

(1) Submission ID#1527495

GeCoPred, a new genomics-based tool for 4CMenB vaccine coverage prediction

Author(s)

Margherita Bodini, n/a

Senior Data Scientist

GSK

Marco Podda, n/a

Visiting Scientist

University of Pisa

Simone Bonechi, n/a

Assistant Professor

University of Siena

Andrea Palladino, n/a

Senior data scientist

GSK

Laura Serino, n/a

Senior Scientist

GSK

Elena Mori, n/a

Senior Scientist

GSK

Alessia Biolchi, n/a

Researcher

GSK Vaccines

Corrado Priami, n/a

Professor

University of Pisa

Alina Sîrbu, n/a

Associate Professor

University of Pisa

Alessandro Muzzi, n/a

Associate Director, Head of Bacterial Computational Genomics Team

GSK Vaccines

Background

The multicomponent meningococcal serogroup B vaccine (4CMenB, Bexsero, GSK), released in 2013, induces immunity against a broad panel of *Neisseria meningitidis* strains. Before its introduction in new countries and for subsequent surveillance purposes, prediction of the vaccine's strain coverage is important to understand efficacy potential. Coverage prediction for *Neisseria meningitidis* strains is hindered by genetically distant new strains when the epidemiology varies by geography. We recently developed a method to classify genomes using their entire genomes based on machine learning. Here, we present an adaptation of this method for genomic coverage prediction (named GeCoPred) that can predict Bexsero coverage potential for every new isolate.

Aim/Methods

The GeCoPred tool was trained with historical correlates of protection data from the Meningococcal Antigen Typing System (MATS) and human Serum Bactericidal Antibody (hSBA) assays. A dataset of more than 1,400 *Neisseria meningitidis* genomes, for which correlate of protection data were available (MATS, hSBA, or both), was used.

Results

On an independent test set, our method had an 85% balanced accuracy when predicting MATS coverage, with positive and negative predictive values of 90% and 80%, respectively. We then tested the method on epidemiologically relevant panels from Argentina, Finland, France, Ireland, the United Kingdom, the United States, and a mixed European panel, and the results were close to the MATS predictions. Similarly, on an independent test set, our method had a 73% balanced accuracy when predicting hSBA coverage with positive and negative predictive values of 80% and 67%, respectively.

Conclusions

We developed a new tool for high-accuracy coverage prediction based on machine learning. GeCoPred is a versatile tool that is applicable more widely than methods based on antigen typing such as genetic MATS (gMATS) and the Meningococcal Deduced Vaccine Antigen Reactivity (MenDeVAR) index as it gives a prediction for every isolate rather than only for a subset of strains. The GeCoPred tool, in future versions, will incorporate additional laboratory data to improve its accuracy. Funding: GlaxoSmithKline Biologicals SA. Trademark: Bexsero is a trademark owned by or licensed to GSK Acknowledgements: Business & Decision Life Sciences platform on behalf of GSK (writer: Esther van de Vosse).