

**530 Raw Samples  
(412 OC, 118 Non-OC)**

**Data preparation and processing**

**Classification models**

GSE12172

GSE14407

GSE9899

GSE37648

GSE18521

GSE18521

GSE10971

RMA Normalization

Transcript Quality Filtering

Batch Effect Removal

Select Feature Subsetting

Train, n=372

Test, n=158

Model Fitting, 5-fold cv

1. Linear Discriminant Analysis (LDA)
2. K Nearest Neighbors (KNN)
3. Support Vector Machines (SVM)
4. Decision Tree (CART)
5. Random Forest (RF)

Model Selection and Evaluation