Generation of a candidate gene list for pooled, shRNA-mediated screening:
To create a list of candidate effectors of mutant KRAS activity for pooled shRNA-mediated silencing, we combined 5 gene sets, including controls. Specifically, the following sources were combined to create the gene list:

1. ***OncoSig predictions***: 22 genes identified by OncoSig analysis[1] as novel candidate effectors of mutant KRAS activity.
2. ***Candidate KRAS interactors***: A set of 23 gene predicted to harbor a RAS binding domain (RBD) and thus identified as novel candidate KRAS physical interactors
3. ***Established KRAS pathway effectors***: These include RALGDS, MAPK1, RASA1, SQSTM1 and AKT1
4. ***Negative controls***: This list includes DUSP13, HIGD2B, MAB21L2, OAS2, SLC20A2, SLC25A34, PDCD2, TP53, NUP205, and ZBTB8B, which were selected at random among genes that were not identified by either OncoSig or as candidate KRAS interactors and also not included in KRAS pathway gene sets.
5. ***Positive control***: TBK1, a gene previously reported as essential in KRAS mutant cancer[2].

The following sections describe the methods used to select genes using OncoSig or as candidate KRAS interactors:

OncoSig Predictions**:** As discussed in Broyde et al[1], the OncoSig Naive Bayes algorithm was used to discover novel members of the KRAS interactome in a LUAD specific context. The final score of the OncoSig algorithm is a likelihood ratio, where higher values indicate a greater probability that a gene encodes for a protein representing a bona fide KRAS effector, cognate binding partner, or upstream modulator. It was used to identify the top 40 candidate proteins in the KRAS pathway. Of these 18 were already published as playing a role in KRAS biology. As a result, the remaining 22 were selected for this screen.

Candidate KRAS interactors: To identify proteins representing likely KRAS structural cognate binding partners, we generated a list of candidate Ras Binding Domain (RBD) containing proteins. First, we analyzed the SMART database[3] to identify all proteins containing at least one domain that was structurally similar to an RBD (i.e., comprising the UBQ fold family). The resulting list of 1,849 proteins comprised a total of 84 distinct domain types. To further refine this list, we used the PrePPI algorithm[4] to assess the structural likelihood of its RBD-like domains to interact with KRAS. Finally, we assembled a list of 20 RBDs known to bind to H/K/NRAS as derived by Rodriguez-Viciana and colleagues[5]. For each candidate RBD-containing protein, we then created a 21-feature representing the PrePPI likelihood ratio score (feature 1) and the Protein Structure Distance between the candidate RBD and each of the 20 established RBDs (features 2 – 21). The latter was computed using the Ska structural alignment program[6].

Using the 21-feature vectors of each candidate RBD-containing protein, we trained a supervised learning algorithm using a Random Forest Classifier[7] using a Positive Gold Standard set including known structural RAS interactors in the PDB and a Negative Gold Standard Set based on proteins with an UBQ domain not reported as RAS-binding in PDB.

The rationale is that while there are almost 2,000 UBQ domain containing proteins, only a small fraction of these is expected to be RAS binding[8,9]. Finally, we used the trained classifier to infer novel RAS-binding proteins from the 1,849 candidate UBQ domain containing ones. Of these only 487 had a likelihood score (LS) greater than zero.

To generate the final list for experimental validation, we observed that the top 100 predictions by the Random Forest classifier included 80% (*i.e.*, 16 out of 22) known RBDs. From all significant predictions (LS > 0.5), we thus removed proteins (a) whose UBQ domain missed the characteristic polar basic residues, which are present on virtually all known RBDs[8], (b) already reported as RAS binding in the PDB, or (c) with exceptionally large structural deviations from standard RBD domains. The remaining 23 proteins were selected for further experimental validation.

References:

1. Broyde, J. *et al.* Oncoprotein-specific molecular interaction maps (SigMaps) for cancer network analyses. *Nat. Biotechnol.* **39**, 215–224 (2021).
2. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
3. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5857–5864 (1998).
4. Zhang, Q. C. *et al.* Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* **490**, 556–560 (2012).
5. Rodriguez-Viciana, P., Sabatier, C. & McCormick, F. Signaling Specificity by Ras Family GTPases Is Determined by the Full Spectrum of Effectors They Regulate. *Mol. Cell. Biol.* **24**, 4943–4954 (2004).
6. Yang, A. S. & Honig, B. An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *J. Mol. Biol.* **301**, 665–678 (2000).
7. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
8. Kiel, C. *et al.* Recognizing and defining true ras binding domains II: In Silico prediction based on homology modelling and energy calculations. *J. Mol. Biol.* **348**, 759–775 (2005).
9. Wohlgemuth, S. *et al.* Recognizing and defining true ras binding domains I: Biochemical analysis. *J. Mol. Biol.* **348**, 741–758 (2005).