

## Introduction to Machine Learning for Genetics and Genomics Short Course Outline

- 1) Basic Concepts
  - a. Classification / Regression
  - b. Training / Test / Validation
  - c. Pre-process (Center / Scale / Data Transformation)
  - d. Model and Tuning Parameters
  - e. Variable Importance
- 2) Training Strategies
  - a. Cross-validation
  - b. Leave-one-out cross validation
  - c. Bootstrapping (Bagging)
- 3) Algorithms
  - a. Regression (Logistic and Linear)
  - b. Decision Trees
  - c. Support Vector Machines
  - d. Neural Networks
  - e. Naïve Bayes
  - f. KNN
- 4) Caret Package in R
- 5) Performance Evaluation
  - a. ROC Curve
  - b. Metrics for Regression
  - c. Metrics for Classification
- 6) Applications to Genomics and Genetics