Investigating the predictive power of machine learning algorithms for antigenic novelty of influenza H3 viruses

Lucy Greenwood, MMath MSc

Supervisor: Ricardo Aguas

Nuffield Department of Medicine, University of Oxford

Contact: lucy.greenwood@reuben.ox.ac.uk

Engineering and Physical Sciences Research Council

General overview

HI assay data

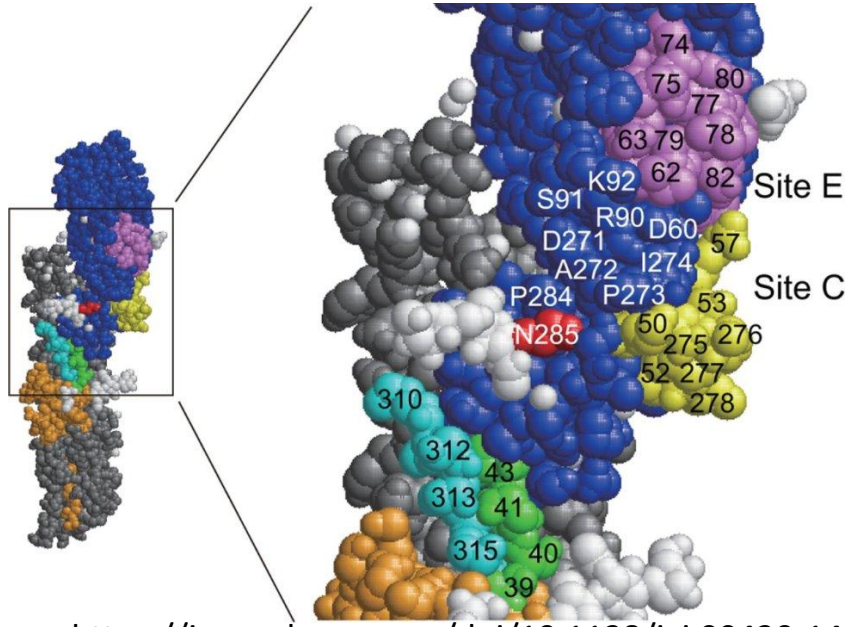
Genetic sequences

Amino acid physicochemical properties

Feature engineering

Genotype to phenotype and antigenic cartography mapping

Genotype to phenotype mapping 1 year ahead



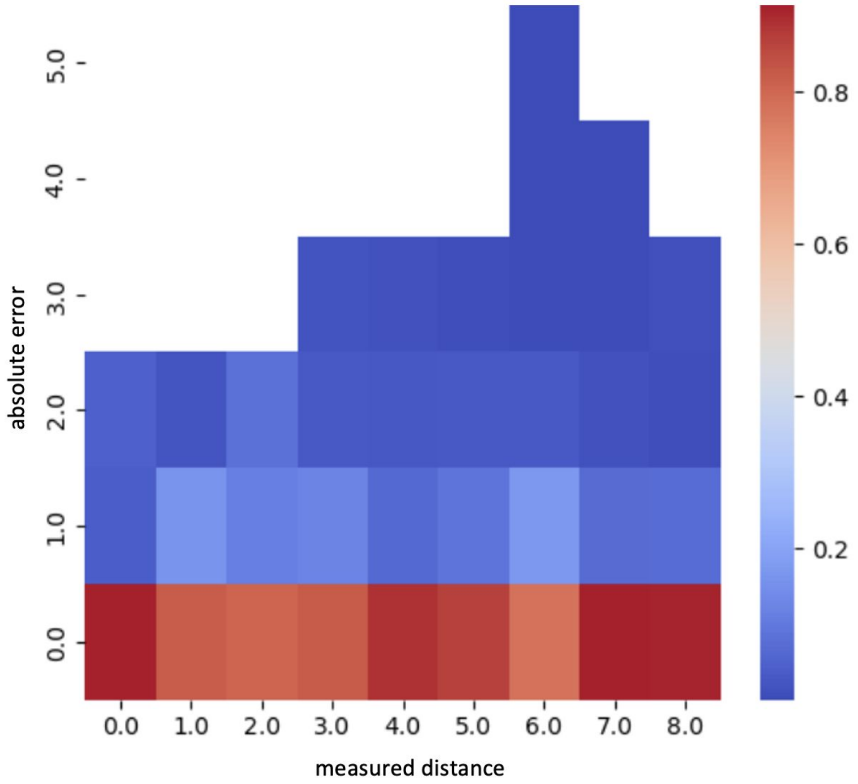
https://journals.asm.org/doi/10.1128/jvi.00420-14

- Property differences
- 3D smoothing methods
  - distance/proximity weight matrices
  - relative solvent accessibility (RSA)
- Epitope weighting
- Positional difference ratios
- Protein language model embeddings

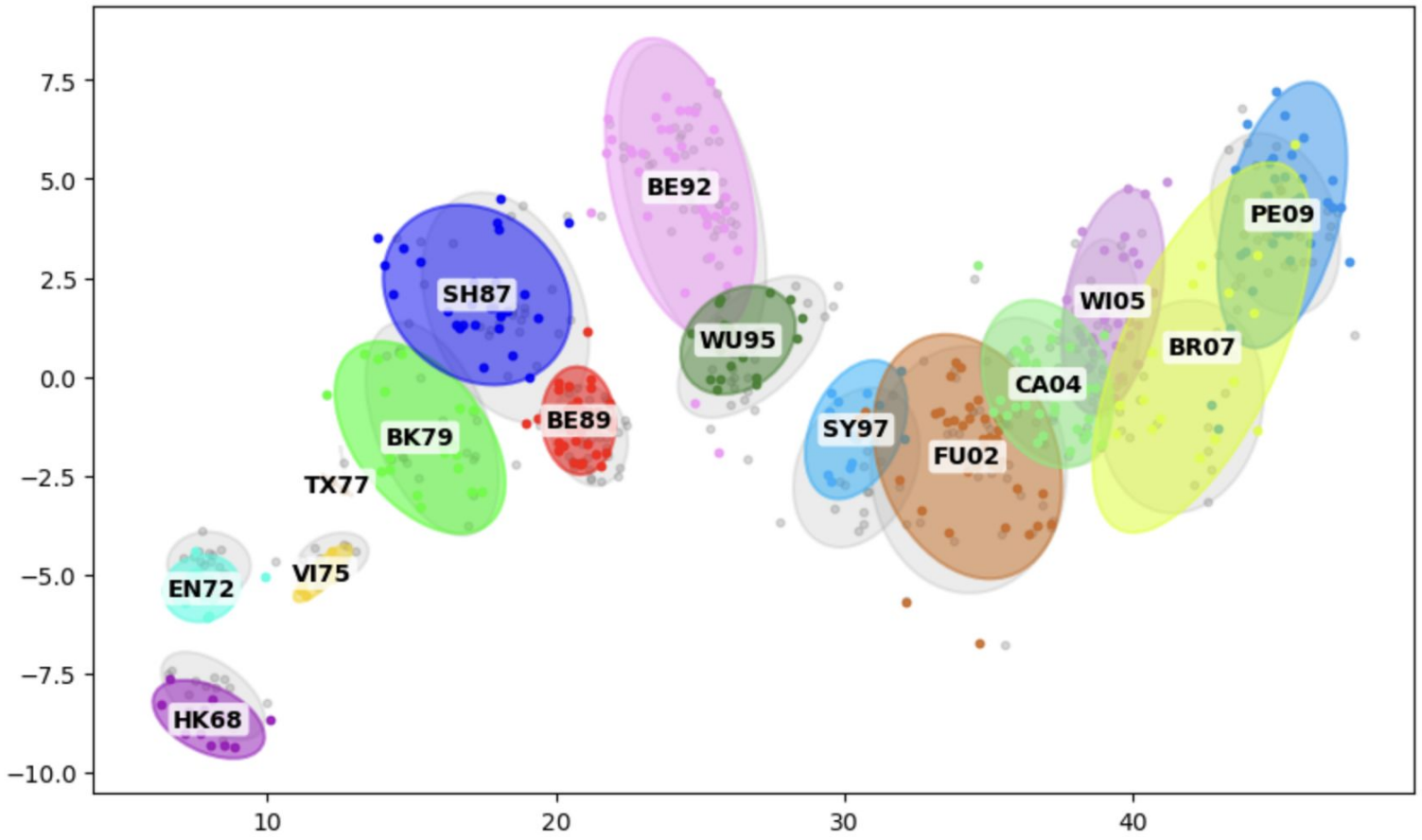
Engineering and Physical Sciences Research Council

HEALTH DATA SCIENCE
BIG DATA INSTITUTE
UNIVERSITY OF OXFORD

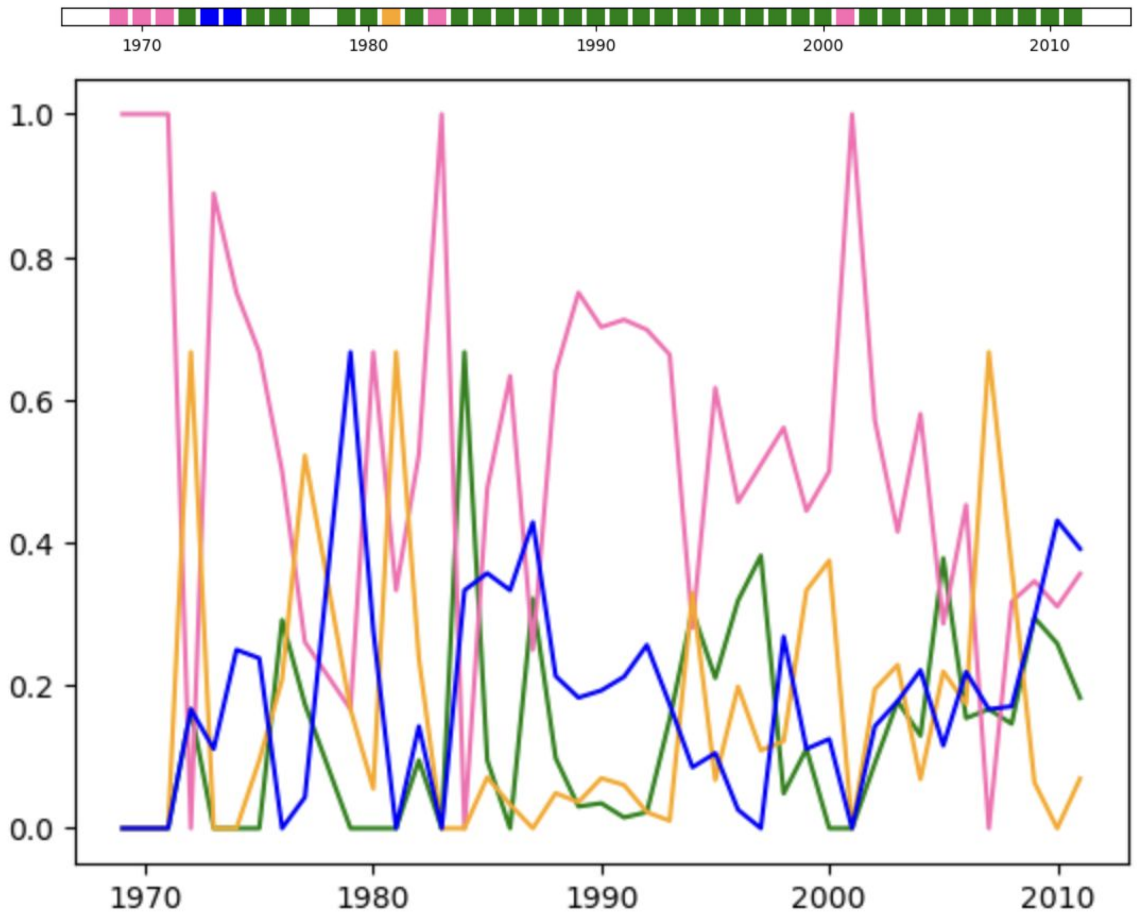
Genotype to phenotype mapping – GP mapping



95.6% within 1 antigenic unit
85.8% exact predictions



GP map – real time implementation



50 positions, context-driven, stochastic gradient descent

Vaccine change threshold set at 4 antigenic units.

Aim: predict when a change in vaccine strain might be needed

PREDICTED DISTANCE

MEASURED DISTANCE

No change predicted or required	Change predicted but not required
Change required but not predicted	Change predicted and required

Engineering and Physical Sciences Research Council

HEALTH DATA SCIENCE
BIG DATA INSTITUTE
UNIVERSITY OF OXFORD

Conclusions and next steps

Conclusions

- A GP map informed on physiochemical property differences can predict antigenic distances between any two viral sequences with a very high degree of certainty.
  - 95.6% within 1 antigenic unit
- Blinded one year ahead predictions indicate good match between measured and predicted distances across contiguous antigenic clusters
- Simulation of real-time implementation as an antigenic novelty detection tool is in the process of being updated

Next steps

- Evolution model and structure identification
- Incorporation into transmission model

lucy.greenwood@reuben.ox.ac.uk