LIC Report for the JDRC

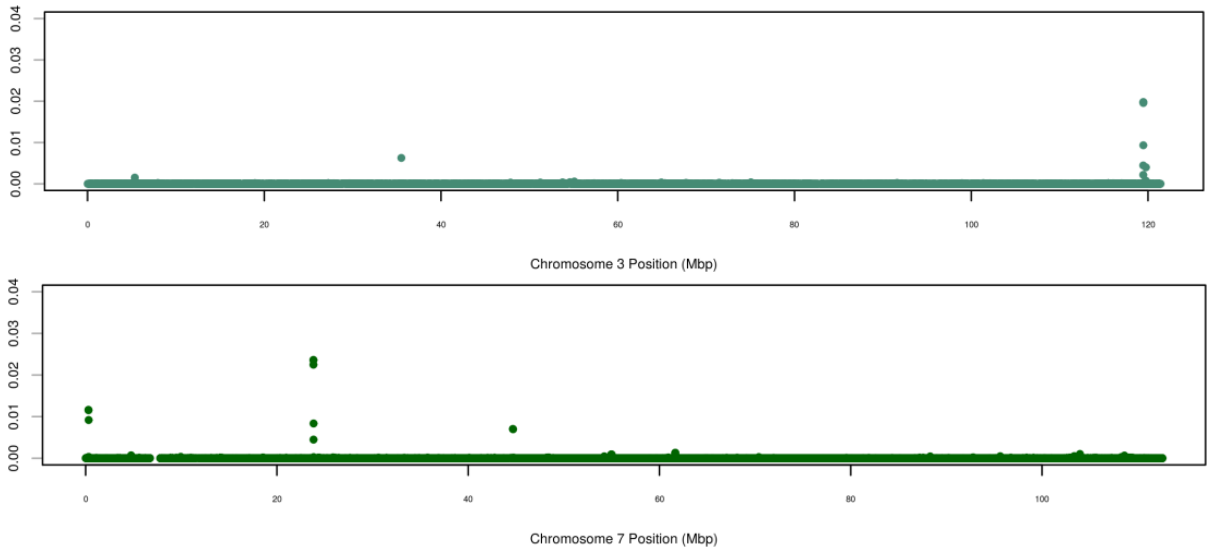
Highlights

- Genomic analysis is complete.
- Genome-wide analyses have identified a number of markers for susceptibility to Johne's disease.
- A predictive model has been developed that classifies an animal's genomic profile as Johne's positive with reasonable accuracy.
- Scientific paper for potential publication is in preparation

Summary of results

Genome-wide analysis has detected multiple signals across the genome associated with Johne's disease. A number of the signal locations have been identified as corresponding to immune-related genes. Of note are the signals on Chromosome 3 and 7 that correspond to the receptor (CSF2RA) and ligand (CSF2) for Colony Stimulating Factor respectively. Previous studies have linked both these genes to infection with *M. avium* subsp. *paratuberculosis* (MAP) or *Mycobacterium bovis*.

Figure 1. Bayes effects for single nucleotide polymorphisms (SNP) on Chromosomes 3 and 7, from a genome-wide association study, calculated using a multi-SNP Bayes B analysis.



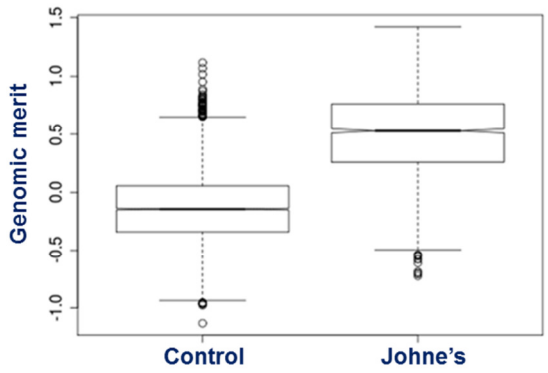
These signals suggest that the study is targeting biologically relevant genetic structures but although deeper investigation of these markers may help further illuminate the biological pathways contributing to Johne's disease, there is arguably more immediate benefit to be gained by using the data to develop a predictive test that could be applied to an animal's genomic profile to predict the susceptibility of the animal (and its progeny) to MAP infection.

A tenfold cross validation study was performed to determine the accuracy with which the data could be used to predict Johne's status based on an animal's genomic profile. The software package GenSel (Fernando and Garrick, Iowa State University), was used to fit a Bayes B model using 1 megabase windows across the genome. The model estimated the genomic merit for the different combinations of SNP in each window for each of the 10 training populations and then used those

estimates to predict the actual Johne's status of the animals in the corresponding 10 test populations.

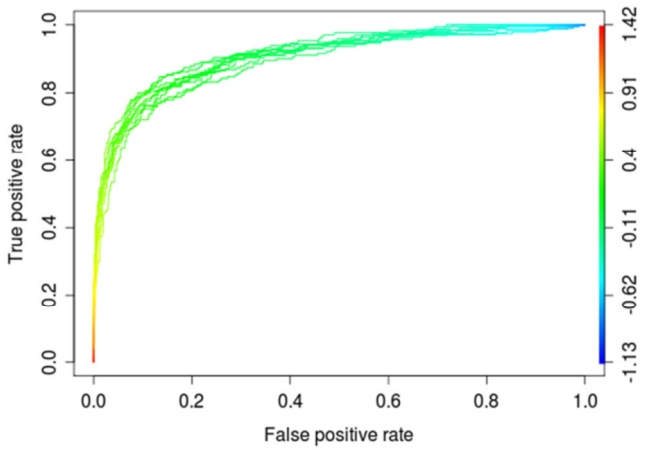
Figure 2 shows a notched boxplot of the genomic merit for Johne's susceptibility predicted by the model for the Control animals, and the Johne's positive animals. The figure suggests that the median genomic merit of the Johne's positive animals is significantly higher than that of the general population, but there is considerable overlap of the two distributions. The general population was likely to include some animals positive for Johne's. The model may well be correctly predicting that the outliers at the top of the control group are in fact Johne's positive, or at least susceptible to MAP infection.

Figure 2. Notched boxplot of the predicted genomic merit of Control and Johne's postive animals for Johne's susceptibility.



The choice of a genomic merit threshold to separate animals into predicted Johne's and control groups, involves trading off the likelihood of correctly identifying Johne's positive animals (true positives) with the likelihood of incorrectly identifying control animals as Johne's positive (false positives). This relationship is shown in the Receiver Operating Characteristics (ROC) curves for the tenfold cross-validation (Figure 3). The area under the ROC curve is equivalent to the probability that the predictive model will correctly rank a randomly chosen positive animal higher than a randomly chosen negative animal. The average area under the ROC curves is 0.906.

Figure 3. Receiver Operating Characteristics curve for tenfold cross-validation. Each line represents one of the 10 training-testing datasets.



Further thought and discussion is required to determine how best to utilise this predictive test to provide the best result for the NZ dairy industry.