# Identifying the molecular drivers of pathogenic aldehyde dehydrogenase missense mutations in cancer and non-cancer diseases

**Qisheng Pan**[1,2], Dana Jessen-Howard[1], David B. Ascher[1,2]

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

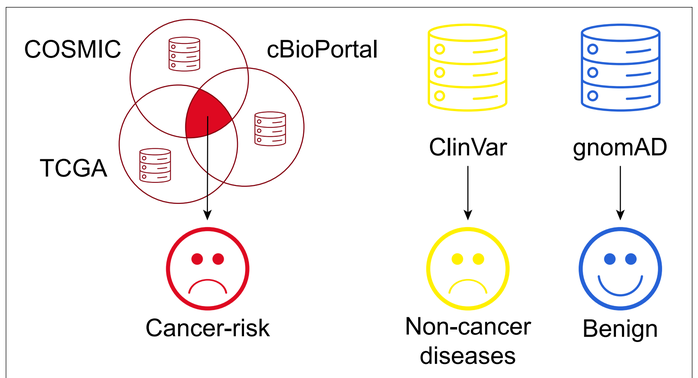
[2]*Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne Victoria 3004, Australia*

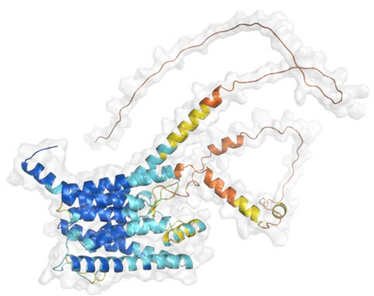
## BACKGROUND

- Aldehyde dehydrogenases (ALDHs) are part of an enzyme superfamily that interacts with endogenous and exogenous aldehyde metabolism.
- **Disruptions** on the activity of ALDHs, however, could result in the accumulation of cytotoxic aldehydes, which have been linked with a wide range of diseases, including both **cancers** as well as **neurological and developmental disorders**.
- We have therefore employed computational biophysical measurements and machine learning to identify potential molecular drivers of pathogenic ALDH missense mutations.

## METHODOLOGY

### Data curation

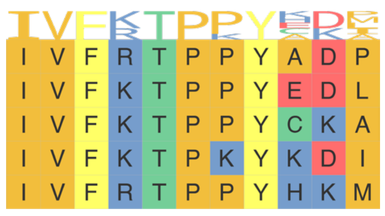


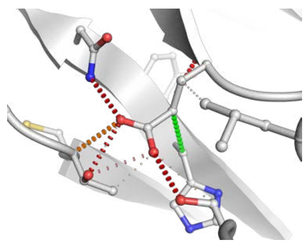
**AlphaFold models**

### Feature generation

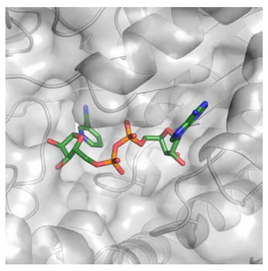
**Conservation**



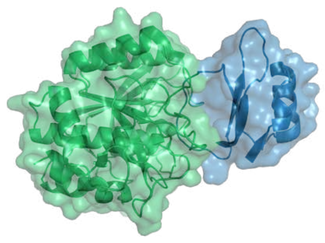
**Residue contacts**



**Protein-ligand binding**

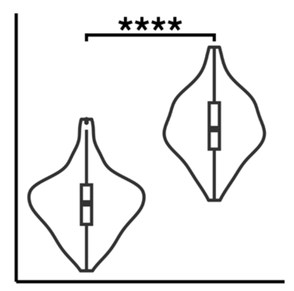


**Protein-protein interaction**

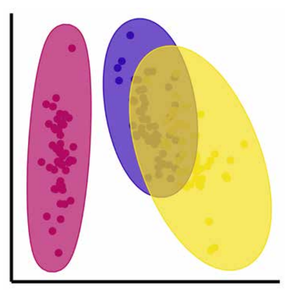


### Data analysis

**Qualitative comparison**




**Dimensionality reduction**



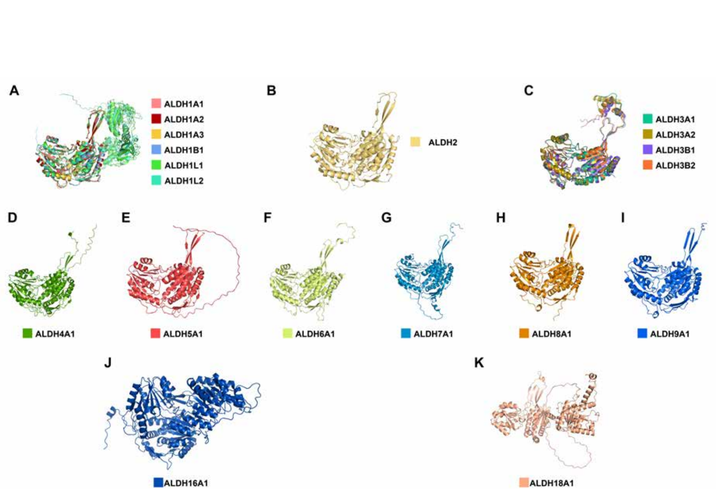
Feature 1
Feature 2
.....
Feature *n*-1
Feature *n*

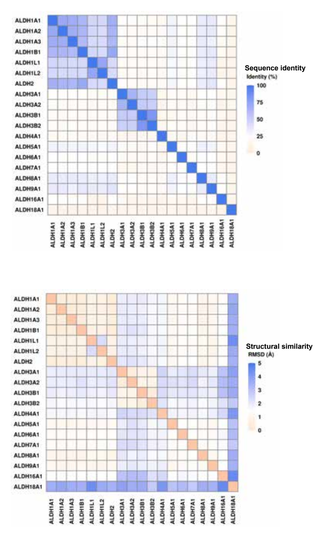
**Machine learning**

Pathogenic
Benign

## RESULTS

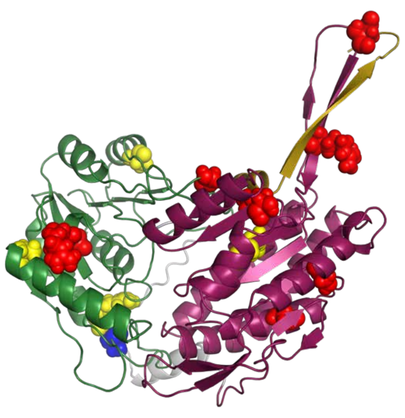
### ► Human ALDH family




- Human ALDH family has 19 isoenzymes shared **similar characterstics on sequences and structures**.

### ► Distribution of mutations on protein structures

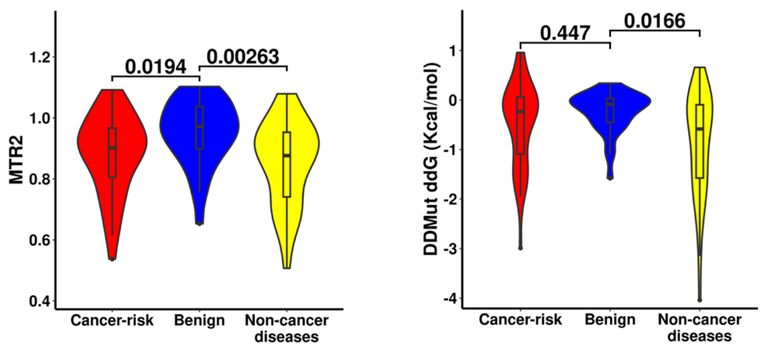


ALDH1A2

NAD+ binding regions
Substrates binding regions
Protein-protein interaction regions
Cancer-risk
Non-cancer diseases
Benign

- There is no obvious pattern of how missense variants with different phenotypes distributed across different functional regions of ALDHs.

### ► Qualitative analysis

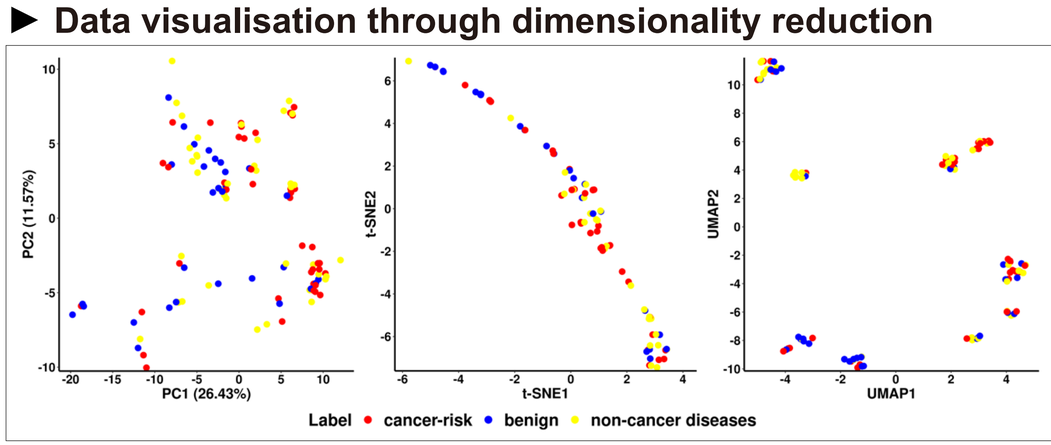



Conservation

Protein stability

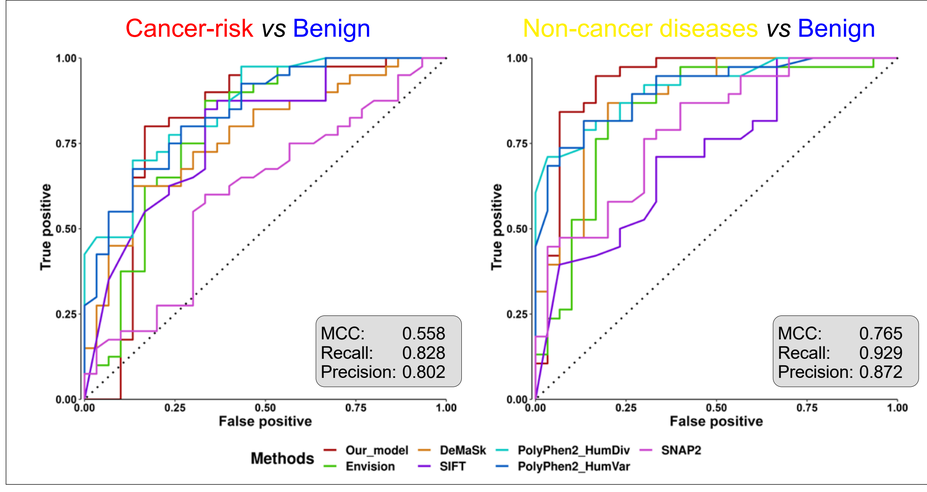
- Statistical tests indicate that **change of protein stability** and **conservation at the mutation site** are two risk factors to diseases.

### ► Data visualisation through dimensionality reduction




- Pathogenic mutations tend to be grouped into **a same cluster**.

### ► Supervised machine learning




- Our models achieved **robust performance** on identifying cancer-risk and non-cancer pathogenic mutations from benign ones.

## CONCLUSION

- Our work provides new biological insights into the pathogenic risk factors in human ALDHs using computational methods, suggesting that **change of protein stability** and **conservation** as key drivers to diseases.
- The molecular drivers found in this work could serve as a resource for further understanding on the **functions** of ALDHs and **the corresponding phenotypes**.

**CONTACT**

Qisheng Pan

QishengPan

*qisheng.pan@uq.net.au*