**Supplementary Tables and Figures**



Figure S1. Classification accuracies of the eight machine learning algorithms trained using the full Cox PH set of 1,768 genes and the top 1000, 500, 100 and 50 genes with the lowest p values from this set. Logistic regression (LR); Decision trees (DT); Naïve bayes (NB); K nearest neighbours (KNN) ; Random forest (RF); XGBoost (XGB); Neural Networks (NN); Support Vector Machine (SVM)



Figure S2. Classification accuracies when subsets of the A) Wilcoxon set of 1,596 genes and B) the Wilcoxon-Cox set of 977 gene sets selected using recursive feature elimination were used to train a Support Vector Machine algorithm.

On the x axis is the number of genes in the subset used for training. On the y axis is the corresponding accuracy of the Support Vector Machine model



Figure S3. Computational time taken for each of the top four ML models to predict relapse status on a test set of 90 patients. These models use the Wilcoxon set of 1596 genes or the Wilcoxon-Cox set of 977 genes with a Support Vector Machine (SVM) or Neural Networks (NN) algorithm



Figure S4. Area Under Receiver Operating Characteristics Curve (AUROC) for the top four ML models: SVM1596, NN1596, SVM1596, NN977.

These models use the Wilcoxon set of 1596 genes or the Wilcoxon-Cox set of 977 genes with a Support Vector Machine (SVM) or Neural Networks (NN) algorithm.

Table S1. Sensitivity and specificity of the top four ML models: SVM1596, NN1596, SVM1596, NN977. These models use the Wilcoxon set of 1596 genes or the Wilcoxon-Cox set of 977 genes with a Support Vector Machine (SVM) or Neural Network (NN) algorithm

|  |  |  |
| --- | --- | --- |
| Model  | Sensitivity | Specificity  |
| SVM1596 | 81.55% | 92.79% |
| NN1596 | 89.23% | 86.27% |
| SVM977 | 91.09% | 78.95% |
| NN977 | 85.29% | 86.77% |



Figure S5. Accuracy comparison of SVM models using the Wilcoxon-Cox set of 977 genes with varying values of the cost hyperparameter selected through a manual grid search for A) cost values between 0 and 5 in 0.5 increments and B) cost values between 0 and 2 in 0.1 increments



Figure S5. Protein-protein interactions of the Wilcoxon-Cox set of 977 genes visualized by Cytoscape.

**Table S2.** List of radiogenes

|  |  |
| --- | --- |
| Gene Symbol | Reference |
| AR | Eschrich et al. CCR 2012 |
| ABL1 | Eschrich et al. CCR 2012 |
| CDK1 | Eschrich et al. CCR 2012 |
| JUN | Eschrich et al. CCR 2012 |
| HDAC1 | Eschrich et al. CCR 2012 |
| IRF1 | Eschrich et al. CCR 2012 |
| PRKCB | Eschrich et al. CCR 2012 |
| RELA | Eschrich et al. CCR 2012 |
| STAT1 | Eschrich et al. CCR 2012 |
| SUMO1 | Eschrich et al. CCR 2012 |
| ASF1B | Speers et al. CCR 2015 |
| ATM | Speers et al. CCR 2015 |
| BATF | Speers et al. CCR 2015 |
| BCAR3 | Speers et al. CCR 2015 |
| BIN1 | Speers et al. CCR 2015 |
| BLM | Speers et al. CCR 2015 |
| BTG2 | Speers et al. CCR 2015 |
| BUB1B | Speers et al. CCR 2015 |
| C4orf19 | Speers et al. CCR 2015 |
| CDT1 | Speers et al. CCR 2015 |
| COBL | Speers et al. CCR 2015 |
| CPE | Speers et al. CCR 2015 |
| DNAJB4 | Speers et al. CCR 2015 |
| DNMT3B | Speers et al. CCR 2015 |
| DTL | Speers et al. CCR 2015 |
| EMP3 | Speers et al. CCR 2015 |
| FZD2 | Speers et al. CCR 2015 |
| GALC | Speers et al. CCR 2015 |
| GINS2 | Speers et al. CCR 2015 |
| HELLS | Speers et al. CCR 2015 |
| HOOK1 | Speers et al. CCR 2015 |
| HOXC13 | Speers et al. CCR 2015 |
| ITPR1 | Speers et al. CCR 2015 |
| KRT86 | Speers et al. CCR 2015 |
| LAMP2 | Speers et al. CCR 2015 |
| MCM10 | Speers et al. CCR 2015 |
| MKI67 | Speers et al. CCR 2015 |
| MSRA | Speers et al. CCR 2015 |
| MTRF1 | Speers et al. CCR 2015 |
| MYO5C | Speers et al. CCR 2015 |
| NFIL3 | Speers et al. CCR 2015 |
| OSMR | Speers et al. CCR 2015 |
| PBX3 | Speers et al. CCR 2015 |
| PDCD4 | Speers et al. CCR 2015 |
| PRC1 | Speers et al. CCR 2015 |
| PRKCA | Speers et al. CCR 2015 |
| PXMP4 | Speers et al. CCR 2015 |
| RAB15 | Speers et al. CCR 2015 |
| RAB25 | Speers et al. CCR 2015 |
| RAD51 | Speers et al. CCR 2015 |
| RBPMS | Speers et al. CCR 2015 |
| SARS | Speers et al. CCR 2015 |
| SDHD | Speers et al. CCR 2015 |
| SERPINE2 | Speers et al. CCR 2015 |
| SLC39A14 | Speers et al. CCR 2015 |
| SPC25 | Speers et al. CCR 2015 |
| VAMP7 | Speers et al. CCR 2015 |
| TK1 | Speers et al. CCR 2015 |
| TSPAN13 | Speers et al. CCR 2015 |
| ZPR1 | Speers et al. CCR 2015 |
| LTN1 | Speers et al. CCR 2015 |
| ADH1B | Tramm et al. CCR 2014 |
| DNALI1 | Tramm et al. CCR 2014 |
| hCG2023290 | Tramm et al. CCR 2014 |
| HLA-DQA1 | Tramm et al. CCR 2014 |
| IGKC | Tramm et al. CCR 2014 |
| OR8G2 | Tramm et al. CCR 2014 |
| RGS1 | Tramm et al. CCR 2014 |

Table S3. List of 16 radiogenes found in the 977 gene set and their aliases. All genes are protein coding

|  |  |  |
| --- | --- | --- |
| Symbol | Gene name | Aliases   |
| *MKI67* | Marker Of Proliferation Ki-67   | *KIA, MIB- , MIB-1, PPP1R105, MKI67* |
| *CDK1* | Cyclin Dependent Kinase 1   | *CDC2, CDC28A, P34CDC2, CDK1* |
| *SPC25* | SPC25 Component Of NDC80 Kinetochore Complex | *AD024, SPBC25,HSpc25* |
| *PRC1* | Protein Regulator Of Cytokinesis 1  | *ASE1, PRC1* |
| *BLM* | BLM RecQ Like Helicase  | *BS, MGRISCE1, RECQ2, RECQL2, RECQL3,BLM* |
| *SLC39A14* | Solute Carrier Family 39 Member 14 | *HCIN, HMNDYT2, LZT-Hs4, NET34, ZIP14, cig19, SLC39A14* |
| *ASF1B* | Anti-Silencing Function 1B Histone Chaperone | *CIA-II* |
| *PDCD4* | Programmed Cell Death 4  | *H731, PDCD4* |
| *DNMT3B* | DNA Methyltransferase 3 Beta | *ICF, ICF1, M.HsaIIIB, DNMT3B* |
| *PRKCA* | Protein Kinase C Alpha | *PRKACA* |
| *RELA* | RELA Proto-Oncogene, NF-KB Subunit | *CMCU, NFKB3, p65, RELA* |
| *DTL* | Denticleless E3 Ubiquitin Protein Ligase Homolog | *CDT2, DCAF2, L2DTL, RAMP, DTL* |
| *GINS2* | GINS Complex Subunit 2 | *HSPC037, PSF2, Pfs2, GINS2* |
| *MSRA* | Methionine Sulfoxide Reductase A | *PMSR, MSRA* |
| *HELLS* | Helicase, Lymphoid Specific | *ICF4, LSH, Nbla10143, PASG, SMARCA6, HELLS* |
| *MCM10* | Minichromosome Maintenance 10 Replication Initiation Factor | *CNA43, DNA43, PRO2249, MCM10* |