA16653 A*01:01:01:10 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA17618 A*01:01:01:11 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA17985 A*01:01:01:12 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA18246 A*01:01:01:13 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA18976 A*01:01:01:14 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA18991 A*01:01:01:15 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA19155 A*01:01:01:16 365 bp	HLA	MAVMARTSGAATTWAGSHSMRYTSVSRGRGRAVGYVDD…
HLA:HLA19049 A*01:01:01:17 365 bp	HLA	MA

sequences | protein[1] | protein[1][["atom"]] | pdb[1] | protein | spike_structure | df | f

Filter

	type	eleno	elety	alt	resid	chain	resno	insert	x	y	z	o	b	segid	elesy
1	ATOM	1	N	NA	ILE	A	1	NA	7.273	-8.286	20.609	1	19.02	NA	N
2	ATOM	2	CA	NA	ILE	A	1	NA	6.579	-9.551	20.253	1	12.47	NA	C
3	ATOM	2	C	NA	ILE	A	1	NA	5.464	-9.319	19.253	1	17.47	NA	C
4	ATOM	4	O	NA	ILE	A	1	NA	5.167	-8.185	18.887	1	21.51	NA	O
5	ATOM	5	CB	NA	ILE	A	1	NA	5.934	-10.244	21.468	1	13.11	NA	C
6	ATOM	6	CG1	NA	ILE	A	1	NA	5.341	-9.181	22.403	1	11.54	NA	C
7	ATOM	7	CG2	NA	ILE	A	1	NA	6.898	-11.249	22.078	1	6.43	NA	C
8	ATOM	8	CD1	NA	ILE	A	1	NA	6.350	-8.423	23.251	1	6.44	NA	C
9	ATOM	9	N	NA	GLU	A	2	NA	4.863	-10.425	18.828	1	14.06	NA	N
10	ATOM	10	CA	NA	GLU	A	2	NA	3.772	-10.466	17.870	1	17.32	NA	C
11	ATOM	11	C	NA	GLU	A	2	NA	3.112	-9.142	17.583	1	12.44	NA	C
12	ATOM	12	O	NA	GLU	A	2	NA	2.230	-8.736	18.331	1	12.00	NA	O
13	ATOM	13	CB	NA	GLU	A	2	NA	2.719	-11.448	18.369	1	27.32	NA	C
14	ATOM	14	CG	NA	GLU	A	2	NA	3.185	-12.868	18.326	1	35.85	NA	C
15	ATOM	15	CD	NA	GLU	A	2	NA	3.086	-13.431	16.933	1	39.43	NA	C
16	ATOM	16	OE1	NA	GLU	A	2	NA	1.949	-13.686	16.505	1	41.58	NA	O
17	ATOM	17	OE2	NA	GLU	A	2	NA	4.125	-13.605	16.265	1	42.33	NA	O