|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number** | **scRNA-seq dataset** | **Bulk dataset** | **MiRNA + SDES Metric** | **MiRNA + MAD Metric** | **Km p-value from MiRNA + SDES** | **MiRNA + SDES Significant KM?** | **Km p-value from MiRNA + MAD** | **MiRNA + MAD Significant KM?** |
| 1. | GSE81861\_CRC\_tumor | TCGA-COAD | 0.8658 (0.01937) [178] | 0.8620 (0.01229) [106] | 5.207e-01 | N | 7.162e-02 | N |
| 2. | GSE81861\_Cell\_line\_hct116 | TCGA-COAD | 0.7340 (0.02430) [16] | 0.7717 (0.01566) [29] | 2.345e-01 | N | 3.06e-01 | N |
| 3. | GSE81861\_CRC\_tumor | TCGA-READ | 0.9131 (0.03869) [36] | 0.8628 (0.05574) [44] | 3.698e-01 | N | 5.063e-01 | N |
| 4. | GSE81861\_CRC\_tumor | TCGA-COAD + TCGA-READ | 0.8518 (0.01525) [211] | 0.8385 (0.02231) [137] | 8.324e-01 | N | 6.849e-01 | N |
| 5. | GSM1599500\_K562 | TCGA-DLBC | 0.8286 (0.05871) [15] | 0.9643 (0.02062) [7] | 1.644e-01 | N | 5.637e-01 | N |

Summary: Based on this data, I conclude using the active genes as the starting point from a previous cox model is overfitting the data. It is also apparent that these overfit c-index values are not robust because when using the KM curves, we do not see a separation of high and low-risk groups. I tested the K562 dataset because it had high-performing c-indexes in our original data, and it was a different cancer from colon or rectal cancer.