InSiGHT 2019, Auckland

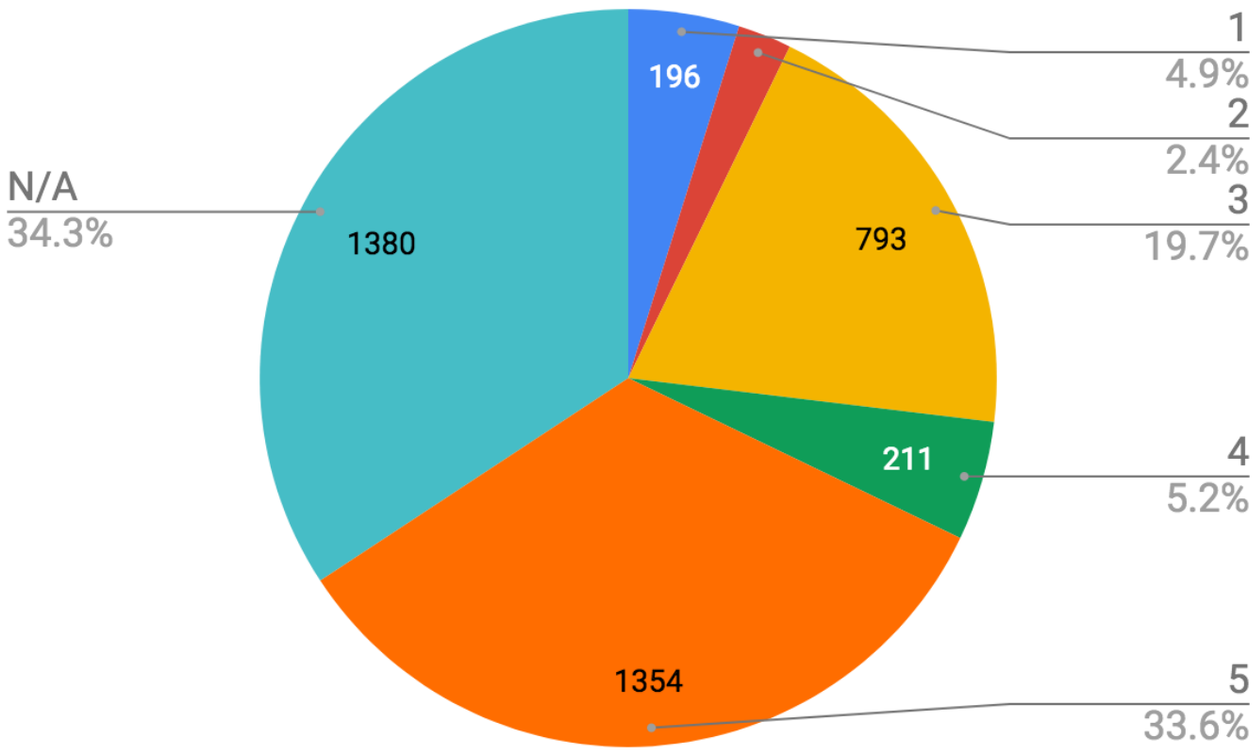
Predictors of class

Using protein structure and function information to predict and understand mismatch repair variant pathogenicity

Bernard Pope
Lead Bioinformatician, Cancer and Clinical Genomics
Victorian Health and Medical Research Fellow
Melbourne Bioinformatics
The University of Melbourne, Australia

The InSiGHT database

% classifications for *MSH2, MSH6, MLH1, PMS2*



Class	Meaning
1	Not pathogenic
2	Likely not pathogenic
3	Uncertain
4	Likely Pathogenic
5	Pathogenic
N/A	Unclassified

- Class 3 and unclassified variants pose challenges for variant interpretation.
- 63% of Class 3 variants in *MSH2, MSH6, MLH1, PMS2* are missense.

Aims

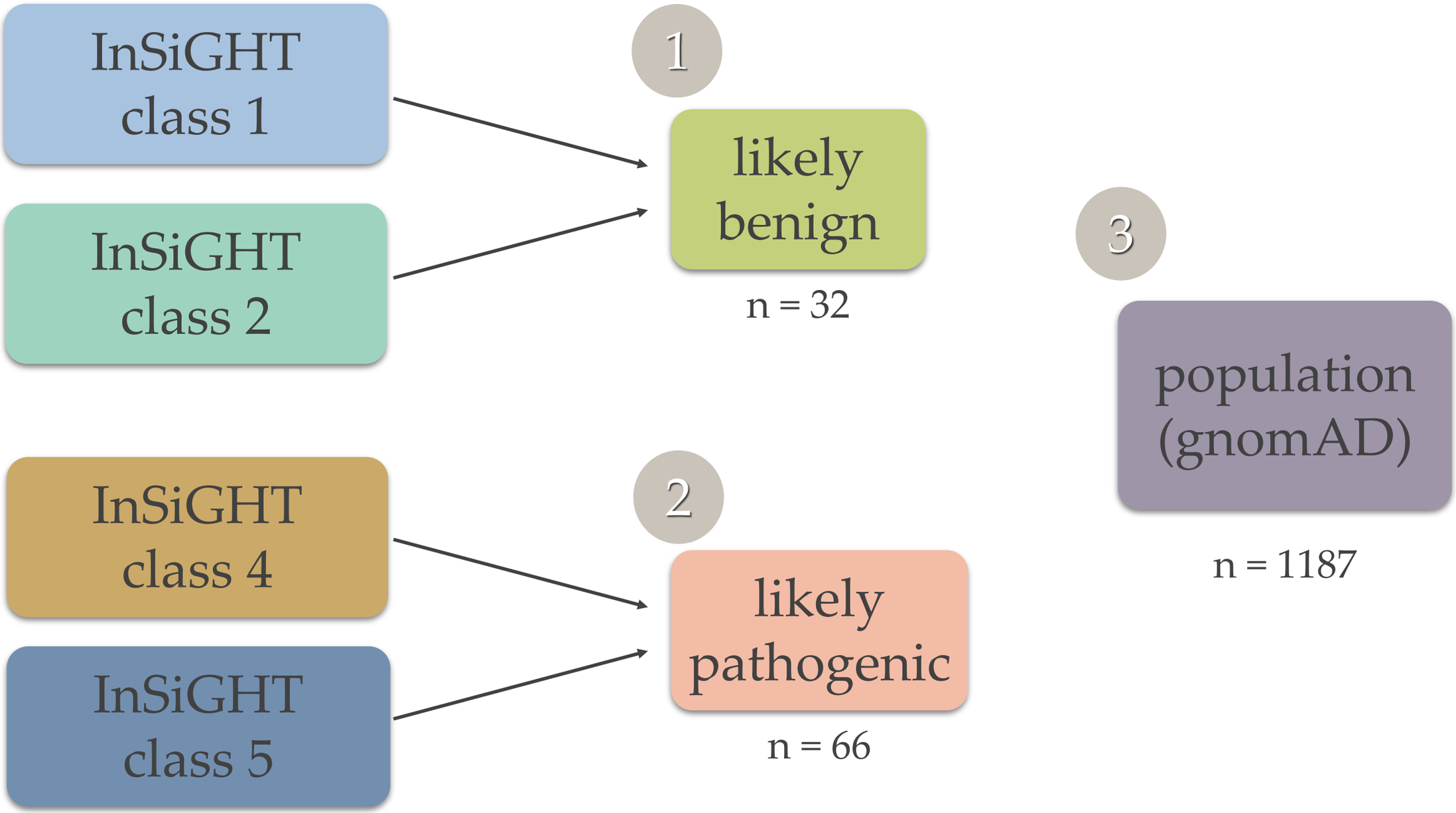
1. Predict important protein structural and functional features of missense MMR variants in *MSH2* and *MSH6*.

2. Train a machine-learning classifier on these features to distinguish pathogenic, benign and population variants.

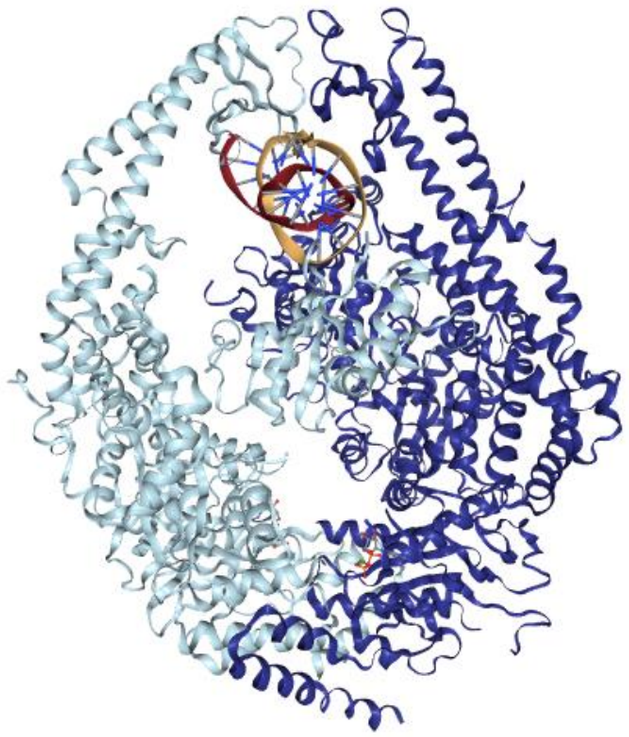
3. Apply the classifier to all *MSH2* and *MSH6* missense variants in InSiGHT.

Methods

Missense variants from *MSH2* and *MSH6* were placed into 3 independent groups



Methods



variant

mCSM suite

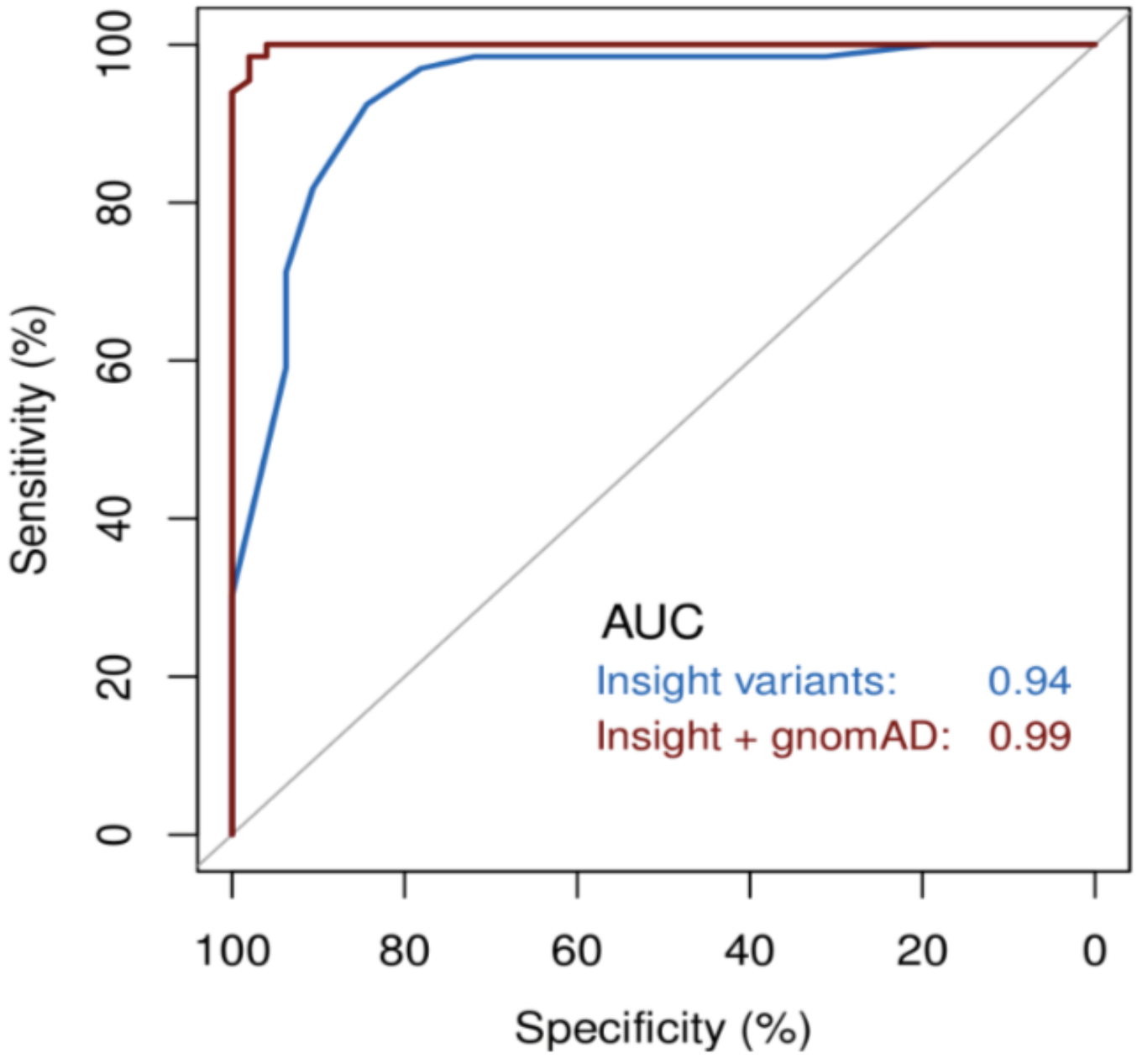
For each variant in the 3 groups:

- Map to the crystal structure for human MutS*α* (*MSH2* + *MSH6*).
- Predict protein structural and functional effects using the mCSM software suite.

Predicted Effects
protein stability
protein-protein binding
protein-DNA binding
protein-ligand binding
mutation tolerance
protein dynamics

Mutation Cutoff Scanning Matrix (mCSM)
http://biosig.unimelb.edu.au/mcsm/

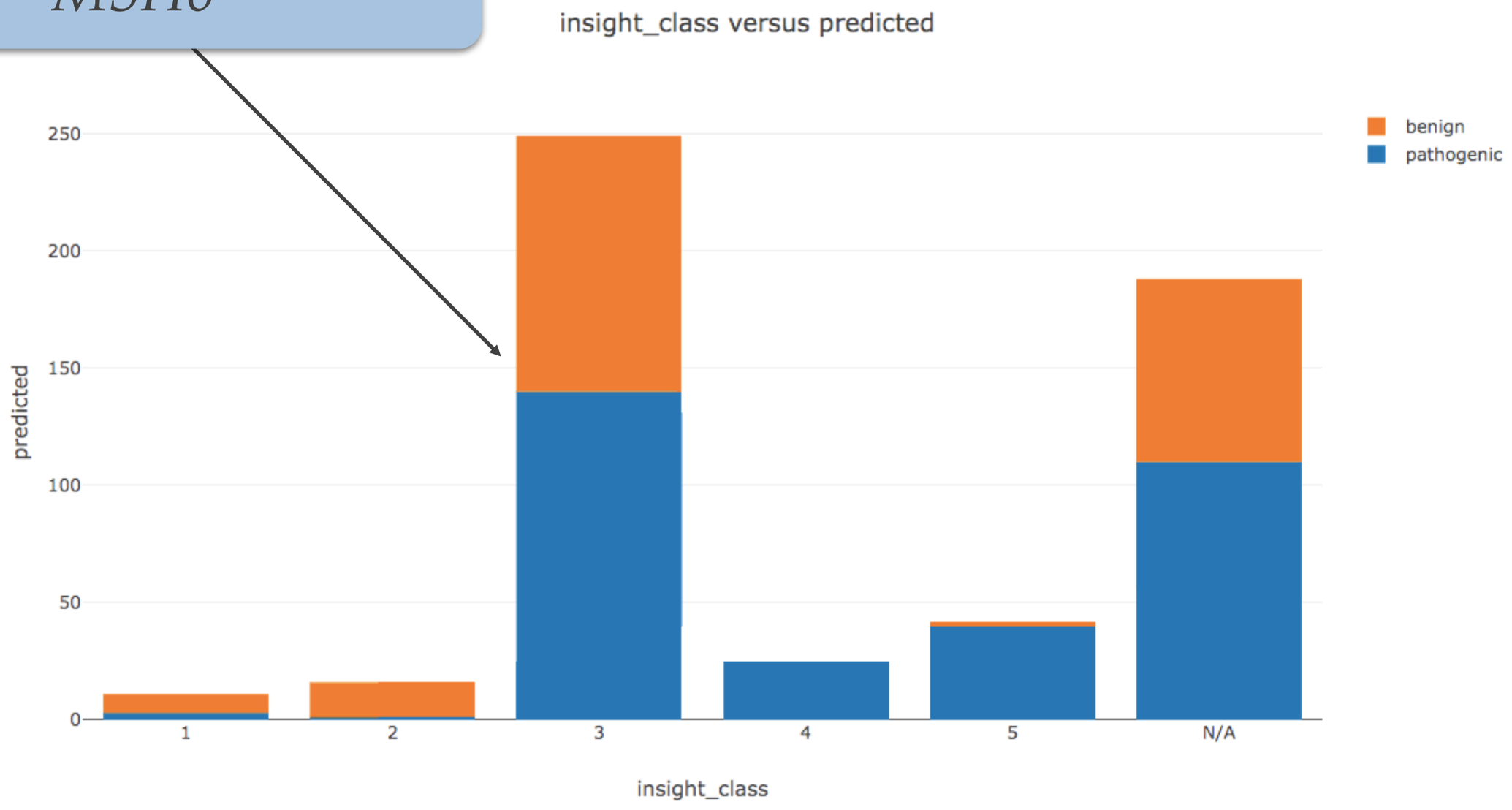
Results



- We trained a random forest classifier to distinguish between each of the 3 groups.
- Predictive performance tested using 10-fold cross validation.
  - Red = likely pathogenic *vs* population.
  - Blue = likely pathogenic vs likely benign.

Results

Re-classification of Class 3 variants in *MSH2* and *MSH6*

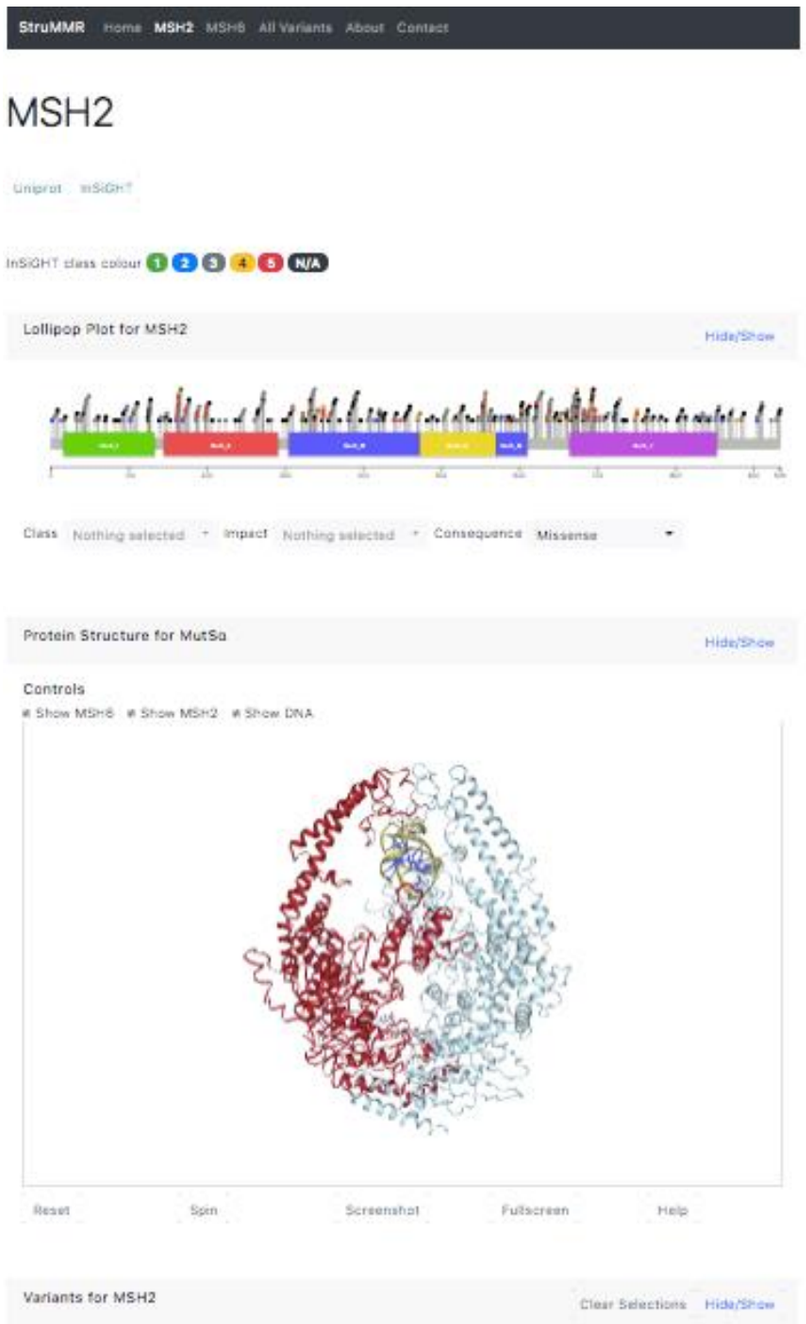


Conclusion

- Results are provided on the new interactive StruMMR website:

  http://biosig.unimelb.edu.au/strummr/

- Future work is to expand the technique to *MLH1* and *PMS2*.



Acknowledgements

Colorectal Oncogenomics Group, UoM

- Daniel Buchanan
- Mark Clendenning
- Eric Joo
- Harindra Jayasekara
- Peter Georgeson
- Romy Walker
- Susan Preston
- Sharelle Joseland

Melbourne Bioinformatics, UoM

- Khalid Mahmood

Structural Biology and Bioinformatics, UoM

- David Ascher
- Douglas Pires
- Michael Silk
- Carlos Rodrigues

On behalf of the InSiGHT Variant Interpretation Committee

- John-Paul Plazzer
- Elke Holinski-Feder
- Maurizio Genuardi
- Finlay Macrae

Extra Results

- PATH variants were situated in protein regions with lower tolerance to missense mutations compared to POP variants ($p = 0.011$)
- PATH variants were predicted to greater disrupt protein function than BEN variants ($p < 10^{-11}$) and POP variants ($p < 10^{-16}$).
- PATH variants were predicted to have a larger effect on protein stability compared to both BEN variants ($p < 10^{-6}$) and POP variants ($p < 10^{-6}$).
- Likely-pathogenic variants were also clustered closer to the ATP and Mg2+ binding sites than BEN variants ($p = 0.05$) and POP variants ($p < 0.0025$).