Harnessing AlphaFold and explainable AI to better characterise human missense variants and diseases

Qisheng Pan[1,2], Stephanie Portelli[1,2], Thanh Binh Nguyen[1,2], David B. Ascher[1,2]

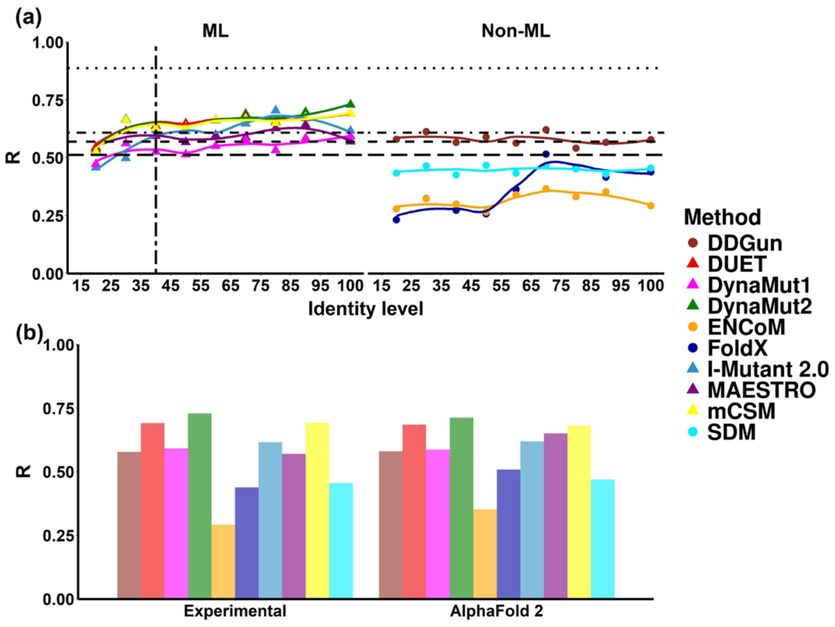
[1]*School of Chemistry and Molecular Bioscience, University of Queensland, Brisbane Queensland 4072, Australia*

[2]*Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne Victoria 3004, Australia*

Introduction

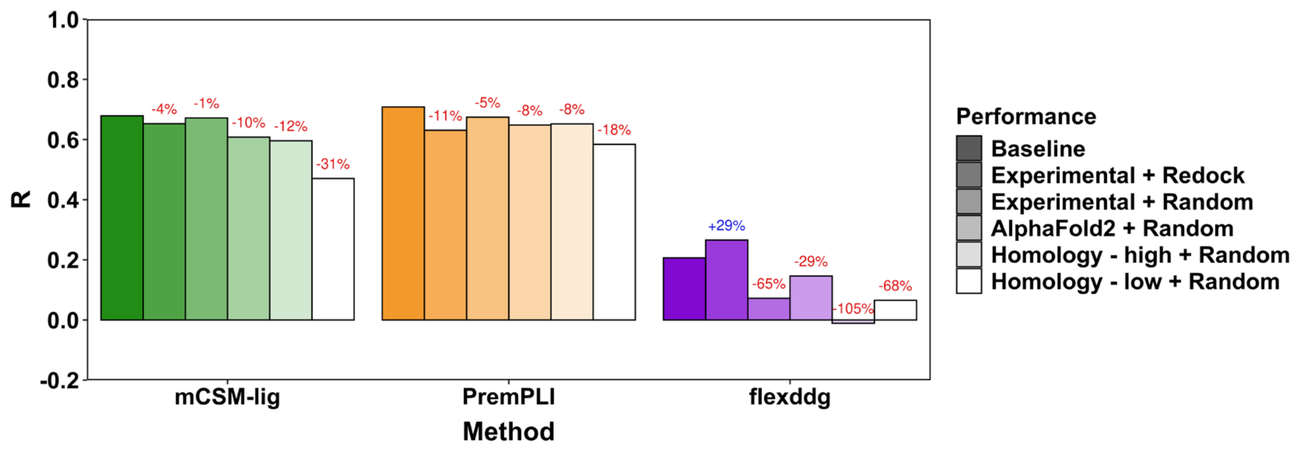
- Pathogenic missense mutations significantly disrupt **protein stability**, **interaction**, and **function**, while benign mutations bring **mild** effect to protein structures.
- Previous variant effect predictors primarily focus on effect of mutation on protein sequence and conservation.
- AlphaFold results in a **wealth** of protein structures, but these predicted structures have **not** been validated to study the effect of mutations.
- In this work, we studied the **structural consequences** caused by mutations, and used these features to develop a machine learning model to classify pathogenic outcome.

1. Homology modelling is reliable to study change of protein stability!



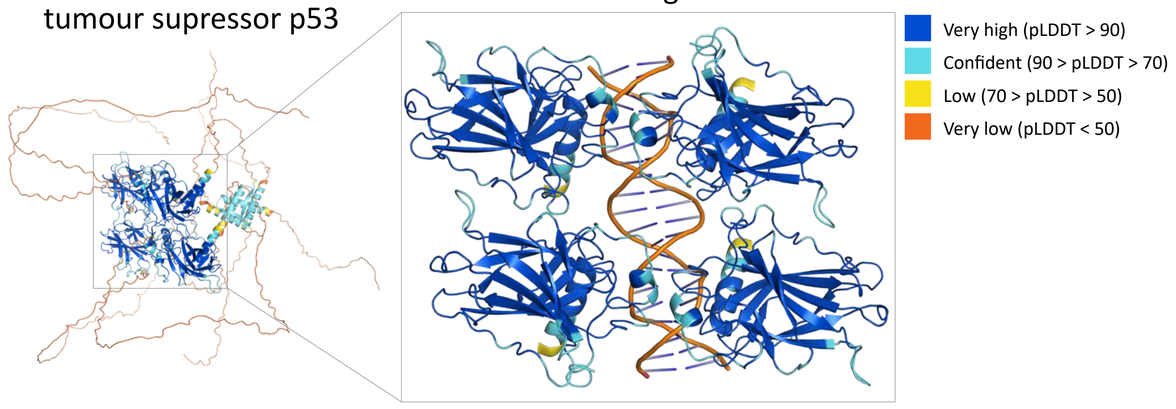
- Machine learning-based methods present **reliable performance** when using homology models with target-template identity down to **40%**.
- Performance on **AlphaFold2** models is **comparable** to performance using experimental structures.

2. Using AlphaFold for ligand interaction is NOT as good as expected!



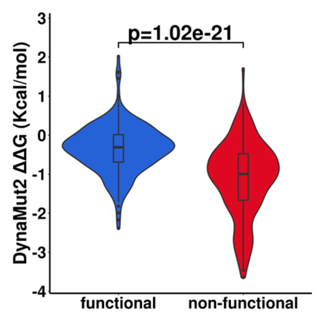
- There is around **5% deterioration** when the input complexes were composed of **experimental** receptors and docked ligands.
- The performance **deteriorated** by **10-20%** when we used **AlphaFold2-based** models as receptors to generate protein-ligand complexes,
- Which is **comparable** to traditional homology modelling-based analysis.

3. Strucutral features enable robust prediction of cancer mutations!

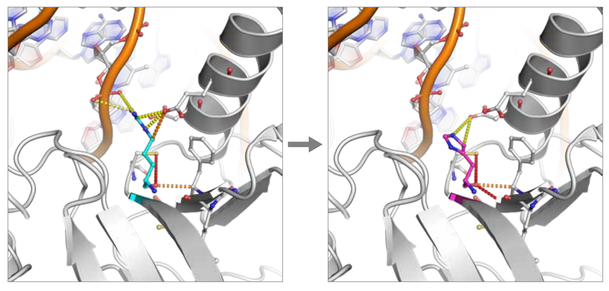


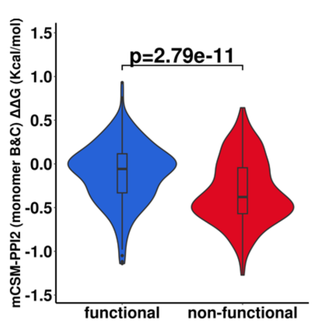
$n_{pathogenic}$ = 283

n_{benign} = 1011



Protein stability



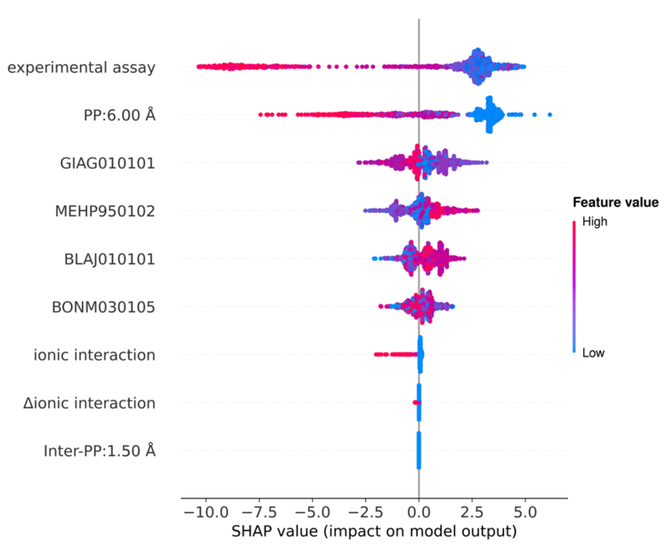


PPI

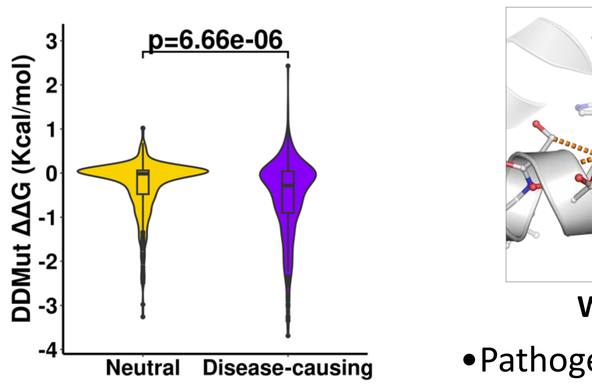
- Oncogenic mutations significantly **destabilise** both monomeric and tetricmeric p53 structure.
- The change of stability is mainly caused by the **loss** of intramolecular interactions.

Method	MCC	
	Blind test	Clinical test
Our model	**0.88**	**0.83**
AlphaMissense	0.78	0.82
SIFT	0.46	0.58
PolyPhen2	0.42	0.47
Experiment	**0.87**	**0.69**

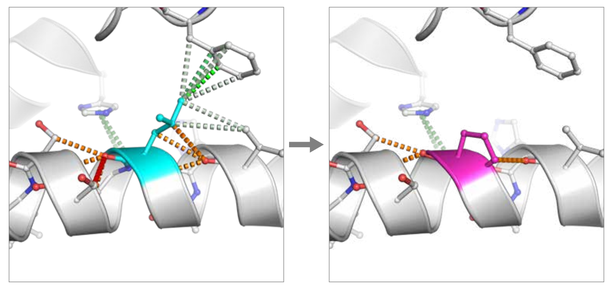
- Our strucutre-based method presents **robust** performance on both blind test and clinical validation.



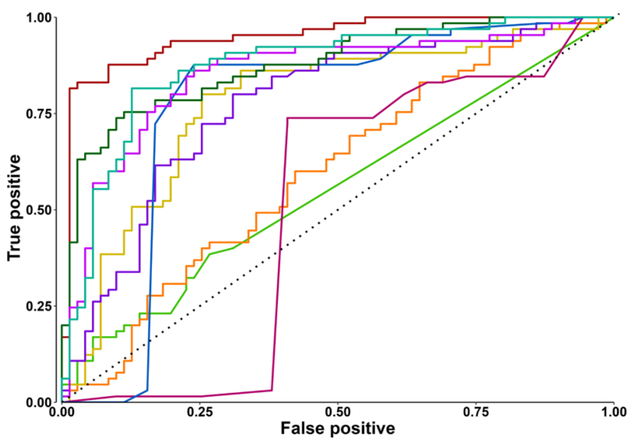
4. Structural features improve accurate identification of Alzheimer's Disease!



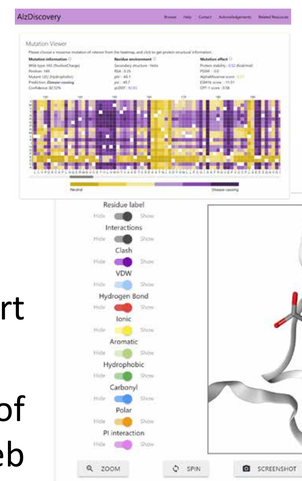
Protein stability



- Pathogenic mutations in AD also decrease **protein stability**, showing some common underlying disease mechanism.



Method	AUC	
	Blind test	Clinical test
Our model	**0.95**	**0.70**
Alz-Disc	0.55	0.54
AlphaMissense	0.88	0.40
SIFT	0.51	0.61
SNAP2	0.77	0.54



https://biosig.lab.uq.edu.au/alzdiscovery/

- Our model **outperformed** other state-of-the-art methods, particularly in clinical test.
- Researchers can **freely** access the predictions of all the possible missense mutations via our web server.

Summary

- It is **generally reliable** to use AlphaFold models to study the structral consequences of missense mutations.
- Characterising missense variants in **structural context** can provide a **molecular evidence** of their effect on protein, and the potential disease association.
- Incorporating **structural features** for machine learning model development **improve** predictive performance and **generalisability** to identify pathogenic variants.

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Baker
HEART & DIABETES INSTITUTE

Find out more about me