Supplementary A: Molecular Classifier Development on Legacy Gene Expression Datasets

# A.5 Class Prediction

## *Table SA7*: Contingency table showing the relationship between predicted labels (row) and the published labels (column) (n=215).

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | AOCS Published Labels | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| AOCS Predicted Labels |  |  |  |  |  |
| C1.MES |  | 76 | 3 | 0 | 1 |
| C2.IMM |  | 4 | 42 | 0 | 1 |
| C4.DIF |  | 0 | 5 | 46 | 1 |
| C5.PRO |  | 3 | 0 | 0 | 33 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.92 (0.87 - 0.95) | < 0.001 |
| kappa | 0.88 (0.83 - 0.94) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.92 | 0.97 | 0.95 | 0.95 | 0.93 | 0.37 | 0.94 |
| C2.IMM | 0.84 | 0.97 | 0.89 | 0.95 | 0.87 | 0.22 | 0.90 |
| C4.DIF | 1.00 | 0.96 | 0.88 | 1.00 | 0.94 | 0.24 | 0.98 |
| C5.PRO | 0.92 | 0.98 | 0.92 | 0.98 | 0.92 | 0.17 | 0.95 |

## *Table SA8*: Contingency table showing the relationship between all-array k-modes clustering and TCGA training (published) labels (n=434).

### A - Confusion Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | TCGA Training Labels | C1.MES | C2.IMM | C4.DIF | C5.PRO |
| All-Array K-modes Clustering Labels |  |  |  |  |  |
| C1.MES |  | 79 | 6 | 2 | 4 |
| C2.IMM |  | 36 | 76 | 19 | 13 |
| C4.DIF |  | 2 | 9 | 101 | 23 |
| C5.PRO |  | 3 | 1 | 3 | 57 |

### B - Overall Metrics

|  |  |  |
| --- | --- | --- |
| Metric |  | P value |
| accuracy | 0.72 (0.68 - 0.76) | < 0.001 |
| kappa | 0.63 (0.57 - 0.68) |  |

### C - By-Class Metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Sensitivity | Specificity | PPV | NPV | F1 | Detection Prevalence | Balanced Accuracy |
| C1.MES | 0.66 | 0.96 | 0.87 | 0.88 | 0.75 | 0.21 | 0.81 |
| C2.IMM | 0.83 | 0.80 | 0.53 | 0.94 | 0.64 | 0.33 | 0.81 |
| C4.DIF | 0.81 | 0.89 | 0.75 | 0.92 | 0.78 | 0.31 | 0.85 |
| C5.PRO | 0.59 | 0.98 | 0.89 | 0.89 | 0.71 | 0.15 | 0.78 |