Utilising machine-learning to inform Neisseria gonorrhoeae molecular assay diagnostic targets.

Authors:
Verich A[1&2], Ramarao-Milne P[2], Reguant Comellas R[2], Alharbi B[3], Trembizki E[3], Applegate TL[1], Bauer DC[2],

[1]Surveillance and Evaluation Research Program, The Kirby Institute UNSW Sydney, 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, The University of Queensland, Queensland, Brisbane, Australia

Background:
Global, rapid emergence of bacterial antimicrobial resistance (AMR) costs billions of USD and contributes to a death toll estimated to reach ~10 million by 2050. Molecular diagnostics characterising AMR profiles can guide effective treatment, reduce inappropriate antibiotic use, and minimise the emergence of resistance. Neisseria gonorrhoeae (NG) is a sexually transmitted bacteria with 'high-risk' for resistance, and propensity to rapidly evolve and acquire AMR mutations. This project aims to apply a novel machine-learning framework for phenotype (resistance) gene prediction, VariantSpark (VS), to NG genotypic and phenotypic datasets; aiming to inform the development of next generation molecular NG AMR and surveillance tools.

Methods:
VS, like traditional genome-wide association study frameworks, annotates additive effects of single genes on phenotype. Notably, it is also capable of characterising properties emerging only from system-based complex interactions rather than individual genes alone. VS is currently being piloted on high-quality, geographically local (N=314) NG whole-genome sequencing (WGS) dataset with equal distribution of tetracycline resistant and susceptible isolates; aiming to investigate emergent properties and validate current tetracycline AMR genes.

VS also excels by optimisation for WGS datasets allowing for emergent interactions analysis in datasets previously too computational demanding. Therefore, following the pilot-study, VS will be applied to large, multi-thousand global NG WGS datasets (e.g., Pathogenwatch) to investigate molecular AMR assay targets including complex resistance phenotypes for other NG antibiotics.

Conclusion:
Overall, as WGS datasets expand in quantity and complexity, machine-learning frameworks like VS provide an opportunity to extract valuable information to combat issues like AMR. These frameworks also have capacity for integration into `real-time` analysis pipelines; allowing for flagging of AMR genes and emergent properties as they appear. Thus, increased integration of machine-learning should occur to inform development of next generation molecular NG AMR diagnostic and surveillance tools; thus, minimising the impact of AMR.

Disclosure of Interest Statement:

There are no conflicts of interest to declare.