Utilising machine learning to inform *Neisseria gonorrhoeae* molecular assay diagnostic targets

Authors & Affiliations: Verich A[1&2], Ramarao-Milne P[2], Reguant R[2], Alharbi B[3], Trembizki E[3], Applegate L T[1], Bauer D[2]

1 - Surveillance and Evaluation Research Program, The Kirby Institute University of New South Wales, New South Wales, Sydney, Australia;
2 - Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, New South Wales, Sydney, Australia;
3 - The University of Queensland Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Queensland, Brisbane, Australia

Aims & Background

- **This pilot study aims to utilise a novel machine learning (ML) platform, VariantSpark (VS)[1], to investigate genetic variations contributing to *Neisseria gonorrhoeae* (NG) antimicrobial resistance (AMR).**
- NG is the second-most prevalent sexually-transmitted bacterial pathogen. It is listed as a WHO 'high-priority' resistance organism, due to its current high AMR distribution and historical ability to develop AMR[2].
- Applying ML to genomic data can improve identification of variants important for surveillance and diagnostics.

Methods

- Downloaded NG sequence and resistance data for 314 patients from a New Zealand publication[3].
- Processed Illumina reads to generate variant call files (VCFs) with NZ_AP023069.1 as reference.
- Input VCFs into VS for random forest analysis and Hail2.0 for Firth's logistic regression analysis as a comparison.
- Utilised tetracycline AMR data because of its equal distribution of resistant and susceptible samples.

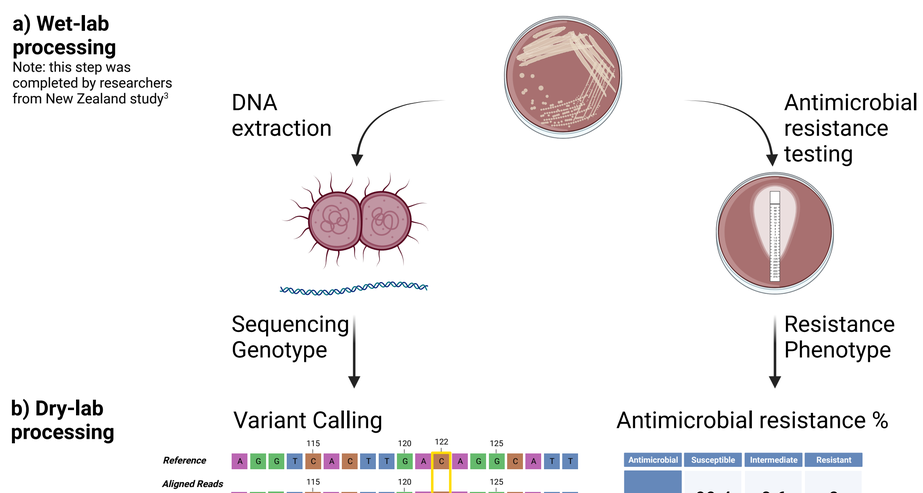

Figure 1. Pipeline for preparing data and executing VariantSpark random forest prediction.
Example numbers shown, splits count per node and variable weight not covered.

Results & Discussion

- VS (Figure 2) provided **high importance scores for known tetracycline allele variants**[4]: *rpsJ* (tetracycline target) and *porB* (membrane pore). However, mtrCDE complex genes (5 genes for membrane efflux-pump) had low scores (<0.05).
- Interestingly, **allele variants in two chaperone proteins** (protein stabilisers), *hscA* and *hemW,* also had high importance scores. While these are not currently known to cause AMR, other chaperones have been linked to NG resistance[5].
- Compared to logistic regression (Figure 3), VS had less 'noise', however both assigned high values to 'hypothetical' proteins.

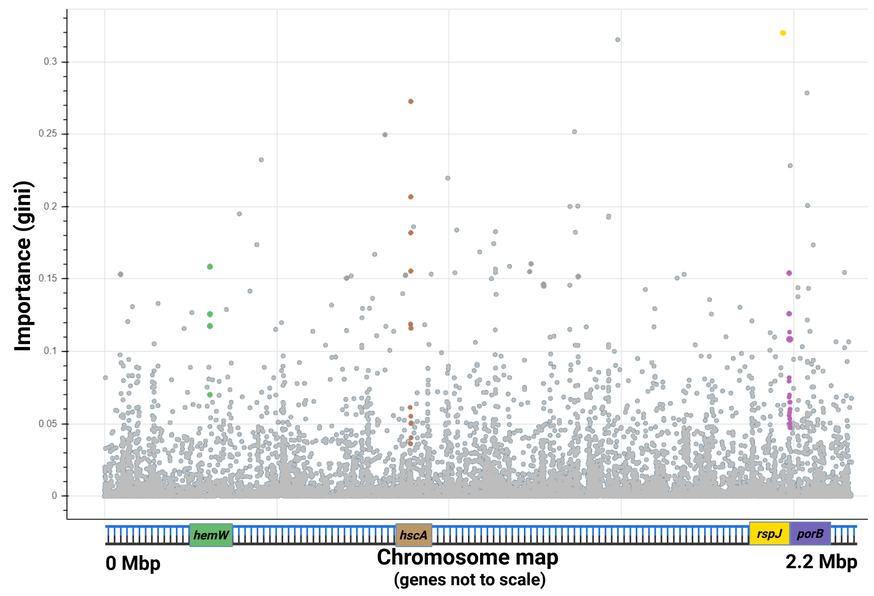

Figure 2. VariantSpark random forest importance score assignments to allele variants.
Each dot signifies a unique allele variant.

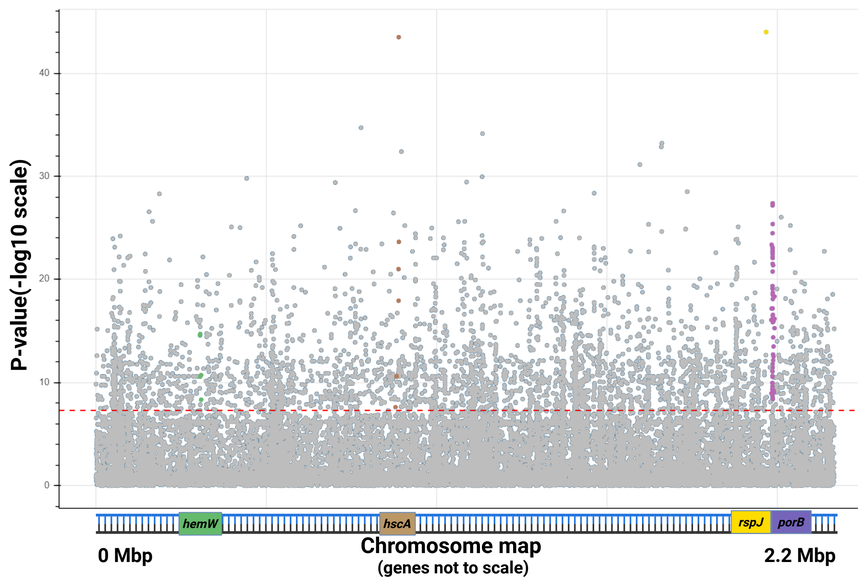

Figure 3. Hail2.0 logistic regression P-value assignments to allele variants.
Each dot signifies a unique allele variant.

In conclusion, machine learning **successfully determined** known and potential **novel resistance genes**.

Future Directions

- Integrate multiple, public NG datasets for VS analysis. Importantly, combining large datasets offers varied resistance profiles, thus allowing prediction for other antimicrobials.
- Incorporate patient data (e.g., gender) into VS analysis to examine these as potential confounding factors.

1 - Bayat A, et al. VariantSpark: Cloud-based machine learning for association study of complex phenotype and large-scale genomic data. GigaScience. 2020 Aug 1;9(8):giaa077
2 - Tacconelli E, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. The Lancet Infectious Diseases. 2018 Mar;18(3):318–27.
3 - Straub C, et al. The antimicrobial resistance landscape of Neisseria gonorrhoeae in New Zealand from November 2018 to March 2019 and the role of sexual orientation in transmission. Microb Genom. 2021 Nov;7(11).
4 - Mlynarczyk-Bonikowska B, et al. Molecular Mechanisms of Drug Resistance and Epidemiology of Multidrug-Resistant Variants of Neisseria gonorrhoeae. International Journal of Molecular Sciences. 2022 Jan;23(18):10499.
5 - Bodoev I, et al. Substitutions in SurA and BamA Lead to Reduced Susceptibility to Broad Range Antibiotics in Gonococci. Genes (Basel). 2021 Aug 25;12(9):1312.
ACKNOWLEDGEMENTS: This research is supported by an Australian Government Research Training (RTP) Scholarship and top-up by UNSW-CSIRO Industry PhD program by the Commonwealth Scientific and Industrial Research Organisation and SpeeDx Pty. Ltd. This work was supported by the Australian Research Council Industrial Transformation Research Hub to Combat Antimicrobial Resistance (IH190100021). Made with BioRender.

VariantSpark

Contact Details