**CBCB**
**SEMINAR**
**03/18/2019**

3:30 PM
**DELAWARE BIOTECHNOLOGY INSTITUTE**
Room 102

**bioinformatics.udel.edu**

# BIOINFORMATICS SEMINAR

## JUNIPER LAKE

*PhD Graduate Student,*
**UNIVERSITY OF DELAWARE**

### GENOME-WIDE ASSOCIATION STUDY AND BLOOD METABOLITE PROFILING OF WOODEN BREAST DISEASE

ooden Breast is a myopathy of fast growing, commercial broilers causing myofiber necrosis, vasculitis (phlebitis), myoregeneration, and fibrosis with extensive fibrillar collagen deposition in the superficial part of the pectoralis major, presenting clinically as palpably firm breast muscle. Rapid growth, high feed efficiency, and large breast muscle yield are predisposing factors, although the etiology of the disease is still poorly understood. This talk will outline the methods that will be used for genome-wide association and blood metabolomics profiling of the disease, as well as discuss potential genetic pathways that may regulate the disease based on the existing literature.

## MADOLYN MACDONALD

*PhD Student, Department of Computer and Information Sciences*
**UNIVERSITY OF DELAWARE**

### DEVELOPMENT OF A NOVEL TOOL FOR GENOME ASSEMBLY QUALITY EVALUATION AND ITS APPLICATION TO ESTABLISH A REFERENCE ASSEMBLY FOR CHO CELLS

Genome assemblies are used widely as references to guide both wet lab and bioinformatics studies. Accordingly, scientists should select the highest quality assembly for their organism of interest (if multiple are available) to be this reference and be aware of any limitations posed by the level of completeness and accuracy of the assembly. I will be presenting on a novel tool, called EvalDNA (Evaluation of De Novo Assemblies), which assists in the model development for quality scoring of genome assemblies and that does not require an existing reference genome for accuracy assessment. EvalDNA calculates a list of quality metrics from an assembled sequence and applies a model created from supervised machine-learning methods to integrate various metrics into a comprehensive quality score. A well-tested, accurate model for scoring mammalian genome sequences is provided as part of EvalDNA. This random forest regression model evaluates an assembled sequence based on continuity, completeness, and accuracy, and was able to explain 86% of the variation in reference-based quality scores within the testing data.

EvalDNA with this mammalian model was used to evaluate several assemblies of the Chinese hamster (CH) genome to help establish a better reference genome for CH and Chinese hamster ovary (CHO) cells. The resulting EvalDNA scores also enabled the quality comparison of the selected CH reference genome to the reference assemblies of other organisms at both the full assembly and chromosome levels.

**College of Engineering**
**CENTER FOR BIOINFORMATICS & COMPUTATIONAL BIOLOGY**