BSP 2023 Abstract

Project Title
Network Representation of Host and Pathogen Interactomes And Machine Learning For *Eimeria* Vaccine Development

Authors:
Roman Baptista[1], Virginia Marugan-Hernandez[1], Kaige Yang[2], Pedro Gomes[2], Damer Blake[1], Fiona Tomley[1], Laura Toni[2], Dong Xia[1]
[1] Pathobiology and Population Sciences, Royal Veterinary College, Hawkshead Lane, North Mymms AL9 7TA, UK.
[2] Department of Electronic and Electrical Engineering, University College London, London, UK.

Funding Bodies
BBSRC London Interdisciplinary Doctoral Programme
The Houghton Trust, Small Project Research Grant

Abstract

Eimeria tenella is an intracellular apicomplexan parasite which can infect the chicken with absolute host specificity, causing a haemorrhagic variant of enteric coccidiosis. This disease can cause diarrhoea, blood loss, malnutrition, and increased risk of secondary infection, with a deleterious effect on egg and meat production. The financial impact of *E. tenella* is estimated at ~£10.4 billion/year worldwide and is predicted to rise as poultry becomes the primary sustainable food source for a growing global human population. The established method of rotating anticoccidial drug regimens has led to widespread resistance to all existing chemoprophylactics, prompting consumer and legislative pressure in both the US and EU to move towards vaccine-based interventions. Identification of subunit vaccine candidates has focused on leveraging *Eimeria* proteins to generate effective immune responses, promising higher efficiency and lower cost-per-animal compared to current live-oocyst vaccines which require the full *in vivo* infective process to occur.

This project seeks to generate an *in silico*, graph-based network (GBN) representation of the host-pathogen protein interactome to streamline candidate prioritisation in vaccine development. Mapped with RNAseq gene expression profiles generated from an infection time course experiment, interactome GBNs have been curated that comprise 7,430 unique proteins and ~200,000 connections for the host, and 1,675 proteins with ~40,000 connections for the pathogen.

This network is being exploited in a variety of ways. Firstly, as a platform for the visualisation and analysis of multi-omics data, including longitudinal expression changes that provides an understanding of the localisation of differentially expressed genes, as well as the foundation for a graph-based deep learning dynamic model of gene expression profiles across time. Secondly, it provides network-specific measures such as node degree and centrality which determine changes in connectivity and importance of groups of nodes, pinpointing candidates for disruption with drug or vaccine-based interventions. Thirdly, embedding of the complete network can be used to highlight nodes which experience temporal changes in local neighbourhood structure, representing pathways or cellular compartments normally affected by the infective process for consideration in treatment design.

A mixture of binding prediction algorithms along with curated interaction data are currently being employed to produce a novel, joint host-pathogen interactome for *E. tenella* infections. The joint network will allow us to highlight interfacing hotspots between organisms with vaccine target potential, supplemented by the VACCEED machine learning framework for target prioritisation based on a reverse vaccinology approach. The aforementioned deep learning dynamic model will be developed in tandem with this joint network, to be used for *in silico* simulation of potential interventions and later validated using downstream wet-lab techniques. The application of this pipeline will be tested and fine-tuned with RNASeq datasets collected from in-house chicken vaccine trials.