Characterisation on the pathogenic effect of the missense mutations of p53 via machine learning

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

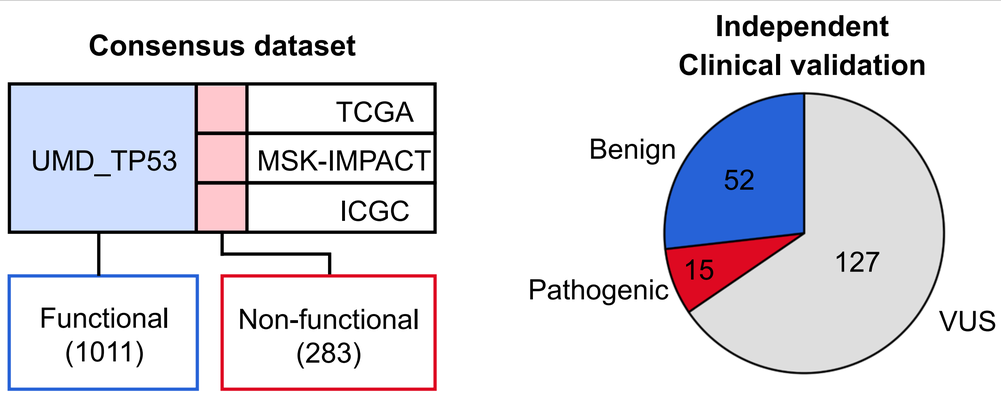
Background

- **Over 50% of cancers** are associated with the missense mutations in tumour suppressor protein p53.
- p53 plays a crucial role in DNA damage-induced activation by repairing erroneous replication and activating cellular apotosis.
- It is time-consuming and labour intensive to experimentally elucidate all the possible effects of **all missense variants**.

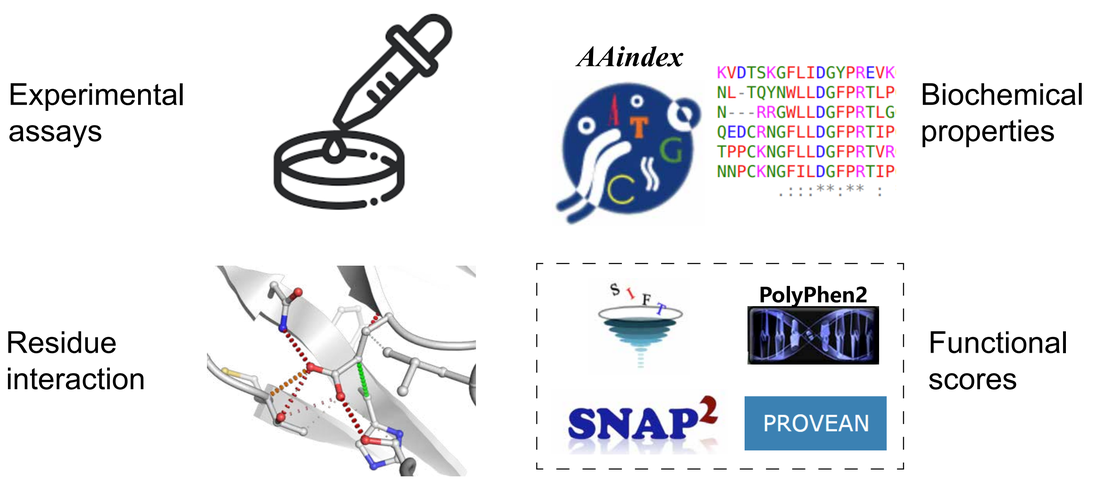
Aim: we aimed to accurately identify deleterious missense mutations in p53 by leveraging computational biophysical tools and employing a machine learning analysis.

Methodology

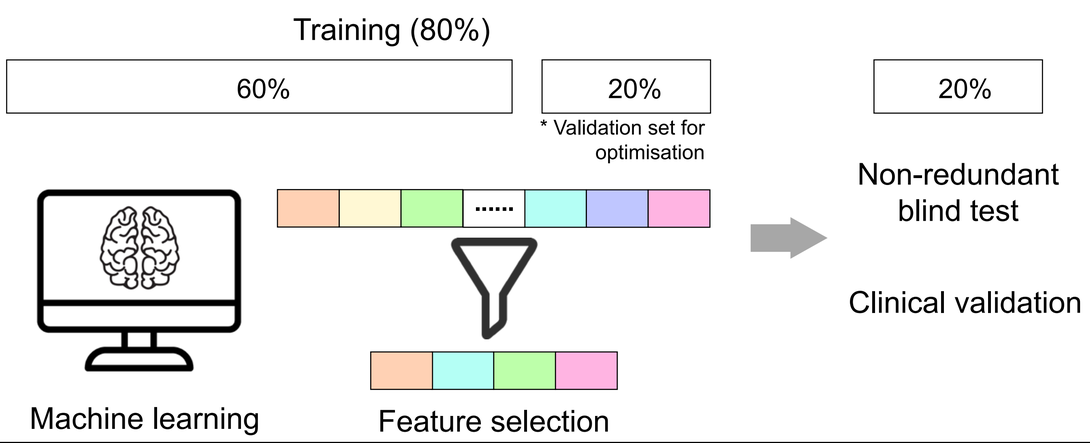
Data curation and preprocessing



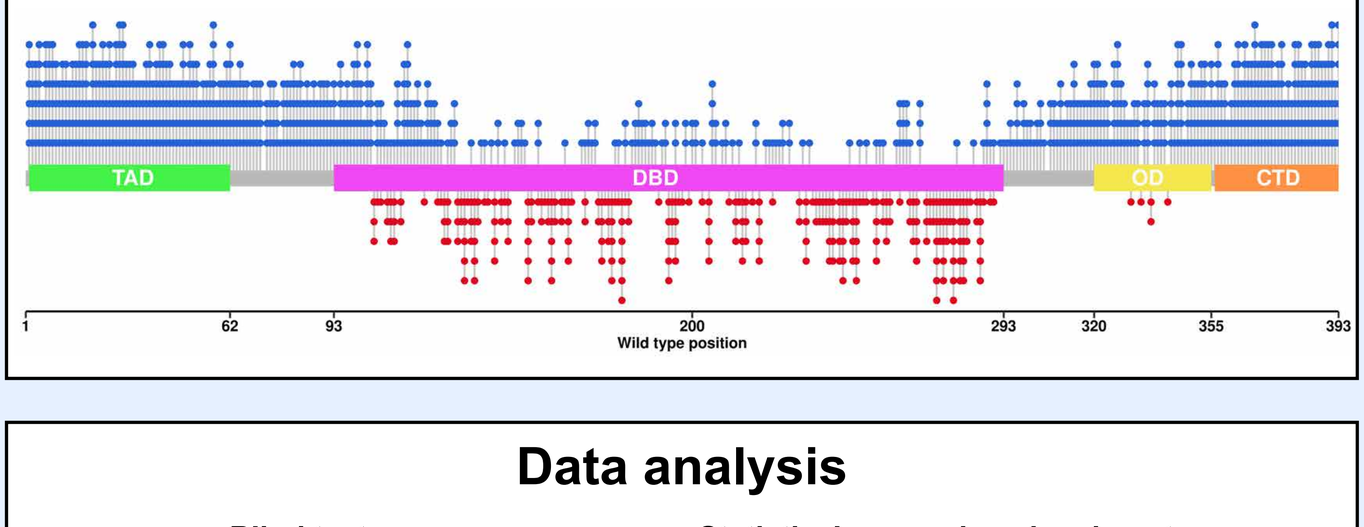
Feature generation



Model development and evaluation



Data distribution



Data analysis

Blind test

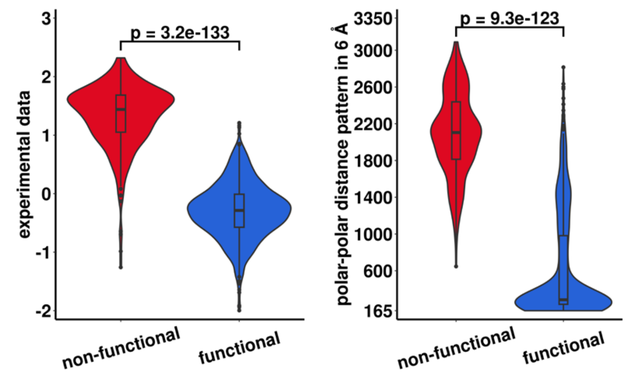
Method	MCC	Recall	Precision
our Model	*0.88*	*0.99*	*0.96*
TP53_PROF	**0.88**	**0.91**	**0.90**
Envision	0.60	0.95	0.89
VARITY	0.72	0.91	0.96
SIFT	0.46	0.59	0.98
PolyPhen2	0.42	0.54	0.98

Clinical validation

Method	MCC	Recall	Precision
our Model	*0.83*	*1.00*	*0.93*
TP53_PROF	**0.83**	**1.00**	**0.93**
Envision	0.58	0.96	0.88
VARITY	0.78	0.98	0.93
SIFT	0.58	0.75	0.98
PolyPhen2	0.47	0.64	0.97

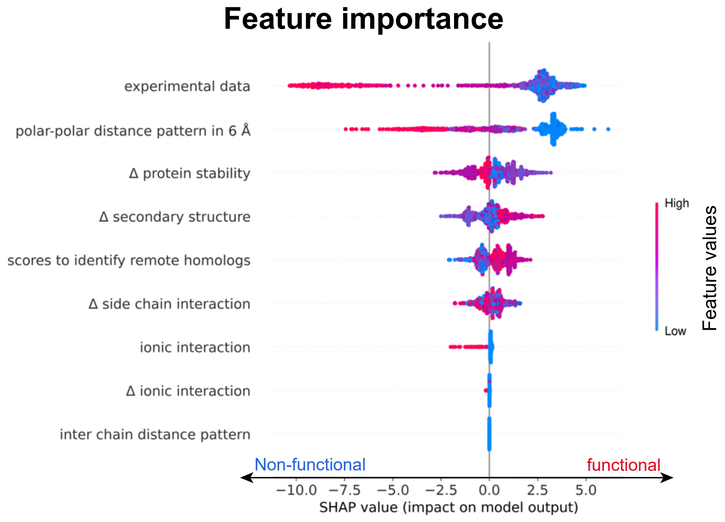
* MCC: Matthew's Correlation Coefficient
* Recall: True positive rate
* Precision:1 - False discovery rate

Statistical comparison by phenotypes

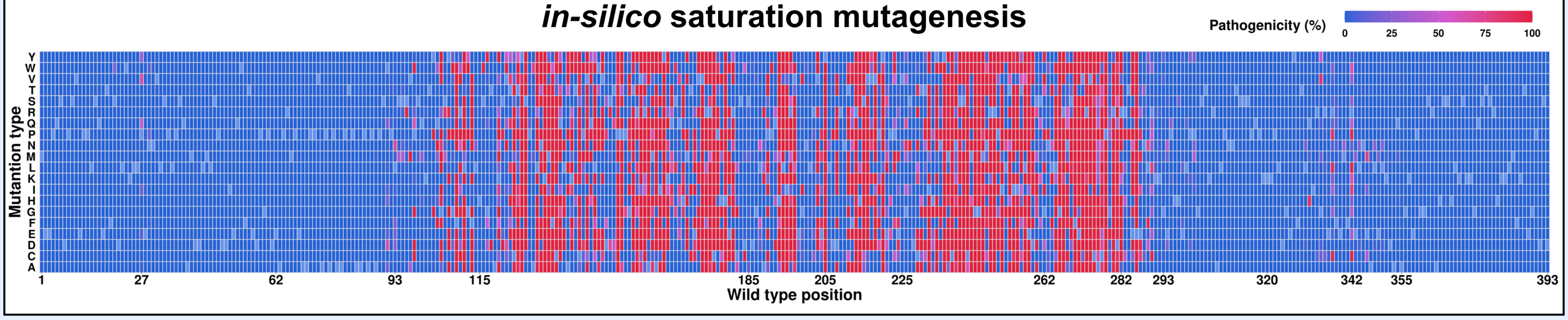


* Wilcoxon signed-rank test

Feature importance



in-silico saturation mutagenesis



Conclusion

- Our structure-based model **accurately characterises** the oncogenic effects of **all possible missense mutations** in p53, with a comparable performance to state-of-the-art methods.
- Feature interpretation reveals that intact p53 function is strongly reliant on **experimental residue activity**, **the number of polar residues within 6 Å**, and **the change of protein stability upon mutation**.
- This work offers **clinical diagnostic utility**, which is crucial for patient monitoring, and the development of personalised cancer treatment.