Cell Reports, Volume *23*

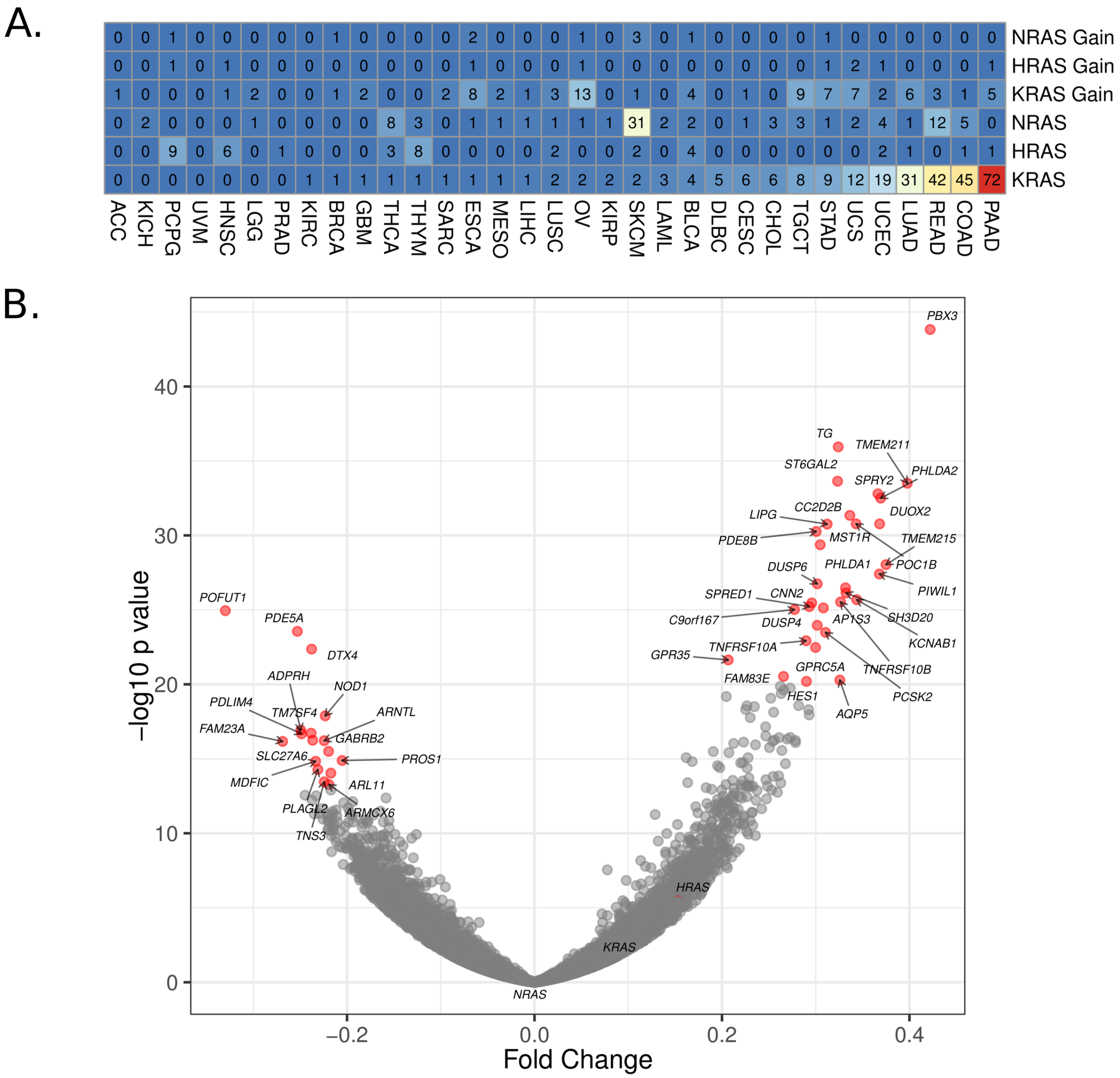
Supplemental Information

Machine Learning Detects

Pan-cancer Ras Pathway Activation

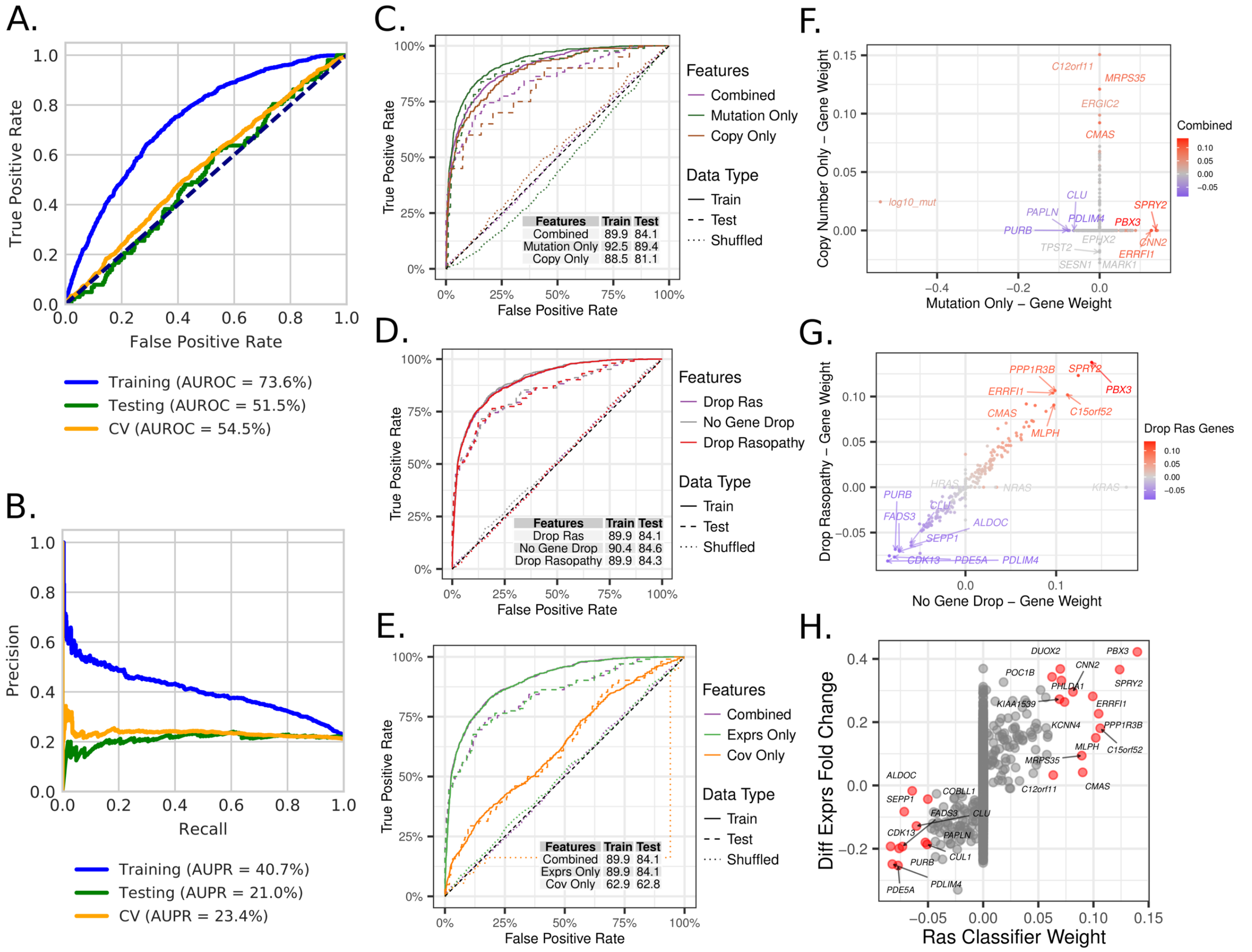
in The Cancer Genome Atlas

Gregory P. Way, Francisco Sanchez-Vega, Konnor La, Joshua Armenia, Walid K. Chatila, Augustin Luna, Chris Sander, Andrew D. Cherniack, Marco Mina, Giovanni Ciriello, Nikolaus Schultz, The Cancer Genome Atlas Research Network, Yolanda Sanchez, and Casey S. Greene



	ACC	KICH	PCPG	UVM	HNSC	LGG	PRAD	KIRC	BRCA	GBM	THCA	THYM	SARC	ESCA	MESO	LIHC	LUSC	OV	KIRP	SKCM	LAML	BLCA	DLBC	CESC	CHOL	TGCT	STAD	UCS	UCEC	LUAD	READ	COAD	PAAD
NRAS Gain	0	0	1	0	0	0	0	0	1	0	0	0	0	2	0	0	0	1	0	3	0	1	0	0	0	0	1	0	0	0	0	0	0
HRAS Gain	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	2	1	0	0	0	1
KRAS Gain	1	0	0	0	1	2	0	0	1	2	0	0	2	8	2	1	3	13	0	1	0	4	0	1	0	9	7	7	2	6	3	1	5
NRAS	0	2	0	0	0	1	0	0	0	0	8	3	0	1	1	1	1	1	1	31	2	2	0	1	3	3	1	2	4	1	12	5	0
HRAS	0	0	9	0	6	0	1	0	0	0	3	8	0	0	0	0	2	0	0	2	0	4	0	0	0	0	0	0	2	1	0	1	1
KRAS	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	2	2	3	4	5	6	6	8	9	12	19	31	42	45	72

Figure S1. *Ras pathway alteration percentages in TCGA PanCanAtlas; related to Figure 2 and Data S1.* **(A)** Percentage of *KRAS*, *HRAS*, and *NRAS* mutations and copy number gains across 33 different cancer-types from TCGA PanCanAtlas. **(B)** Differentially expressed genes between Ras aberrant and Ras wild-type PanCanAtlas tumors. Analysis is controlled for cancer-type.



Features	Train	Test
Combined	89.9	84.1
Mutation Only	92.5	89.4
Copy Only	88.5	81.1

Features	Train	Test
Drop Ras	89.9	84.1
No Gene Drop	90.4	84.6
Drop Rasopathy	89.9	84.3

Features	Train	Test
Combined	89.9	84.1
Exprs Only	89.9	84.1
Cov Only	62.9	62.8

Figure S2. *Benchmarking PanCancer Ras Classifiers; related to Figure 2.* **(A)** Receiver operating characteristic (ROC) curve and **(B)** Precision recall (PR) curve for a null model trained on a randomly shuffled RNAseq matrix. Also provided are the area under the ROC (AUROC) and area under the PR (AUPR) curves for training, testing, and cross validation sets. **(C)** ROC curve for three models predicting: 1) Ras mutations only; 2) Ras copy number gains only; 3) Combined data (model in Figure 2). The AUROC is provided for both training and testing sets. **(D)** ROC/AUROC across train and test sets for dropping different genes from the RNAseq matrix. The Drop Ras model is the model provided in Figure 2. **(E)** ROC/AUROC across train and test sets for using expression data or covariates only. The combined model is the model provided in Figure 2. In all ROC curves, the dashed navy line represents a hypothetical random guess classifier. Gene coefficients for the models presented in **(F)** panel C and in **(G)** panel D. The points are colored by the model presented in Figure 2. **(H)** Differential fold change for tumors with active Ras against tumors with wild-type Ras compared against the Ras classifier gene coefficients provided in Figure 2. Red points correspond to labelled genes.

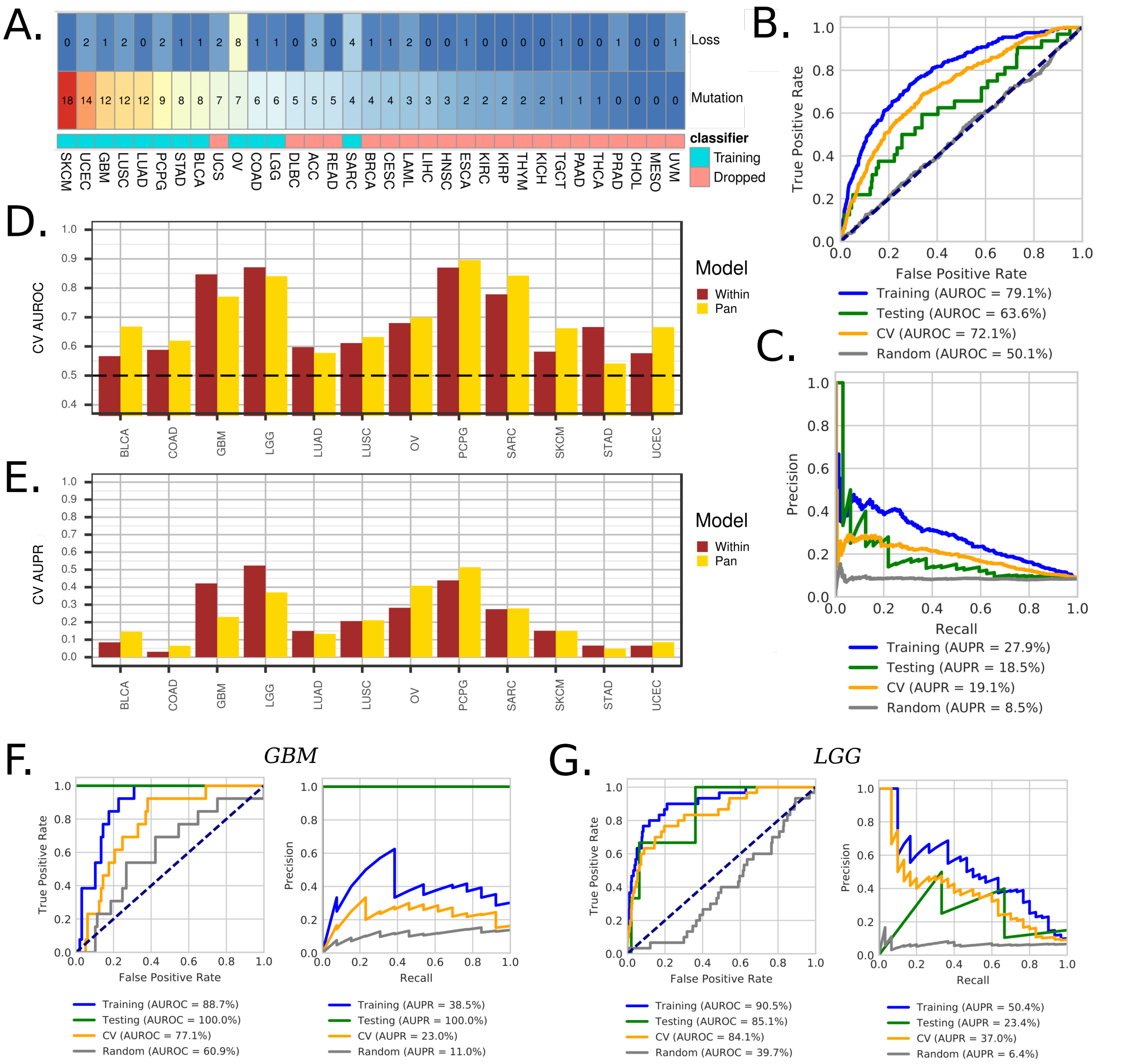


Figure S3. *Pan-cancer NF1 classification performance; related to Figures 2 and 4A.* **(A)** Cancer-type specific percentages of *NF1* inactivation by copy number loss and deleterious mutation. The colored squares indicate if the cancer type was included in model training. **(B)** Receiver operating characteristic (ROC) curve and Area under the ROC curve (AUROC) given for training, testing, and cross-validation (CV) sets. **(C)** Precision Recall (PR) Curve and corresponding area under the PR (AUPR) curve for each evaluation set. Cancer-type specific CV **(D)** AUROC and **(E)** AUPR for the *NF1* pan-cancer model compared to separate models trained on each cancer type independently. ROC and PR curves for predicting *NF1* inactivation in **(F)** GBM and **(G)** LGG using the pan-cancer model. The grey lines represent predictions made on a shuffled gene expression matrix.

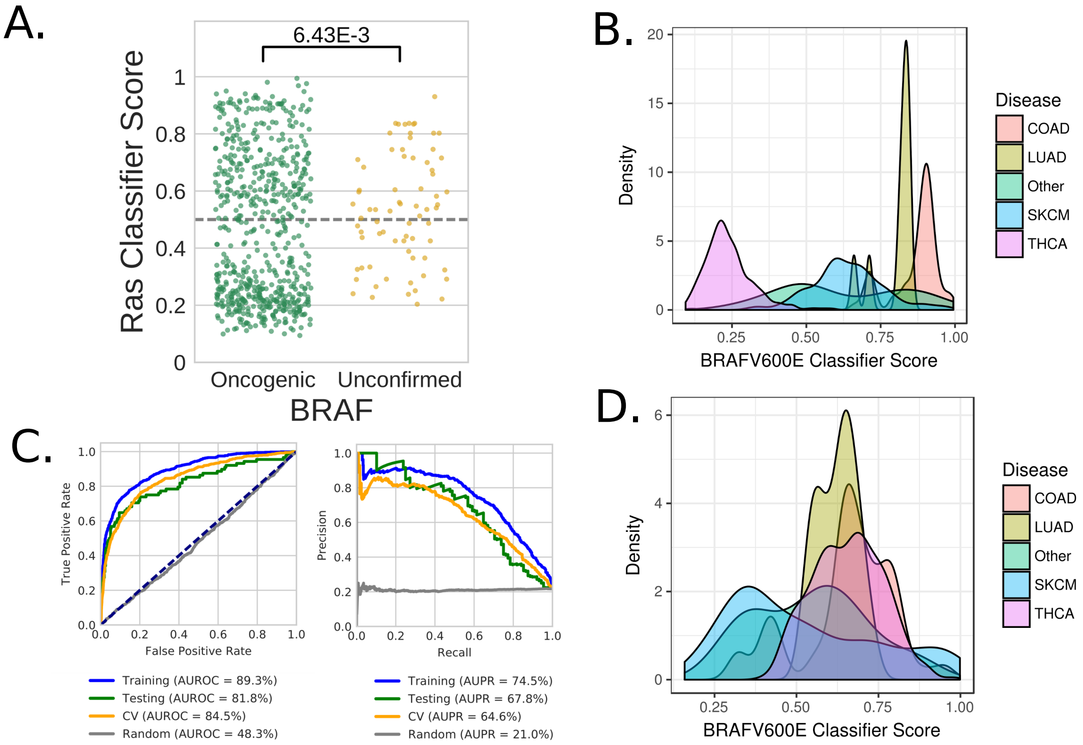


Figure S4. *Predicting BRAF with the Ras Classifier; related to Figure 2.* **(A)** Predictions for tumors with oncogenic or unconfirmed variants in *BRAF* given by the Ras classifier evaluated in Figure 2. **(B)** Ras classifier scores assigned to samples with *BRAF* V600E mutations stratified by cancer type. A score above 0.5 indicates a prediction of activated Ras. **(C)** Ras classifier evaluation after removing THCA and SKCM from training. ROC and PR curves for the Ras classifier without THCA and SKCM does not indicate reduced performance. The grey lines represent predictions made on a shuffled gene expression matrix. **(D)** Ras classifier without THCA and SKCM classify *BRAF* V600E as Ras wildtype in THCA, but not in SKCM.