Supplementary B: NanoString Sample Processing, Prediction, and Modelling

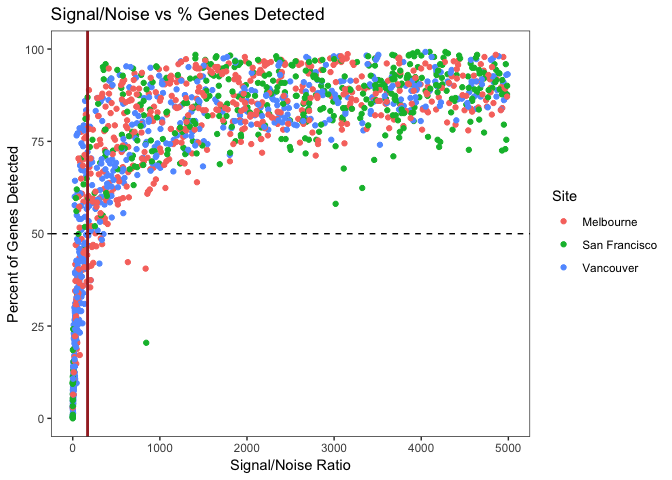
# B.6 Pre-processing, Batch Effect, QC and Normalization of the NanoString Data

## QC Table

* **Sample Quality** fails if either **Limit of Detection** or **Signal to Noise** fail
* **Overall QC** fails if either **Smallest Positive Control**, **Imaging**, or **Linearity** fail

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Total | AOC | USC | Vancouver |
| **Total Samples Run** | 5258 (100.0%) | 984 (18.7%) | 1766 (33.6%) | 2508 (47.7%) |
| Imaging Failures | 11 (0.2%) | 0 (0.0%) | 7 (0.4%) | 4 (0.2%) |
| Linearity Failures | 15 (0.3%) | 0 (0.0%) | 15 (0.8%) | 0 (0.0%) |
| Smallest PC Failures | 1 (0.0%) | 0 (0.0%) | 1 (0.1%) | 0 (0.0%) |
| Limit of Detection Failures | 289 (5.5%) | 70 (7.1%) | 27 (1.5%) | 192 (7.7%) |
| Signal to Noise Failures | 327 (6.2%) | 80 (8.1%) | 34 (1.9%) | 213 (8.5%) |
|  |  |  |  |  |
| **Sample Quality Failures** | 346 (6.6%) | 92 (9.3%) | 35 (2.0%) | 219 (8.7%) |
| **Overall QC Failures** | 22 (0.4%) | 0 (0.0%) | 18 (1.0%) | 4 (0.2%) |

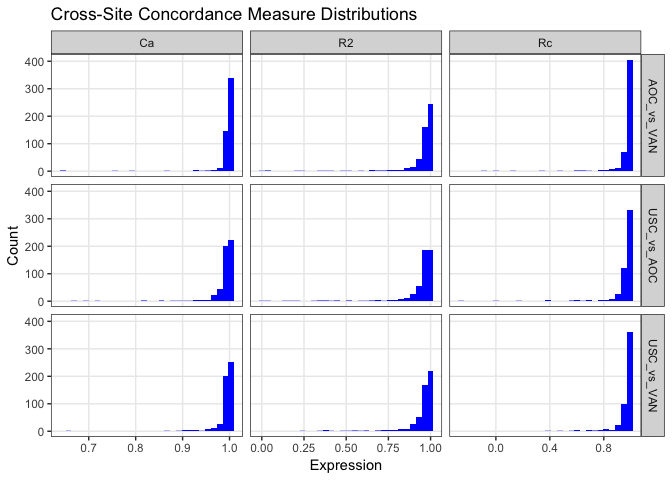
## Signal/Noise vs % Genes Detected



## Reference Pools Over Time

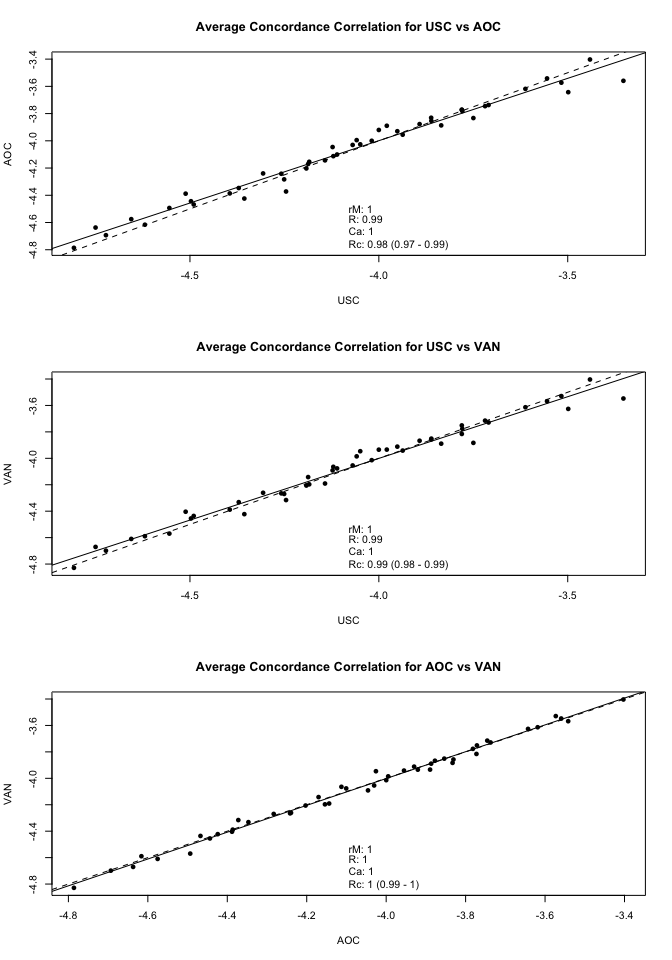


## Data Consistency of Cross-Site Controls Across Sites



Concordance Measures:

* Ca: coefficient of accuracy
* R2: squared Pearson’s correlation coefficient
* Rc: Lin’s concordance correlation



# B.7 Subtype Assignment in the NanoString Data

## Agreement between All-Array and TCGA

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Predicted TCGA | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted Array |  |  |  |  |  |
| C1.MES |  | 922 | 94 | 17 | 44 |
| C2.IMM |  | 149 | 619 | 148 | 41 |
| C4.DIF |  | 0 | 170 | 923 | 37 |
| C5.PRO |  | 21 | 15 | 63 | 566 |

### B - Overall Metrics

Cohen’s Kappa measures inter-rater agreement while taking into account the expected values and agreement by random chance.

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.79 (0.78 - 0.8) | < 0.001 |
| kappa | 0.72 (0.7 - 0.74) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.84 | 0.94 | 0.86 | 0.94 | 0.85 | 0.28 | 0.89 |
| C2.IMM | 0.69 | 0.88 | 0.65 | 0.90 | 0.67 | 0.25 | 0.79 |
| C4.DIF | 0.80 | 0.92 | 0.82 | 0.92 | 0.81 | 0.30 | 0.86 |
| C5.PRO | 0.82 | 0.97 | 0.85 | 0.96 | 0.84 | 0.17 | 0.90 |

## Predictive Entropy

### Mann-Whitney U Test

The null hypothesis tested within each data source is that a randomly sampled entropy score from samples with prediction agreement will be less than or greater than a randomly sampled entropy score with prediction disagreement with equal probability.

#### All-Array Entropies

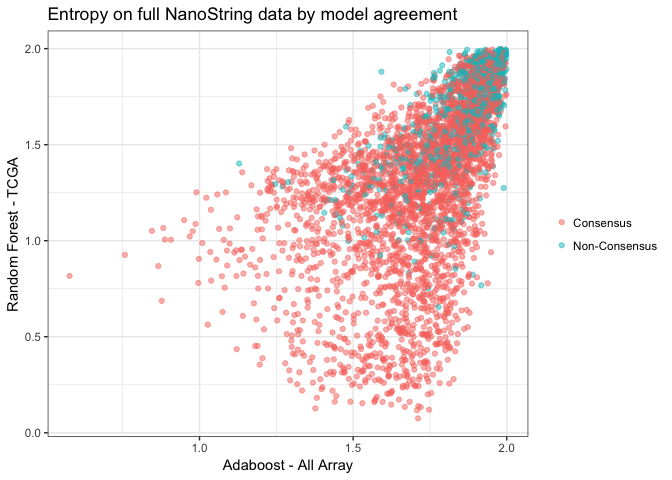
|  |  |  |  |
| --- | --- | --- | --- |
| statistic | p.value | method | alternative |
| 643554 | 1.846e-92 | Wilcoxon rank sum test with continuity correction | two.sided |

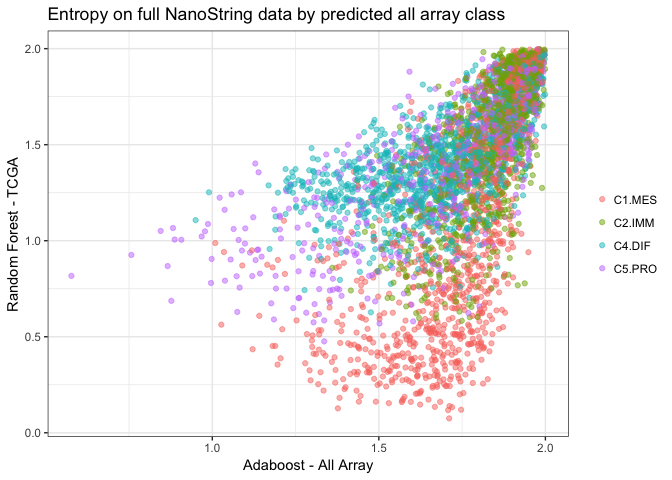
#### TCGA Entropies

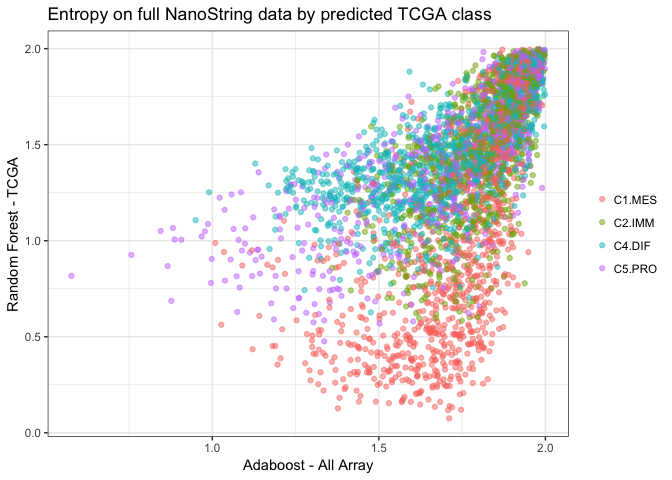
|  |  |  |  |
| --- | --- | --- | --- |
| statistic | p.value | method | alternative |
| 525466 | 4.339e-134 | Wilcoxon rank sum test with continuity correction | two.sided |

### Entropy Plots on the entire NanoString data

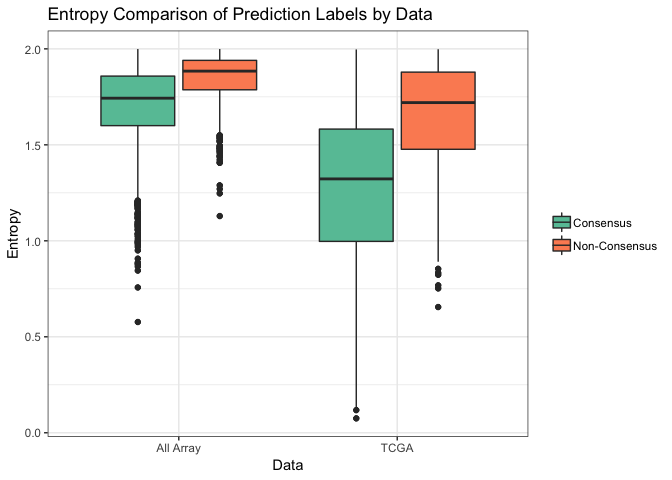
#### Scatterplot Comparison

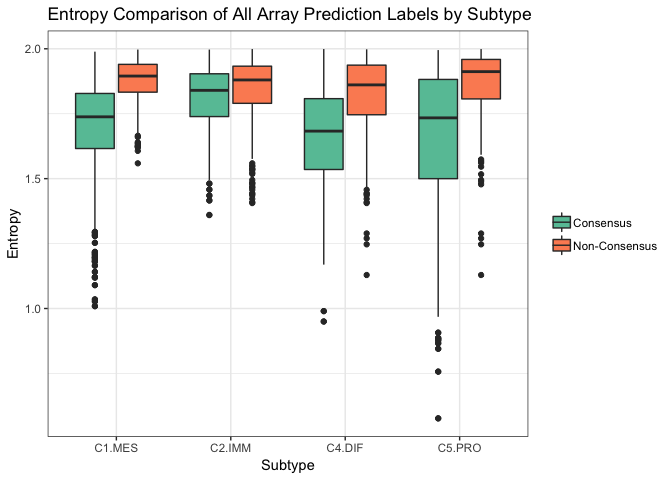


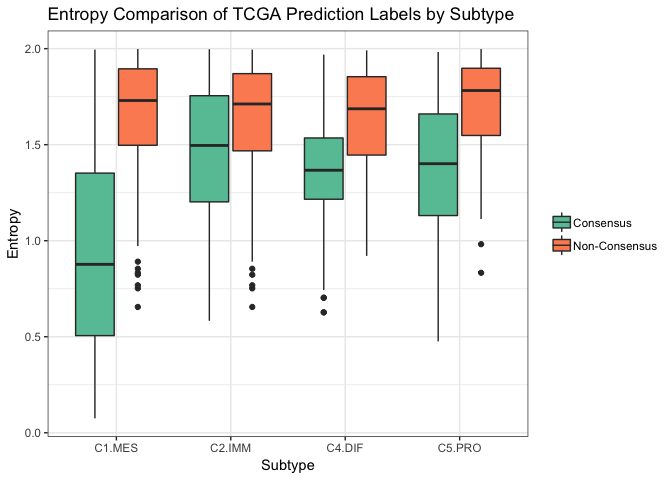




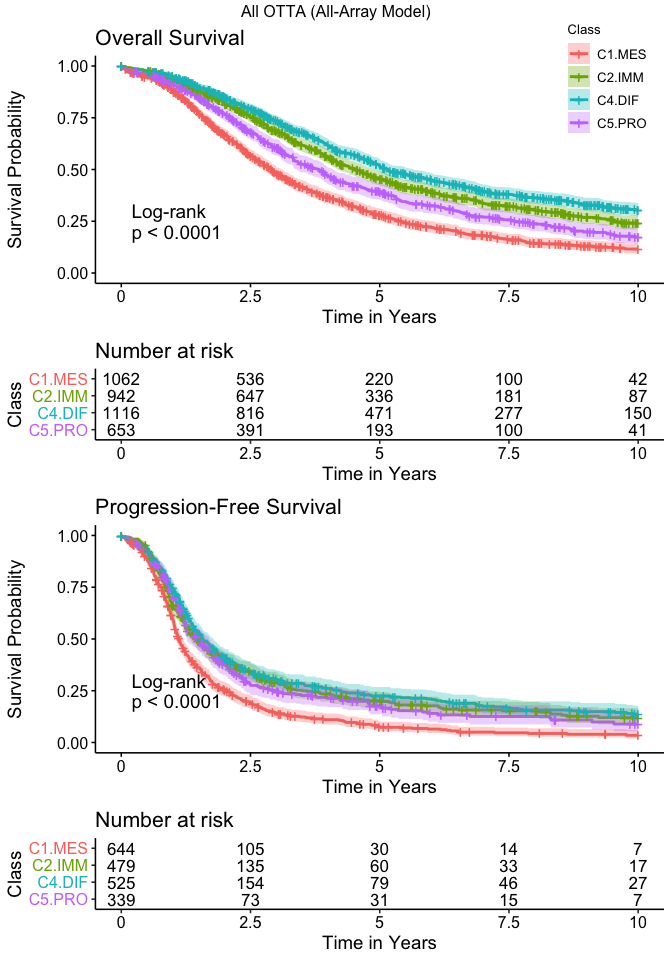
#### Boxplot Comparison

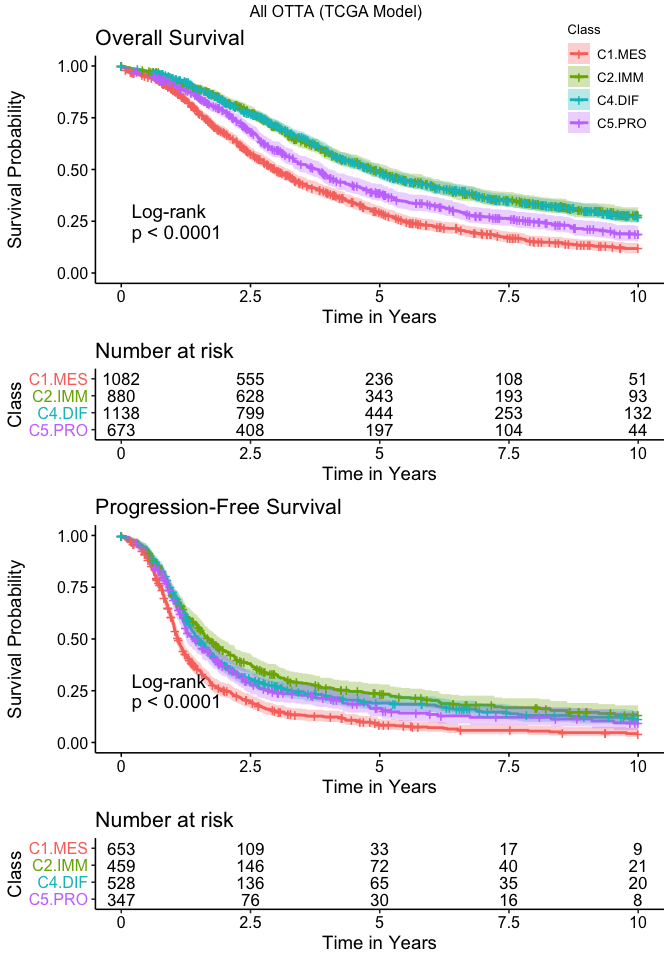


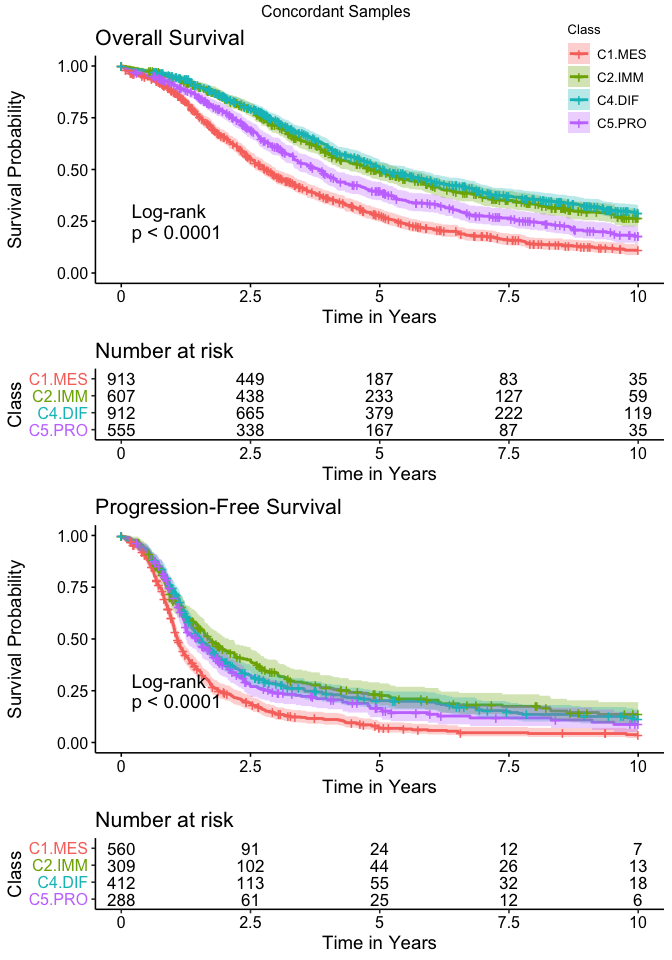


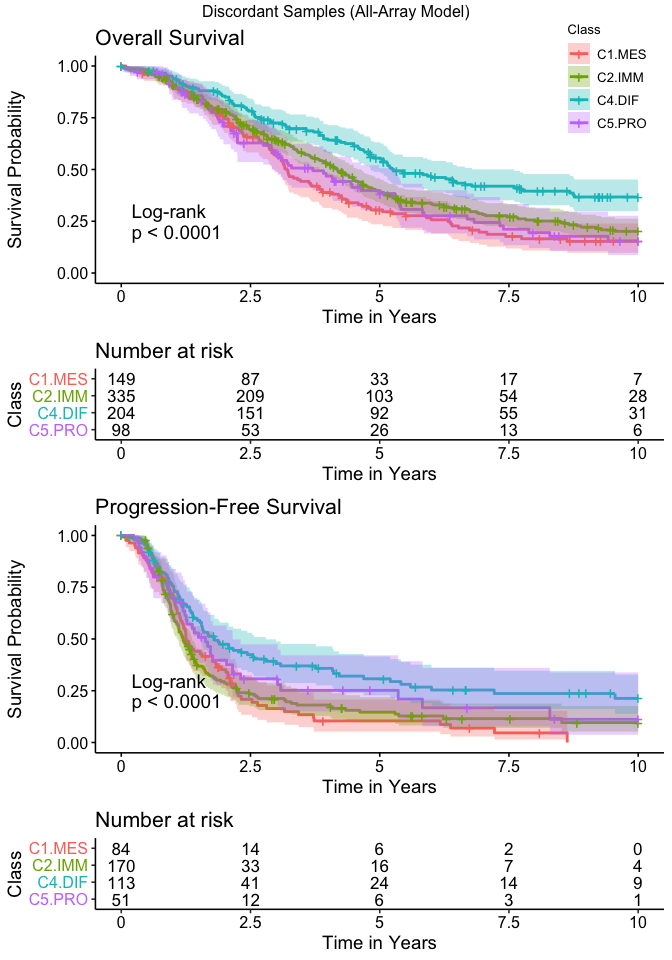


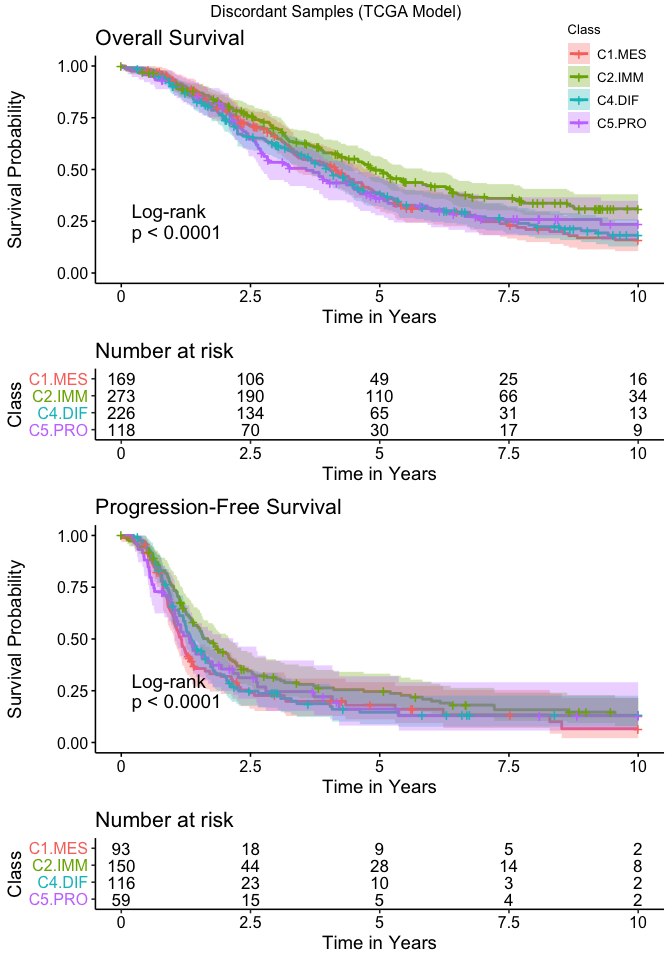
## Survival











# B.8 Platform Portability of the Array Classifier

## *Table SB4:* All-array model portability comparing predictions on NanoString to predictions on Array.

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Predicted from Array | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted from NanoString |  |  |  |  |  |
| C1.MES |  | 21 | 2 | 1 | 3 |
| C2.IMM |  | 2 | 13 | 4 | 0 |
| C4.DIF |  | 2 | 5 | 20 | 2 |
| C5.PRO |  | 3 | 2 | 1 | 16 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.72 (0.62 - 0.81) | < 0.001 |
| kappa | 0.63 (0.51 - 0.75) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.75 | 0.91 | 0.78 | 0.90 | 0.76 | 0.28 | 0.83 |
| C2.IMM | 0.59 | 0.92 | 0.68 | 0.88 | 0.63 | 0.20 | 0.76 |
| C4.DIF | 0.77 | 0.87 | 0.69 | 0.91 | 0.73 | 0.30 | 0.82 |
| C5.PRO | 0.76 | 0.92 | 0.73 | 0.93 | 0.74 | 0.23 | 0.84 |

## *Table SB5:* TCGA model portability comparing predictions on NanoString to predictions on Array.

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Predicted from Array | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted from NanoString |  |  |  |  |  |
| C1.MES |  | 21 | 2 | 1 | 2 |
| C2.IMM |  | 1 | 7 | 2 | 1 |
| C4.DIF |  | 1 | 7 | 19 | 1 |
| C5.PRO |  | 0 | 1 | 1 | 18 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.76 (0.66 - 0.85) | < 0.001 |
| kappa | 0.68 (0.56 - 0.8) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.91 | 0.92 | 0.81 | 0.97 | 0.86 | 0.31 | 0.92 |
| C2.IMM | 0.41 | 0.94 | 0.64 | 0.86 | 0.50 | 0.13 | 0.68 |
| C4.DIF | 0.83 | 0.85 | 0.68 | 0.93 | 0.75 | 0.33 | 0.84 |
| C5.PRO | 0.82 | 0.97 | 0.90 | 0.94 | 0.86 | 0.24 | 0.89 |

## *Table SB6:*: Comparing consensus samples predictions with published labels.

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Published Labels | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Consensus Labels |  |  |  |  |  |
| C1.MES |  | 20 | 1 | 0 | 0 |
| C2.IMM |  | 0 | 10 | 1 | 0 |
| C4.DIF |  | 0 | 1 | 17 | 1 |
| C5.PRO |  | 0 | 0 | 0 | 16 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.94 (0.85 - 0.98) | < 0.001 |
| kappa | 0.92 (0.84 - 1) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 1.00 | 0.98 | 0.95 | 1.00 | 0.98 | 0.31 | 0.99 |
| C2.IMM | 0.83 | 0.98 | 0.91 | 0.96 | 0.87 | 0.16 | 0.91 |
| C4.DIF | 0.94 | 0.96 | 0.89 | 0.98 | 0.92 | 0.28 | 0.95 |
| C5.PRO | 0.94 | 1.00 | 1.00 | 0.98 | 0.97 | 0.24 | 0.97 |

# B.9 Data Breakdown for Downstream Analysis

## Labelled Samples with All-Array and TCGA Models

Summary of All-Array and TCGA studies, cases, and gold-standard labels by dataset

|  |  |  |  |
| --- | --- | --- | --- |
| cut | studies | n | agreement |
| cut1 | 8 | 1550 | 1191 |
| cut2 | 5 | 1032 | 837 |
| cut3 | 4 | 894 | 719 |
| cut4 | 3 | 353 | 283 |

# B.10 Development of a Clinically Applicable Assay

## LOSO plots, boxplots, heatmaps

![](data:application/pdf;base64,)![](data:application/pdf;base64,)![](data:application/pdf;base64,)![](data:application/pdf;base64,)![](data:application/pdf;base64,)![](data:application/pdf;base64,)

## *Table SB7* Comparison of final model predictions with consensus labels from the training set (group 1)

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Consensus Label | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted with Final Model |  |  |  |  |  |
| C1.MES |  | 391 | 0 | 0 | 0 |
| C2.IMM |  | 0 | 256 | 0 | 0 |
| C4.DIF |  | 0 | 0 | 319 | 0 |
| C5.PRO |  | 0 | 0 | 0 | 169 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 1 (1 - 1) | < 0.001 |
| kappa | 1 (1 - 1) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 1 | 1 | 1 | 1 | 1 | 0.34 | 1 |
| C2.IMM | 1 | 1 | 1 | 1 | 1 | 0.23 | 1 |
| C4.DIF | 1 | 1 | 1 | 1 | 1 | 0.28 | 1 |
| C5.PRO | 1 | 1 | 1 | 1 | 1 | 0.15 | 1 |

## *Table SB8* Comparison of final model predictions with consensus labels from the confirmation set (group 2)

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Consensus Label | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted with Final Model |  |  |  |  |  |
| C1.MES |  | 191 | 4 | 0 | 0 |
| C2.IMM |  | 1 | 147 | 1 | 6 |
| C4.DIF |  | 0 | 2 | 285 | 10 |
| C5.PRO |  | 1 | 1 | 4 | 164 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.96 (0.95 - 0.98) | < 0.001 |
| kappa | 0.95 (0.93 - 0.97) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.99 | 0.99 | 0.98 | 1.00 | 0.98 | 0.24 | 0.99 |
| C2.IMM | 0.95 | 0.99 | 0.95 | 0.99 | 0.95 | 0.19 | 0.97 |
| C4.DIF | 0.98 | 0.98 | 0.96 | 0.99 | 0.97 | 0.36 | 0.98 |
| C5.PRO | 0.91 | 0.99 | 0.96 | 0.98 | 0.94 | 0.21 | 0.95 |

## *Table SB9* Comparison of final model predictions with consensus labels from the first validation set (group 3)

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Consensus Label | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted with Final Model |  |  |  |  |  |
| C1.MES |  | 211 | 4 | 0 | 4 |
| C2.IMM |  | 3 | 145 | 8 | 2 |
| C4.DIF |  | 0 | 1 | 210 | 7 |
| C5.PRO |  | 1 | 0 | 4 | 119 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.95 (0.93 - 0.97) | < 0.001 |
| kappa | 0.94 (0.91 - 0.96) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.98 | 0.98 | 0.96 | 0.99 | 0.97 | 0.30 | 0.98 |
| C2.IMM | 0.97 | 0.98 | 0.92 | 0.99 | 0.94 | 0.22 | 0.97 |
| C4.DIF | 0.95 | 0.98 | 0.96 | 0.98 | 0.95 | 0.30 | 0.96 |
| C5.PRO | 0.90 | 0.99 | 0.96 | 0.98 | 0.93 | 0.17 | 0.95 |

## *Table SB10* Comparison of final model predictions with consensus labels from the second validation set (group 4)

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Consensus Label | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted with Final Model |  |  |  |  |  |
| C1.MES |  | 97 | 2 | 0 | 1 |
| C2.IMM |  | 1 | 47 | 1 | 3 |
| C4.DIF |  | 0 | 2 | 67 | 5 |
| C5.PRO |  | 0 | 0 | 1 | 56 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.94 (0.91 - 0.97) | < 0.001 |
| kappa | 0.92 (0.89 - 0.96) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.99 | 0.98 | 0.97 | 0.99 | 0.98 | 0.35 | 0.99 |
| C2.IMM | 0.92 | 0.98 | 0.90 | 0.98 | 0.91 | 0.18 | 0.95 |
| C4.DIF | 0.97 | 0.97 | 0.91 | 0.99 | 0.94 | 0.26 | 0.97 |
| C5.PRO | 0.86 | 1.00 | 0.98 | 0.96 | 0.92 | 0.20 | 0.93 |

## *Table SB11* Comparison of final model predictions with consensus labels from the overlap set

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Consensus Label | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| Predicted with Final Model |  |  |  |  |  |
| C1.MES |  | 25 | 0 | 0 | 0 |
| C2.IMM |  | 0 | 8 | 0 | 0 |
| C4.DIF |  | 0 | 0 | 23 | 0 |
| C5.PRO |  | 0 | 0 | 0 | 20 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 1 (0.95 - 1) | < 0.001 |
| kappa | 1 (1 - 1) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 1 | 1 | 1 | 1 | 1 | 0.33 | 1 |
| C2.IMM | 1 | 1 | 1 | 1 | 1 | 0.11 | 1 |
| C4.DIF | 1 | 1 | 1 | 1 | 1 | 0.30 | 1 |
| C5.PRO | 1 | 1 | 1 | 1 | 1 | 0.26 | 1 |

## Entropy Analysis for Final Predictions

### Mann-Whitney U Test

|  |  |  |  |
| --- | --- | --- | --- |
| statistic | p.value | method | alternative |
| 314477 | < 0.001 | Wilcoxon rank sum test with continuity correction | two.sided |

### Boxplot Comparisons

