

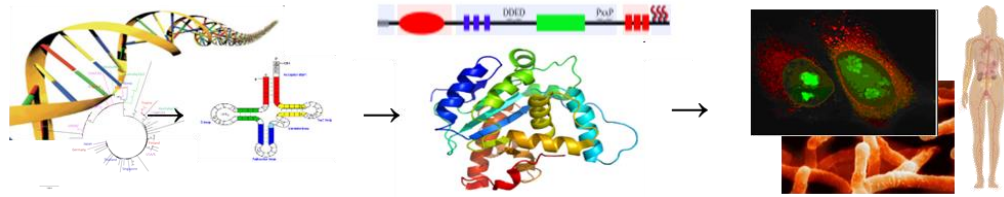
Bioinformatics Institute (BII)

- **Biomedical data hub**
- **Biomedical data analytics**

Sebastian Maurer-Stroh
ED BII
sebastianms@bii.a-star.edu.sg



The new BII 2.0 in RIE2025

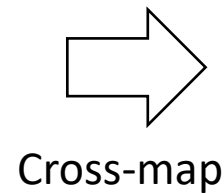


BII is at the interface between disciplines and often connects horizontally

Division by Data Type

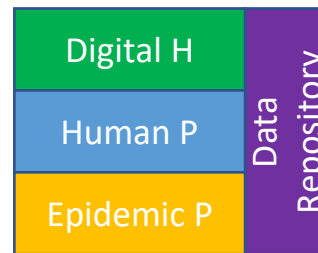


Old



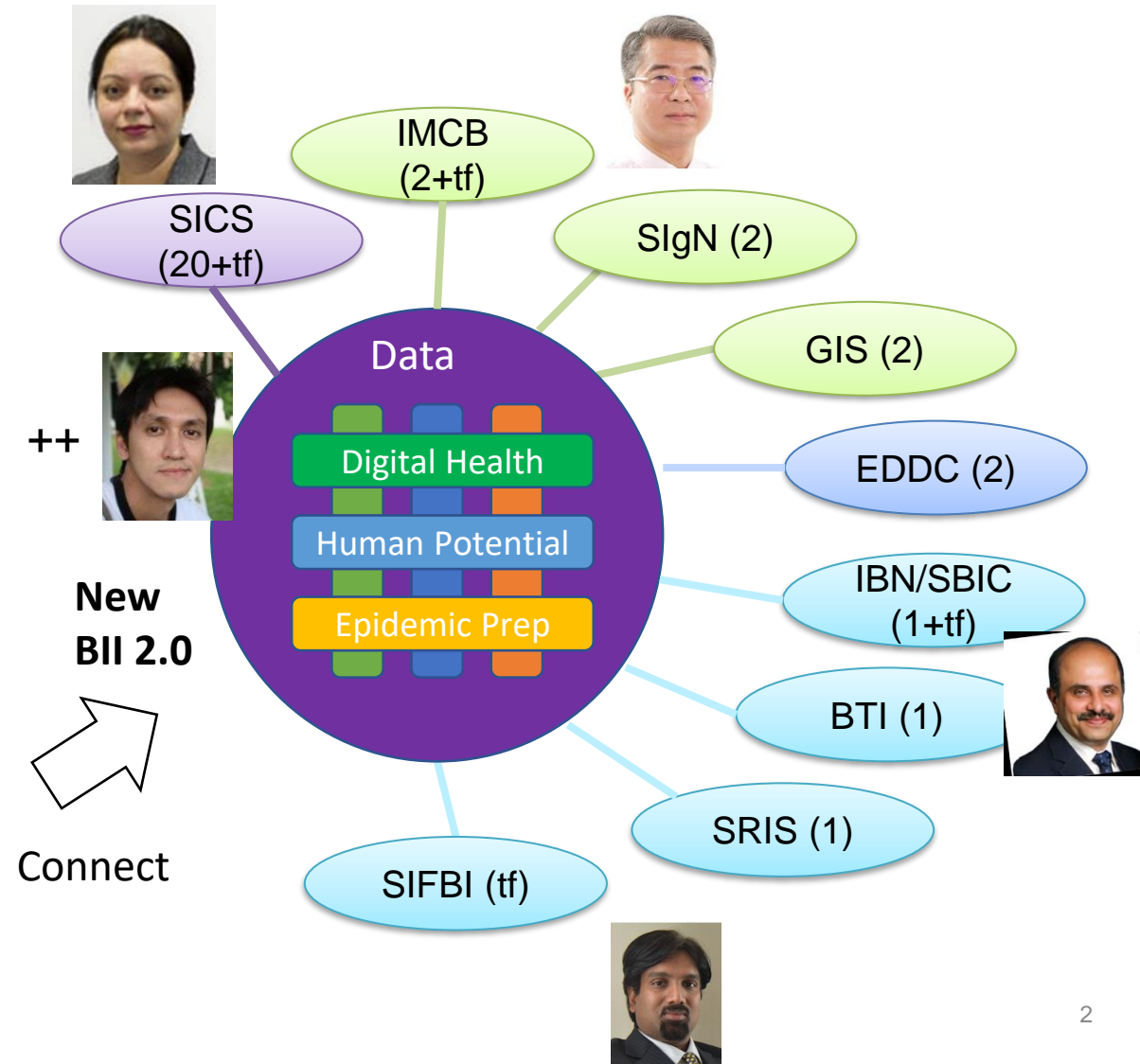
Cross-map

Horizontal by Strategic Area



Focus

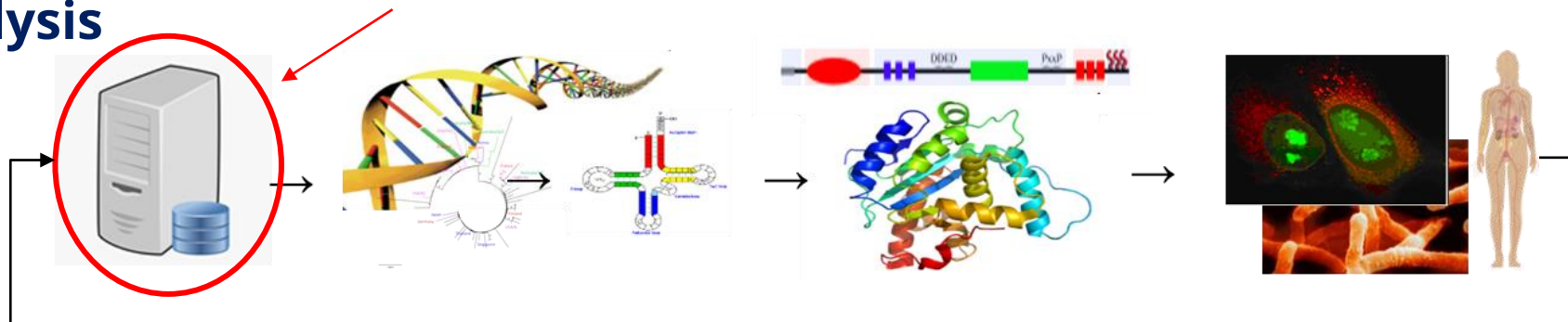
New collaborative Biomedical Data Analysis (BMDA) hub (joint) around the data repository



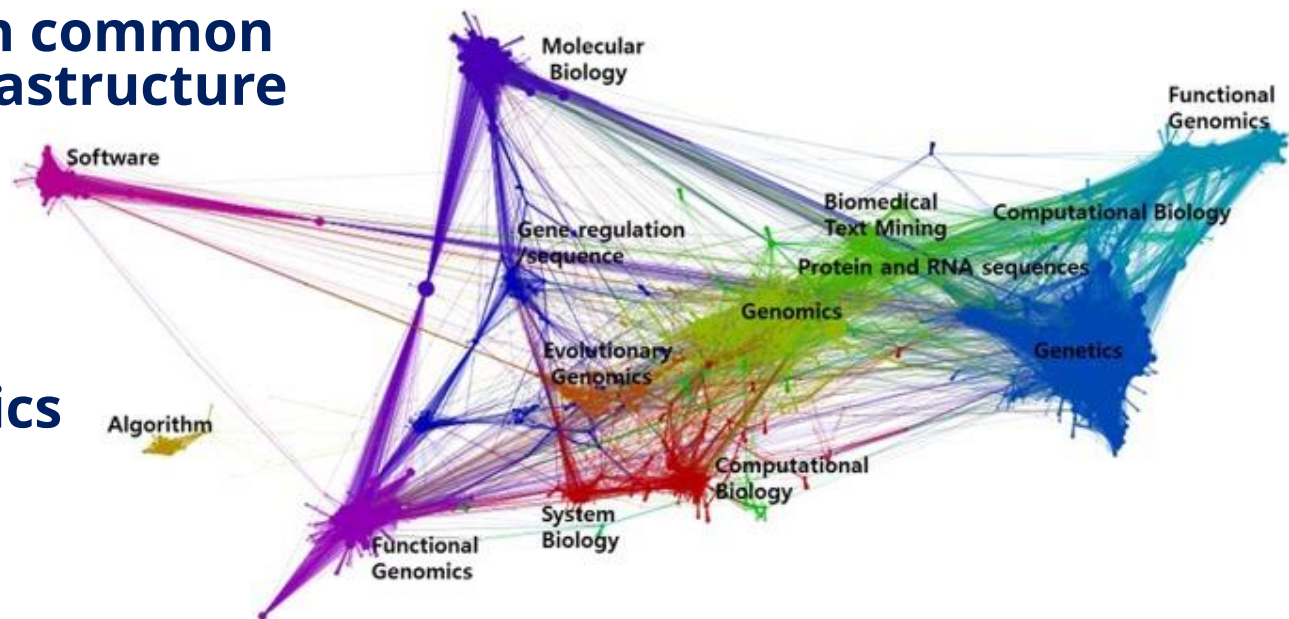


Big Hairy Audacious Goals (BHAG) for BII 2.0

- ❑ To become Singapore's trusted center for Life Science/Medical Research Databases and Data Analysis



- ❑ To become the joint bioinformatics collaboration workspace with common life science data analysis infrastructure for A*STAR



- ❑ To formalize the bioinformatics education/outreach function

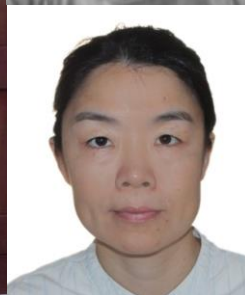


CREATING GROWTH, ENHANCING LIVES



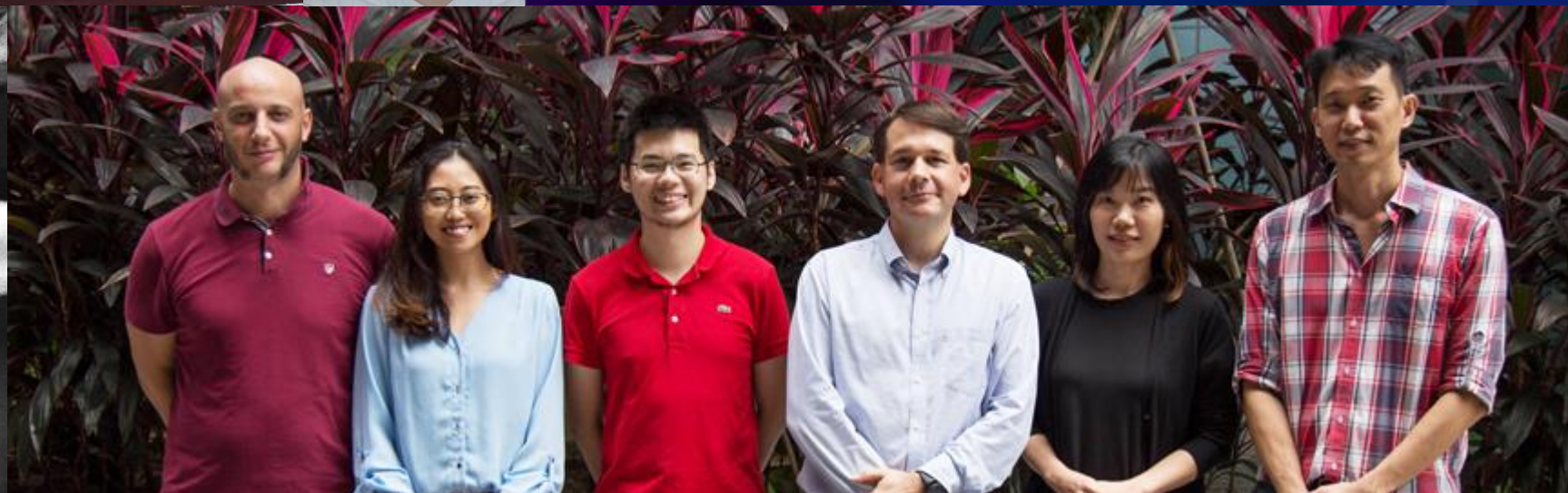
THANK YOU

www.a-star.edu.sg



BIOMOLECULAR SEQUENCE TO FUNCTION

- **Protein Sequence Analysis**



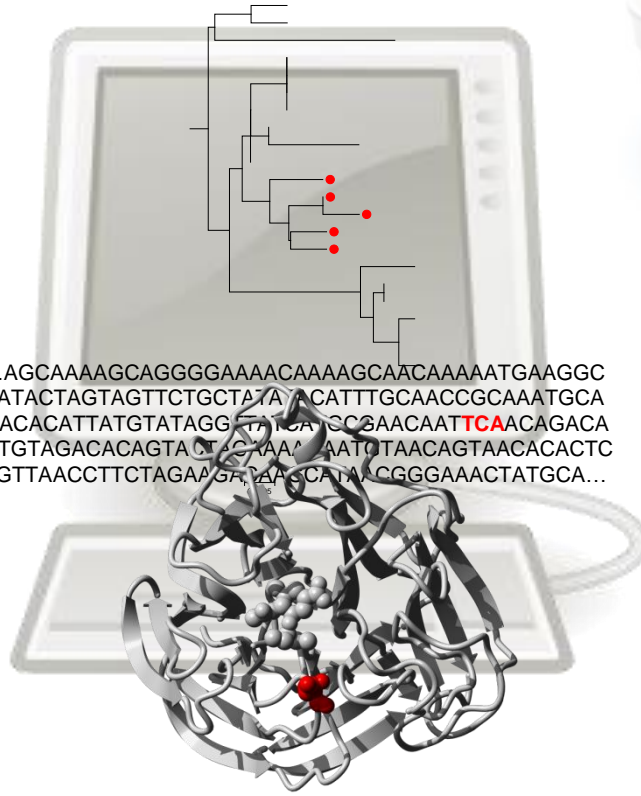


Viruses



Restricted

Computational Sequence and Structure Analysis

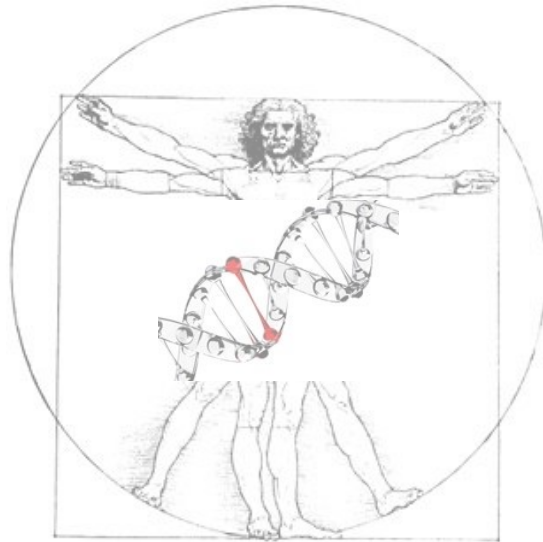


Food flavors



Drugs

Human variation



Product Safety



We are working at the interface between sequence and structure

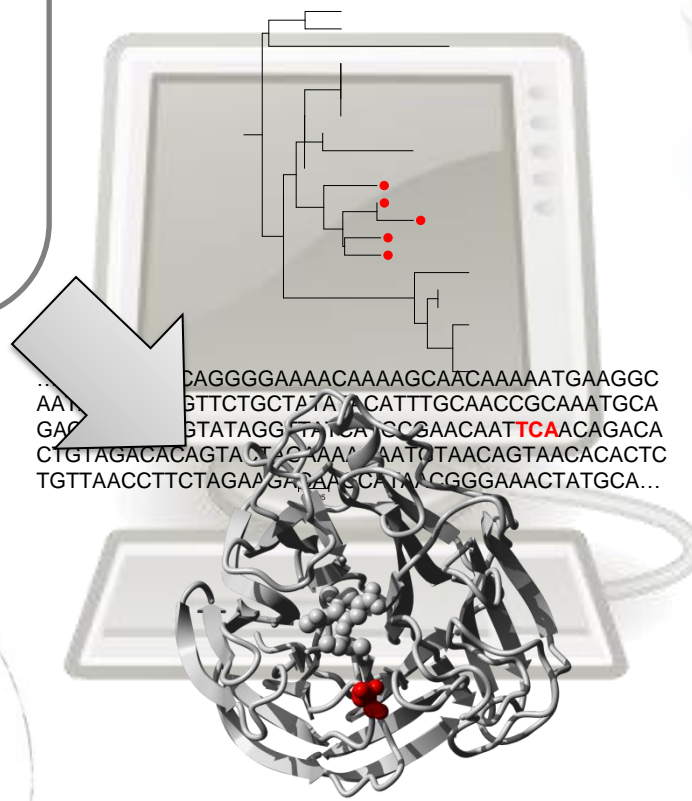


GIS



Restricted

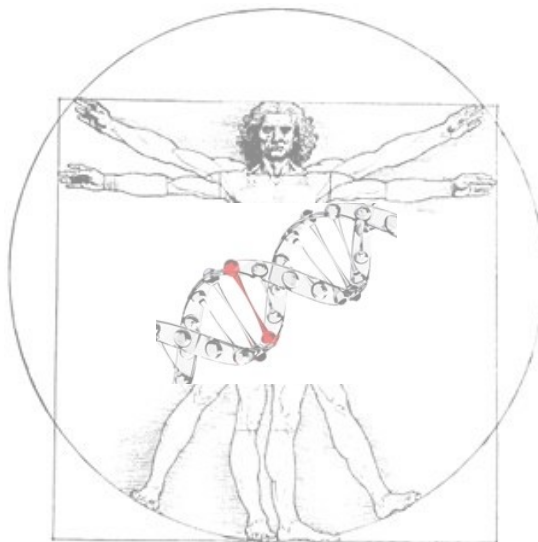
Computational
Sequence and
Structure Analysis



Food flavors



Human variation

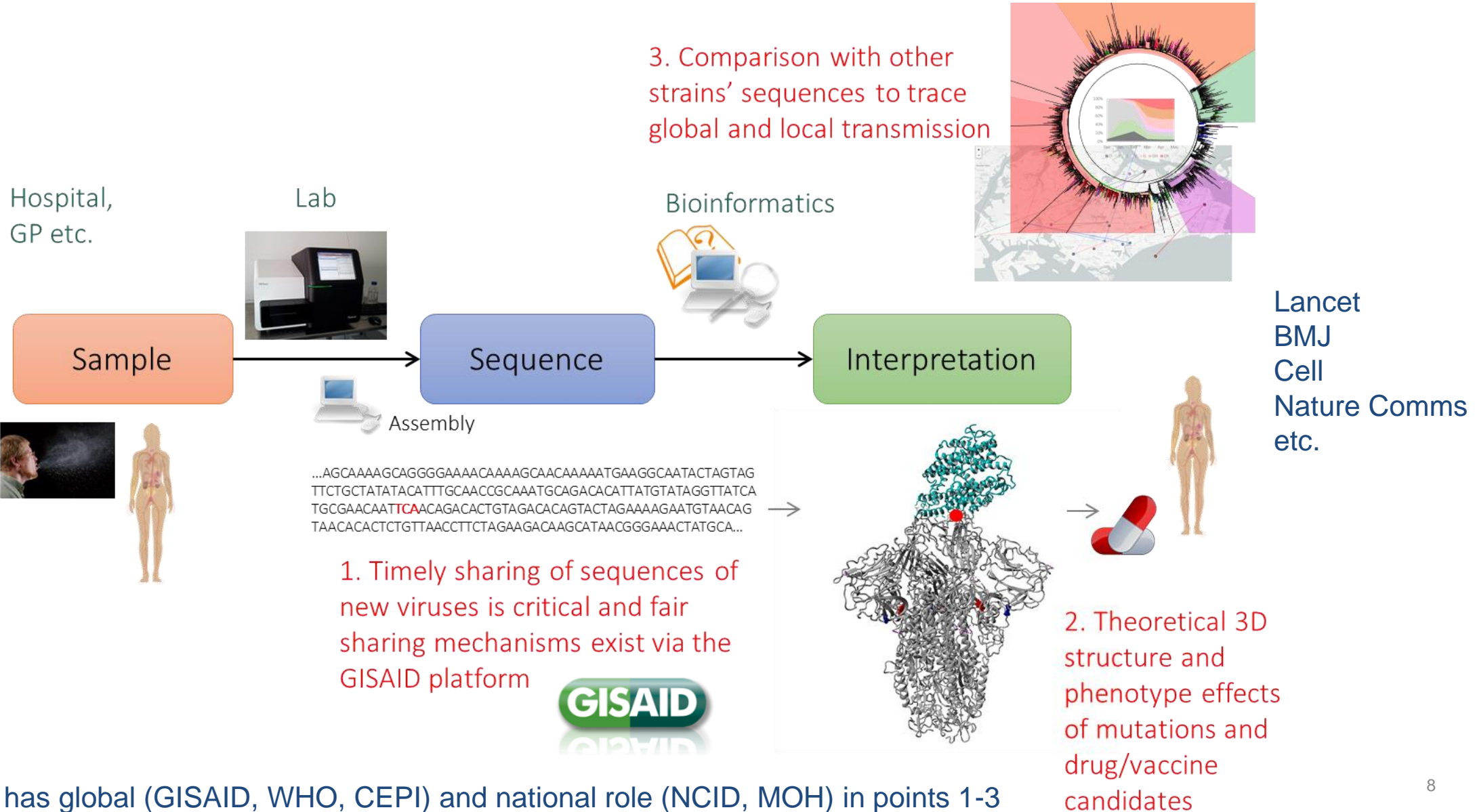


Product Safety



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

Example COVID-19: critical mass data + analysis = impact



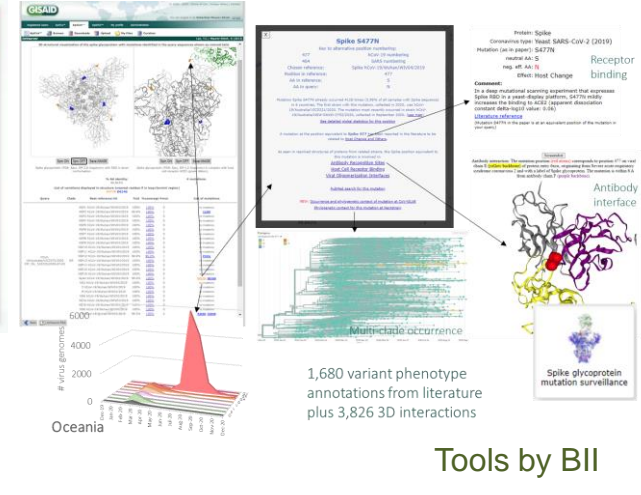
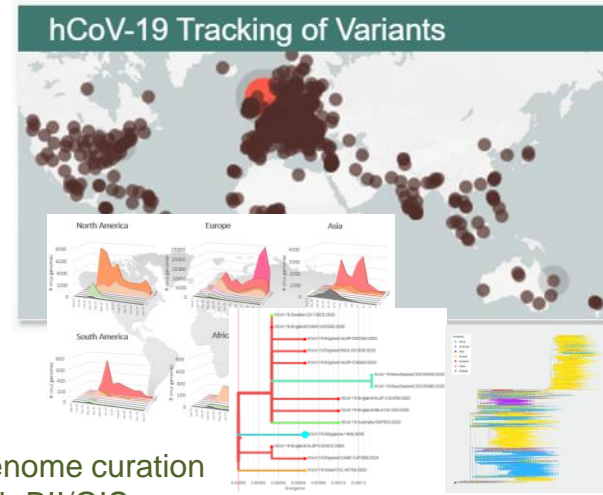
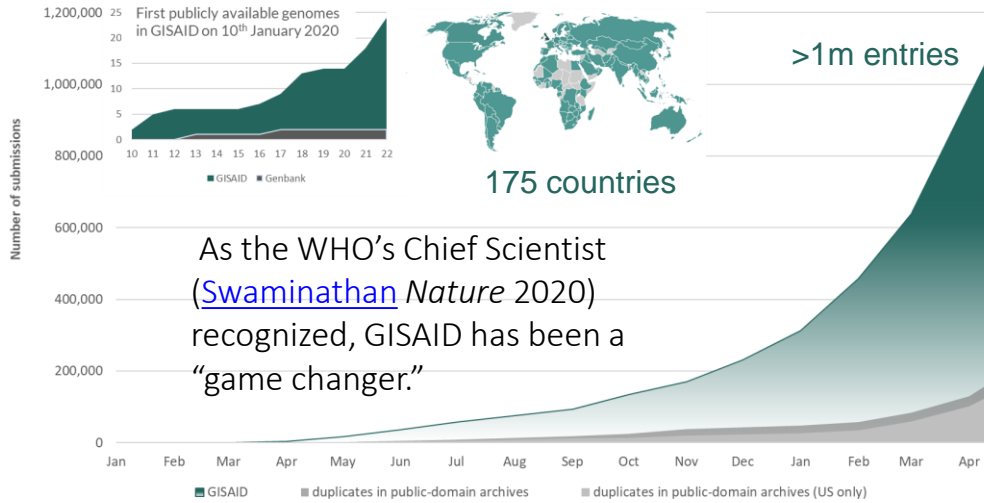
BII/GIS has global (GISAID, WHO, CEPI) and national role (NCID, MOH) in points 1-3



COVID-19: Impact of A*STAR's bioinformatics capabilities via GISAID



CREATING GROWTH, ENHANCING LIVES



The NEW ENGLAND JOURNAL of MEDICINE

THE STRAITS TIMES SINGAPORE

Singapore approves Pfizer-BioNTech Covid-19 vaccine in Asia first: What you need to know about the shot

RESEARCH SUMMARY

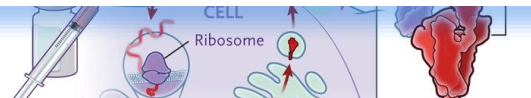
Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

F.P. Polack et al. DOI: 10.1056/NEJMoa2034577

CLINICAL PROBLEM

Safe and effective vaccines to prevent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and Covid-19 are urgently needed. No vaccines against betacoronaviruses are currently available, and mRNA-based vaccines have not been widely tested.

investment of resources. The development of BNT162b2 was initiated on January 10, 2020, when the SARS-CoV-2 genetic sequence was released by the Chinese Center for Disease Control and Prevention and disseminated globally by the GISAID (Global Initiative on Sharing All Influenza Data) initiative.

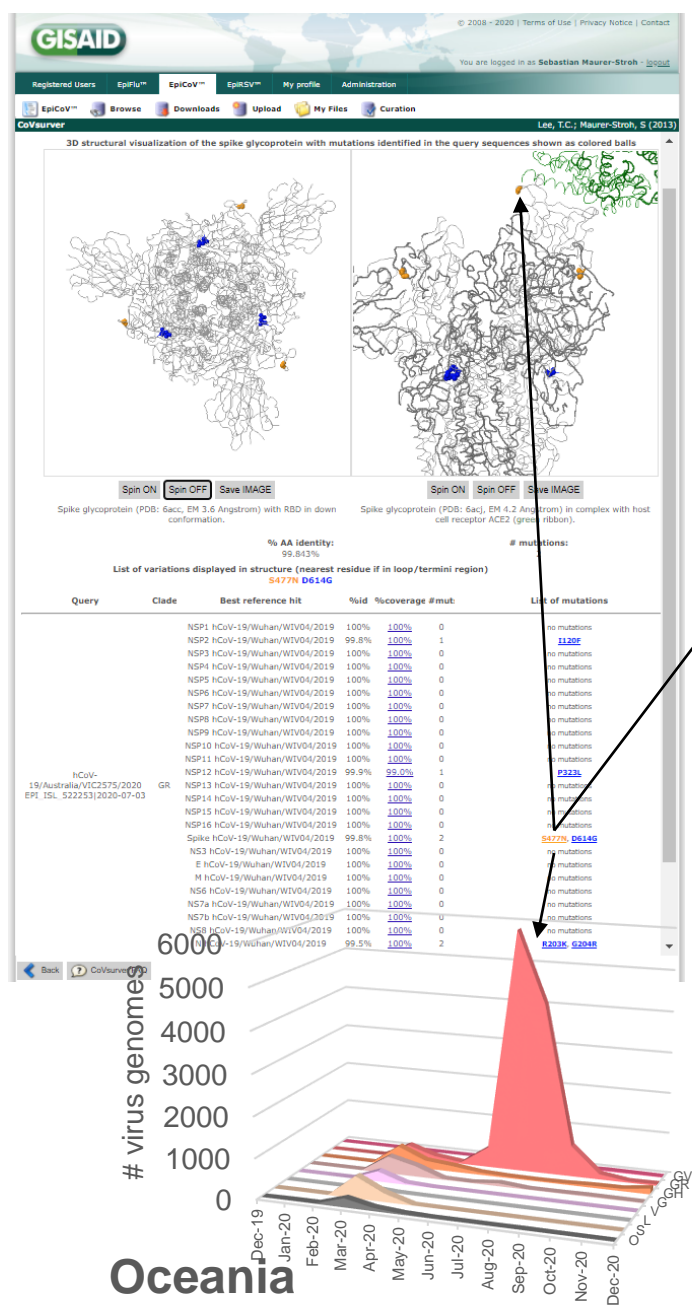




CREATING GROWTH, ENHANCING LIVES

Restricted CoVsurver

Raphael, Joses, Yani, Ashar



Spike S477N

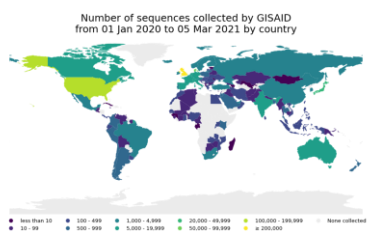
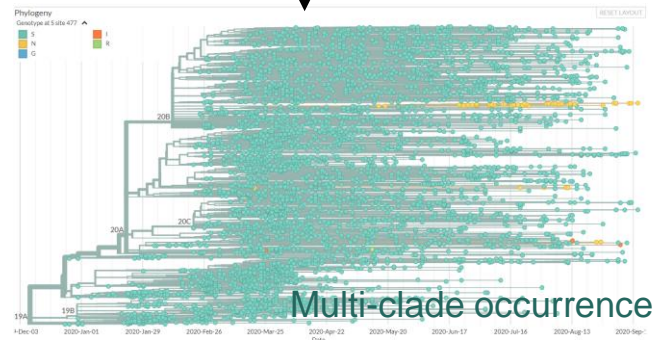
Key to alternative position numbering:
 477 hCoV-19 numbering
 464 SARS numbering
 Chosen reference: Spike hCoV-19/Wuhan/WIV04/2019
 Position in reference: 477
 AA in reference: S
 AA in query: N

Mutation Spike S477N already occurred 4128 times (3.98% of all samples with Spike sequence) in 6 countries. The first strain with this mutation, collected in 2020, was hCoV-19/Australia/VIC5321/2020. The mutation most recently occurred in strain hCoV-19/Australia/NSW-SAVID-2752/2020, collected in September 2020. (see map)
[See detailed global statistics for this position](#)

A mutation at the position equivalent to **Spike 477** has been reported in the literature to be related to **Host Change and Others**.

As seen in resolved structures of proteins from related strains, the Spike position equivalent to this mutation is involved in:
[Antibody Recognition Sites](#)
[Host Cell Receptor Binding](#)
[Viral Oligomerization Interfaces](#)
[PubMed search for this mutation](#)

NEW: [Occurrence and phylogenetic context of mutation at CoV-GLUE](#)
[Phylogenetic context for this mutation](#)

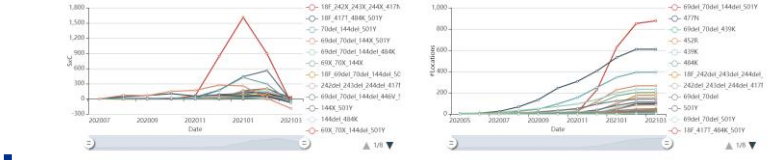
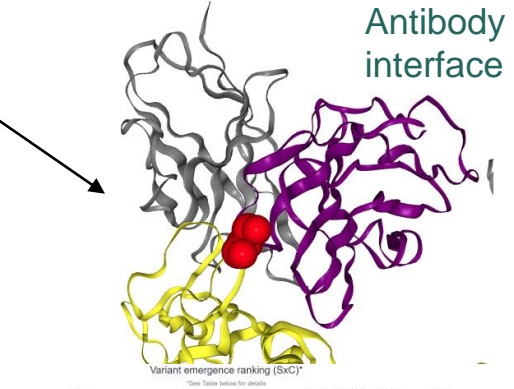


Protein: Spike
 Coronavirus type: Yeast SARS-CoV-2 (2019)
 Mutation (as in paper): S477N
 neutral AA: S
 neg. eff. AA: N
 Effect: Host Change

Receptor binding

Comment:
 In a deep mutational scanning experiment that expresses Spike RBD in a yeast-display platform, S477N mildly increases the binding to ACE2 (apparent dissociation constant delta-log10 value: 0.06)
[Literature reference](#)
 (Mutation S477N in the paper is at an equivalent position of the mutation in your query)

Antibody interaction: The mutation position (red atoms) corresponds to position 477 on viral chain E (yellow backbone) of protein entry 6xn, originating from Severe acute respiratory syndrome coronavirus 2 and with a label of Spike glycoprotein. The mutation is within 6 Å from antibody chain F (purple backbone).



MetaData of variants in a 3-month period

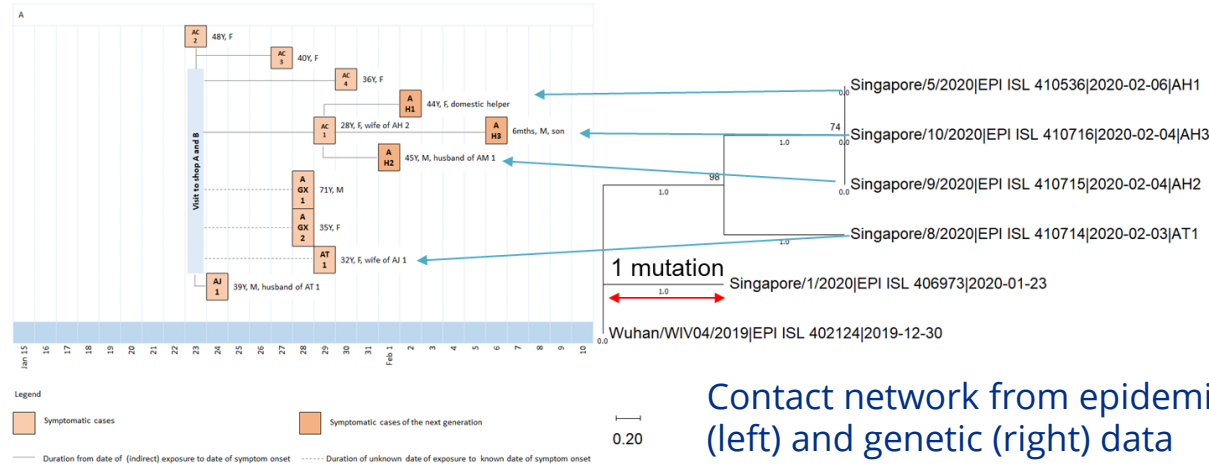
February 2021

Variant	#Genomes	#Top Location	#Top Clade	#Top Lineage	#DeltaLog(B)	#AaChange(sC)	#(SxS)
S477N	14421	98741 england	14589 cny	14448 B.1.1.7	225	4	892
S477N	1592	311 thailand	1207 GH	1052 B.1.351	88	7	988
S477N	1358	185 thailand	1241 GH	1052 B.1.351	48	8	204
S477N	620	169 thailand	588 GH	619 P.1	51	4	204
S477N	570	83 thailand	648 GH	608 B.1.1.7	50	4	200
S477N	483	94 england	477 G	488 B.1.525	39	4	156

Emerging variants

COVID-19: Impact of A*STAR's bioinformatics capabilities

From contact tracing...

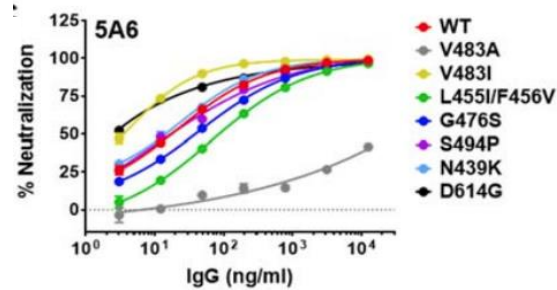
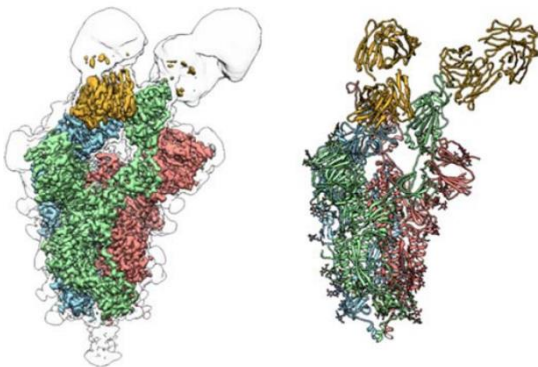


Contact network from epidemiological (left) and genetic (right) data

Characterization of first 3 local clusters, led by MOH

- BII conducted genomic epidemiology analyses
- Findings published in **The Lancet** (Pung et al 2020)

...to therapeutics



A*STAR-led team discovered anti-COVID-19 therapeutic antibodies

- SigN led characterization of antibodies
- BII assessed mutations that could affect binding
- Findings accepted for publication in **Cell** (Wang et al 2021 – in press)



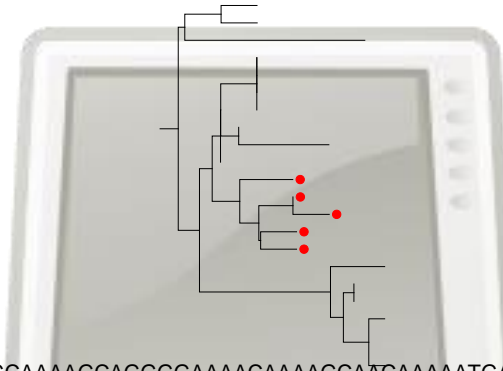


Viruses



Restricted

Computational Sequence and Structure Analysis



Food flavors



Drugs

...AGCAAAAGCAGGGGAAAACAAAAGCAACAAAATGAAGGC
AATACTAGTAGTTCTGCTATAACATTGCAACCGCAAATGCA
GACACATTATGTATAGGATACACCCBAACAATTCAACAGACA
CTGTAGACACAGTACTCAAAATAATTAACAGTAACACCTC
TGTTAACCTTCTAGAAATTAACATACGGGAAACTA

Human variation



Product safety



Proteins can cause allergy if similar to other allergens.



P&G – BII Allergenicity Project

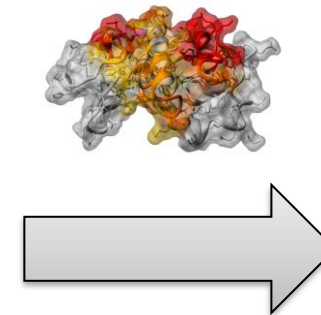


Protein-Antibody



MGVFN YETETTSVIP AARLFKAFILDGDNLFPK
 VAPQA ISSVENIEGNGGPGTIKKISFPEGFPFKY
 VKDRVDEVDHTNFKYNYSVIEGGPIGDTLEKIS
 NEIKIVATPDGGSILKISNKYHTKGDHEVKAEQ
 VKASKEMGETLLRAVESYLLAHS DAYN

Protein Sequence



Allergenicity Potential

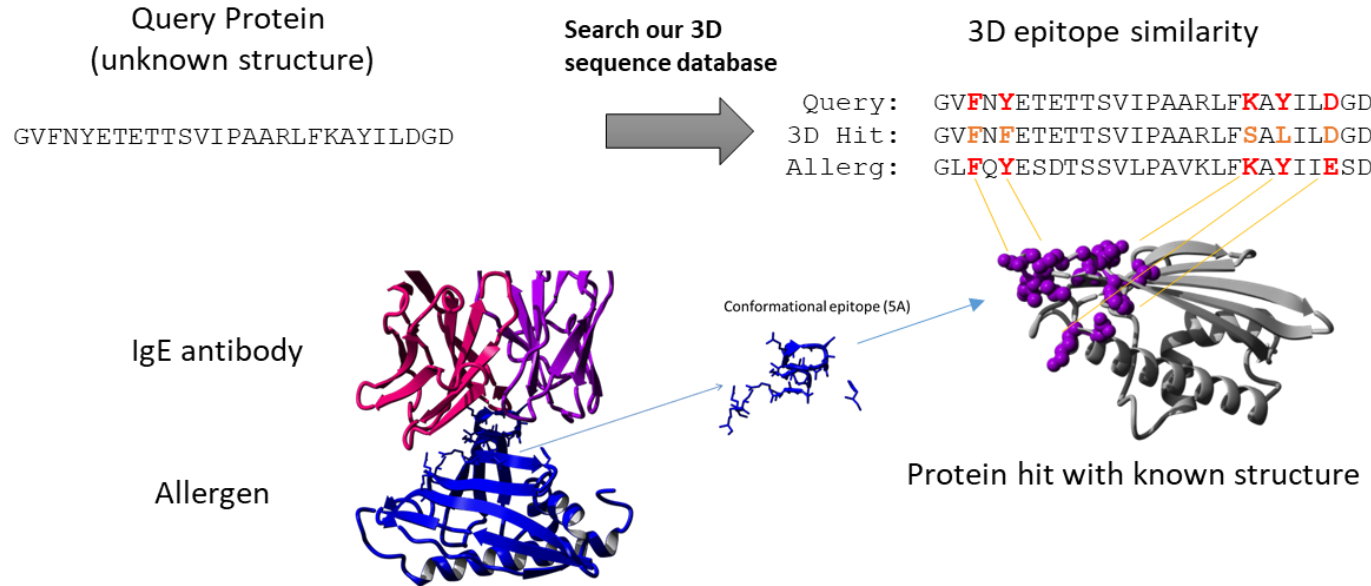
HIGH

LOW

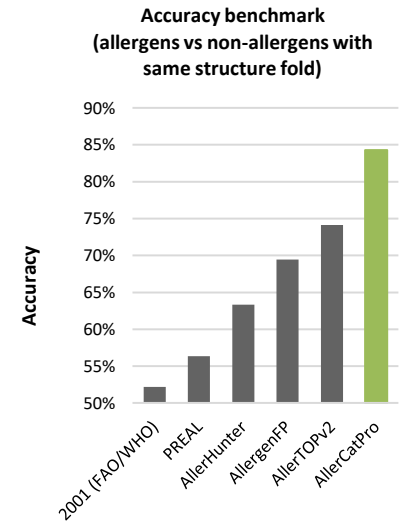


BII tool AllerCatPro
 predicts if a protein is similar to known allergens

Computational prediction of protein allergenicity: from viruses to shampoo and food safety



AllerCatPro predicts if a protein is similar to known allergens



The impact (BII+Ben Smith SIFBI):

- ✓ Industry:
 - ✓ 4 RCAs long-term work with P&G, active use by P&G in product development safety pipeline
 - ✓ Gluten extension work with The Coca Cola Company
 - ✓ Project pipeline: next 2 companies in preparation
- ✓ Supporting 30-by-30 goal by providing tool to assess risk of alternative/novel protein sources for food
 - ✓ Project with CSIRO/JCU on seafood allergy and risk of insect proteins as food source
 - ✓ Consulting for AVA/SFA on Novel Foods, ongoing safety assessments in SFA expert panel



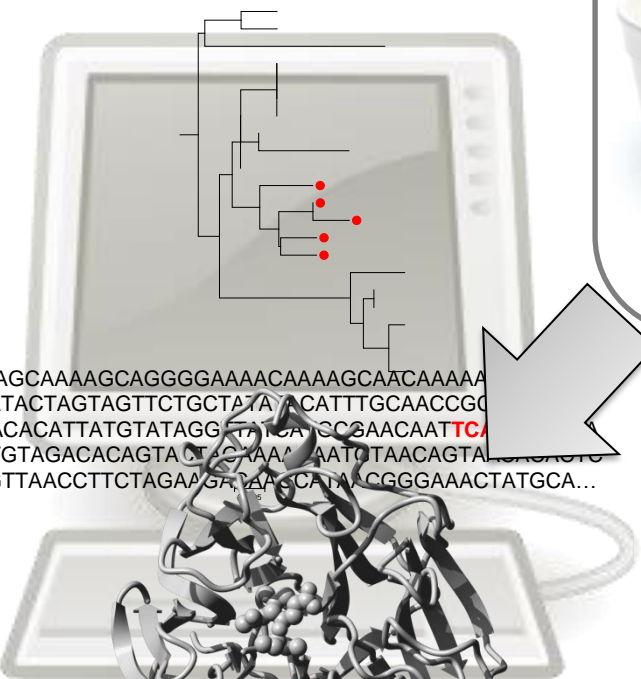


Viruses



Restricted

Computational
Sequence and
Structure Analysis



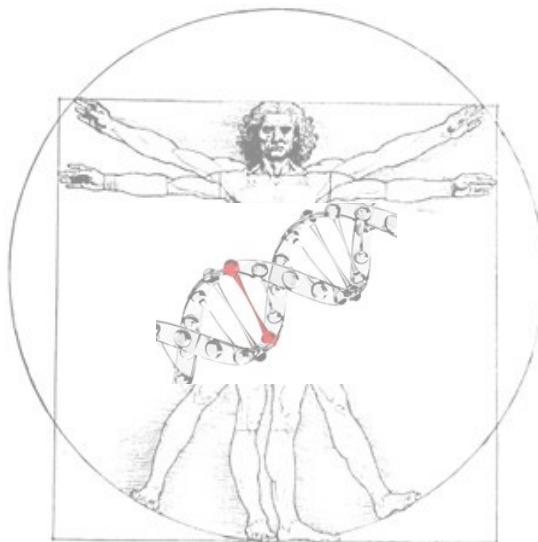
```
...AGCAAAAGCAGGGGAAAACAAAAGCAATCAAAA  
AATACTAGTAGTTCTGCTATAACATTGCAACCGG  
GACACATTATGTATAGGATACACCCBAACAATTC  
CTGTAGACACAGTACTACAAAATAATTAACAGT  
TGTTAACCTTCTAGAAATACATACCGGGAAACTATGCA...
```

Food flavors



Drugs

Human variation



Product Safety

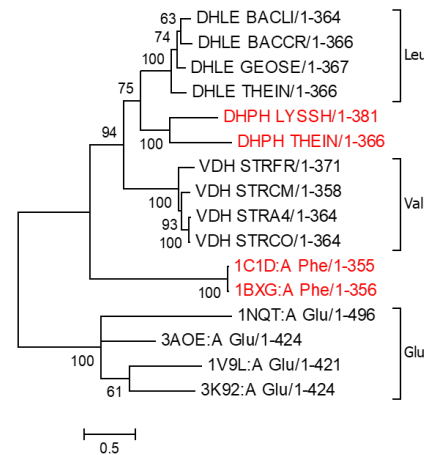
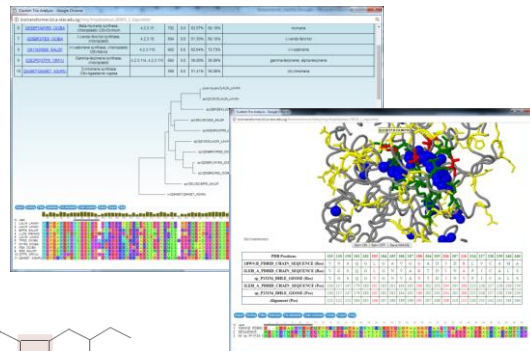
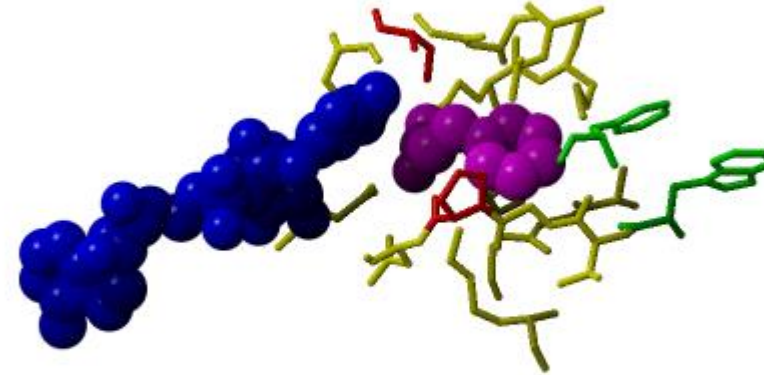
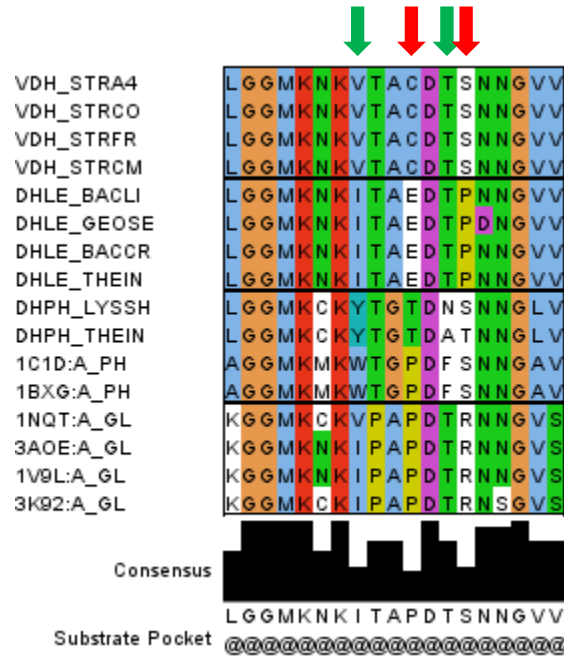


Enzymes can be used to naturally produce food flavours and fragrances but also drugs!

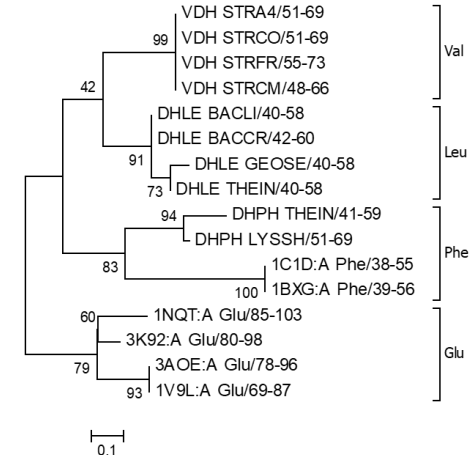
Moving from full sequence similarity to substrate binding pocket similarity



CREATING GROWTH, ENHANCING LIVES



Full length sequence tree

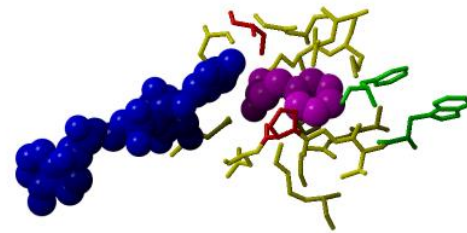


Substrate binding pocket profile tree

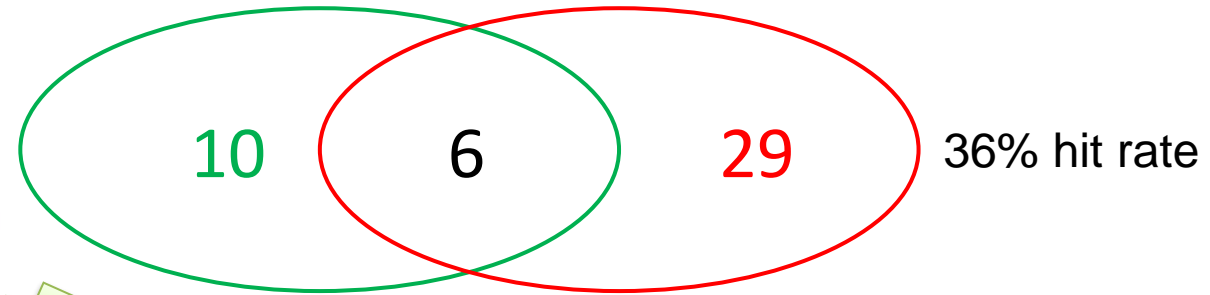


Iterative computational/experimental refinement cycle

First screen



Positive/negative examples

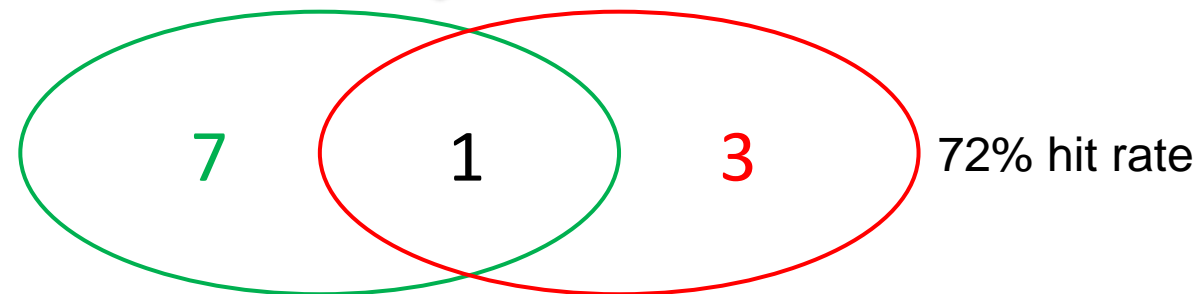


Substrate binding pocket profile to score candidates

	ELVF	BCAT	AOX	max
	7.25		6.40	7.25
	-3.92	4.89		4.89
	-5.21	1.82	13.08	13.08
		5.00		5.00
		2.09		2.09
		1.82		1.82
		1.82	6.44	6.44
		1.68	6.91	6.91
		1.68		1.68
		1.68		1.68
		1.68		1.68
		1.00		1.00
			8.26	8.26

Confidential

Second screen



Computational modelling of mutation effects on enzyme function

Kinetics

EVB Model

Binding

Docking scores
Binding free energy

Stability

Yasara + FoldX

Aggregation

Tango
Waltz

Allostery

Allosigma

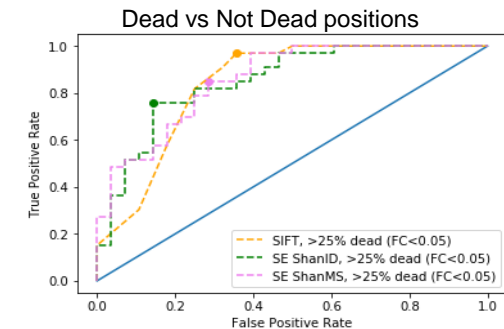
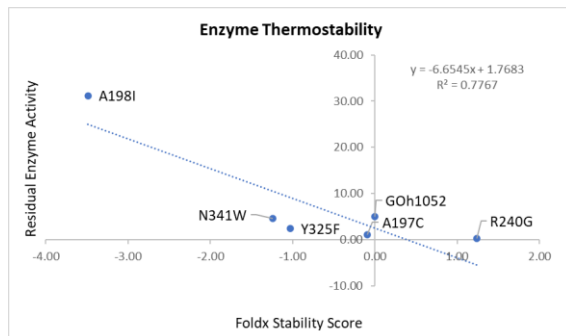
Sequence Conservation

SIFT
Shannon entropy



Enzyme-Ligand Composition Extraction

PHARMA INNOVATION
PROGRAMME SINGAPORE (PIPS)



SIFT: AUC=0.84, Best Threshold=0.70
ShanID: AUC=0.85, Best Threshold=3.27
ShanMS: AUC=0.86, Best Threshold=1.76





Viruses



Restricted

Computational Sequence and Structure Analysis



Food flavors



Human variation

gsk
MSD
Pfizer

GIS

Individual mutations can cause disease or affect drug efficacy

Product Safety

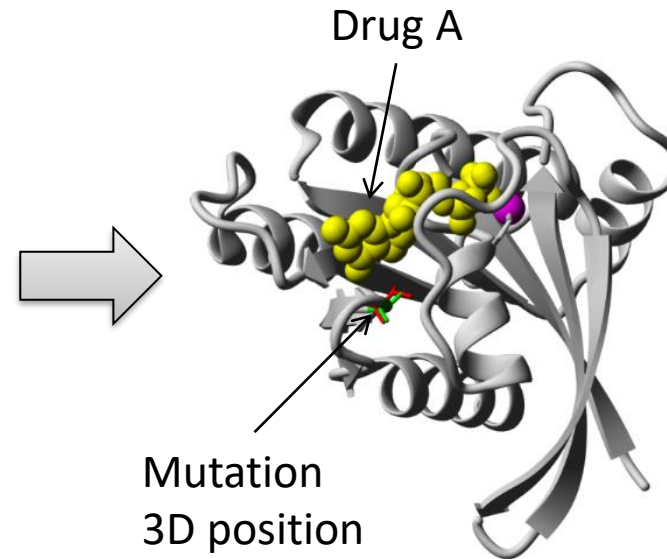


SNPdrug3D database for Precision Medicine

```

...AGCAAAAGCAGGGGAAAACAAAAGCAA
CAAAAATGAAGGCAATACTAGTAGTTCTG
CTATATACATTTGCAACCGCAAATGCAGA
CACATTATGTATAGGTTATCATGCGAACAA
TTCAACAGACACTGTAGACACAGTACTAG
AAAAGAATGTAACAGTAACACACTCTGTTA
ACCTTCTAGAAGACAAGCATAACGGGAAA
CTATGCA...
    
```

Sequence SNP/variant X



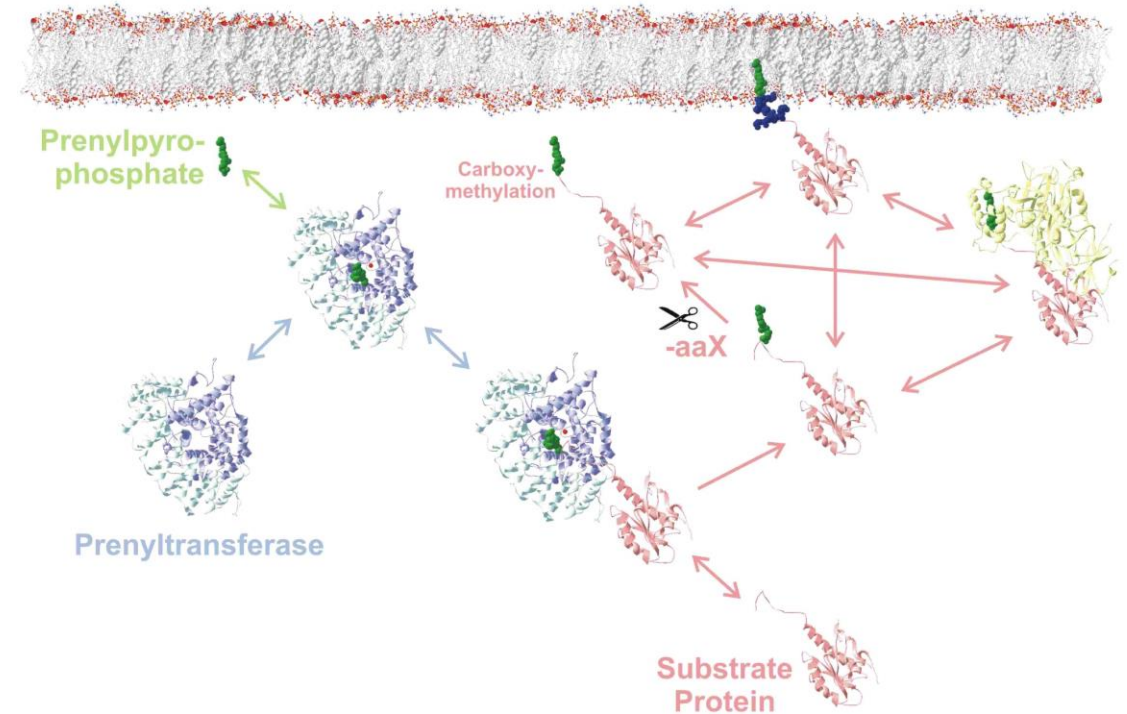
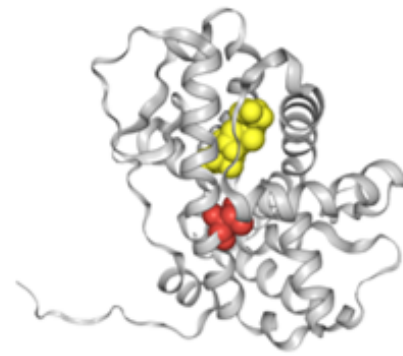
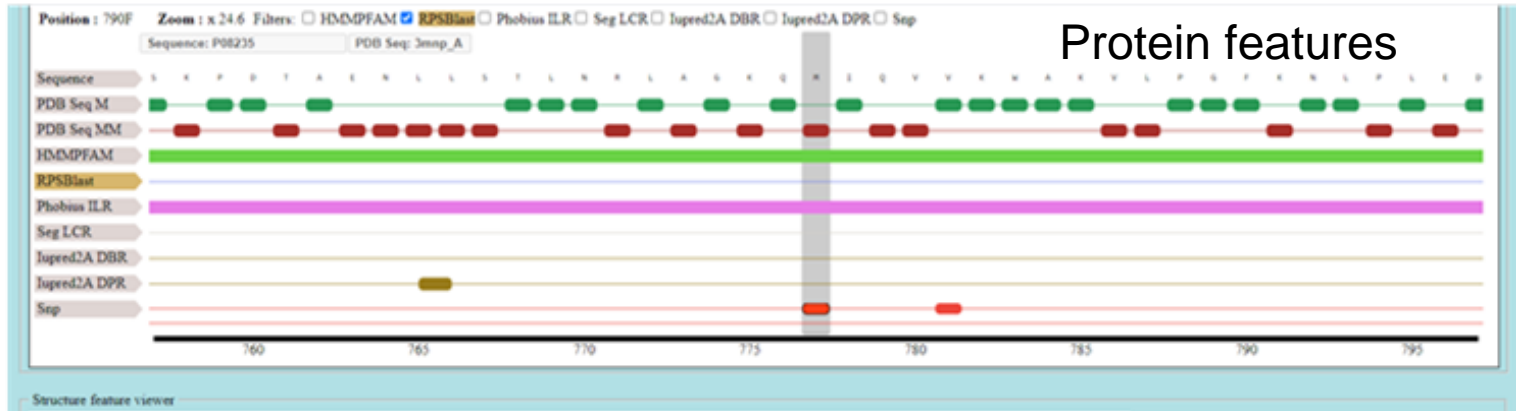
You have variant X that could affect response to drug A so I suggest drug B!

Good, thanks!



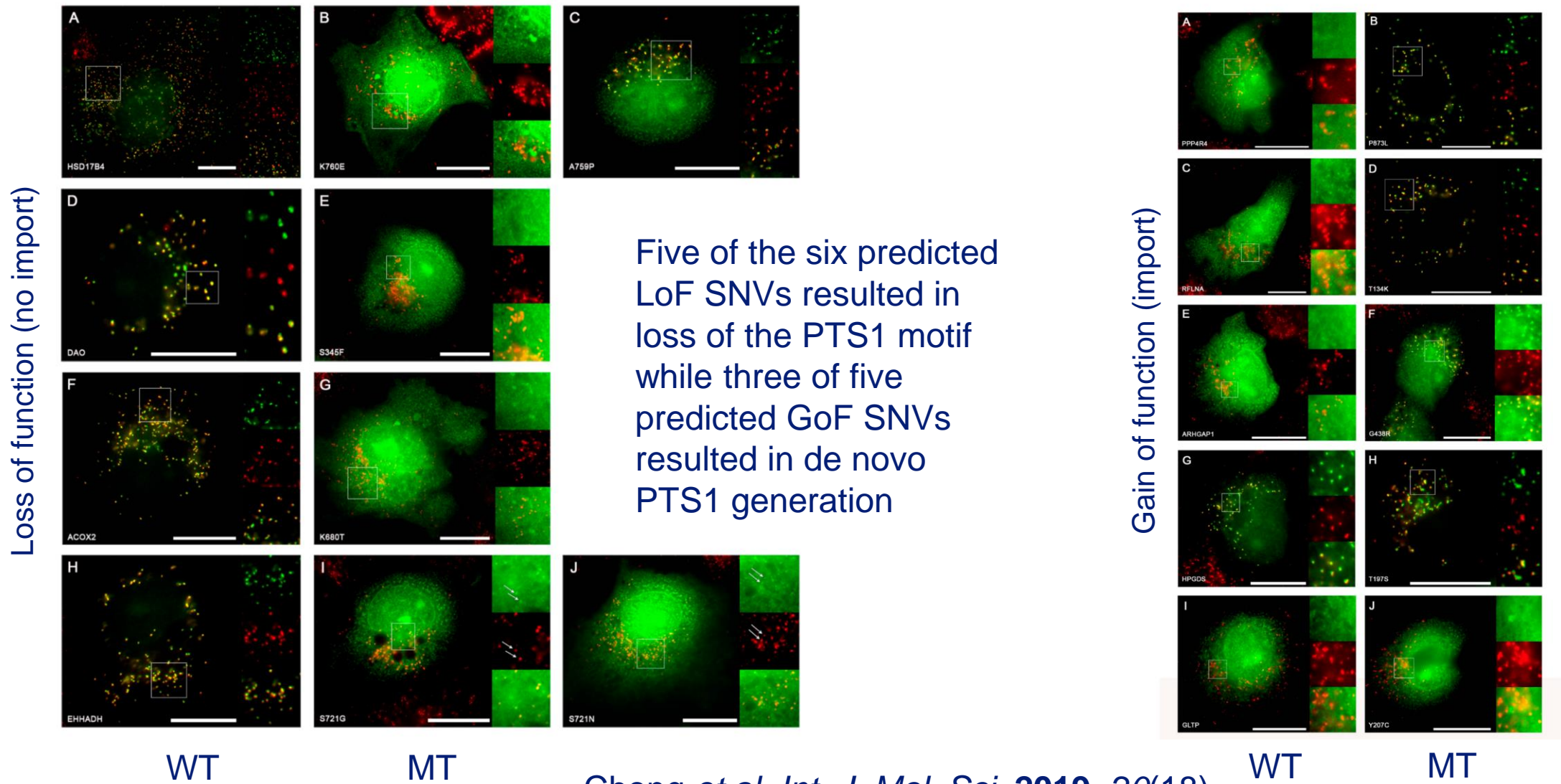
User	Value
Scientist	SNP prioritization VUS annotation
Clinician scientist	SNP prioritization VUS annotation Known and new PGx SNPs
Clinician	Known PGx SNPs

Functional evaluation of coding variants – Targeting motifs



Functional evaluation of coding variants – Targeting motifs

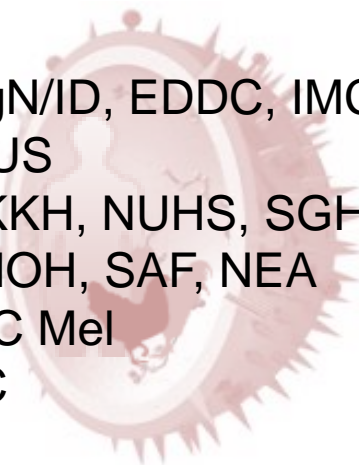
Systematic computational screen for SNPs in GnomAD affecting peroxisomal import





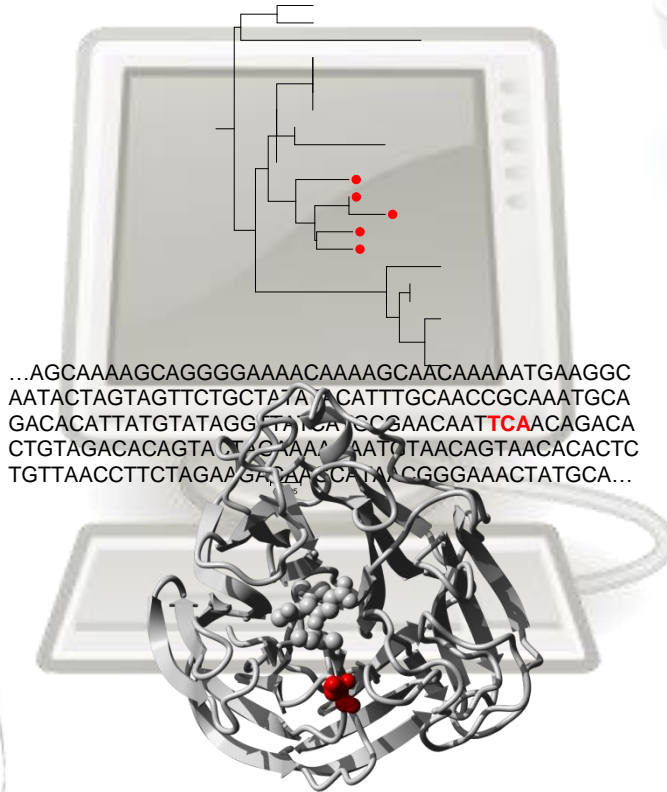
Viruses

GIS, SIgN/ID, EDDC, IMCB
 Duke-NUS
 TTSH, KKH, NUHS, SGH
 NPHL/MOH, SAF, NEA
 WHO CC Mel
 US CDC
 GISAID
 Saiba, Sanofi



Restricted

Acknowledgements



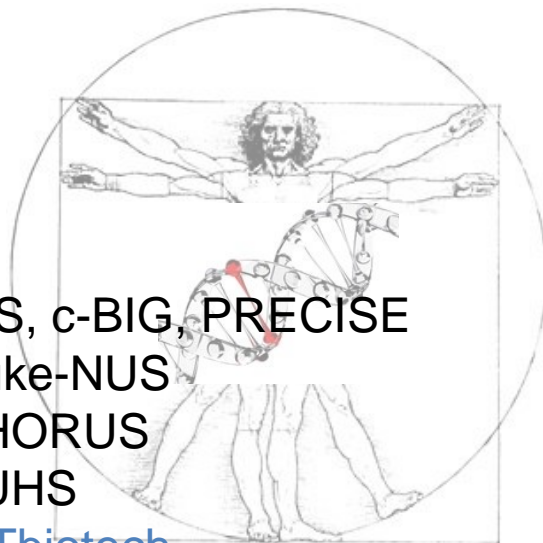
Food flavors

FNCC SIFBI
 Nestle
 Ferrero
 Firmenich
 Evolva
 Drugs
 GSK
 MSD
 Pfizer



Human variation

GIS, c-BIG, PRECISE
 Duke-NUS
 CHORUS
 NUHS
 AITbiotech



Product Safety

SRIS
 FNCC IFCS
 SFA
 US EPA
 U. Manchester
 U. Trier
 CSIRO
 P&G



Thank You!