Evaluate testing data (multi-class) - xgboost

John Doe

2019-01-30

Contents

Labels: 0: Basal 1: LumA 2: LumB 3: Her2

```
## user input
project_home <- "~/EVE/examples"
project_name <- "xgboost_multi_outCV_test"
```

0. Load Data

```
## 300 of samples were used
```

```
## 100 of full features
```

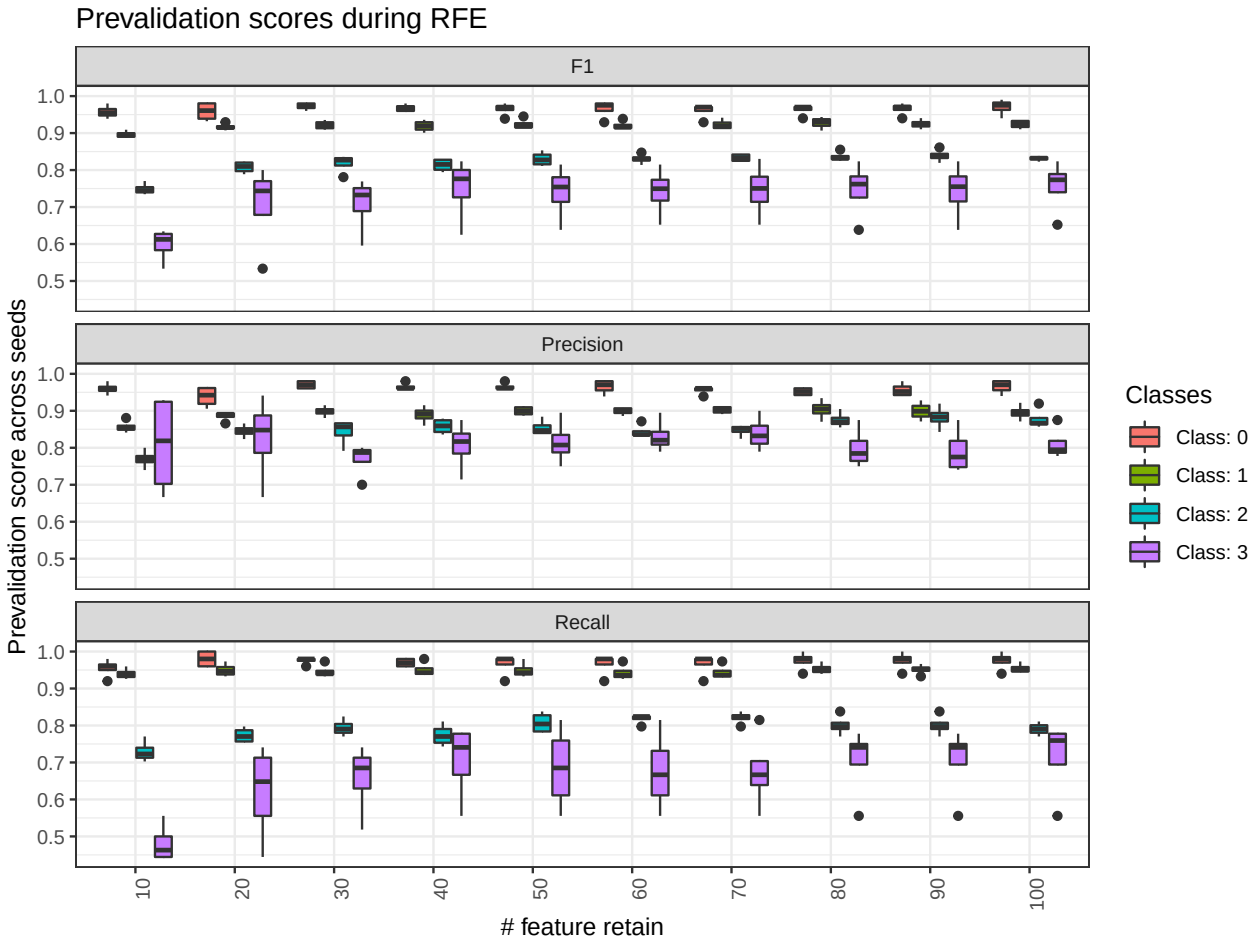
```
## 4 runs, each run contains 3 CVs.
```

```
## Labels:
```

```
##
##   0   1   2   3
##  50 149  74  27
```

1. Scores

1.1 Scores per Class

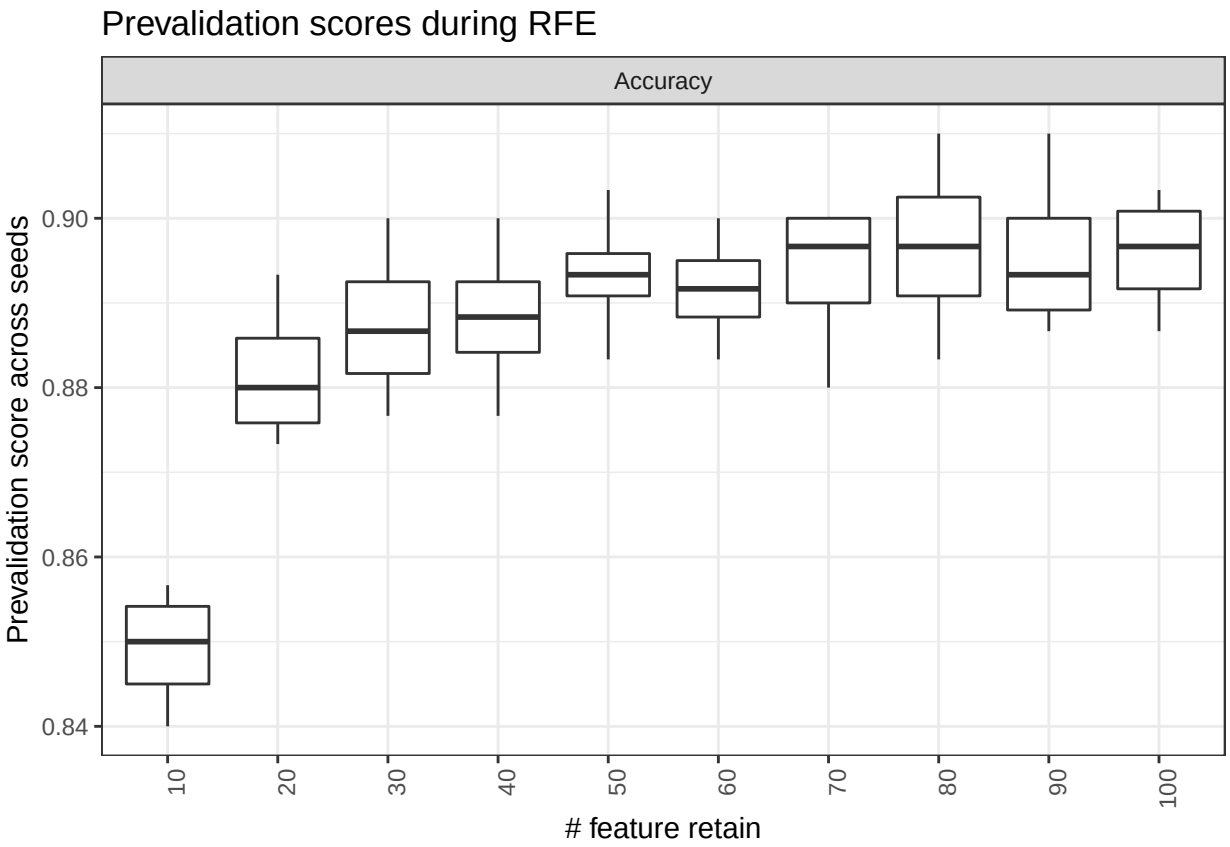


Confusion Matrix

```
## confusion matrix at feature size = 100
## sum across 4 seeds
```

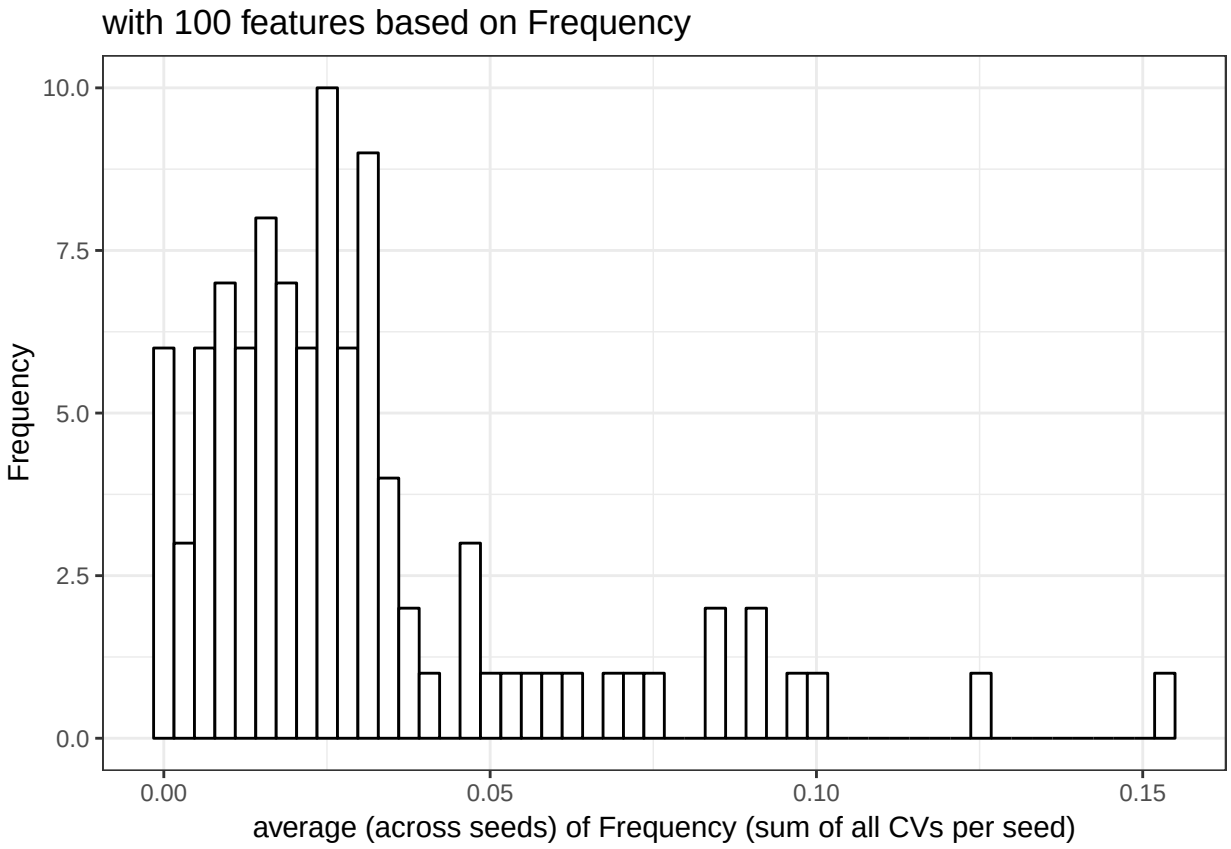
```
##            Reference
## Prediction   0   1   2   3
##          0 195   5   0   2
##          1   2 569  49  16
##          2   0  20 234  13
##          3   3   2  13  77
```

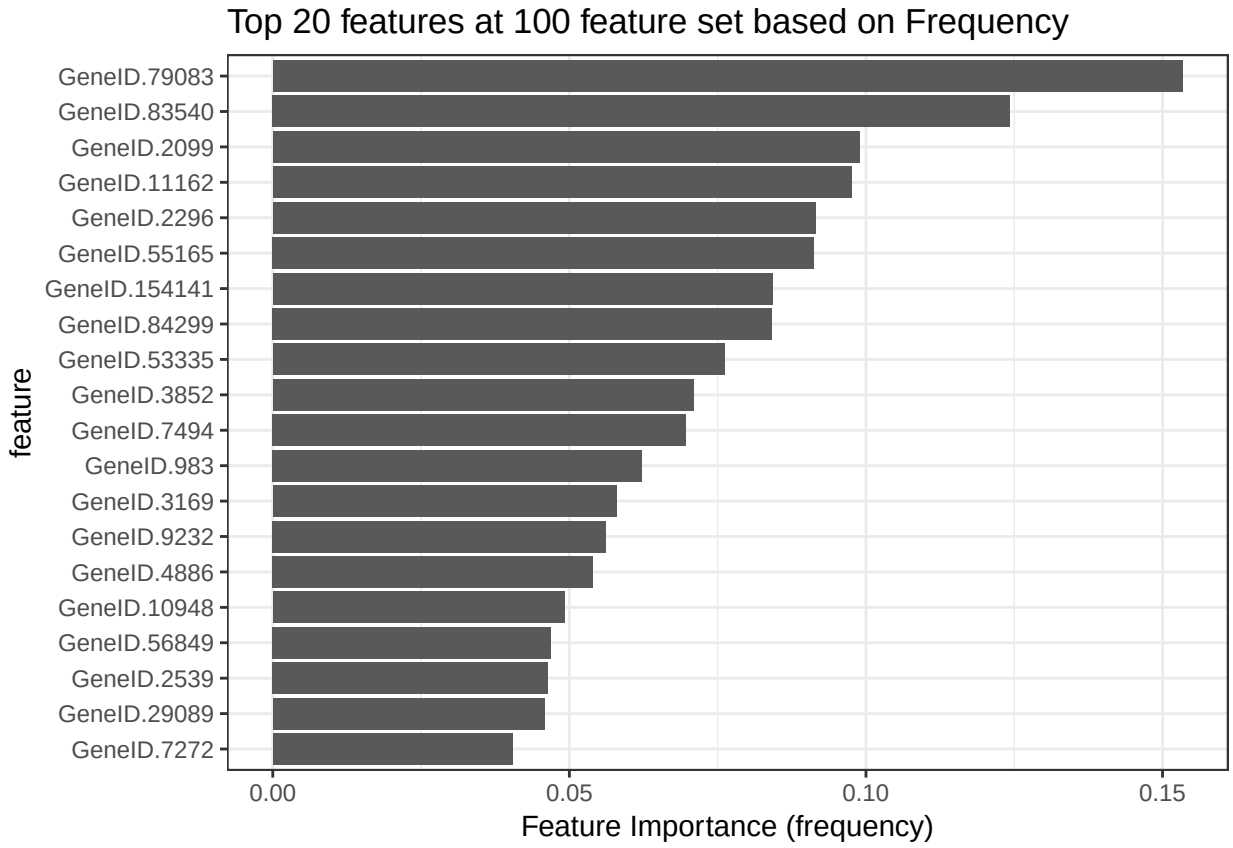
1.2 Average score

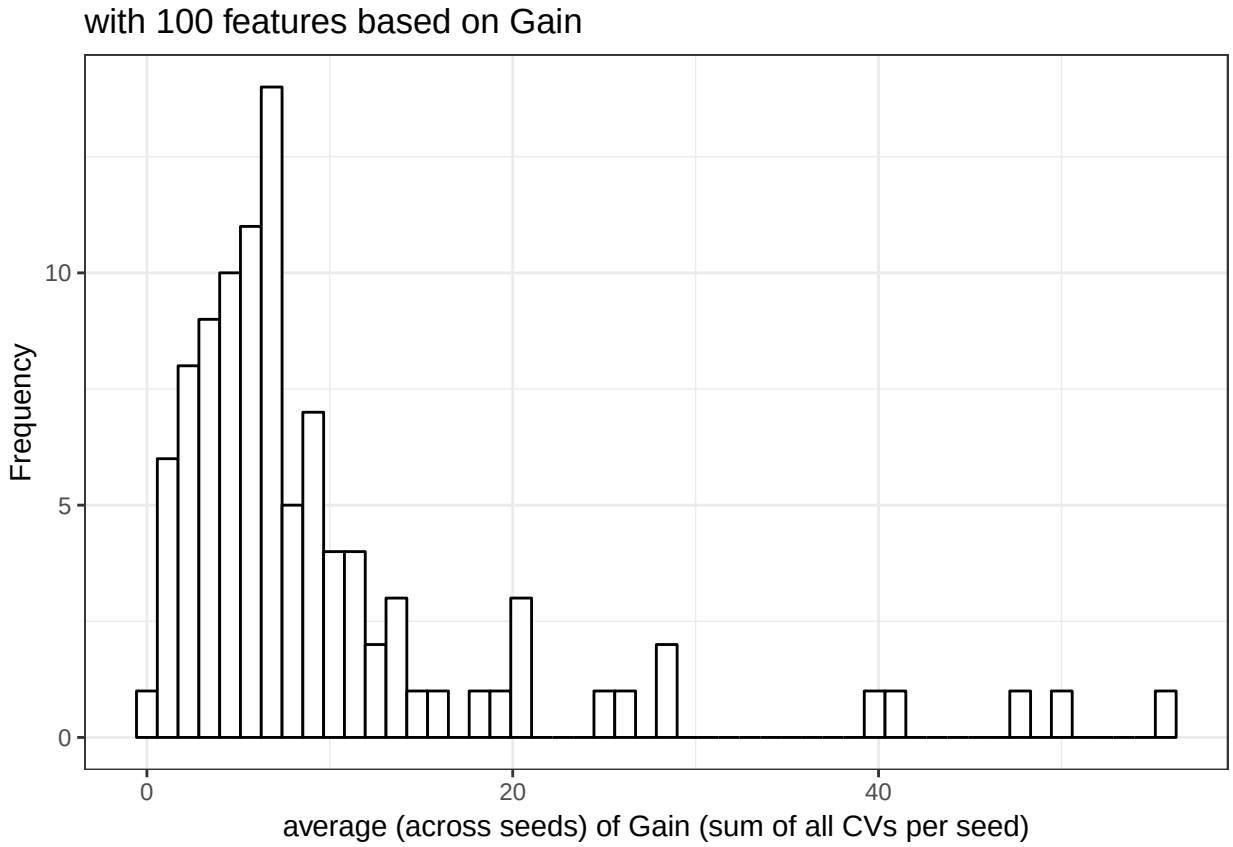


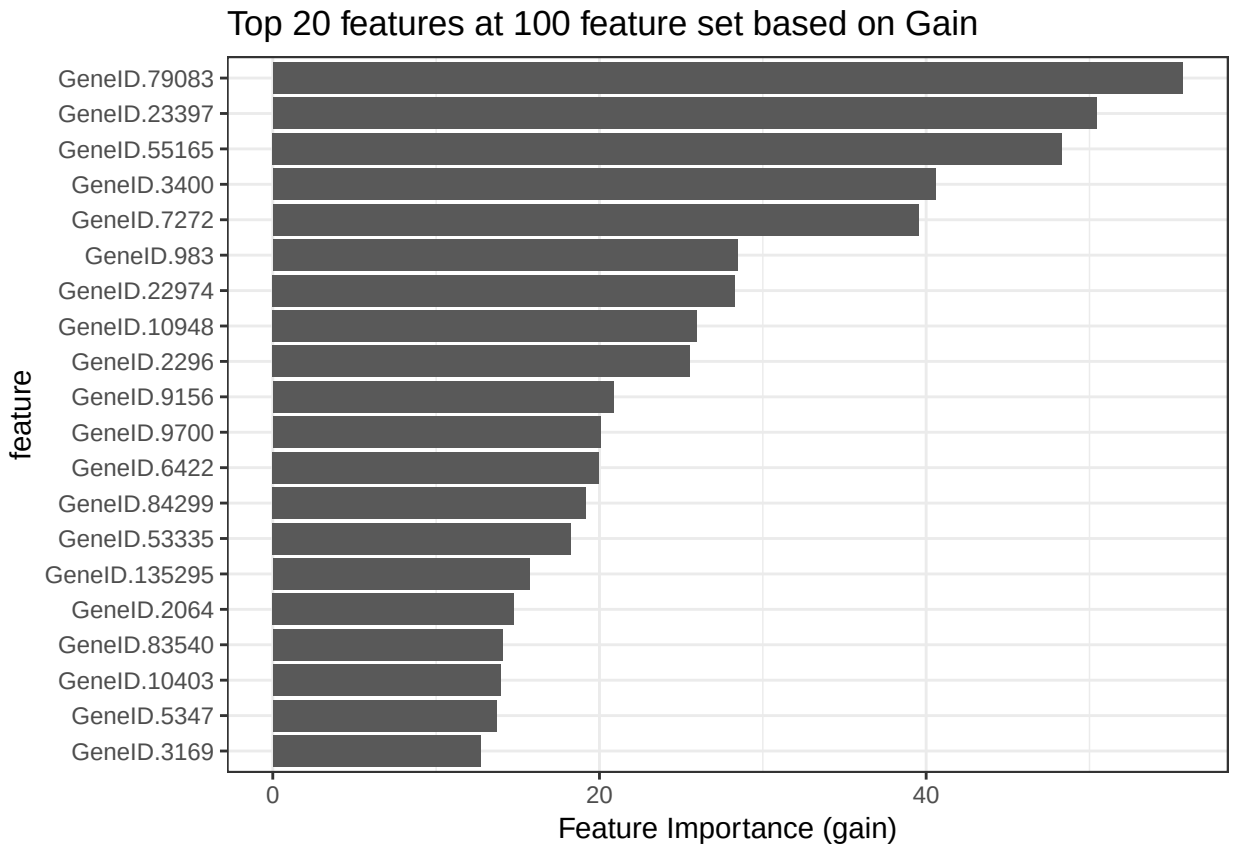
metrics	size.max	median.max	size.min	median.min
Accuracy	70	0.897	10	0.850
F1	100	0.873	10	0.803
Precision	70	0.887	10	0.850
Recall	80	0.864	10	0.778
average scor	es across c	lasses		

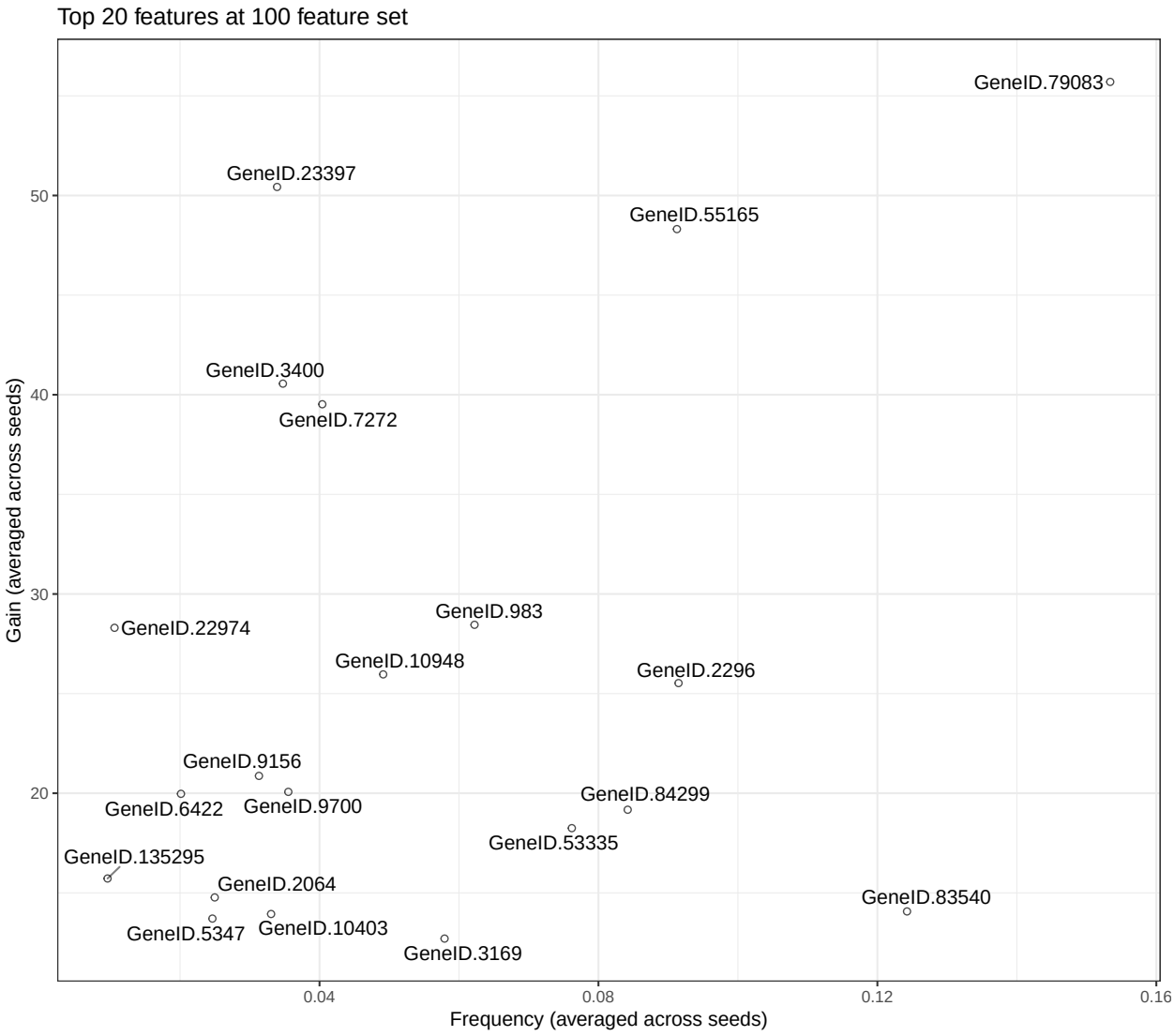
2. Important Features

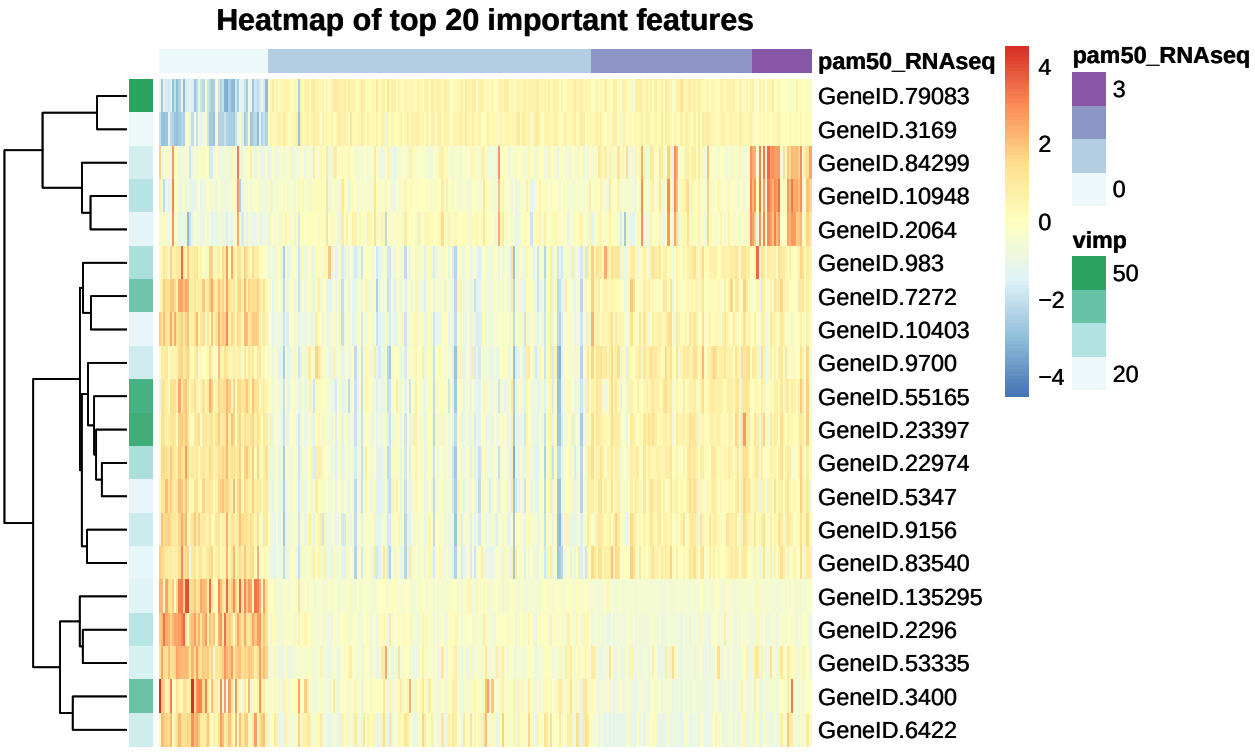



Top 20 features at 100 feature set based on Frequency
GeneID.79083
GeneID.83540
GeneID.2099
GeneID.11162
GeneID.2296
GeneID.55165
GeneID.154141
GeneID.84299
GeneID.53335
GeneID.3852
GeneID.7494
GeneID.983
GeneID.3169
GeneID.9232
GeneID.4886
GeneID.10948
GeneID.56849
GeneID.2539
GeneID.29089
GeneID.7272
feature
0.00
0.05
0.10
0.15
Feature Importance (frequency)


with 100 features based on Gain
Frequency
10
5
0
0
20
40
average (across seeds) of Gain (sum of all CVs per seed)


Top 20 features at 100 feature set based on Gain
feature
GeneID.79083
GeneID.23397
GeneID.55165
GeneID.3400
GeneID.7272
GeneID.983
GeneID.22974
GeneID.10948
GeneID.2296
GeneID.9156
GeneID.9700
GeneID.6422
GeneID.84299
GeneID.53335
GeneID.135295
GeneID.2064
GeneID.83540
GeneID.10403
GeneID.5347
GeneID.3169
0
20
40
Feature Importance (gain)


Top 20 features at 100 feature set
GeneID.79083
GeneID.23397
GeneID.55165
GeneID.3400
GeneID.7272
GeneID.983
GeneID.22974
GeneID.10948
GeneID.2296
GeneID.9156
GeneID.6422
GeneID.9700
GeneID.84299
GeneID.53335
GeneID.135295
GeneID.2064
GeneID.83540
GeneID.5347
GeneID.10403
GeneID.3169
Gain (averaged across seeds)
50
40
30
20
0.04
0.08
0.12
0.16
Frequency (averaged across seeds)



3. Hyper-parameters