The genetic interaction network of mutationally activated

Joshua Cook1,2, Giorgio Melloni3, Peter J. Park3,4\*, Kevin M. Haigis1,2,5,6\*

Mutational activation of the *KRAS* oncogene promotes initiation and/or progression of cancer in a variety of tissues. Though the mutant variants seemingly exert similar biological outputs, the biochemical properties and downstream signaling of each is distinct and highly context-dependent. As such, the genetic interactions associated with *KRAS* mutants are likely to vary according to the specific allele and the tissue-of-origin of the cancer. To explore this concept, 13,492 samples were collated from four tumor types with the highest frequency of mutation in *KRAS*: colorectal adenocarcinoma, lung adenocarcinoma, multiple myeloma, and pancreatic adenocarcinoma. Each cancer had a distinct spectrum of *KRAS* activating mutations that could not be predicted by the prevalence of known mutagenic mechanisms. Moreover, each allele was associated with a distinct comutation network that was also tissue-specific. Analyzing genetic dependencies highlighted cellular functions and individual genes that were or were not required for tumors with specific *KRAS* alleles. Overall, this analysis demonstrates that the *KRAS* alleles have distinct genetic interactions likely linked to their biological differences that can be further investigated as therapeutic targets.

1. Department of Cancer Biology, Dana Farber Cancer Institute, Boston, Massachusetts. 2. Department of Medicine, Harvard Medical School, Boston, Massachusetts. 3. Department of Medical Informatics, Harvard Medical School, Boston, Massachusetts, USA. 4. Ludwig Center at Harvard, Boston, MA 02115, USA. 5. Broad Institute, Cambridge, Massachusetts, USA. 6. Harvard Digestive Disease Center, Harvard Medical School, Boston, Massachusetts.

\*corresponding author(s): Kevin M. Haigis (khaigis@bidmc.harvard.edu) Peter J. Park (peter\_park@hms.harvard.edu)

# 

Sitting at a critical signaling junction between extracellular growth receptors and pro-growth pathways, *KRAS* is one of the most commonly mutated genes in cancer.1,2 However, it is only found frequently mutated in just a few cancers: colorectal adenocarcinoma (COAD), lung adenocarcinoma (LUAD), multiple myeloma (MM), and pancreatic adenocarcinoma (PAAD) being among the cancers with the highest rate of oncogenic *KRAS* mutations. When mutated at one if its four hotspot codons, 12, 13, 61, and 146, *KRAS* is thought to hyperactivate many downstream effector pathways, for instance, the MAPK and PI3K-Akt signaling pathways.3 However, the mutations found in *KRAS* vary substantially across cancers, pointing to significant differences in signaling behavior that complement the environment of the specific cellular context.

Previous studies have documented substantial differences in the biochemistry and signaling properties of the common KRas variants (extensively reviewed by4,5). KRas operates as a molecular switch, activating downstream pathways when bound to GTP, but inactive when GDP-bound following the hydrolysis of the -phosphate. This reaction is catalyzed by a GTPase-activating protein (GAP) and the exchange of the GDP molecule for a new GTP molecule is facilitated by a guanine nucleotide exchange factor (GEF).1 Mutations to any of the four hotspot codons causes an increase in downstream pathway activation by increasing the steady-state amount of GTP-bound KRas. More specifically, mutations to codons 12, 13, and 61 reduce the rate of intrinsic and GAP-mediated hydrolysis, and mutants at 13 and 61, but not 12, also enhance the rate of GDP exchange.6,7 Further, A146 mutations do not alter the rate of GTP hydrolysis, but cause hyperactivation almost entirely though an increased rate of GDP exchange.8–11 Additional biochemical, structural, and signaling distinctions have been identified between different mutant alleles at the same amino acid position.5,6,11–18

Likely as a consequence of their distinct properties, associations have been uncovered between the specific *KRAS* mutation status of a patient’s cancer and its drug-response or clinical outcome.5,19 For instance, a retrospective study indicated that COAD tumors with a *KRAS* G13D allele are sensitive to anti-EGFR therapies, a treatment generally discouraged for *KRAS*-mutant tumors.20 It has recently been proposed, via computational and experimental means, that differential interaction kinetics between KRas G13D and the Ras GAP NF-1 explain this effect.21–23 Another example is that the *KRAS* G12D allele is associated with reduced overall survival in advanced PAAD when separately compared to patients with WT *KRAS*, *KRAS* G12R, or *KRAS* G12V.24 So far, the hypothesis has been that the different biological properties of the mutant *KRAS* alleles is the cause of these clinical distinctions. However, it is also possible that allele-specific genetic interactions drive the varying clinical outcomes.

For the reasons noted above, understanding the heterogeneous properties of the *KRAS* alleles is likely essential to effectively treating *KRAS*-driven cancers. The prominent reductionist strategy of treating all *KRAS* mutants the same has so far proven insufficient. Thus, the current study describes genetic interactions found in tumors with the common *KRAS* alleles in COAD, LUAD, MM, and PAAD. The origins of *KRAS* mutations were studied to assess the extent to which latent mutational processes determined the frequency of the mutant alleles. Further, comutation networks were constructed for each *KRAS* allele and interrogated to identify properties of the alleles. Finally, allele-specific genetic dependency interactions were analyzed to identify potential drug targets. Integrating these two forms of genetic interactions highlighted the distinct effects of each *KRAS* allele on the genetic landscape, and thus behavior, of the tumor.

# Results

## *KRAS* alleles are non-uniformly distributed across cancers.

This study utilized publicly available sequencing data from COAD, LUAD, MM, and PAAD. There were whole exome or genome data available for 1,536 COAD (1,280 after removing hypermutated samples), 891 LUAD, 1,201 MM, and 1,395 PAAD samples. In addition, there were targeted-sequencing data available for 3,329 COAD (2,865 after removing hypermutated samples), 4,160 LUAD, 61 MM, and 919 PAAD sample. More information on the sequencing data is available in the Methods and Supplemental Data.

In the data collected for this study, *KRAS* was most frequently mutated at four “hotspots.” Of these hotspots, codon 12 mutations accounted for 78.9 % of all mutations followed by codon 13 (10.5 %), 61 (7.5 %) and 146 (3.1 %), when adjusted for the incidence of the cancer (Fig. [1](#fig:mutational-signatures-main)a, Supplementary Fig. [7](#sfig:expanded-kras-allele-distribution)). Further, there was substantial variability of the alleles found at these hotspots across the *KRAS*-driven cancers (Fig. [1](#fig:mutational-signatures-main)b, Supplemental Fig. [7](#sfig:expanded-kras-allele-distribution)). Notably, the most variation in *KRAS* alleles was found in MM, and it was the only cancer where a non-G12 allele was the most frequent. COAD had a unique enrichment of G13D and A146T alleles while PAAD was unique in its high frequency of G12R mutations.

## The *KRAS* alleles have different mutagenic origins.

Most *KRAS* mutations were caused by single nucleotide variants. An exception was the exceedingly rare, though transformative *in vitro*,1 G12F allele (0.4% of all *KRAS* mutations) caused by the dinucleotide substitution c.34\_35GG>TT. Glycine 12 and 13 can be transformed to six different amino acids (A, C, D, R, S, and V) through single nucleotide changes in the first two guanine residues. Glutamine 61 can be mutated to six other amino acids (E, H, K, L, P, and R) and a stop codon via a single nucleotide mutation. Alanine 146 can become one of six other amino acids (E, G, P, S, T, and V) from mutations to a single nucleotide.

One explanation for the distinct allelic frequencies across cancer types is that tissue-specific mutational processes drive the selection for mutations. To explore this hypothesis, the active mutational processes in individual tumor samples were elucidated using mutational signatures.25 Briefly, all single-nucleotide mutations can be represented by the combination of the six possible pyrimidine to purine base substitutions (C>T, C>A, C>G, T>A, T>C, T>G) and all possible 3’ and 5’ flanking bases. This composes a mutational spectrum with 96 possible trinucleotide contexts. The signatures were discovered using non-negative matrix factorization and measured in each sample using non-negative least squares regression (see methods for the complete details). The distribution of the levels of each mutational signature were generally uniform within a cancer, regardless of the *KRAS* allele of the tumor (Supplementary Fig. [8](#sfig:mutational-signatures-summary)a, b, c). An apparent exception was for microsatellite instable (MSI) tumors, where the signatures associated with that characteristic dominated (these samples were only included in Supplementary Fig. [8](#sfig:mutational-signatures-summary)a and b, but were removed for the rest of the study). For each cancer with a *KRAS* mutation, the probability that the allele was caused by each detectable mutational process was calculated. The average of these probabilities for each allele are shown in Fig. [1](#fig:mutational-signatures-main)c. Most of the common *KRAS* mutations in COAD, MM, and PAAD, were likely caused by mutations in “clock-like” signatures 1 and 5, mutations believed to accumulate with age26 (Supplementary Fig. [8](#sfig:mutational-signatures-summary)b). LUAD was the only cancer with *KRAS*-mutant samples enriched for a mutational signature of exogenous cause: the *KRAS* G12A/C/V and G13C mutations were primarily attributable to mutations caused by tobacco smoke (signature 427) while *KRAS* G12D mutations were most likely attributable to clock-like mutations (Fig. [1](#fig:mutational-signatures-main)c, d). Signature 8, of unknown etiology, had a substantial probability of causing some of the *KRAS* alleles across all four cancers.

There were some interesting links between specific alleles and mutational signatures. For example, in COAD and PAAD, signature 18, likely caused by damage from reactive oxygen species,28,29 was strongly associated with G12C mutations (Fig. [1](#fig:mutational-signatures-main)c, d). This corroborated the previous finding that *KRAS* G12C mutations were more frequent in patients with MUTYH-Associated Polyposis,28 a recessive autosomal disease caused by biallelic loss-of-function mutations to the gene encoding the DNA glycosylase, *MUTYH*, responsible for clearing 8-oxoguanine:A mismatches that can cause the G12C mutation. In MM, signature 9, associated with mutations introduced by polymerase repair of activation-induced deaminase (AID) activity,25,30,31 was strongly linked with Q61H (Fig. [1](#fig:mutational-signatures-main)c, d), the most common *KRAS* mutation in that cancer.

## The frequency of most *KRAS* alleles cannot be attributed to the prevalence of detected mutagens.

Whether mutational signatures represent the mechanism driving *KRAS* allelic diversity between tissues was further analyzed by calculating the predicted frequency of each allele based on the frequency of mutations in the same trinucleotide context throughout the genome (Fig. [1](#fig:mutational-signatures-main)e, f). The first null hypothesis tested was that, assuming the cancer would acquire a *KRAS* mutation, any of the common alleles was sufficient. Thus, the frequency of the *KRAS* alleles would be determined by the mutational processes alone (Fig. [1](#fig:mutational-signatures-main)e). The second null hypothesis was similar, but restricted to just codon 12 mutations (Fig. [1](#fig:mutational-signatures-main)f). The predicted frequencies were compared against the observed allele frequencies. The alleles above the diagonal line were predicted to be more frequent than observed, while those below the line were more frequently observed than predicted by the mutational signatures. In COAD, G13D was predicted to be the most frequent allele, and G12D/V mutations were considerably underestimated. However, if only codon 12 mutants were considered, the null hypothesis could not be rejected for G12D and G12V (binomial test, p < 0.05, triangles). It should be noted that the predictions have large 95 % confidence intervals that suggest a high amount of variation amongst the samples. Inversely, the frequencies of G12S and A146T mutations were significantly overestimated in COAD (binomial test, p < 0.05, circles). In LUAD and MM, most of the frequencies were predicted in the correct order, though the frequencies were significantly different from the predicted values in most cases (binomial test, p < 0.05, circles). This is especially apparent in MM when only considering the codon 12 mutants. In MM, the most frequent allele, Q61H, was dramatically underestimated with a predicted frequency of 22 % but actual frequency of 45 % of *KRAS* mutations. In PAAD, all of the alleles were observed at a significantly different frequency than predicted by mutational signatures for both null hypotheses. Thus, while it was likely that the active mutational processes in a tissue contributed to determining which *KRAS* mutation was gained, they were unlikely to be deterministic. This observation suggests that the particular biologic properties of the alleles drive their selection, warranting further investigation into their genetic interactions.

## The *KRAS* alleles have distinct comutation networks.

We reasoned that if biological selection is driving allele *KRAS* selection in cancer, then distinct functions of each mutant form of K-RAS would be reflected in the cooperating genetic events. An increased frequency of comutation with another gene suggests a cooperative effect, whereas a reduced frequency of comutation suggests that either the second event is functionally redundant or introduces an inhibitory effect on growth. The extreme of the latter effect is commonly known as "mutual exclusivity." For instance, in COAD, *APC* comutation enhances the effects of oncogenic *KRAS*-induced hyperactivation of the Wnt signaling pathway, essential for the growth of cancer stem cells in the intestinal crypts.32–35 Alternatively, the activation of *BRAF* via a V600E mutation was demonstrated to induce senescence in the presence of a *KRAS* mutant, and, thus, the two are rarely found in the same tumor.35–37

The result of the comutation analysis on COAD tumors was a weakly connected network of the *KRAS* alleles with only a few genes linking the alleles together (Fig. [2](#fig:coad-comutation-main)a). These linking genes tended to be well-studied oncogenes such as *BRAF*, *APC*, and *TP53*. Contrary to a common assumption, while *KRAS* and *TP53* are frequently found mutated in the same tumor, there is a detectable reduction in comutation between *TP53* with *KRAS* G12D and G13D compared to the rest of the alleles (Fig. [2](#fig:coad-comutation-main)a, b). Consistent with the idea that each allele is functionally distinct, there were a substantial number of genes detected to comutate with just one *KRAS* allele. To gain functional insight into the network, genes known to physically interact with KRas,14 signal up- or downstream of KRas,38,39 or are known oncogenes40,41 were extracted (Fig. [2](#fig:coad-comutation-main)b). Several alleles had reduced comutation with *NRAS* and *BRAF* and increased comutation with *APC* and *PIK3CA*.32,34–37,42–47 Some novel interactions included increased comutation of *PORCN* with *KRAS* A146T, *MTOR* with G12C, and *SMAD4* with G12V. Further, several of the alleles showed enrichment for cellular functions in their comutation networks (Fig. [2](#fig:coad-comutation-main)c). One of the strongest effects was an enrichment in the G12D comutation network of interactors with *YWHAZ*, a 14-3-3 scaffolding protein implicated in modulating many interactions including Arhgef7 activity on Rac1 in regulating membrane dynamics for activities such as phagocytosis and cell adhesion.48 Also, genes involved in the Hippo and Wnt signaling, key pathways in COAD, were enriched in the comutation networks of *KRAS* G12V. The comutation network of the G13D allele was enriched for genes implicated in apoptosis and senescence.

The *KRAS* allele-specific comutation network uncovered in LUAD was far larger than that of COAD (Fig. [3](#fig:luadmm-comutation-main)a). This was likely caused by the higher mutation rate in this cancer (Supplemental Fig. [9](#sfig:mutation-burden-of-cancer-samples)b), increasing the power to detect both increased and reduced comutation interactions. Clinical data from TCGA49 were collected to assess the effect of these comutation events on the overall survival (OS) of patients with LUAD. There was no detectable difference in survival between patients with *KRAS* mutant and WT tumors (log-rank test, p-value: 0.290), though, when stratified by *KRAS* allele, the strongest contrast was between G12C and WT tumors (log-rank test, p-value: 0.052) (Supplementary Fig. [10](#sfig:luad-comutation-supplementary)). Thus, for each gene found to comutate with G12C, the patients were divided into four groups: those with WT *KRAS* and interacting gene (grey), *KRAS* G12C and WT interacting gene (black), *KRAS* WT and mutated interacting gene (pink), or *KRAS* G12C and mutant interacting gene. Several were found to have statistically significant differences between these groups (Cox proportional hazards, p-value < 0.05), four of which are shown in Fig. [3](#fig:luadmm-comutation-main)b. *CHRNB4*, *VN1R2*, and *ZNF445* were found to have increased rates of comutation with G12C, and patients with these comutations tended to have worse overall survival than the other groups. Alternatively, *ZNF804A* had reduced comutation with the G12C allele, and the patients with both *KRAS* G12C and mutated *ZNF804A* tended to have better overall survival than the other groups. While these are relatively rare events and consequently the survival data is sparse, these trends lend support to the cooperative or inhibitory effect of these *KRAS* allele-specific comutation interactions.

There were several intriguing cellular processes enriched in the LUAD networks for each allele (Fig. [3](#fig:luadmm-comutation-main)c, d). Interestingly, some of the enrichments were driven by *both* increased and reduced comutation interactions, while others were driven by increased *or* reduced comutation interactions (Fig. [3](#fig:luadmm-comutation-main)d). For example, *KRAS* G12C had increased and reduced comutation interactions with many genes encoding proteins that interact with Myc ("PPI of MYC (TF)"), compared to how reduced comutation interactions drove the enrichment of focal adhesion genes with *KRAS* G12D (Fig. [3](#fig:luadmm-comutation-main)d, Supplemental Fig. [10](#sfig:luad-comutation-supplementary)).

In MM, the *KRAS* comutation network was sparse and highly disconnected (Fig. [3](#fig:luadmm-comutation-main)e). In fact, the only three alleles that shared a detectable interaction were *KRAS* G12D, Q61L, and Q61R, which all demonstrated strongly reduced comutation with *NRAS*, a gene previously reported to be mutually exclusive with *KRAS* in this and other cancers.50 A major limitation of this study was that MM is known to be highly heterogeneous, often with multiple subclonal populations experiencing parallel evolution.50–56 Thus, some comutation events were, in fact, mutations acquired by distinct populations in a single patient, potentially obfuscating comutation interactions. Due to this caveat, focusing on genes known to be recurrently mutated in MM reduces the chance of highlighting a false positive. Previous studies have indicated that mutations in *ACTG1*, *ATR*, *BRAF*, *CYLD*, *DIS3*, *FAM46C*, *NBEA*, *NRAS*, *PRDM1*, *RB1*, *SOX21*, *TP53* and *TRAF3* tend to drive progression of MM41,50 and several showed patterns of comutation with the *KRAS* alleles. *NRAS* had reduced comutation with *KRAS* G12D, Q61L, and Q61R, but one of the highest rates of comutation (18.5 %) with *KRAS* Q61H, the most common *KRAS* mutation in MM (Fig. [3](#fig:luadmm-comutation-main)f). Interestingly, this was just below the rate of *NRAS* mutation in *KRAS* WT tumors (23.6 %), suggesting that the signaling of the Q61H allele is fundamentally different from the other *KRAS* mutations in MM, especially G12D. Indeed, the Q61H allele showed a reduced frequency of comutation with *BRAF* (1.2 %) and *TRAF3* (3.7 %), whereas the G12D allele had a comutation frequency of 9.4 % with both of those genes, again suggesting varying signaling properties of these two common *KRAS* alleles in MM.

The *KRAS* allele comutation network found in the PAAD tumor samples demonstrated that many genes had detectable comutation interactions with multiple alleles (Supplemental Fig. [11](#sfig:paad-comutation)a). Most of these interactions were reduced comutation interactions. There were numerous genes that had opposing comutation interactions with different alleles. Four of these were direct interactors of *KRAS*,14 signal through *KRAS*,38,39 or are known oncogenes40,41 (Supplemental Fig. [11](#sfig:paad-comutation)b, c). Notably, while *TP53* tended to comutate with *KRAS* G12V, it was at a significantly lower rate than expected by random chance, given the overall mutation rate of *TP53* and the mutational burden of the tumors (Supplemental Fig. [11](#sfig:paad-comutation)c). There were many notable cellular functions and processes enriched in the comutation networks of the *KRAS* alleles (Supplemental Fig. [11](#sfig:paad-comutation)d) including a strong enrichment of genes involved in calcium ion transport in the network of G12V (Supplemental Fig. [11](#sfig:paad-comutation)e). Also, the PPI of SMAD1-3 was enriched in the comutation networks of multiple alleles, though the genes that caused the enrichment varied for each protein (Supplemental Fig. [11](#sfig:paad-comutation)f). For instance, the comutation events of *ACVR1B* with *KRAS* were primarily with Q61H, whereas those with *FLNA* were mostly with G12R. These subtle differences suggest that specific and nuanced alterations of SMAD signaling best complement a given *KRAS* allele in PAAD.

## *KRAS* allele-specific genetic dependencies reveal potential synthetic lethal vulnerabilities.

The perturbations necessary to drive cancer expose vulnerabilities not present in the native cell. For example, the microsatellite instability that often leads to cancer simultaneously makes the inhibition of Werner syndrome ATP-dependent helicase (WRN) lethal to the tumor cells.57,58 As the *KRAS* alleles have measurably different signaling behavior and genetic interactions, they likely have specific genetic vulnerabilities. To this end, data from a genome-wide, CRISPR-Cas9 knock-out screen of cancer cell lines59,60 were used to identify genes with *KRAS* allele-specific genetic dependencies.

For COAD, there were only a sufficient number of cell lines (at least 3) with *KRAS* G12D, G12V, and G13D mutations and WT *KRAS* for this analysis. Measuring for gene set enrichment revealed strong patterns in differential dependency of various cellular processes (Fig. [4](#fig:coadluad-dependency-main)a). For example, many genes comprising Complex I of the electron transport chain had a greater lethal effect when knocked out in cell lines with *KRAS* G12V mutations than *KRAS* G12D, G13D or WT cell lines (Fig. [4](#fig:coadluad-dependency-main)b). Alternatively, the *KRAS* G13D cell lines were less affected when genes involved in the complement immune pathway were targeted (Fig. [4](#fig:coadluad-dependency-main)b). To discover individual genes as potential drug-screen targets, each gene was tested for differential genetic dependency with the cell lines grouped by *KRAS* allele. The resulting 77 genes were hierarchically clustered into six groups by their dependency scores (Figure [4](#fig:coadluad-dependency-main)c, d). Genes in clusters 1 and 3 tended to have reduced genetic dependency scores in cell lines with *KRAS* G12D and G12V, respectively, compared to the other cell lines. Alternatively, clusters 4, 5, and 6 demonstrated increased dependency in *KRAS* G13D, G12V, and G12D cell lines, respectively, compared to the other cell lines. Genes in cluster 2 tended to have a stronger dependency in *KRAS* WT cell lines. One notable gene with allele-specific associations was the kinetochore-associated protein (*KNTC1*), a regulator of the mitotic checkpoint,61–63 which demonstrated moderate to strong lethal effects when knocked out in almost every cell line except for those with a *KRAS* G12V allele. Also, isocitrate dehydrogenase (*IDH1*), a component of the citric acid cycle in cellular carbon metabolism,64 was found to have a negligible effect on growth reduction when knocked-out in *KRAS* G12D cell lines, an intermediate effect in G12V cell lines, and the strongest effect in G13D and WT cell lines. Interestingly, a negative regulator of the MAPK pathway,65 *WDR26*, was found to be almost essential for the *KRAS* G12D cell lines, though resulted in more moderate growth reduction when knocked out in the other cell lines. Increased expression of *WDR26* has previously been implicated in driving breast cancer by serving as a scaffolding protein in the PI3K-Akt pathway,66 though there does not appear to be a link between RNA levels and genetic dependency in the COAD cell lines (Supplementary Fig. [12](#sfig:coad-dep-wdr26)a).

The same analysis was conducted on the LUAD cell lines with WT *KRAS* or *KRAS* G12C or G12V mutations. Gene set enrichment analysis of the dependency scores highlighted a reduced dependency on the genes in the p53 hypoxia pathway (Fig. [4](#fig:coadluad-dependency-main)f) and an increased dependency for genes in the Bard1 pathway (Fig. [4](#fig:coadluad-dependency-main)g) for G12C cell lines. The former enrichment suggests that these cells would react less severely to inhibition HIF1- and other components regulating the p53 response to hypoxic stress.67–69 However, the latter suggests they are more susceptible to inhibitors of postreplication DNA-damage repair mechanisms because the enrichment was driven by many components of the Fanconi anemia (FA) pathway responsible for resolving DNA interstrand crosslink.70 A gene-by-gene analysis revealed allele-specific genetic dependencies in 583 genes that were further clustered into four groups (Fig. [4](#fig:coadluad-dependency-main)h). The second group was enriched in genes that encode for proteins in the integrin signaling pathway including *RAP2C*, *MAPK8* (JNK1), *MAPK9* (JNK2), *ITGB5* ( integrin), *ELMO2*, and *PTK2* (FAK) (Supplemental Data). Further, cluster 4 contained a large component of the cellular protein-protein interaction network (PPIN) centered around components of the ubiquitination complex including *UBC*, *UBA3*, *UBE2F*, *RNF19B*, *STAMBP*, *USP14*, and *FEM1A* (Fig. [4](#fig:coadluad-dependency-main)i). Notably, the gene encoded by *STAMBP*, STAM-binding protein, is a negative regulator of the PI3K-Akt pathway, and MAPK6 (ERK3), a component of the MAPK signaling pathway was included in this PPI subnetwork.

For the genetic dependency analysis of PAAD, the only *KRAS* alleles with a sufficient number of cell lines were G12D, G12R, and G12V. A gene set enrichment analysis revealed substantial differences in the dependencies of critical cellular pathways (Fig. [5](#fig:paad-dependency-main)a). For instance, the G12D cell lines tended to be more dependent on the pathway involving the G12/G13 subunits of G protein-coupled receptors, which regulate the actin cytoskeleton during proliferation71–73 (Fig. [5](#fig:paad-dependency-main)b). Moreover, the G12R cell lines demonstrated a greater dependency on the genes at the DNA-damage checkpoint between the G2 and M phases of mitosis (Fig. [5](#fig:paad-dependency-main)c). This enrichment is driven by two of the three components of the MRE11-RAD50-NBN (MRN) complex, *MRE11* and *NBN*. Interestingly, the cell lines with *KRAS* G12V mutations were more sensitive to knock-out of genes in the Hedgehog pathway (Fig. [5](#fig:paad-dependency-main)d). Of the genes in this pathway with the greatest level of enrichment, several are specifically involved in cellular adhesion and migration including *CRMP1*, *NRCAM*, *CDK6*, and *NRP1*. More than 90 individual genes demonstrated *KRAS* allele-specific genetic dependency (Fig. [5](#fig:paad-dependency-main)e). Interestingly, the deletion of *JUN* had no effect in *KRAS* G12V cell lines, while it slowed growth in G12D and G12R cell lines. *JUN* encodes the transcription factor c-Jun which was previously demonstrated to play a role in the transcriptional repression of the tumor suppressors *CDKN2A* and *CDKN2B* in *KRAS*-driven COAD.74 This is consistent with the reduced dependency of G12V cell lines on c-Jun NH2 terminal kinases (JNK) activation from the gene set analysis (of note, the enriched gene set did not include *JUN*, itself; Fig. [5](#fig:paad-dependency-main)a, Supplemental Fig. [13](#sfig:paad-dependency-JUN)a). Leading this enrichment was *MAPK8* (JNK-1) which, when knocked out, lead to increased growth in every case, though most strongly in G12V cell lines (Supplemental Fig. [13](#sfig:paad-dependency-JUN)b). These data suggest a reduced dependency on the activation of c-Jun via JNK signaling, potentially pointing to a tumor suppression pathway with greater potency in PAAD expressing *KRAS* G12V.

## An integrated analysis of allele-specific comutation and genetic dependencies.

Integrating the results from the allele-specific comutation analysis with those from the dependency analysis provided further insight into the distinctions between the *KRAS* alleles. Surprisingly, there was little overlap between the genes found to comutate with an allele and those with differential dependency - the only overlap was found with in the genes resulting from analysis of *KRAS* G12C in LUAD (Fig. [6](#fig:results-integration-main)a). One of these genes was *STK11*, the gene encoding STK11 (LKB1), a tumor suppressor that controls the activity of AMP activated protein kinases (AMPK) to regulate cellular processes including metabolism, apoptosis, and the DNA-damage response.75,76 The high rate of comutation between *KRAS* and *STK11* has been documented previously, though not specifically with *KRAS* G12C. Previous studies have indicated unique biological properties of LUAD tumors with mutations in both *KRAS* and *STK11*, including distinct expression profiles,77 worse clinical prognosis,78,79 and reduced response to immunotherapy.80 The results presented here may suggest a unique synergism between the G12C mutant and *STK11* loss-of-function mutations. Further analysis of the comutation and dependency networks may provide deeper insight into this clinically significant association. Importantly, due to the strong influence of smoking-induced mutations to the prevalence of *KRAS* G12C in LUAD, there was no apparent, nor statistically detectable, difference in the types or locations of *STK11* mutations between G12C-mutant samples and the rest of the LUAD tumor samples (Fig. [6](#fig:results-integration-main)b). Thus, it is unlikely that the genetic associations found here were driven by latent mutational processes, but instead they were determined by the selective advantage that concomitant *KRAS* G12C and *STK11* mutations have in a nascent tumor.

To further mine the genetic interactions found for each *KRAS* allele, functional information was incorporated by annotating the nodes of the PPIN with the associations. For each allele in each cancer, the geodesic distance (the number of nodes on the shorted path) between the genes with allele-specific interactions was on average shorter than between nodes selected randomly on the network (Fig. [6](#fig:results-integration-main)c). This observation indicate that, instead of being randomly distributed throughout the PPIN, there were functional patterns in the identified genes. To inspect these cellular functions, subnetworks of the PPIN of the proteins with allele-specific interactions were extracted and compared. For COAD, the largest components of each allele’s subnetwork were centered around KRas and the other oncogenes most frequently mutated in the cancer: p53, BRaf, PI3K, Apc, and NRas (Fig. [6](#fig:results-integration-main)d highlighted in red). Surrounding these nodes were others that had either comutation or dependency interactions with multiple *KRAS* alleles (in brown). Finally, there were communities of nodes with associations to only a single *KRAS* allele, some representing distinct biological functions. For instance, the G12D allele was again associated with cell adhesion and motility via the module consisting of Myosin-6 and 9 (MYH6/9) two components of lamanin (LAMA1/2), and neural cell adhesion molecule L1 (L1CAM) signaled to via 6 integrin (ITGA6), PI3K (PIK3CA), and PIP5K1- (PIP5KA). This subnetwork succinctly demonstrated the overall pattern observed in the analysis of the *KRAS* allele comutation and genetic dependency interactions: the strongest interactions were common within a cancer type, though each allele had distinguishing features pointing to distinct biological properties of the mutation.

# Discussion

This study presents a genetic interaction analysis of the common oncogenic *KRAS* alleles in COAD, LUAD, MM, and PAAD. Measuring the levels of mutational signatures revealed that the cancer-specific distribution of *KRAS* mutations was not determined by the active mutational processes in the tumor samples. This result suggests that the biological properties of the *KRAS* alleles, within the context of the tissue of origin, is an important factor in the positive selection of a *KRAS* mutation during the evolution of a tumor. To investigate these properties, we conducted statistical tests to determine patterns of comutating genes and genetic dependencies for each *KRAS* allele in each cancer. The former identified genes that comutated with specific *KRAS* alleles at an unexpectedly high frequency, suggesting they were alterations that cooperated with the *KRAS* allele to promote positive selection in the tumor. On the other hand, some genes comutated with a *KRAS* allele less frequently than expected by random chance, suggesting they were either functionally redundant mutations or introduced an inhibitory effect on the tumor’s progression. Finally, functional interactions were identified between *KRAS* alleles and cellular processes and individual genes. Together, these findings provide further support to the mounting evidence that the various oncogenic *KRAS* mutations are not biologically redundant, but instead have distinct properties relevant to the treatment of *KRAS* mutant tumors.

In 2018, Pantsar *et al.* used the transition:transversion mutation ratio to demonstrate that the distribution of codon 12 *KRAS* mutations was not random.18 The current study employed mutational signatures to estimate the frequency of all *KRAS* mutations under the null hypothesis that the alleles were functionally equivalent. Even this more nuanced model of the random distribution of mutations was unable to explain the frequency of *KRAS* alleles across the four cancers studied. While this use of mutational signatures was not without precedent,81,82 it had limitations, primarily the assumption that the mutational signatures completely described the mutagenic forces acting on the genome. This constraint was mitigated by the fact that the signatures were extracted from single base substitutions, the same type of mutations that cause the *KRAS* alleles. Still, there are known artifacts from the sample-processing procedure and NMF that prevent the mutational signatures from perfectly modeling the mutational processes in a tumor.26,83,84

The comutation analysis for this study was limited to relatively small somatic mutations while copy number alteration (CNA) and complex structural variants, such as those caused by chromothripsis, were not included. CNA were not included in this analysis determining the copy number of genes and effects of catastrophic chromosomal events is still a current area of study and is primarily limited to WGS data. Thus, it is not clear that including these data would add clarification to this study or, rather, the additional complexity would make the results more difficult to resolve. For the same reason, the comutation analysis did not include the prediction of a mutation’s impact on the encoded RNA or protein; while continually improving, the current tools are known to be imperfect with high rates of false positives and negatives, often contradicting each other.85–90

The analysis of a genome-wide, CRISPR-Cas9, loss-of-function screen of 89 human COAD, LUAD, and PAAD cell lines revealed cellular processes and individual genes with differing degrees of essentiality associated with specific *KRAS* alleles. The goal of this analysis was to identify potential allele-specific drug targets, progressing the concept of precision medicine to incorporate the specific *KRAS* mutation in a patient’s cancer, the relevance of which is discussed below. We also conducted an analysis of the integrated results from the comutation and dependency analyses. Specifically, we focused on the interaction between *KRAS* G12C and *STK11* because of the previously documented importance of this tumor suppressor, but it is just one instance of the greater pattern where slight variations to the oncogenic properties of mutant *KRAS* has far reaching effects. We hypothesize that the differing properties of the KRas mutants result in changes to the behaviour of the local signaling network that propagate out to more distal regions of the cellular PPIN. These network effects "prime" diverse cellular processes to become comutated or essential for the tumor.

The first claim is evidenced by the fact that the different mutant forms of KRas demonstrate distinct kinetic properties regarding GTP hydrolysis and GDP exchange, further complicated by the various preferences of the alleles to interact with GAPs and GEFs.6,7 For instance, KRas G12C exhibited a higher intrinsic rate of GTP hydrolysis than the other common G12 and G13 alleles, though in the presence of P120GAP, the alleles had more similar rates, with KRas G12A, G12D, and G13D all hydrolyzing GTP faster than G12C.6 Further, our lab and others have demonstrated the A146T mutant had a higher rate of exchange of GDP for GTP than both WT KRas and the G12D mutant.8–11 This mutant form also had a greater response to interactions with the GEF SOS1 compared to the two other proteins.

The effects of these differential kinetics likely result in differences in the immediate signaling behavior of the mutant KRas protein.5,19 For instance, Q61L/R mutations achieved greater activation of the MAPK pathway than G12D/V mutations in LUAD cell lines.91 Further, expression of G12C mutant KRas resulted in greater concentration of phospho-ERK compared to G12D or G12V.92 There are also examples of the KRas mutant proteins having different binding preferences. While all of the G12 and Q61 mutants had greater affinity for the Ras Binding Domain (RBD) of Raf than WT KRas, the range extended from a 120 % to a 734 % increase over the WT affinity.6 Also, LUAD cell lines expressing KRas G12D demonstrated increased activation of Akt via phosphorylation at T308 and S473, while KRas G12C had increased interactions with RalA and RalB. Other physical properties of KRas, such as dimerization, nanoclustering, and interactions with the cell membrane, have yet to be studied in an allele-specific fashion, though additional distinctions may be found.

The long-range interactions between *KRAS* alleles and other cellular functions are more difficult to assess given the complexity of the interactome. One possible investigatory mechanism has been to examine survival data. Comparing the progression free survival of patients with LUAD treated with platinum-based chemotherapy revealed a greater response by patients whose tumors harbored *KRAS* G12V mutations.93,94 This agreed with earlier *in vitro* experiments of the effects of cisplatin on cell lines with ectopic overexpression of KRas G12C, G12D, and G12V.95 The reduced efficacy of these DNA crosslinking drugs was later linked to increased activity of base excision repair clearing the intermediate cisplatin-nucleotide adduct beforethe final crosslinked product could form.96 In our analysis, we uncovered an increased reliance of *KRAS* G12C LUAD cell lines on the Bard1 pathway, primarily driven by genes from the FA pathway, whose primary function is to resolve DNA interstrand crosslinks.70 While Caiola *et al.* did not find a link between the activity of the FA pathway in G12C cell lines, our results, using a different experimental system, suggest that this DNA repair mechanism is also implicated in the reduced response of *KRAS* G12C mutant tumors to cisplatin. Another example of using survival analysis to identify large-scale differences between *KRAS* alleles was that patients with PAAD tumors carrying *KRAS* G12D mutations had worse overall survival than those with *KRAS* G12R, G12V, or WT;24 this trend was also recapitulated using TCGA PAAD data (data not shown). This trend was maintained for those patients who did or did not receive chemotherapy, and the mechanistic cause has yet to be resolved.

Another approach that has been used to identify large-scale distinctions between *KRAS* alleles is to create the alleles in otherwise isogenic model systems and measure the effects experimentally. Using this approach, Hobbs *et al.* demonstrated that while all of their PAAD cell lines were dependent upon macropinocytosis, only those with *KRAS* G12D or G12V, but not G12R, were dependent upon *KRAS* expression for this cellular process.12 They experimentally demonstrated that this was due to the unique conformation of Switch II of KRas G12R interfering with interactions with PI3K, but were unable to elucidate the mechanism behind how this occluded interaction resulted in the allele-specific dependency of macropinocytosis. Still, this an example of an experimentally verified interaction between a specific *KRAS* allele and a seemingly separate cellular process due to a perturbation in the local KRas signaling network. Our lab recently studied the proteomic and phosphoproteomic effects of KRas G12D compared to KRas A146T in neoplastic and tumorigenic contexts in the colon and pancreas.11 This analysis revealed differential activation of various signaling pathways including increased activation of p90RSK in *KRAS* A146T colonic epithelium compared to *KRAS* G12D. More nuanced differences were also uncovered including changes in the targets of activated ERK2.

Taken together, it is clear that the oncogenic KRas mutants have different effects on the local signaling network including differential GTPase and GDP-exchange kinetics, binding affinity, and signaling intensity. As such, we also found evidence for the different local effects of the various KRas mutants represented by the clusters of allele-specific interactions within the immediate neighborhood of KRas in the PPIN (Fig [6](#fig:results-integration-main)d). Further, there is substantial experimental evidence that these relatively minor changes have significant and diverse global impacts on the *KRAS* mutant cell. This study highlighted many novel interactions beyond the KRas signaling pathways by identifying allele-specific patterns of comutation and genetic dependency. Thus, as supported by the analyses of patient outcomes, it has become apparent that not only are the *KRAS* alleles biologically distinct, they impose specific impacts on the behavior of the tumor.

# Methods

## Cancer sample data sources and acquisition

Whole genome sequencing (WGS), whole exome sequencing (WES), and targeted gene panel sequencing data were collected of colorectal adenocarcinoma (COAD), lung adenocarcinoma (LUAD), multiple myeloma (MM), and pancreatic adenocarcinoma (PAAD). WES and WGS data were downloaded from cBioportal,97,98 which included relevant projects from The Cancer Genome Atlas (TCGA)47,49,99 and other smaller studies. Additional data were acquired from the International Cancer Genome Consortium (ICGC) for pancreatic cancer100 and colorectal cancer. MM WES data were gathered from the Multiple Myeloma Research Foundation (MMRF)-CoMMpass online repository.101 Panel data for multiple cancers were retrieved from AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE).102 GENIE data are an aggregation of several different panels ranging from 30 to 600 genes. *KRAS* was included in all of the libraries. A detailed list of all cancer studies can be found in Supplementary Data.

## Hypermutated sample cutoff

Some of the COAD samples had 5 to 10-times more mutations than average, often due to microsatellite instability (MSI). A Gaussian mixed model was used to find the optimal cutoff based on available WGS and WES data. The top 17 % and 21 % of samples were considered hypermutants in WGS and WES, respectively. The same 17 % cutoff was applied to the gene panel data. Hypermutants were not excluded from the identification of mutational signatures because signature 6 (marked as "MSI") is caused by MSI.

## Tissue gene expression filter

A conservative filter for tissue-specific gene expression was used to remove genes not expressed in the tissues of study. Normal tissue gene expression data was gathered from the GTEx Portal (12/03/2018)103 and The Human Protein Atlas (HPA, 12/03/2018),104,105 and tumor expression data was collected from MMRF-CoMMpass (01/14/2019), TCGA-COAD, TCGA-LUAD, and TCGA-PAAD.47,49,99,101 Supplementary Data indicates the number of samples per tissue in the GTEx and tumor gene expression data. A gene was considered “expressed” in a tissue if it met at least one of the following criteria: 1) a median expression level of at least 1 TPM across all samples of the tissue in GTEx, 2) indicated as expressed at at least 1 TPM in the HPA data set for the tissue, 3) expressed with a median level of 1 batch-normalized raw counts (using RSEM) in the corresponding tumor RNA-sequencing data.

## Predicting effect of mutations on gene or protein function

The effect of a mutation on the function of a gene or encoded protein was predicted using SIFT,106,107 PolyPhen2,108 LRT,86 MutationTaster,109 MutationAssessor,110,111 FATHMM,112 MetaSVM,85 and MetaLR.85 Known clinically significant mutations were acquired from ClinVar.113 Annotations were applied using ANNOVAR.114 A mutation was marked as deleterious if it was predicted as such by at least one of the prediction methods or declared as such by CinVar.

## Protein-Protein Interaction Network (PPIN)

The PPIN used throughout the study was the combination of interactions from STRING,115,116 HINT,117 and BioPlex.118 The full edge list is available in the Supplementary Data.

## Identifying mutational signatures

The genome-wide mutations of a sample can be deconvolved into mutational signatures that represent endogenous or exogenous mutagenic processes.25 Single nucleotide variants (SNVs) from exomes or genomes were divided into 96 types, according to the 6 mutations of a pyrimidine (C>A, C>G, C>T and T>A, T>C, T>G) and the 16 possible combinations of 3’ and 5’ adjacent bases. The MATLAB119 implementation of Non-Negative Matrix Factorization (NMF) algorithm, SigProfiler,25 was used to discover the underlying mutational patterns that are common across tumors. Mutational signatures were discovered separately for each tumor type and the optimal number of signatures was determined based on silhouette width and Frobenius error.

The spectrum of the signatures discovered by NMF were matched to the COSMIC catalog.120 For the signatures for which none of the 30 signatures in COSMIC catalog was found to be compatible, we referred to more recent studies in literature and expanded upon the COSMIC catalog. In particular, there were multiple subtypes of signature 7 (reported previously in.84,121 Further, the analysis revealed a signature that was predominantly C>A but not a subtype of signature 7. This signature 38 was previously reported to be caused by indirect UV exposure.84 Three versions of the signature associated to POLE mutations, signature 10, were discovered (previously reported in.84 These three POLE signatures differed in the C>A, C>T or C>G parts of the mutational spectrum. In LUAD, a signature with mutations of type C[C>A]N and T[C>A]N attributable to 8-oxo-guanine84 was found. One signature that we discovered COAD did not have a good match in any specific signature in literature, although it resembled a signature previously reported to be caused by SBSA122 and signatures 34 and 41 in reference.84 This signature was not adjusted to resemble those previously reported because the results from different studies were not in strong agreement. This signature, referred to as "N," did not contribute to *KRAS* mutations. Three of the signatures discovered via NMF were likely to be artifacts123 and were removed from downstream analysis. Signatures present at very low levels were removed. The levels of each signature in each tumor sample were calculated using Non-Negative Least Squares.124

## Probability of *KRAS* mutations from mutational signatures

For each sample harboring a *KRAS* mutation, the probability of occurrence given the mutational signatures present was calculated by considering the weight of the base change among the 96 possibilities and the relative contribution of the signature to the mutations in the sample. Thus, the probability of a tumor sample acquiring the *KRAS* mutation from signature from all signatures can be calculated using Eq. [[eq:kras\_mutation\_from\_signature]](#eq:kras_mutation_from_signature).

## Predicting *KRAS* allele frequency by mutational signatures

The mutational signatures are linear combinations of the 96-dimension spectrum of possible mutations (see "Identifying mutational signatures" above). Thus, assuming the null hypothesis that the prevalence of active mutational processes alone determines the frequency of *KRAS* alleles in a cancer, the predicted frequency of each *KRAS* allele can be calculated as the frequency of the same mutation across the entire genome. The 95 % confidence intervals were bootstrapped.125

## Comutation with *KRAS* alleles

A one-tailed Fisher’s exact test of independence was used to identify increased frequency of comutation between *KRAS* alleles and other mutated genes. Only comutation partners with at least three comutation events were considered. Further, only genes with a mutation frequency of at least 1 % or a comutation frequency of at least 10 % were considered.

The Row-Column Test for Exclusivity (RC-test) was used to identify reduced frequency of comutation between *KRAS* alleles and other mutated genes.126 This is a permutation-based test that finds the probability of observing the actual number of mutually exclusive events given than the number of time the genes is mutated in all samples is fixed and the number of mutations in each sample is fixed. Thus, the test conditions on both the frequency of mutation of the gene and the mutational burden of the samples. For this reason, only WGS and WES data could be used for this analysis. Only genes with a mutational frequency of at least 2 % and 10 mutually exclusive events were considered.

## Functional enrichment

The R interface to the online *Enrichr* tool was used to identify enriched gene sets in the comutation networks and allele-specific synthetic lethal clusters.127–129 Gene sets from the following sources provided by Enrichr were used: BioCarta (2016), GO Biological Process (2018), KEA (2015), KEGG (2019), Panther (2016), PPI Hub Proteins, Reactome (2016), Transcription Factor PPIs, and WikiPathways (2019).

## Survival analysis

Overall survival data of patients with COAD, LUAD, and PAAD were acquired from TCGA (47,49,99) and of patients with MM from the MMRF-CoMMpass.101 The log-rank test was used when comparing the overall survival of two groups and the likelihood test was used when comparing the overall survival of more than two groups. When the patients were stratified into one of four groups depending on the *KRAS* mutation and mutation status of a comutating gene in their tumor (Fig. [3](#fig:luadmm-comutation-main)), a group was only included in the model if there were at least 5 samples. The ’survival’ package in R was used for computing these statistics.130

## Modeling of cancer cell line genetic dependencies

Genetic dependency data was downloaded from the online DepMap portal (https://depmap.org/portal/download/) (2019 quarter 3). Cell lines with multiple activating *KRAS* mutations or an activating mutation in *BRAF*, *EGFR*, or *NRAS* were removed from the data set. For each cancer, only cell lines with a *KRAS* allele found in at least 3 cell lines were included in the study. The only exception to this was the removal of the LUAD cell lines with *KRAS* G13D mutations because this allele is exceedingly rare in LUAD. This is supported by the fact that knocking out *KRAS* in these cell lines had an equivalent effect than when the gene was knocked out in *KRAS* WT cell lines: the average ( std. dev.) dependency score for G13D LUAD cell lines was -0.55 0.26, compared to that of *KRAS* WT cell lines: -0.55 0.28. The rest of the *KRAS* mutant samples demonstrated a far greater dependency on *KRAS*: -1.26 0.33. The genetic dependency score is often linked to the expression of the gene. Thus, if the RNA expression of the gene could explain the dependency score (linear model, p-value < 0.01), the gene was not tested for *KRAS* allele-specific genetic dependency. Of the remaining genes, an ANOVA was used to measure if the mean dependency scores for the cell lines grouped by *KRAS* allele were different. These genes are declared as deferentially dependent. For these genes, pairwise Student’s t-tests were used to compare the dependency scores of each group. These contrasts were used to decide with which *KRAS* allele(s) a gene shows differential dependency (Benjamini-Hochberg FDR adjusted p-value < 0.05).

## Gene Set Enrichment Analysis (GSEA) of genetic dependency

The GSEA tool (version 3.0) was acquired from the online GSEA portal (https://www.gsea-msigdb.org/gsea/index.jsp). Gene sets were acquired through MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). The analysis used the Hallmark and C2 gene sets and permuted the genes 10,000 times for the statistical test. All other settings were set to default values.

## Code availability

All code is available at https://github.com/jhrcook/comutation. See the README for the organization of the code and how to run the analysis. Python131 and R132 were used for most of the analyses.

# Acknowledgements

The Acknowledgements should contain text acknowledging non-author contributors. Acknowledgements should be brief, and should not include thanks to anonymous referees and editors or effusive comments. Grant or contribution numbers may be acknowledged.

These data were generated as part of the Multiple Myeloma Research Foundation Personalized Medicine Initiative.

# Author contributions

Each author’s contribution to the work should be described briefly, on a separate line, in the Author Contributions section.

# Competing interests

A competing interests statement is required for all papers accepted by and published in *Scientific Data*. If there is no conflict of interest, a statement declaring this must still be included in the manuscript.

1. Barbacid, M. ras genes. *Annual review of biochemistry* **56**, 779–827 (1987).

2. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **174**, 1034–1035 (2018).

3. Simanshu, D. K., Nissley, D. V. & McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17–33 (2017).

4. Miller, M. S. & Miller, L. D. RAS mutations and oncogenesis: Not all RAS mutations are created equally. *Frontiers in Genetics* **2**, 1–9 (2012).

5. Li, S., Balmain, A. & Counter, C. M. A model for RAS mutation patterns in cancers: finding the sweet spot. *Nature reviews. Cancer* **18**, 767–777 (2018).

6. Hunter, J. C. *et al.* Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. *Molecular cancer research : MCR* **13**, 1325–35 (2015).

7. Smith, M. J., Neel, B. G. & Ikura, M. NMR-based functional profiling of RASopathies and oncogenic RAS mutations. *Proceedings of the National Academy of Sciences* **110**, 4574–4579 (2013).

8. Feig, L. A. & Cooper, G. M. Relationship among guanine nucleotide exchange, GTP hydrolysis, and transforming potential of mutated ras proteins. *Molecular and cellular biology* **8**, 2472–8 (1988).

9. Edkins, S. *et al.* Recurrent KRAS codon 146 mutations in human colorectal cancer. *Cancer biology & therapy* **5**, 928–32 (2006).

10. Janakiraman, M. *et al.* Genomic and biological characterization of exon 4 KRAS mutations in human cancer. *Cancer research* **70**, 5901–11 (2010).

11. Poulin, E. J. *et al.* Tissue-Specific Oncogenic Activity of KRASA146T. *Cancer discovery* **9**, 738–755 (2019).

12. Hobbs, G. A. *et al.* Atypical KRASG12R Mutant Is Impaired in PI3K Signaling and Macropinocytosis in Pancreatic Cancer. *Cancer discovery* (2019) doi:[10.1158/2159-8290.CD-19-1006](https://doi.org/10.1158/2159-8290.CD-19-1006).

13. Yu, T. A new dynamic correlation algorithm reveals novel functional aspects in single cell and bulk RNA-seq data. *PLoS computational biology* **14**, e1006391 (2018).

14. Kovalski, J. R. *et al.* The Functional Proximal Proteome of Oncogenic Ras Includes mTORC2. *Molecular cell* **73**, 830–844 (2019).

15. Ihle, N. T. *et al.* Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. *Journal of the National Cancer Institute* **104**, 228–39 (2012).

16. Spoerner, M., Wittinghofer, A. & Kalbitzer, H. R. Perturbation of the conformational equilibria in Ras by selective mutations as studied by 31P NMR spectroscopy. *FEBS letters* **578**, 305–10 (2004).

17. Smith, M. J. & Ikura, M. Integrated RAS signaling defined by parallel NMR detection of effectors and regulators. *Nature chemical biology* **10**, 223–30 (2014).

18. Pantsar, T. *et al.* Assessment of mutation probabilities of KRAS G12 missense mutants and their long-timescale dynamics by atomistic molecular simulations and Markov state modeling. *PLoS computational biology* **14**, e1006458 (2018).

19. Haigis, K. M. KRAS Alleles: The Devil Is in the Detail. *Trends in cancer* **3**, 686–697 (2017).

20. De Roock, W. *et al.* Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. *JAMA* **304**, 1812–20 (2010).

21. McFall, T. *et al.* A systems mechanism for KRAS mutant allele-specific responses to targeted therapy. *Science signaling* **12**, 8288 (2019).

22. Rabara, D. *et al.* KRAS G13D sensitivity to neurofibromin-mediated GTP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 22122–22131 (2019).

23. Zafra, M. P. *et al.* An in vivo KRAS allelic series reveals distinct phenotypes of common oncogenic variants. *bioRxiv* **14**, 847509 (2019).

24. Bournet, B. *et al.* KRAS G12D Mutation Subtype Is A Prognostic Factor for Advanced Pancreatic Adenocarcinoma. *Clinical and translational gastroenterology* **7**, e157 (2016).

25. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–21 (2013).

26. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nature genetics* **47**, 1402–7 (2015).

27. Alexandrov, L. B. *et al.* Mutational signatures associated with tobacco smoking in human cancer. *Science (New York, N.Y.)* **354**, 618–622 (2016).

28. Viel, A. *et al.* A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine* **20**, 39–49 (2017).

29. Pilati, C. *et al.* Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *The Journal of pathology* **242**, 10–15 (2017).

30. Rogozin, I. B. *et al.* DNA polymerase mutational signatures are found in a variety of different types of cancer. *Cell cycle (Georgetown, Tex.)* **17**, 348–355 (2018).

31. Petljak, M. & Alexandrov, L. B. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis* **37**, 531–40 (2016).

32. Janssen, K.-P. *et al.* APC and oncogenic KRAS are synergistic in enhancing Wnt signaling in intestinal tumor formation and progression. *Gastroenterology* **131**, 1096–109 (2006).

33. Fearon, E. R. & Wicha, M. S. KRAS and cancer stem cells in APC-mutant colorectal cancer. *Journal of the National Cancer Institute* **106**, djt444 (2014).

34. Sakai, E. *et al.* Combined Mutation of Apc, Kras, and Tgfbr2 Effectively Drives Metastasis of Intestinal Cancer. *Cancer research* **78**, 1334–1346 (2018).

35. Jauhri, M. *et al.* Prevalence and coexistence of KRAS, BRAF, PIK3CA, NRAS, TP53, and APC mutations in Indian colorectal cancer patients: Next-generation sequencing-based cohort study. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **39**, 1010428317692265 (2017).

36. Seth, R. *et al.* Concomitant mutations and splice variants in KRAS and BRAF demonstrate complex perturbation of the Ras/Raf signalling pathway in advanced colorectal cancer. *Gut* **58**, 1234–41 (2009).

37. Cisowski, J., Sayin, V. I., Liu, M., Karlsson, C. & Bergo, M. O. Oncogene-induced senescence underlies the mutual exclusive nature of oncogenic KRAS and BRAF. *Oncogene* **35**, 1328–33 (2016).

38. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research* **45**, D353–D361 (2017).

39. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research* **44**, 457–62 (2016).

40. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *British journal of cancer* **91**, 355–8 (2004).

41. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature reviews. Cancer* **18**, 696–705 (2018).

42. Sensi, M. *et al.* Mutually exclusive NRASQ61R and BRAFV600E mutations at the single-cell level in the same human melanoma. *Oncogene* **25**, 3357–64 (2006).

43. Kennedy, A. L. *et al.* Activation of the PIK3CA/AKT pathway suppresses senescence induced by an activated RAS oncogene to promote tumorigenesis. *Molecular cell* **42**, 36–49 (2011).

44. Wang, G. M. *et al.* Single copies of mutant KRAS and mutant PIK3CA Cooperate in immortalized human epithelial cells to induce tumor formation. *Cancer Research* **73**, 3248–3261 (2013).

45. Green, S., Trejo, C. L. & McMahon, M. PIK3CA(H1047R) Accelerates and Enhances KRAS(G12D)-Driven Lung Tumorigenesis. *Cancer research* **75**, 5378–91 (2015).

46. Yeang, C.-H., McCormick, F. & Levine, A. Combinatorial patterns of somatic gene mutations in cancer. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **22**, 2605–22 (2008).

47. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–7 (2012).

48. Angrand, P.-O. *et al.* Transgenic mouse proteomics identifies new 14-3-3-associated proteins involved in cytoskeletal rearrangements and cell signaling. *Molecular & cellular proteomics : MCP* **5**, 2211–27 (2006).

49. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–50 (2014).

50. Lohr, J. G. *et al.* Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer cell* **25**, 91–101 (2014).

51. Melchor, L. *et al.* Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. *Leukemia* **28**, 1705–15 (2014).

52. Lionetti, M. *et al.* Molecular spectrum of BRAF, NRAS and KRAS gene mutations in plasma cell dyscrasias: implication for MEK-ERK pathway activation. *Oncotarget* **6**, 24205–17 (2015).

53. Keats, J. J. *et al.* Clonal competition with alternating dominance in multiple myeloma. *Blood* **120**, 1067–76 (2012).

54. Corre, J., Munshi, N. & Avet-Loiseau, H. Genetics of multiple myeloma: another heterogeneity level? *Blood* **125**, 1870–6 (2015).

55. Lohr, J. G. *et al.* Genetic interrogation of circulating multiple myeloma cells at single-cell resolution. *Science translational medicine* **8**, 363ra147 (2016).

56. Xu, J. *et al.* Molecular signaling in multiple myeloma: association of RAS/RAF mutations and MEK/ERK pathway activation. *Oncogenesis* **6**, e337 (2017).

57. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).

58. Chan, edmond M. *et al.* WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* **568**, 551–556 (2019).

59. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564–576 (2017).

60. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature genetics* **49**, 1779–1784 (2017).

61. Chan, G. K. T., Jablonski, S. A., Starr, D. A., Goldberg, M. L. & Yen, T. J. Human Zw10 and ROD are mitotic checkpoint proteins that bind to kinetochores. *Nature cell biology* **2**, 944–7 (2000).

62. Scaërou, F. *et al.* The ZW10 and Rough Deal checkpoint proteins function together in a large, evolutionarily conserved complex targeted to the kinetochore. *Journal of cell science* **114**, 3103–14 (2001).

63. Kops, G. J. P. L. *et al.* ZW10 links mitotic checkpoint signaling to the structural kinetochore. *The Journal of cell biology* **169**, 49–60 (2005).

64. Geisbrecht, B. V. & Gould, S. J. The human PICD gene encodes a cytoplasmic and peroxisomal NADP(+)-dependent isocitrate dehydrogenase. *The Journal of biological chemistry* **274**, 30527–33 (1999).

65. Goto, T., Matsuzawa, J., Iemura, S.-i., Natsume, T. & Shibuya, H. WDR26 is a new partner of Axin1 in the canonical Wnt signaling pathway. *FEBS letters* **590**, 1291–303 (2016).

66. Ye, Y., Tang, X., Sun, Z. & Chen, S. Upregulated WDR26 serves as a scaffold to coordinate PI3K/ AKT pathway-driven breast cancer cell growth, migration, and invasion. *Oncotarget* **7**, 17854–69 (2016).

67. Goda, N. *et al.* Hypoxia-inducible factor 1alpha is essential for cell cycle arrest during hypoxia. *Molecular and cellular biology* **23**, 359–69 (2003).

68. Sermeus, A. & Michiels, C. Reciprocal influence of the p53 and the hypoxic pathways. *Cell death & disease* **2**, e164 (2011).

69. Jing, X. *et al.* Role of hypoxia in cancer therapy by regulating the tumor microenvironment. *Molecular cancer* **18**, 157 (2019).

70. Ceccaldi, R., Sarangi, P. & D’Andrea, A. D. The Fanconi anaemia pathway: new players and new functions. *Nature reviews. Molecular cell biology* **17**, 337–49 (2016).

71. Worzfeld, T., Wettschureck, N. & Offermanns, S. G(12)/G(13)-mediated signalling in mammalian physiology and disease. *Trends in pharmacological sciences* **29**, 582–9 (2008).

72. Siehler, S. Regulation of RhoGEF proteins by G12/13-coupled receptors. *British journal of pharmacology* **158**, 41–9 (2009).

73. Suzuki, N., Hajicek, N. & Kozasa, T. Regulation and physiological functions of G12/13-mediated signaling pathways. *Neuro-Signals* **17**, 55–70 (2009).

74. Serra, R. W., Fang, M., Park, S. M., Hutchinson, L. & Green, M. R. A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *eLife* **3**, e02313 (2014).

75. Momcilovic, M. & Shackelford, D. B. Targeting LKB1 in cancer - exposing and exploiting vulnerabilities. *British journal of cancer* **113**, 574–84 (2015).

76. Korsse, S. E., Peppelenbosch, M. P. & Veelen, W. van. Targeting LKB1 signaling in cancer. *Biochimica et biophysica acta* **1835**, 194–210 (2013).

77. Skoulidis, F. *et al.* Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. *Cancer discovery* **5**, 860–77 (2015).

78. La Fleur, L. *et al.* Mutation patterns in a population-based non-small cell lung cancer cohort and prognostic impact of concomitant mutations in KRAS and TP53 or STK11. *Lung Cancer* **130**, 50–58 (2019).

79. Bange, E. *et al.* Impact of KRAS and TP53 Co-Mutations on Outcomes After First-Line Systemic Therapy Among Patients With STK11-Mutated Advanced Non-Small-Cell Lung Cancer. *JCO precision oncology* **3**, 1–11 (2019).

80. Skoulidis, F. *et al.* STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. *Cancer discovery* **8**, 822–835 (2018).

81. Dietlein, F. *et al.* Identification of cancer driver genes based on nucleotide context. *Nature genetics* **52**, 208–218 (2020).

82. Bergstrom, E. N. *et al.* Generating realistic null hypothesis of cancer mutational landscapes using SigProfilerSimulator. *bioRxiv* 2020.02.13.948422 (2020) doi:[10.1101/2020.02.13.948422](https://doi.org/10.1101/2020.02.13.948422).

83. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nature reviews. Genetics* **17**, 175–88 (2016).

84. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

85. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human molecular genetics* **24**, 2125–37 (2015).

86. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome research* **19**, 1553–61 (2009).

87. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* **7**, 575–6 (2010).

88. Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Human mutation* **32**, 358–68 (2011).

89. Wei, P., Liu, X. & Fu, Y.-X. Incorporating predicted functions of nonsynonymous variants into gene-based analysis of exome sequencing data: a comparative study. *BMC proceedings* **5 Suppl 9**, S20 (2011).

90. Flanagan, S. E., Patch, A.-M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic testing and molecular biomarkers* **14**, 533–7 (2010).

91. Pershing, N. L. K. *et al.* Rare codons capacitate Kras-driven de novo tumorigenesis. *The Journal of clinical investigation* **125**, 222–33 (2015).

92. Li, S. *et al.* Assessing Therapeutic Efficacy of MEK Inhibition in a KRASG12C-Driven Mouse Model of Lung Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **24**, 4854–4864 (2018).

93. Cserepes, M. *et al.* Subtype-specific KRAS mutations in advanced lung adenocarcinoma: a retrospective study of patients treated with platinum-based chemotherapy. *European journal of cancer (Oxford, England : 1990)* **50**, 1819–1828 (2014).

94. Marabese, M. *et al.* KRAS mutations affect prognosis of non-small-cell lung cancer patients treated with first-line platinum containing chemotherapy. *Oncotarget* **6**, 34014–22 (2015).

95. Garassino, M. C. *et al.* Different types of K-Ras mutations could affect drug sensitivity and tumour behaviour in non-small-cell lung cancer. *Annals of oncology : official journal of the European Society for Medical Oncology* **22**, 235–7 (2011).

96. Caiola, E. *et al.* Base excision repair-mediated resistance to cisplatin in KRAS(G12C) mutant NSCLC cells. *Oncotarget* **6**, 30072–87 (2015).

97. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**, pl1 (2013).

98. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401–4 (2012).

99. Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer cell* **32**, 185–203 (2017).

100. Scarlett, C. J., Salisbury, E. L., Biankin, A. V. & Kench, J. Precursor lesions in pancreatic cancer: morphological and molecular pathology. *Pathology* **43**, 183–200 (2011).

101. Walker, B. A. *et al.* A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia* **33**, 159–170 (2019).

102. AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer discovery* **7**, 818–831 (2017).

103. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

104. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)* **347**, 1260419 (2015).

105. Uhlén, M. *et al.* Transcriptomics resources of human tissues and organs. *Molecular systems biology* **12**, 862 (2016).

106. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* **4**, 1073–81 (2009).

107. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nature protocols* **11**, 1–9 (2016).

108. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248–9 (2010).

109. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods* **11**, 361–2 (2014).

110. Reva, B., Antipin, Y. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome biology* **8**, R232 (2007).

111. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research* **39**, e118 (2011).

112. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human mutation* **34**, 57–65 (2013).

113. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic acids research* **46**, D1062–D1067 (2018).

114. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164 (2010).

115. Mering, C. von *et al.* STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic acids research* **33**, 433–7 (2005).

116. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**, D607–D613 (2019).

117. Das, J. & Yu, H. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC systems biology* **6**, 92 (2012).

118. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425–440 (2015).

119. MATLAB. *Version 7.10.0 (r2010a)*. (The MathWorks Inc., 2010).

120. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research* **47**, D941–D947 (2019).

121. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).

122. Lee-Six, H. *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).

123. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic acids research* **41**, e67 (2013).

124. Gulhan, D. C., Lee, J. J.-K., Melloni, G. E. M., Cortés-Ciriano, I. & Park, P. J. Detecting the mutational signature of homologous recombination deficiency in clinical samples. *Nature genetics* **51**, 912–919 (2019).

125. Canty, A. & Ripley, B. *Boot: Bootstrap functions (originally by angelo canty for s)*. (2019).

126. Leiserson, M. D. M., Reyna, M. A. & Raphael, B. J. A weighted exact test for mutually exclusive mutations in cancer. *Bioinformatics (Oxford, England)* **32**, i736–i745 (2016).

127. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* **14**, 128 (2013).

128. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **44**, 90–7 (2016).

129. Jawaid, W. *EnrichR: Provides an r interface to ’enrichr’*. (2019).

130. Therneau, T. M. *A package for survival analysis in s*. (2015).

131. Van Rossum, G. & Drake Jr, F. L. *Python tutorial*. (Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995).

132. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2019).