|  |  |  |  |
| --- | --- | --- | --- |
| Analyses | Logistic regression | SVM Poly | Random forest |
| All (29k probes) | 0.54 ±0.06 | 0.59 ±0.05 | 0.51 ±0.05 |
| GLL (117±10 probes) | 0.51 ±0.05 | 0.54 ±0.06 | 0.51 ±0.06 |
| Liu et al. (11 genes 15 probes) | 0.62 ±0.05 | 0.62 ±0.05 | 0.55 ±0.05 |
| Riester et al. (7 genes 12 probes) | 0.48 ±0.05 | 0.50 ±0.05 | 0.50 ±0.05 |
| Tucker et al. 2 genes 2 probes) | 0.45 ±0.04 | 0.49 ±0.04 | 0.53 ±0.05 |
| TCGA 4 molecular subtypes (3 features) | 0.50 ±0.04 | 0.52 ±0.04 | 0.54 ±0.05 |

Supplement Table 3: Results of distinct biostatistical models and methods to validate and evaluate gene signatures predicting residual disease. Included N=188 RD≤1 cases and N=78 cases RD>1; mean (AUC) ± 1 STD over 10 folds GLL: Generalized local learning; SVM Poly: support vector machine with polynomial kernel