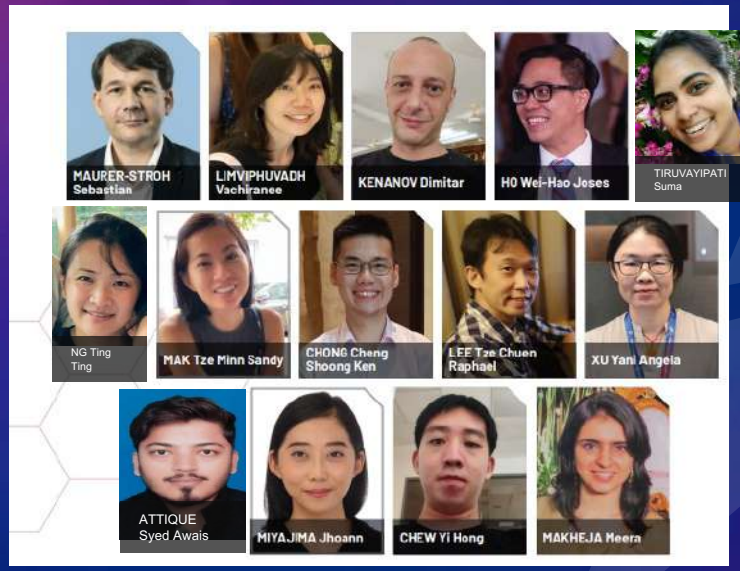
Agency for Science, Technology and Research
SINGAPORE

Bioinformatics Institute (BII)

Protein Sequence Analysis

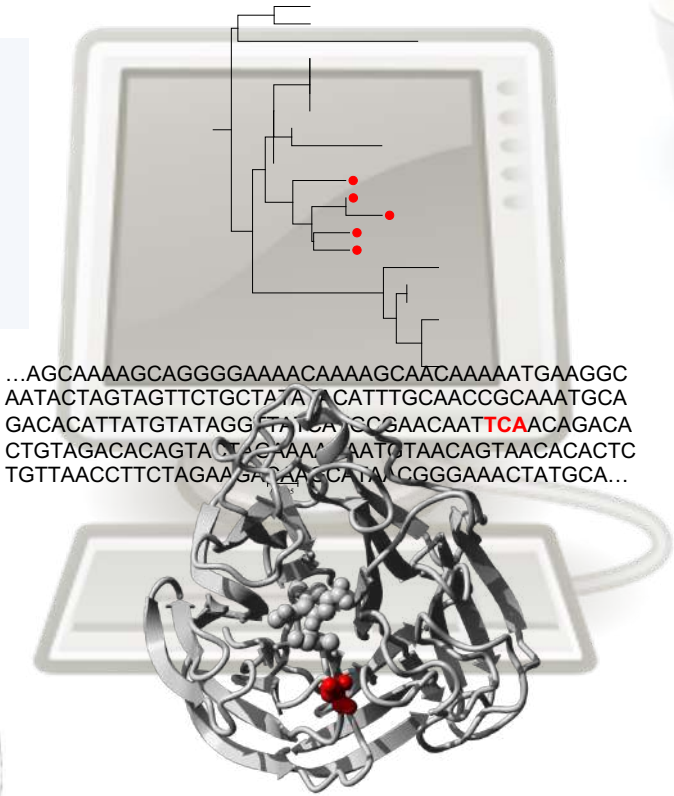


CREATING GROWTH, ENHANCING LIVES

Viruses

Nature Methods. 2023 Apr;20(4):512-522.
Emerg Infect Dis. 2023 Apr;29(4):778-781.
Nature Commun. 2022 Nov 16;13(1):7003.
JAMA Netw Open. 2022 Aug 1;5(8):e2228900.
Clin Infect Dis. 2022 Aug 24;75(1):e1128-e1136.
EMBO Mol Med. 2022 Mar 7;14(3):e15227.
... +10 other publications

Computational Sequence and Structure Analysis



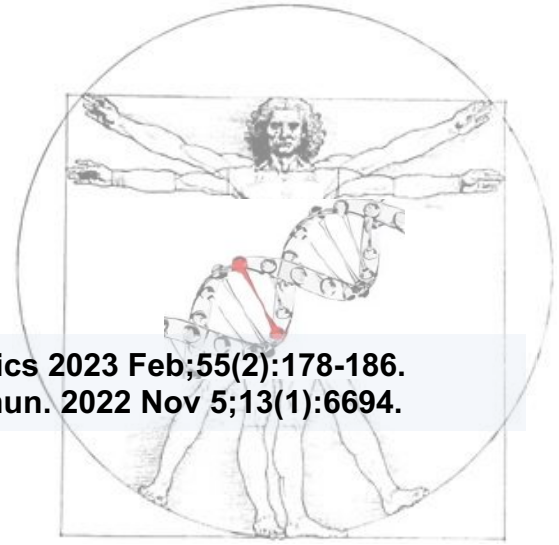
We are working at the interface between sequence and structure

Food flavors

Drugs

Patent application: BII/P/13305/00/SG
Patent application: ISCE2/P/13296/00/SG

Human variation



Nature Genetics 2023 Feb;55(2):178-186.
Nature Commun. 2022 Nov 5;13(1):6694.

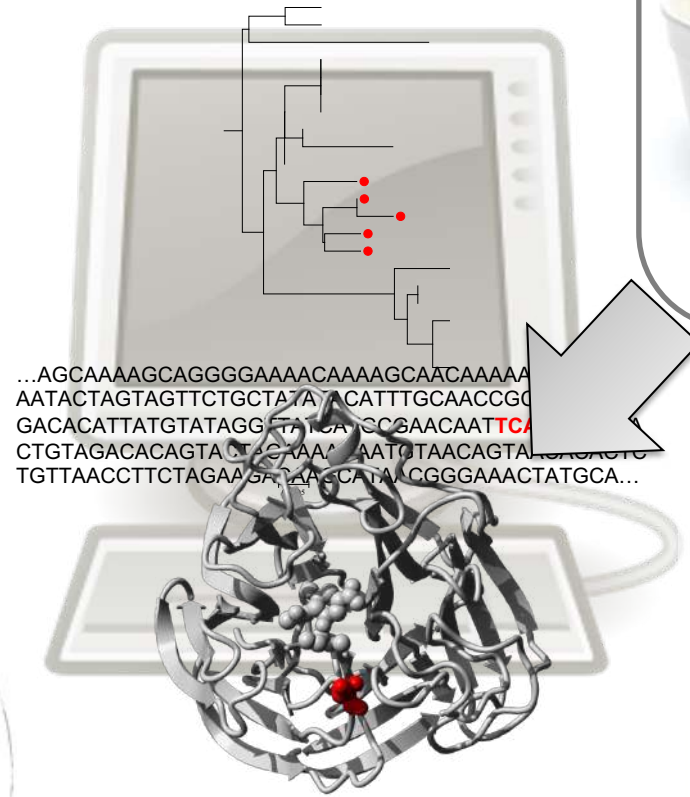
Product Safety

J Proteomics. 2022 Oct 30;269:104724.
Nucleic Acids Res. 2022 Jul 5;50(W1):W36-W43.

CREATING GROWTH, ENHANCING LIVES

Viruses

Computational Sequence and Structure Analysis



Food flavors

Nestlé

FERRERO

Drugs

gsk

Pfizer

MSD

Human variation

Enzymes can be used to naturally produce food flavours and fragrances but also drugs!

Product Safety

AI (CNN) prediction of mutation effects on enzyme function

CREATING GROWTH, ENHANCING LIVES

PHARMA INNOVATION PROGRAMME SINGAPORE (PIPS)

Kinetics

EVB Model

Binding

Docking scores

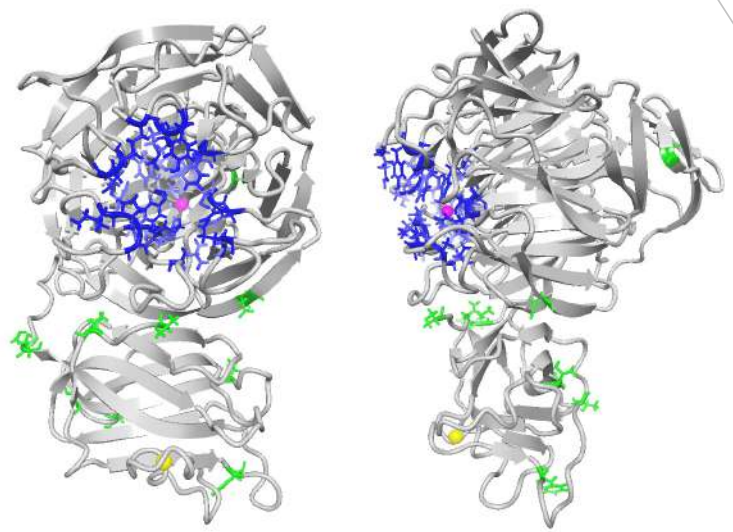
r=0.9

Binding free energy

Stability

Yasara + FoldX

12 of 18 correct predictions of increased stability (incl. top 3)



Aggregation

Tango

Waltz

1 predicted mutation set increases solubility 2-fold

Allostery

Allosigma

7 of 17 correct predictions of long range binding effect

Sequence Conservation

SIFT

Shannon entropy

Jump from 12% good sites to 63% good sites

CREATING GROWTH, ENHANCING LIVES

Viruses

Computational Sequence and Structure Analysis

Food flavors

Drugs

Human variation

Product safety

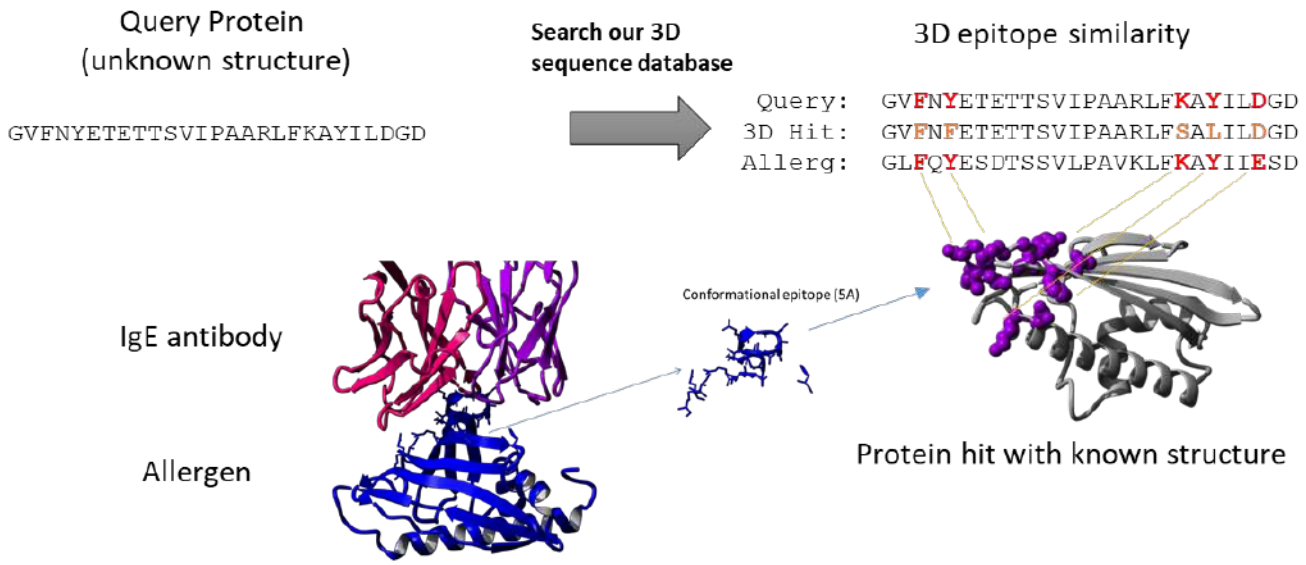
Proteins can cause allergy if similar to other allergens.

Singapore Food Agency

Bioinformatics Institute
BII

Computational prediction of protein allergenicity potential: AllerCatPro

Vachi, Minh



AllerCatPro predicts if a protein is similar to known allergens

CREATING GROWTH, ENHANCING LIVES

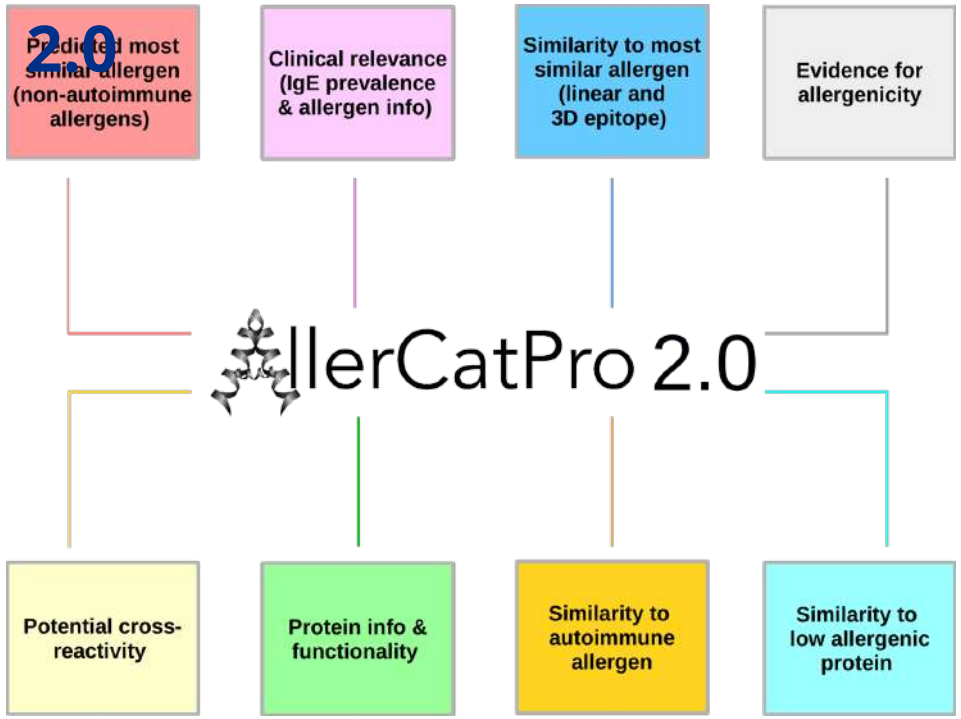
Minh Vachi Andreas

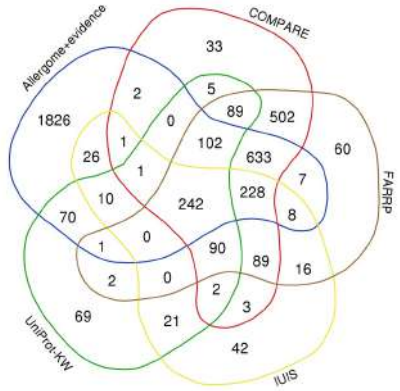
- Using the most comprehensive datasets of reliable proteins associated with allergenicity i.e. **4,979 protein allergens, 162 low allergenic proteins and 165 autoimmune allergens.**

Global Product Stewardship
Safety • Sustainability • Regulatory • Technical Relations

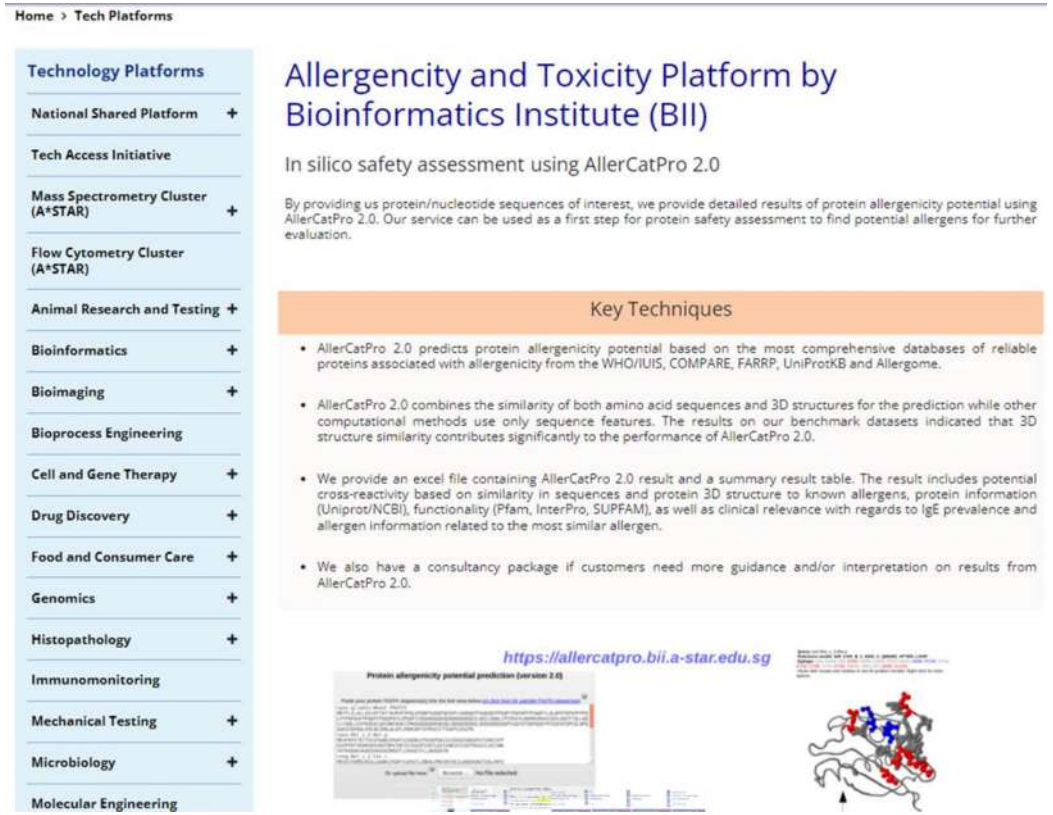
What's new in AllerCatPro

RSC Partnership Agreement on Scientific Services since 28 Feb.2023





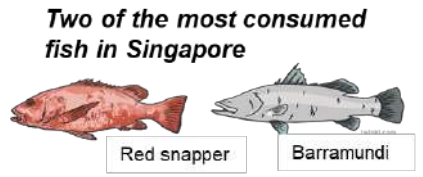
Merged database
(WHO/IUIS, COMPARE, FARRP, UniProtKB, Allergome)

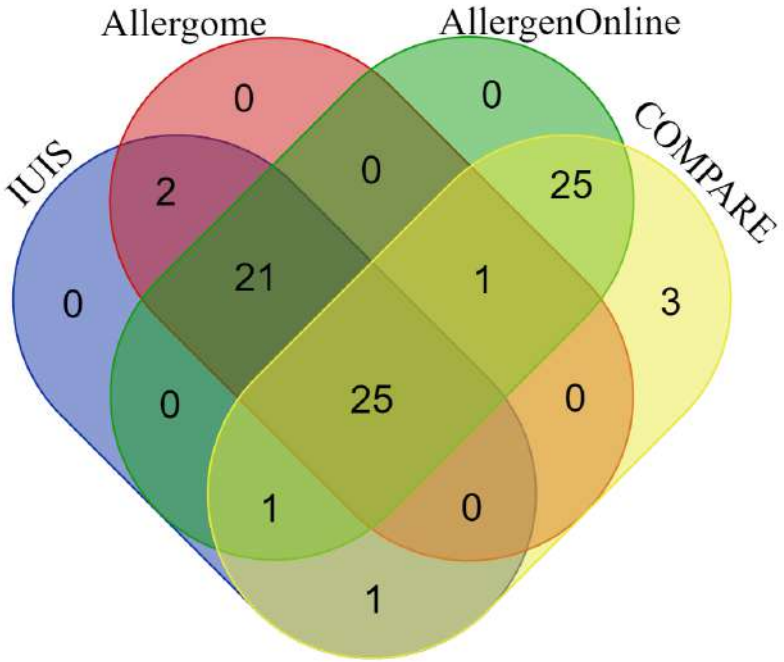


Nguyen MN, Krutz NL, Limviphuvadh V, Lopata AL, Gerberick GF, Maurer-Stroh S. (2022). AllerCatPro 2.0: a web server for predicting protein allergenicity potential. *Nucleic Acids Res*.

Vachi Thimo Andreas

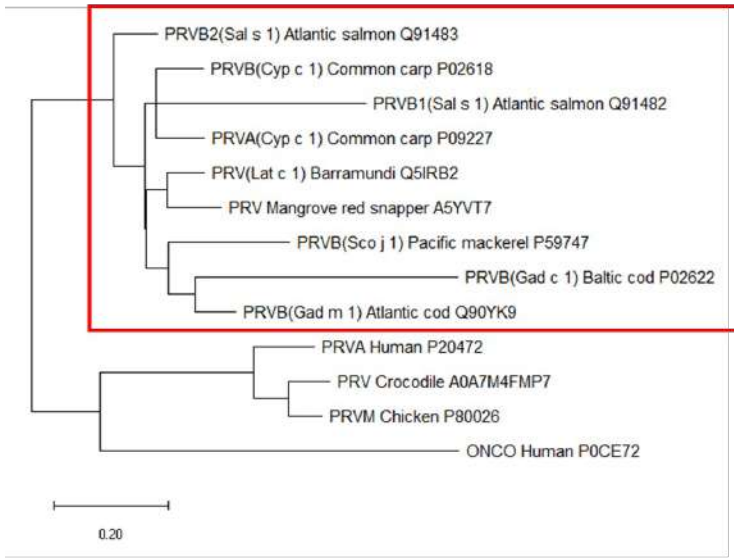
Understanding allergens in alternative food from urban aquaculture fish to underpin food safety





Compiled all known fish allergens from the four most common and well-known allergen databases (n=79)

Parvalbumin



	%identity, linear 80 aa window	%identity, 3D epitope	#pos-IEDB	#pos-IEDB-2M	#neg-IEDB	#neg-IEDB-2M
PRVB2(Sal s 1) Atlantic salmon Q91483	100	100	2	15	33	55
PRVB(Cyp c 1) Common carp P02618	100	100	2	17	5	41
PRVB1(Sal s 1) Atlantic salmon Q91482	100	100	25	37	0	9
PRVA(Cyp c 1) Common carp P09227	100	100	3	23	15	44
PRV(Lat c 1) Barramundi Q5IRB2	100	100	2	26	11	44
PRV Mangrove red snapper A5YVT7	88.8	100	0	17	5	28
PRVB(Sco j 1) Pacific mackerel P59747	100	100	1	22	2	30
PRVB(Gad c 1) Baltic cod P02622	100	100	12	20	6	15
PRVB(Gad m 1) Atlantic cod Q90YK9	100	100	29	40	36	50
PRVA Human P20472	91.2	100	0	1	0	4
PRV Crocodile A0A7M4FMP7	90	94.1	0	0	0	4
PRVM Chicken P80026	100	100	0	0	0	4
ONCO Human P0CE72	55	76.9	0	0	0	0

Created in 2022

- In the future, %relative intensity of IgE binding of parvalbumin of 12 species of fish can be added to phylogenetic tree in order to predict cross-reactivity.
- Family-specific threshold optimization will be implemented to improve AllerCatPro allergenicity assessment and predict cross-reactivity of putative proteins.

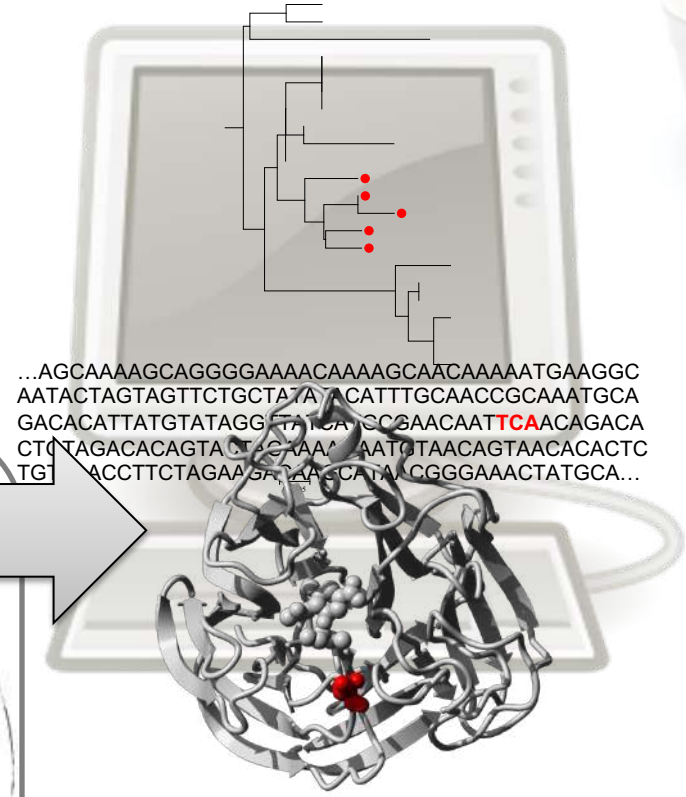
The National Research Foundation, Singapore and A*STAR under the Singapore Food Story R&D Programme [W22W3D0003]

JAMES COOK UNIVERSITY Brisbane | Cairns | Singapore | Townsville
TROPICAL FUTURES INSTITUTE
Singapore Immunology Network SIgN
NANYANG TECHNOLOGICAL UNIVERSITY
UMAMI Meats

CREATING GROWTH, ENHANCING LIVES

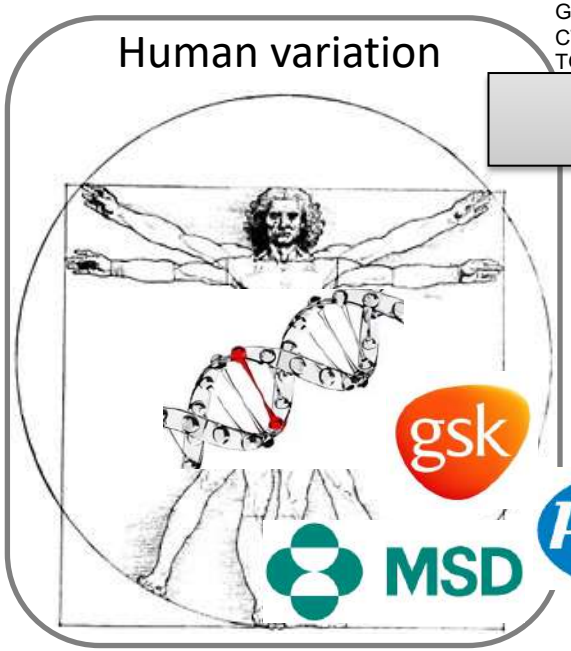
Viruses

Computational Sequence and Structure Analysis



Food flavors

Human variation

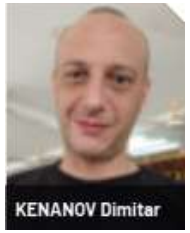


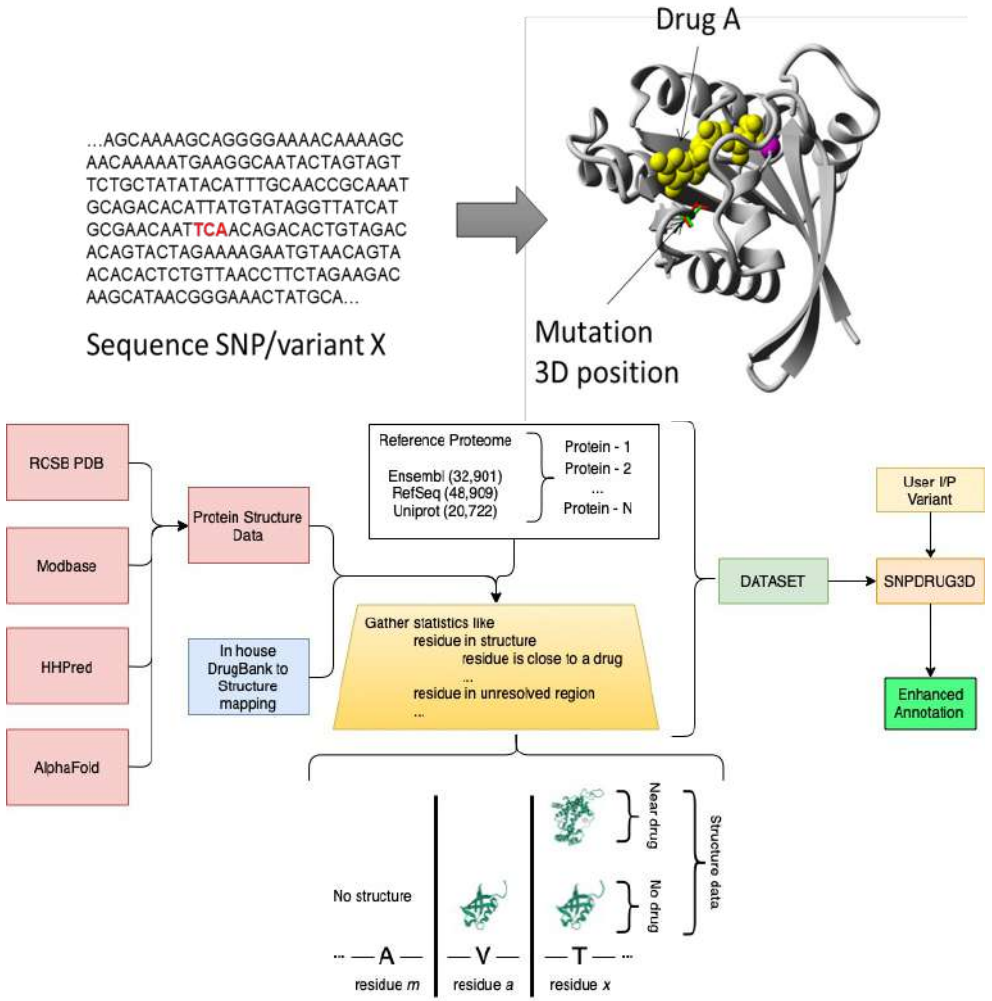
GIS

Individual mutations can cause disease or affect drug efficacy

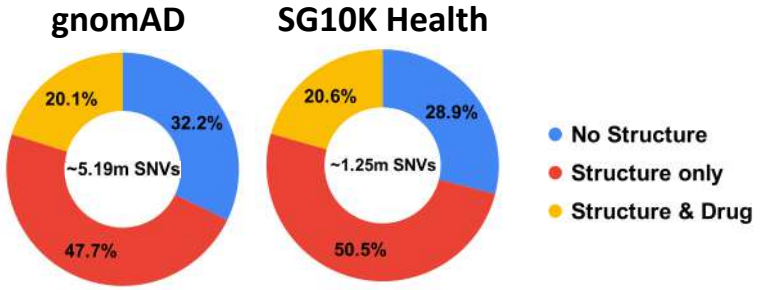
Product Safety

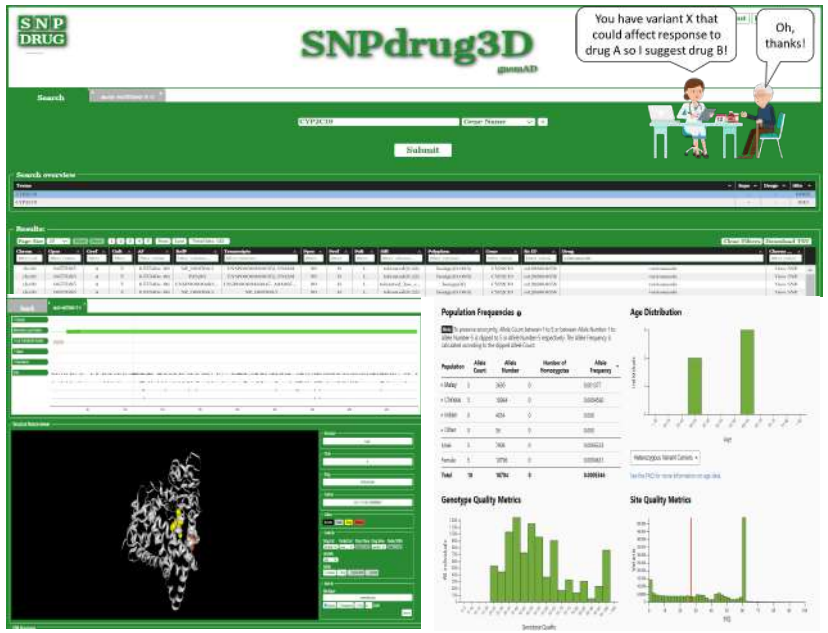
SNPdrug3D – placing genetic variants in a protein structural context





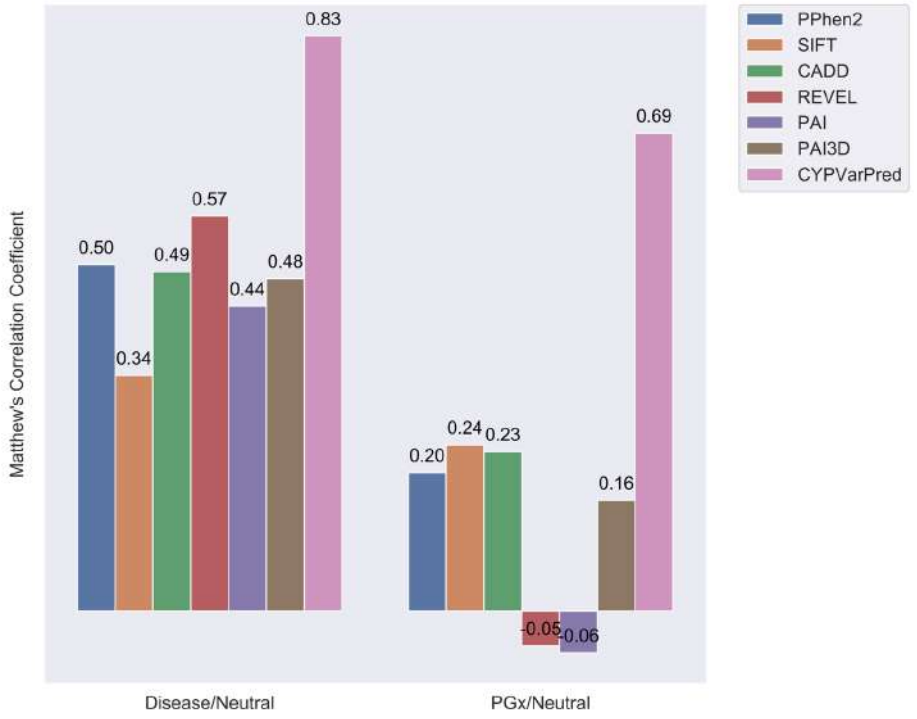
Mapping of genetic variation data to protein structures and sequences



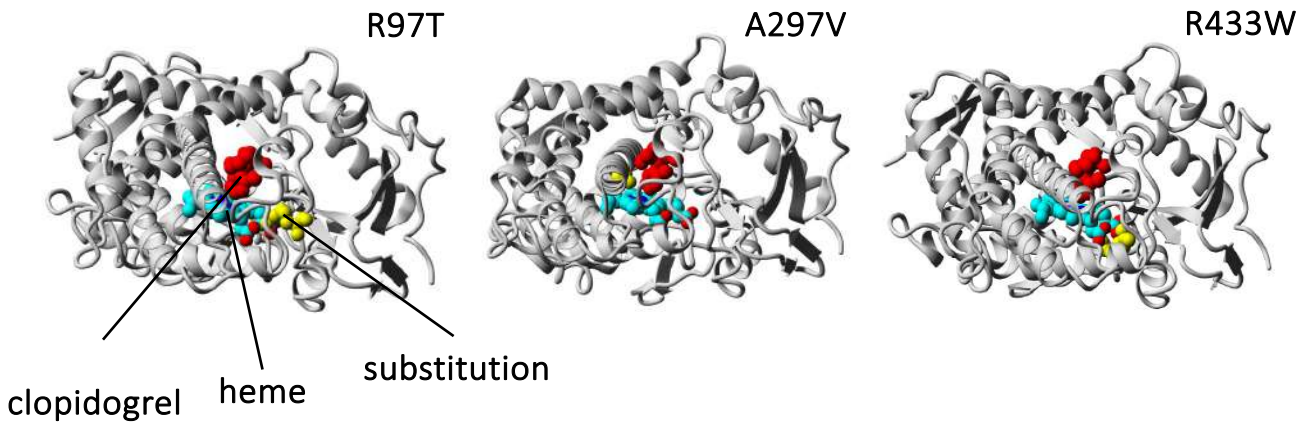


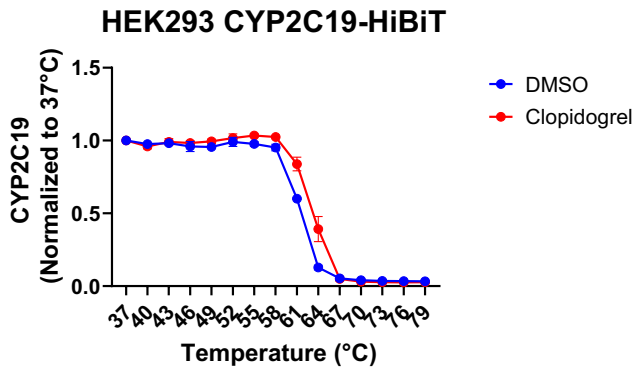
5.8 million SNVs in 20,442 genes from SG10K Health and gnomAD populations mapped to 202,299 protein structures (experimental or AI-generated) containing 5962 drugs

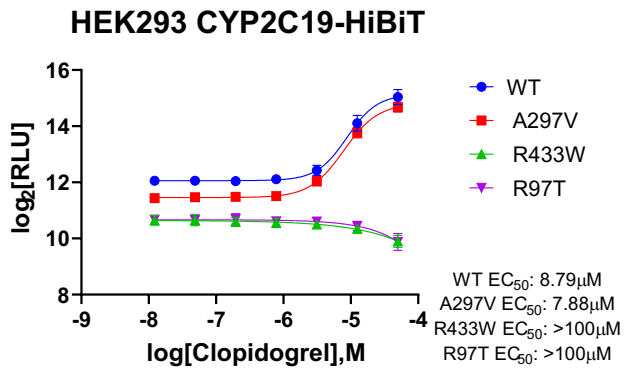
SNPdrug3D – applications



Drug metabolizing enzyme (CYP2C19)







1) Derive features using SNPdrug3D data to build PGx-inclusive variant pathogenicity predictors. Predictor ('CYPVarPred') outperformed all other tested predictors in prediction of PGx CYP variants.

2) Identify variants in protein targets that may affect protein-drug binding. Here, R433W and R97T (novel) but not A297V disrupted CYP2C19-clopidogrel binding according to cellular thermal shift assay results.

SNPdrug3D – clinical use case

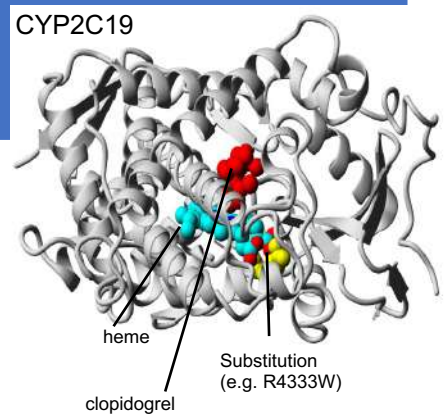
"Several patients I treated with clopidogrel did not respond to it, are there any common mutations in the population that could affect drug response hence allowing me to stratify who should get this drug?"

Search

- Searching for "clopidrogel" shows multiple CYP genes with mutations near the drug
- Filtering for the common drug metabolizing gene CYP2C19 lists possible relevant mutations

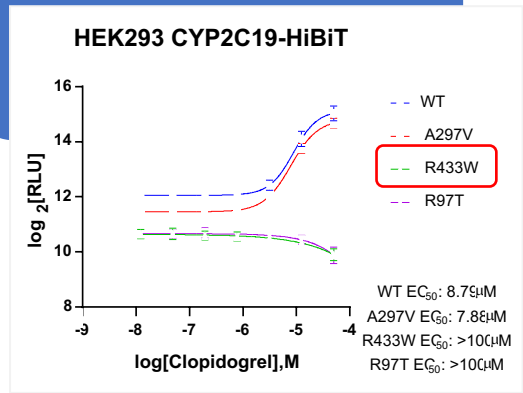
Identify

- Sorting by Allele Frequency shows R433W as top variant
- R433W is predicted to impact CYP2C19-clopidogrel binding. More common in the Malay population (AF = 0.2%)



Validate

- Search if reported in literature
- R433W impaired binding of clopidogrel to CYP2C19 directly.
- Other functional assays can also be performed



Cellular thermal shift assay

Clinical implications

- Stratify patients

Patients with one normal copy and defective copy of the defective CYP2C19 allele (i.e. with R433W mutation) are intermediate metabolizers (IM), producing less active metabolites of the drug.

- Dose adjustments

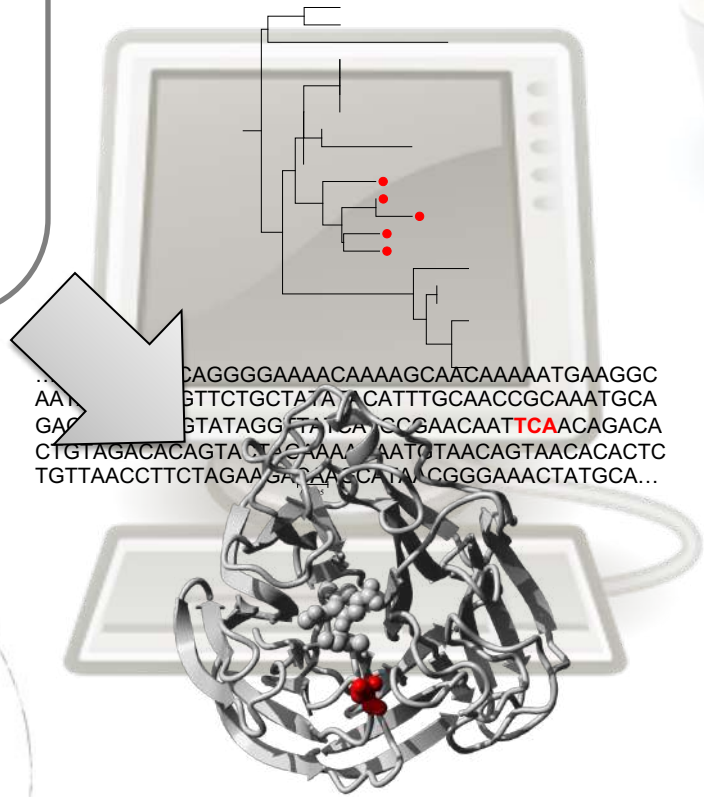
Avoid standard dose of 75mg, use alternative antiplatelet agents (e.g. prasugrel) if possible[1].

[1]CPIC guidelines. Lee et al., 2022. Clin Pharmacol Ther.

GIS

Viruses

Computational Sequence and Structure Analysis



Food flavors

Drugs

Human variation

Product Safety

Long history of flu work but COVID put us into the spotlight

CREATING GROWTH, ENHANCING LIVES

Example COVID-19: data + analysis = impact

One A*STAR: BII works with ID Labs, SIgN, GIS, EDDC, DxDhub, IHPC, I2R

Hospital, GP etc.

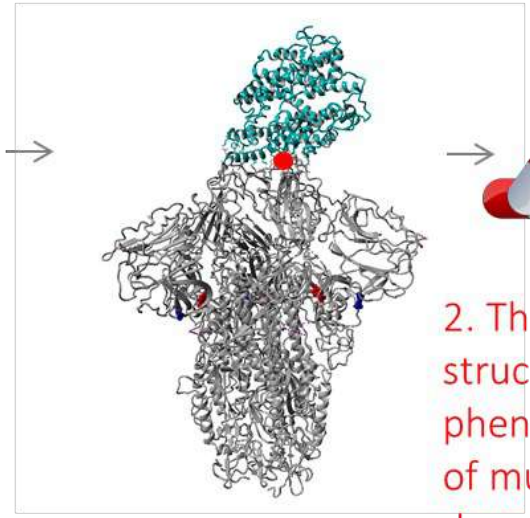
Lab

Bioinformatics

Sample → Sequence → Interpretation

Assembly

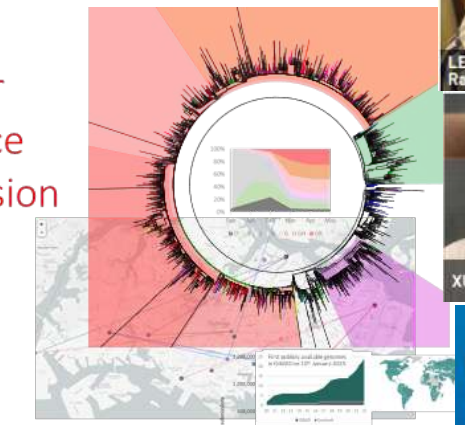
...AGCAAAAGCAGGGGAAAACAAAAGCAACAAAAATGAAGGCAATACTAGTAG
TTCTGCTATATACATTTGCAACCGCAAATGCAGACACATTATGTATAGGTTATCA
TGCGAACAATTCAACAGACACTGTAGACACAGTACTAGAAAAGAATGTAACAG
TAACACACTCTGTTAACCTTCTAGAAGACAAGCATAACGGGAAACTATGCA...

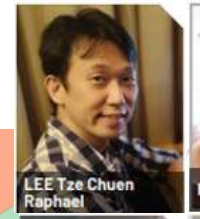


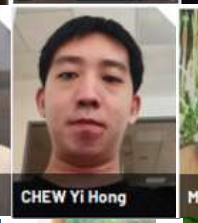
1. Timely sharing of sequences of new viruses is critical and fair sharing mechanisms exist via the GISAID platform

2. Theoretical 3D structure and phenotype effects of mutations and drug/vaccine candidates

3. Comparison with other strains' sequences to trace global and local transmission







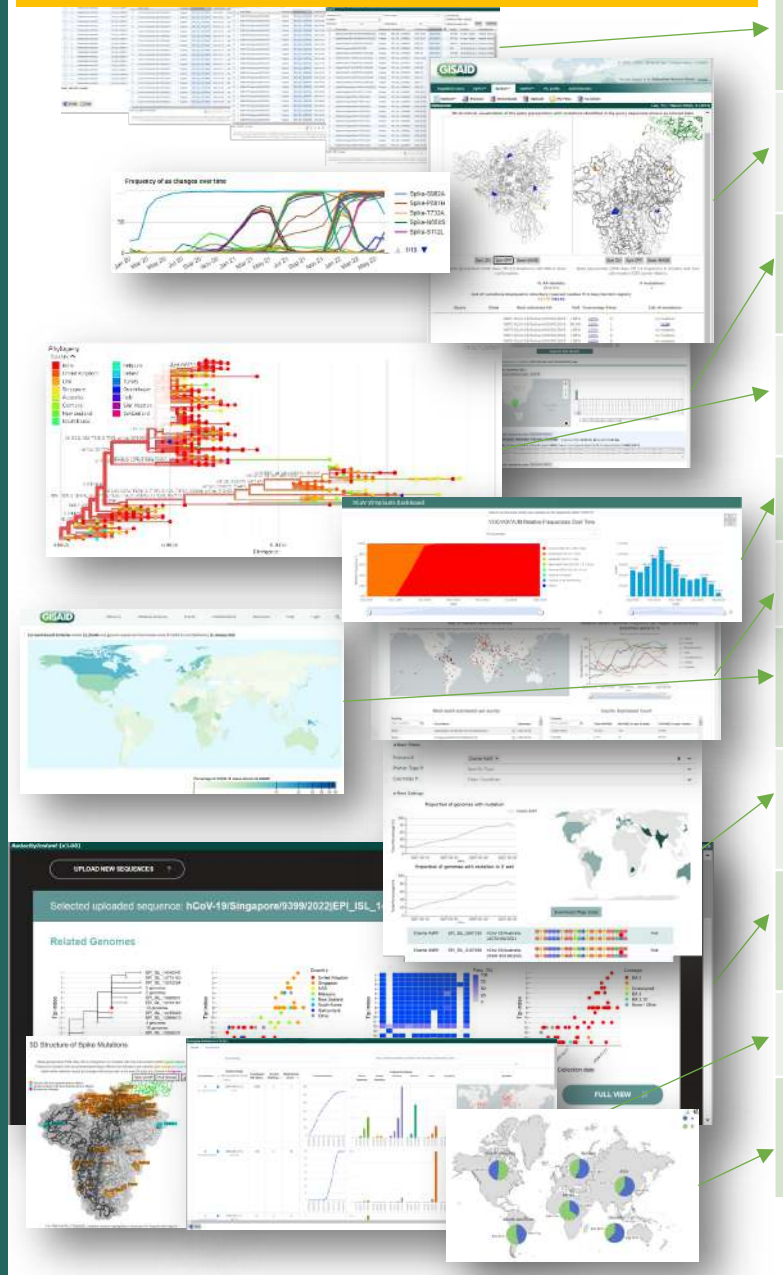
WHO's Chief Scientist (Swaminathan *Nature* 2020) recognized GISAID as "game changer."

Mentioned by PM during National Day Rally, Minister in Parliament etc.

MIT Tech review breakthrough technology 2022

BII/GIS has global (GISAID, WHO, CEPI) and national role (NCID, MOH) in points 1-3

Current status of key tools being expanded to more pathogens



Tool	Purpose	EpiFlu	EpiCoV	EpiRSV	EpiPox
Browse, Search, Download	Data accessibility	Yes	Yes	Yes	Yes
Flu/CoV/RSV/Pox-Surver	Interpret mutation effects	Yes	Yes	Yes	Yes
BLAST	Genomic Epidemiology - Detail	Yes	Yes	Yes	No
Phylodynamics	Genomic Epidemiology - overview	Yes	Yes	Yes	Yes
Variant Frequency	Major variant tracking	Yes	Yes	Yes	Yes
Variant Tracker	Major variant tracking	Yes	Yes	Yes	Yes
Submission Tracker	Surveillance capacity monitoring	Yes	Yes	Yes	Yes
PrimerChecker	Surveillance capacity monitoring	Yes	Yes	Yes	Near Future
AudacityInstant	Genomic Epidemiology - Detail	Yes	Yes	Near Future	Near Future
EmergingVariants	Emerging variant tracking	Yes	Yes	Yes	Yes
EpiCharts (beta)	Graphical overview of user selected data	Yes	Yes	Yes	Yes

September 2022

Yes … new in Nov 22 Yes … new in Feb 23 Near Future … later in 2023

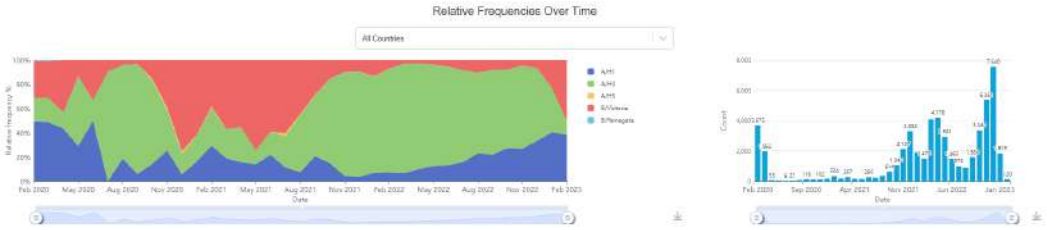
A scale-free view of virus evolution

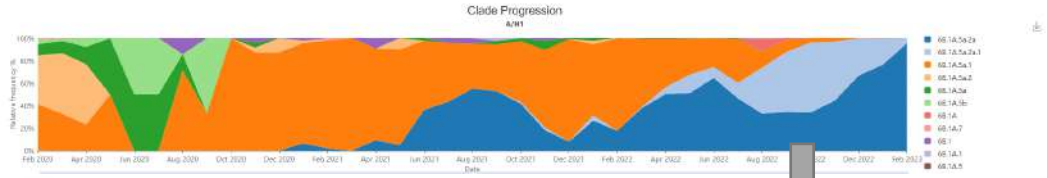
Subtype

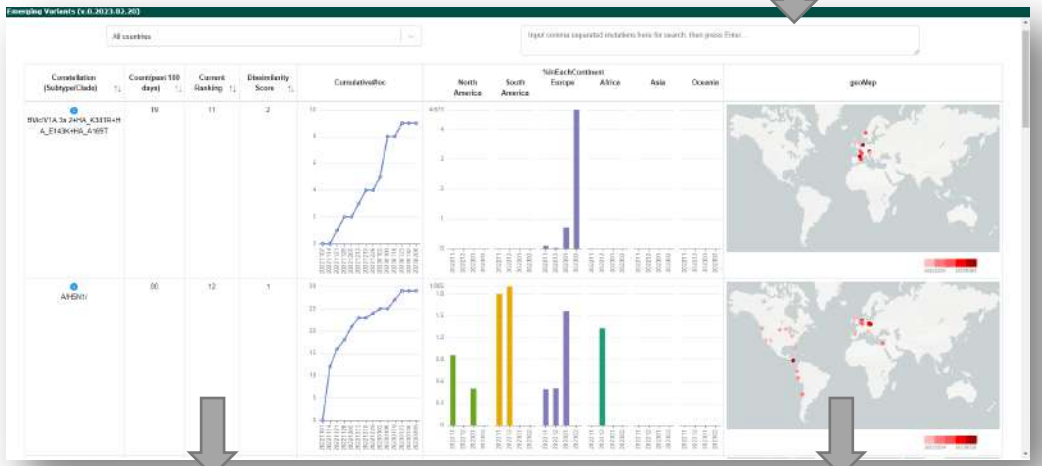
Clade

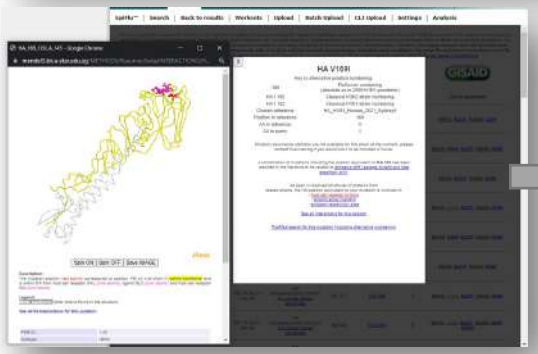
Emerging Variant
(unique set of mutations)

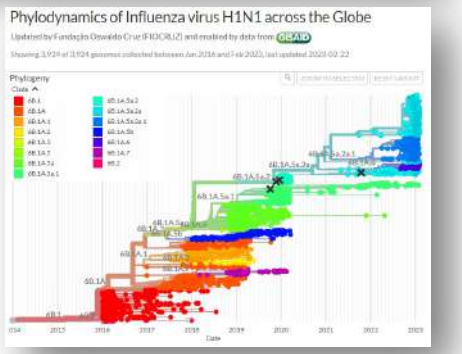
Individual mutation





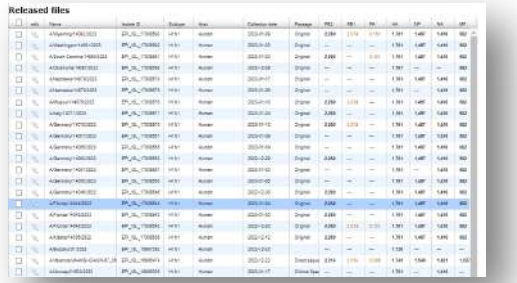




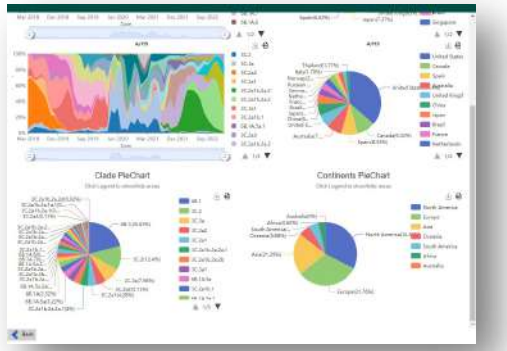


Connected levels in GISAID's integrated tool ecosystem

Single entry detail



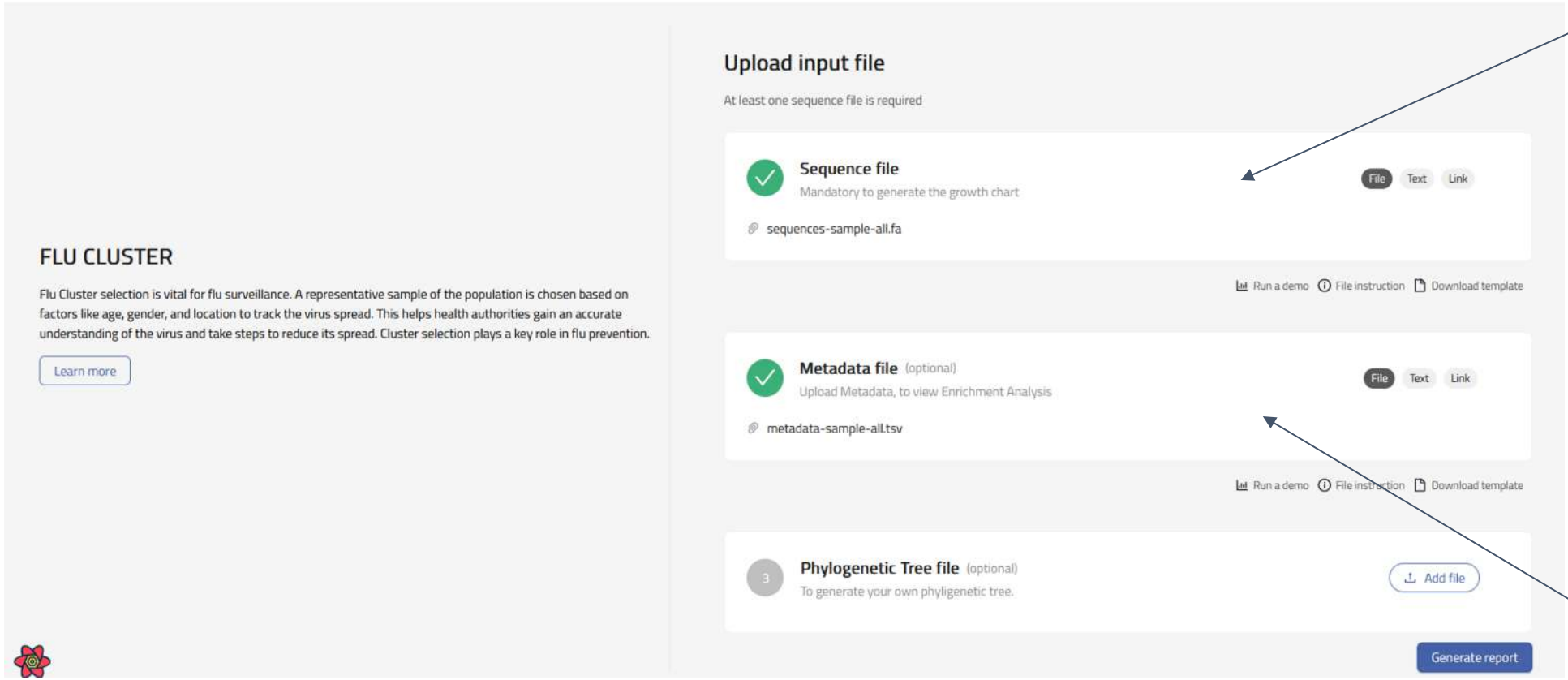
Set of entries in the database

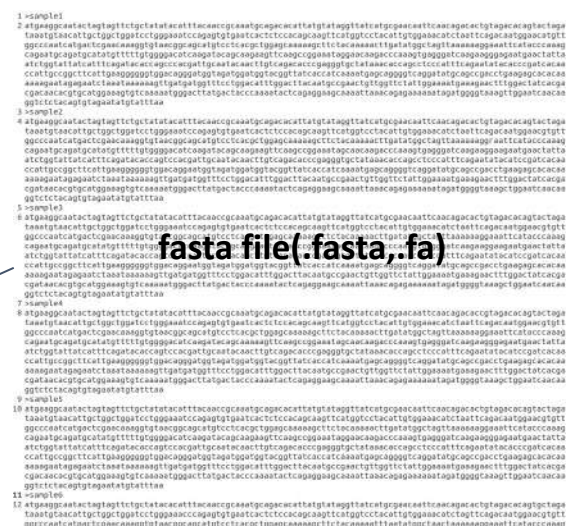


Set of entries graphical summary

FluClusterAI: Landing page

New, in preparation!





fasta file(.fasta,.fa)

	A	C	D	E	F	G
1	sampleID	age	vaccine status	patient status	drug foldchange	HI titer
2	sample1	45		ICU	0.2	256
3	sample2	38		ICU	0.7	128
4	sample3	30			1.3	256
5	sample4	27		ICU	1.5	32
6	sample5	57		ICU		128
7	sample6	97			1.8	
8	sample7	57		ICU	0.9	256
9	sample8	81		ICU	1.7	256
10	sample9	81		ICU	1.5	128
11	sample10	22				64
12	sample11	17		ICU	1.3	32
13	sample12	30		ICU	1.9	
14	sample13	75	vaccine breakthrough		1.9	256
15	sample14				0.9	256
16	sample15	26	vaccine breakthrough			64
17	sample16	87	vaccine breakthrough		1.7	128
18	sample17	83	vaccine breakthrough		0.9	64
19	sample18	1	vaccine breakthrough		1.4	
20	sample19	90			0.1	32
21	sample20	68	vaccine breakthrough			128
22	sample21	46	vaccine breakthrough		0.9	32
23	sample22	17	vaccine breakthrough		1.8	128
24	sample23	46	vaccine breakthrough		1	32
25	sample24	0			0.5	
26	sample25	24	vaccine breakthrough			128
27	sample26	29	vaccine breakthrough		0.9	128
28	sample27	40	vaccine breakthrough		0	32
29	sample28				0.1	64
30	sample29	38			0.2	128
31	sample30	13				
32	sample31	29			0.6	128
33	sample32	48			1.1	32
34	sample33	57			0.6	256
35	sample34	57			0.6	128
36	sample35	46				32
37	sample36	46			0.6	
38	sample37	41			0.6	32
39	sample38	35			1	128
40	sample39	34			1.9	64
41	sample40	65				128
42	sample41	33			0.6	64

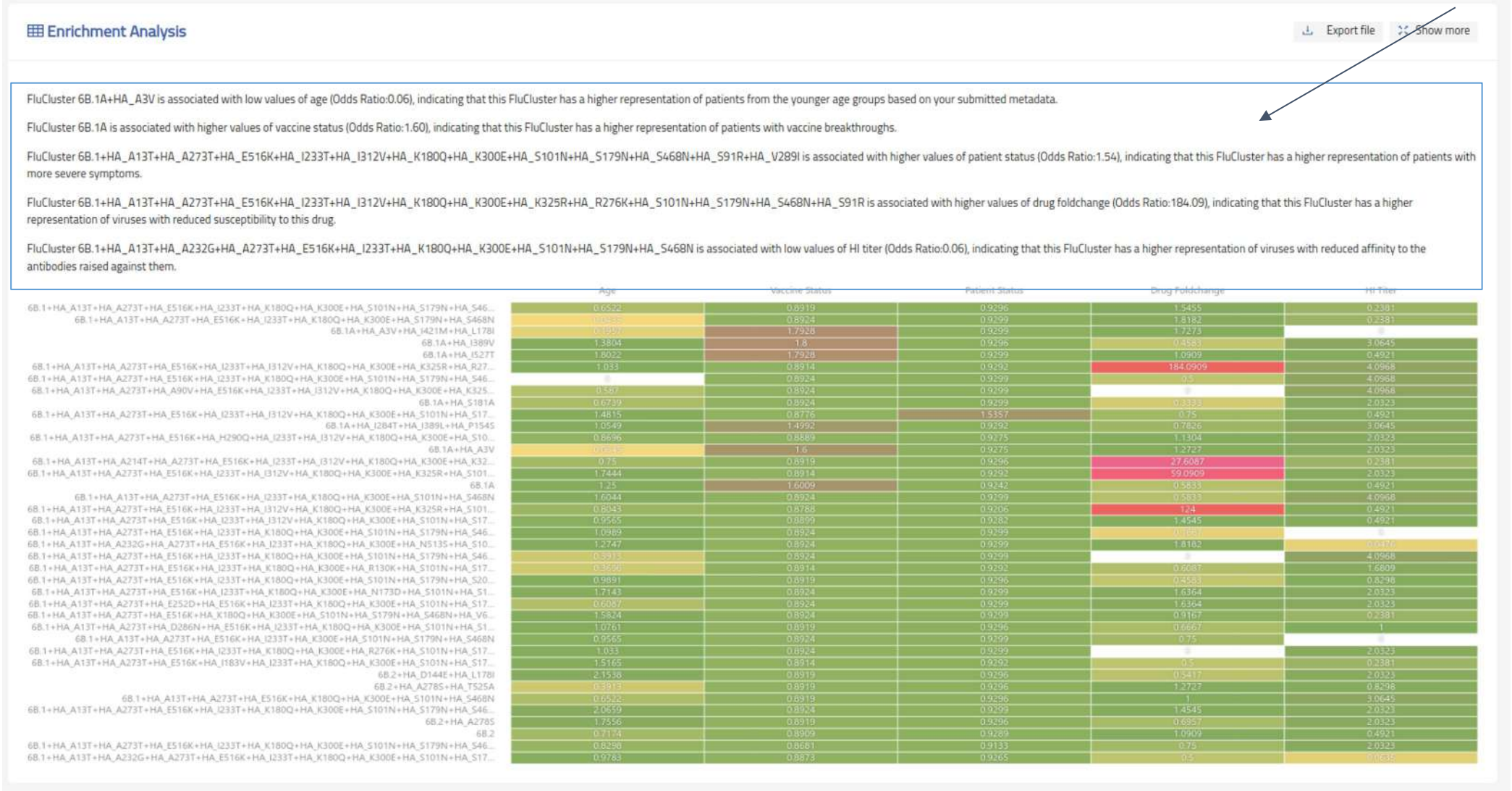
metadata file(tsv/csv/json)

- User upload sequencing fasta files and metadata files(in tsv/csv/json formats) in the landing page.

(Sample dataset: NPHL Singapore dataset H1N1pdm09i 2016-2017)

FluClusterAI: Meta-Enrichment Analysis

Descriptive paragraphs are automatically generated and under representation of metadata.



- Odds Ratio of FluCluster against all matched constellations is used to identify enrichment of specific phenotype with certain FluCluster based on user supplied metadata.

Question to confirm:

- Currently, missing metadata are left uncolored (white). Do let us know if you prefer us to color them as green (insignificant).

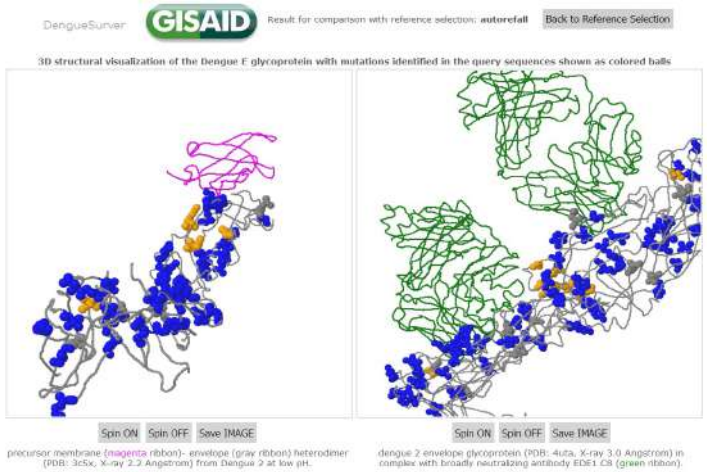
New, in preparation!

DengueSurver & ChikSurver: Tools to judge relevance of mutations that can affect viral fitness

DengueSurver

ChikSurver

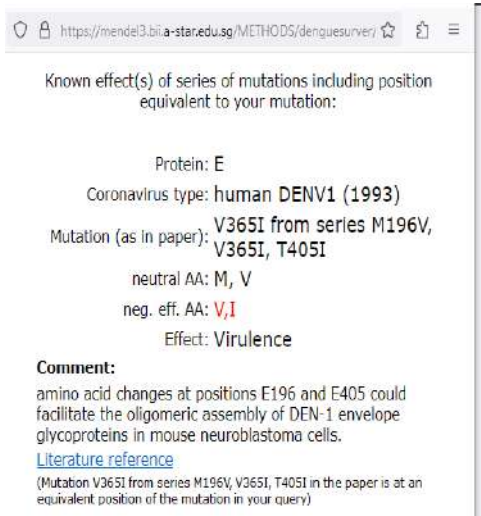
New, in preparation!

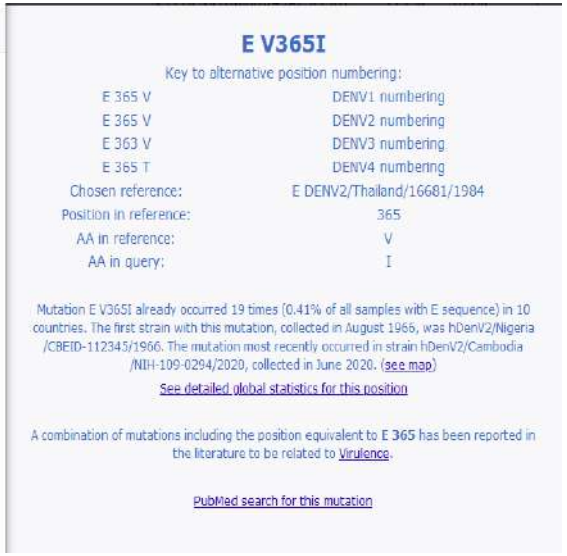


https://mendel3.bii.a-star.edu.sg/METHODS/denguesurver/current/
https://mendel3.bii.a-star.edu.sg/METHODS/chiksurver/current/

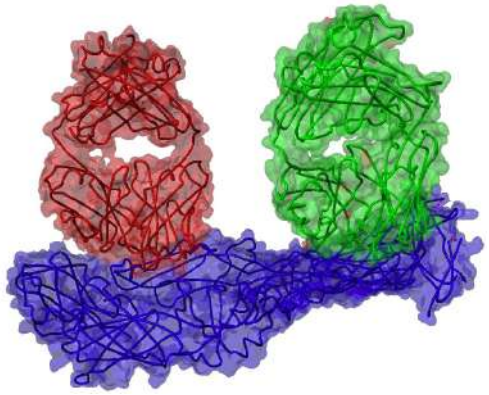
Reference Genomes

	DengueSurver	ChikSurver
# ref genomes	4	1
# genotype ref genomes	45	3





3D Structure Interactions



	DengueSurver	ChikSurver
# PDB structures	197	?

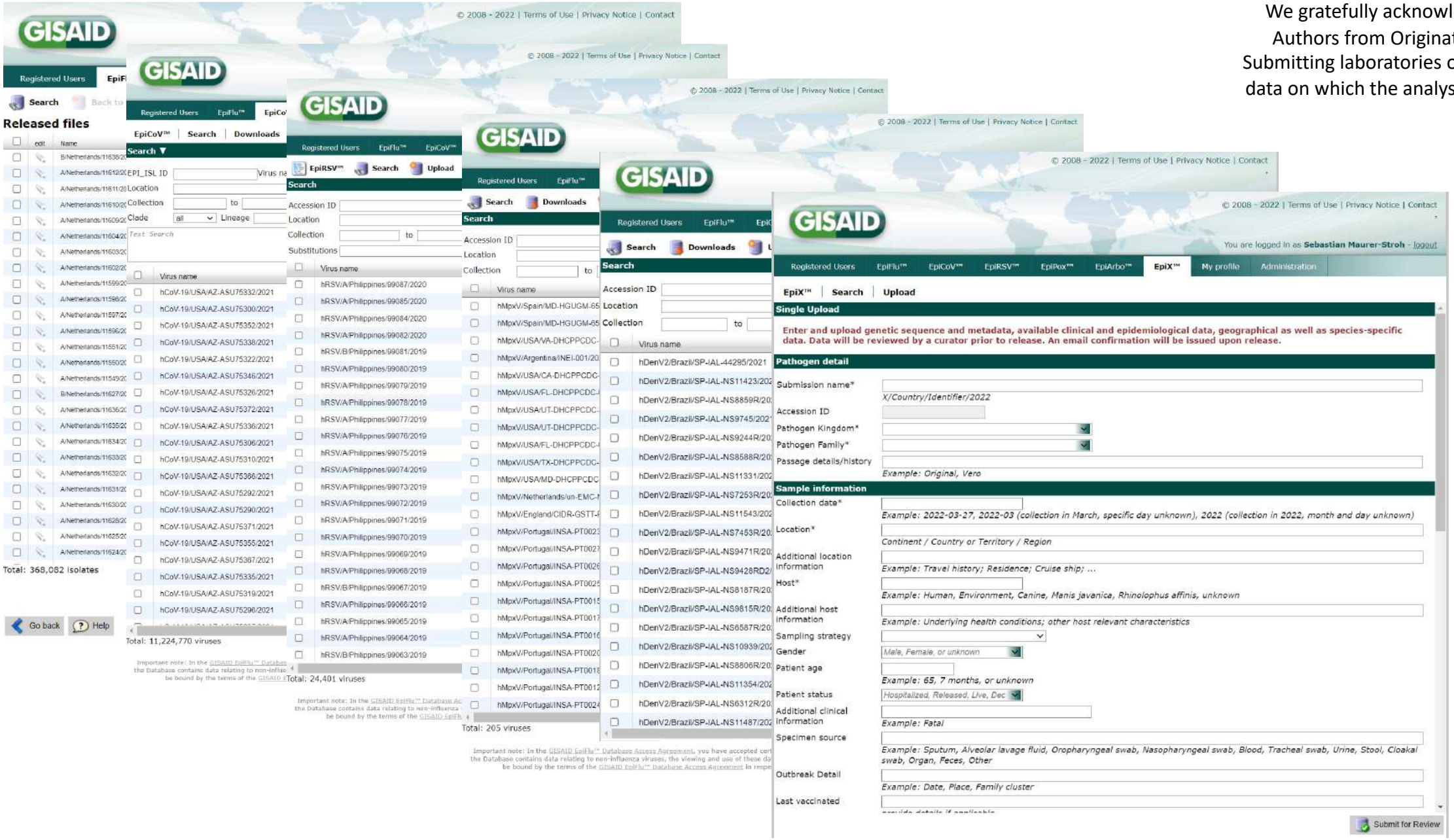
Literature

	DengueSurver	ChikSurver
Drug resistance	2	0
Virulence	14	9
Antigenic drift / escape mutant	7	0
Host specificity change/shift	0	79
Other (enzyme activity, affects protein accumulation/stability/function)	38	18
total literature entries	**61**	**96**

GISAID

EpiFlu™ | EpiCoV™ | EpiRSV™ | EpiPox™ | EpiArbo™ | **EpiX™**

Extension to other priority pathogens



We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.

THANK YOU!

Agency for Science, Technology and Research
SINGAPORE

Bioinformatics Institute (BII)

Protein Sequence Analysis