

The origin, distribution, and genetic interactions of *KRAS* alleles across cancer types

Joshua H. Cook^{1,2,3,7}, Giorgio E. M. Melloni^{3,7}, Doga C. Gulhan³, Peter J. Park^{3,4*}, Kevin M. Haigis^{1,2,5,6*}

June 22, 2020

1. Department of Cancer Biology, Dana Farber Cancer Institute, Boston, Massachusetts. 2. Department of Medicine, Brigham & Women's Hospital, 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. 7. These authors contributed equally.

*corresponding authors:

Kevin M. Haigis (kevin_haigis@dfci.harvard.edu)

Peter J. Park (peter_park@hms.harvard.edu)

Mutational activation of *KRAS* promotes the initiation and progression of cancers, especially in the colorectum, pancreas, lung, and blood plasma, with varying prevalence of specific activating missense mutations. Although epidemiological studies connect specific alleles to clinical outcomes, the mechanisms underlying the distinct clinical characteristics of mutant *KRAS* alleles are unclear. Here, we analyzed 13,492 samples from these four tumor types to examine allele- and tissue-specific genetic properties associated with oncogenic *KRAS* mutations. The prevalence of known mutagenic mechanisms partially explained the observed spectrum of *KRAS* activating mutations, but the prevalence of most alleles across the different cancers could not be explained in this way, suggesting that biological selection underlies the tissue-specific frequencies of mutant alleles. Consistent with experimental studies that have identified distinct signaling properties associated with each mutant form of K-RAS, a genetic analysis revealed that each *KRAS* allele was associated with a distinct and tissue-specific comutation network. Moreover, we identified genetic dependencies associated with specific mutant *KRAS* alleles. Overall, this analysis demonstrates that the genetic interactions associated with oncogenic *KRAS* mutations are allele- and tissue-specific, underscoring the complexity that drives their clinical behaviors.

Located 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 frequently mutated in only a handful of cancers, with the highest frequencies in colorectal adenocarcinoma (COAD), lung adenocarcinoma (LUAD), multiple myeloma (MM), and pancreatic adenocarcinoma (PAAD). Importantly, the mutations found in *KRAS* vary substantially across cancers, pointing to significant differences in signaling behavior of the alleles that complement the environment of the specific cellular context [3, 4].

When mutated at one of its four hotspot codons – 12, 13, 61, or 146 – activated K-RAS protein is thought to hyperactivate many downstream effector pathways, for instance, the MAPK and PI3K-AKT signaling pathways [1]. Previous studies have documented substantial differences in the biochemical and signaling properties of the common K-RAS variants (reviewed by [5, 6]). K-RAS normally operates as a molecular switch, activating downstream pathways when GTP-bound, but inactive when GDP-bound following the hydrolysis of the γ -phosphate. This reaction is catalyzed by GTPase-activating proteins (GAPs), while the exchange of the GDP for a new GTP is facilitated by guanine nucleotide exchange factors (GEFs) [7]. Activating *KRAS* mutations result in elevated downstream pathway activation by increasing the steady-state concentration of GTP-bound K-RAS. More specifically, mutations to codons 12, 13, and 61 reduce the rate of intrinsic and/or GAP-mediated hydrolysis, and mutations at 13 and 61, but not 12, also enhance the rate of nucleotide exchange [8, 9]. Alternatively, 146 mutations do not alter the rate of GTP hydrolysis, but cause hyperactivation through an increased rate of GDP exchange [10, 11, 12, 4]. Additional biochemical, structural, and signaling distinctions have been identified between different mutant alleles, including between those at the same amino acid position [13, 8, 4, 14, 15, 16, 17, 18, 19, 20].

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 [3, 6]. For instance, a retrospective meta-analysis suggested that COAD tumors with a *KRAS* G13D allele are sensitive to anti-EGFR therapies, a treatment generally discouraged for *KRAS*-mutant tumors [21]. It has recently been proposed, via computational and experimental means, that differential interaction kinetics between K-RAS G13D and the Ras GAP NF-1 explain this effect [22, 23, 24]. Another example is that the *KRAS* G12D allele is associated with reduced overall survival in advanced PAAD when compared to patients with WT *KRAS*, *KRAS* G12R, or *KRAS* G12V [25]. Thus 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 essential to effectively treating *KRAS*-driven cancers. Here we describe *KRAS* genetic interactions found in COAD, LUAD, MM, and PAAD. We studied the origins of *KRAS* mutations to assess the extent to which latent mutational processes determined the allelic distribution. Further, we constructed comutation networks for each *KRAS* allele and interrogated to identify properties of the alleles. Finally, we analyzed allele-specific genetic dependencies to identify potential therapeutic 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. We believe that an allele-specific and tissue-specific analysis such as this is necessary to fully understand the nature of the most potent oncogenes.

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 (including 256 hypermutated samples), 891 LUAD, 1,201 MM, and 1,395 PAAD samples. In addition, there were targeted-sequencing data available for 3,329 COAD (including 464 hypermutated samples), 4,160 LUAD, 61 MM, and 919 PAAD samples. More information on the data is available in the Methods and Supplementary Tables 1 and 2.

KRAS was most frequently mutated by single nucleotide substitutions at one of four "hotspots": codons 12, 13, 61, and 146. (Fig. 1a; Supplementary Table 3). 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.

Of these hotspots, codon 12 mutations accounted for 81.7% of all mutations in the dataset, followed by codon 13 (8.4%), 61 (7.3%) and 146 (2.5%). Adjusting for the yearly incidence of each cancer, the distribution of mutations was 76.8%, 11.4%, 8.1%, and 3.7% at codons 12, 13, 61, and 146, across the four cancers. PAAD had the greatest frequency of *KRAS* mutations at 86.3%, followed by COAD (41.4%), LUAD (35.3%), and MM (21.9%)

(Fig. 1b). Further, there was substantial variability of the alleles found at these hotspots across the four *KRAS*-driven cancers (Fig. 1a). 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. At codon 12, LUAD had an enrichment for G12C mutations. COAD had a unique enrichment of G13D and A146T alleles, while PAAD was distinct in its high frequency of G12R mutations.

The *KRAS* alleles have different mutagenic origins.

One explanation for the distinct allelic frequencies across cancer types is that tissue-specific mutational processes determine the distribution. To explore this hypothesis, we elucidated the active mutational processes in the tumor samples using mutational signatures [26] (Supplementary Tables 4 and 5; the signature numbers refer to those in the catalog [27]). 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 comprising this spectrum were discovered in whole exome and whole genome sequencing data using non-negative matrix factorization and measured in each sample using non-negative least squares regression (see Methods; Supplementary Fig. 1).

As expected, the distributions of the levels of each mutational signature were highly variable across tumor types. The most common in COAD, MM, and PAAD, were the "clock-like" single base substitution (SBS) signatures SBS1 and SBS5, which are believed to accumulate with age [28]. LUAD was enriched for a mutational signature of exogenous cause, tobacco smoke carcinogens (SBS4). Within each cancer type, the relative abundance of the mutational signatures were generally uniform, regardless of the *KRAS* allele (Fig. 1c). An exception was for cancers with microsatellite instability (MSI), in which defective DNA mismatch repair and other related signatures dominated (Supplementary Fig. 1a and b). Thus, for each cancer, the allelic frequency of *KRAS* was not primarily caused by differential activity of mutational processes in individual tumors.

Each mutational process is not equally likely to cause each *KRAS* allele. The probability that the allele for an individual tumor was caused by any detectable mutational process was calculated using the trinucleotide context of the *KRAS* mutation and the relative activity of the mutational signature in that tumor (Fig. 1d). In general, such probabilities reflected the underlying distribution of signatures, as seen in the similarities between Figures 1c and

1d, suggesting that the specific set of mutational processes present in a tumor was highly influential in determining the eventual *KRAS* mutation.

There were notable exceptions to this trend, however. For example, in COAD and PAAD, SBS18 (navy blue bars), likely caused by damage from reactive oxygen species [29, 30], was strongly associated with G12C mutations (Fig. 1d). This corroborated the previous finding that *KRAS* G12C mutations were more frequent in patients with MUTYH-Associated Polyposis [29], an autosomal recessive disease form of COAD 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 LUAD, the *KRAS* G12A/C/V mutations were primarily attributable to mutations caused by tobacco smoke, whereas *KRAS* G12D mutations were most likely attributable to clock-like mutations (Fig. 1d). In MM, SBS9, associated with mutations introduced by polymerase η repair of activation-induced deaminase (AID) activity [26, 31, 32], was strongly linked with Q61H (Fig. 1c, d), the most common *KRAS* mutation in that cancer. Finally, SBS8, of unknown etiology, had a substantial probability of causing some of the *KRAS* alleles, particularly G12V, across all four cancers. SBS17, also of uncertain etiology though linked to oxidative stress in other cancers [33], was the main cause for Q61H mutations in PAAD.

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

The extent to which mutational signatures represent the mechanism driving *KRAS* allelic diversity was further analyzed by calculating the predicted frequency of each allele in each tumor sample based on the frequency of mutations in the same trinucleotide context throughout the genome (Fig. 2; Supplementary Table 6). The null hypothesis tested was that, assuming the cancer would acquire a *KRAS* mutation, any one of the common alleles (found in greater than 3% of the tumor samples for a given cancer) was sufficient, and thus, the frequency of the *KRAS* alleles would be determined by the mutational processes alone. The average predicted frequencies across the samples of each cancer were compared against the observed allele frequencies. The alleles below the diagonal line were predicted to be more frequent than observed, while those above the line were more frequently observed than predicted. In COAD, G13D was predicted to be the most frequent allele. G12D/V mutations were considerably underestimated (Chi-squared test, $p < 0.05$, triangles); while the frequencies of G12S and A146T mutations were significantly overestimated in COAD (Chi-squared test, $p < 0.05$, triangles). In LUAD, the frequencies

of the G12A/D/V alleles were accurately predicted, though the frequency of the most common allele, G12C, was substantially underestimated. The high frequency of this allele has been attributed to its association with SBS4 caused by tobacco smoke (Fig. 1c, d), but our observation suggests that there is additional biological pressure promoting this mutation in LUAD. The frequencies of the *KRAS* alleles were best predicted in MM, with an exception for the most frequent allele, Q61H, which was dramatically underestimated with a predicted frequency of 15.0% but actual frequency of 35.7% of *KRAS* mutations. In PAAD, all of the alleles were observed at a significantly different frequency than predicted by mutational signatures.

The correlations between the observed and predicted allele frequencies for each cancer were low, indicating that the frequencies of the *KRAS* alleles were not strictly determined by the prevalence of their respective causative single nucleotide substitutions. PAAD had the strongest concordance between the predicted and observed allele frequencies with a Pearson correlation coefficient of 0.75, while the other three cancers had coefficients around 0.50. Interestingly, the codon 12 mutations demonstrated greater agreement between the observed and predicted frequencies in COAD, LUAD, and PAAD with correlation coefficients of 0.82, 0.83, and 0.84, respectively.

In addition to estimating the distribution of the common *KRAS* alleles for each cancer, we conducted a similar analysis considering all of the alleles found frequently in at least one of the cancer types (Supplementary Fig. 2; Supplementary Table 7). The alleles never or rarely found in each cancer were predicted to occur at frequencies ranging from 1.5% (for Q61L in PAAD) to 10.5% (for Q61K in LUAD), indicating that these alleles are not rare because their causative mutations do not occur, but instead because of weak oncogenic fitness in the tissue. For instance, the fact that *KRAS* A146T is exceedingly rare in PAAD, though predicted to represent 8.9% of *KRAS* mutations, agrees with the previous determination that forced expression of *KRAS* A146T does not induce pancreatic intraepithelial neoplasia (PanIN) [4]. Thus, while it is likely that the active mutational processes in a tissue contributed to which *KRAS* mutation was gained, they were not completely deterministic. Instead, the particular biological 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 *KRAS* allele selection in cancer, then distinct functions of each mutant form of K-RAS would be reflected in cooperating genetic events. An increased frequency of comutation with

another gene suggests a cooperative effect, whereas a reduced frequency of comutation (compared to random) suggests that the second event is functionally redundant or that it introduces an inhibitory effect. 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 [34, 35, 36, 37]. Alternatively, in LUAD, the mutational activation of *EGFR* was demonstrated to be cytotoxic in the presence of a *KRAS* mutant, and, thus, the two are rarely found in the same tumor [38, 39].

The comutation interactions between each *KRAS* allele and every other mutated gene were investigated using a one-sided Fisher's exact test of association to identify increased rates of comutation and a test for mutual exclusivity proposed by Leiserson *et al.* [40] to identify reduced rates of comutation (Supplementary Table 8). 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. 3a). These linking genes tended to be well-studied cancer genes 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. 3b).

Consistent with the idea that each allele is functionally distinct, a substantial number of genes comutated with just one *KRAS* allele. To gain functional insight into the network, genes known to physically interact with K-RAS [16], signal up- or downstream of K-RAS [41, 42], or are known oncogenes [43, 44] were extracted (Fig. 3b). Several *KRAS* alleles had reduced comutation with *NRAS* and *BRAF* and increased comutation with *APC* and *PIK3CA*, interactions that have been previously documented [45, 37, 46, 47, 34, 36, 48, 49, 50, 51, 52]. 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. 3c). 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 the activity of Rho guanine nucleotide exchange factor 7 on RAC1 in phagocytosis and cell adhesion [53]. 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. Additional genes of interest that had comutation interactions with *KRAS* G12D are shown in Fig. 3d and e. These include

increased comutation with *AMER1*, a negative regulator of Wnt signaling [54, 55]

The *KRAS* allele-specific comutation network uncovered in LUAD was far larger than that of COAD (Supplementary Fig. 3). This was likely caused by the higher mutation frequency in this cancer, increasing the statistical power to detect both increased and reduced comutation interactions. As in the network derived from COAD, many of these genes were involved in integral K-RAS signaling pathways, including an increased comutation interaction between *KRAS* G12A and *MAP2K3*, a reduced comutation interaction between *KRAS* G12D and *ERBB4*, and a very strong increased rate of comutation between *KRAS* G12C and *STK11* (Supplementary Fig. 3). There were several intriguing cellular processes enriched in the LUAD networks for each allele (Fig. 3c). For example, *KRAS* G12C had comutation interactions with many genes encoding proteins that interact with Myc ("PPI of MYC (TF)"), and the G12D comutation network was enriched with interactions with focal adhesion genes (Fig. 3).

Conducting this analysis in MM was hampered by the fact that this cancer is known to be frequently multi-clonal [44, 56]. As such, some detectable comutation events were mutations acquired by distinct populations in a single patient, potentially obfuscating true comutation interactions. Due to this caveat, limiting the analysis to genes known to be recurrently mutated in MM reduced the chance of highlighting a false positive [56]. From this limited scope, it was discovered that *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 (Supplementary Fig. 4). 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.

The *KRAS* allele comutation network found in the PAAD tumor samples demonstrated that many genes had detectable comutation interactions with multiple alleles, primarily of reduced comutation (Supplementary Fig. 5). There were numerous genes that had opposing comutation interactions with different alleles. Of these, four interact with or signal through K-RAS [16, 41, 42] or are known oncogenes [43, 44]: *TP53*, *RNF43*, *MAP2K4*, and *RBM10* (Supplementary Fig. 5). 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. There were many notable cellular functions and processes enriched in the comutation networks of the *KRAS* alleles (Fig. 3c) including the protein-protein interaction networks (PPIN) of SMAD1-3 and TGF- β signaling. While these SMAD gene sets were related, the underlying comutation interactions that drove the enrichment were different for each *KRAS* allele (Fig. 3f). 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 cell-of-origin. 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 behaviors and genetic interactions, they likely have specific genetic vulnerabilities. To this end, we used data from a genome-wide, CRISPR/Cas9 knock-out screen of cancer cell lines [59, 60] to identify genes with *KRAS* allele-specific genetic dependencies. The analysis was restricted to *KRAS* alleles for which there were at least 3 different cell lines with the mutation, limiting the following investigation to only COAD and PAAD cell lines. Allele-specific enrichments for signaling pathways and cellular processes were identified using Gene Set Enrichment Analysis (GSEA), and individual genes demonstrating differential genetic dependency by *KRAS* allele were identified using ANOVA and one-versus-all *t*-tests.

For COAD, there was a sufficient number of cell lines with WT *KRAS* or G12D, G12V, and G13D mutations for this analysis. Measuring for gene set enrichment revealed strong patterns in differential dependency of various cellular processes (Fig. 4a). For example, genes involved in ERBB4 signaling tended to have a weaker lethal effect when knocked out in cell lines with *KRAS* G12V mutations than in *KRAS* G12D, G13D, or WT cell lines (Fig. 4b). Similarly, the *KRAS* G13D cell lines were less affected when genes involved in oxidative phosphorylation were targeted (Fig. 4c). To discover individual genes with allele-specific interactions, each gene was tested for differential genetic dependency with the cell lines grouped by their *KRAS* allele. The resulting 62 genes were hierarchically clustered into 4 groups by their dependency scores (Figure 4d; Supplementary Table 9). Genes in cluster 2 tended to have stronger genetic dependency in cell lines with *KRAS* G12V, while those in cluster 3 demonstrated weaker dependency in G12D cell lines. Four notable genes with allele-specific associations are displayed in Fig. 4e. First, knocking-out *LIN7C*, a gene that maintains the asymmetric distribution of membrane proteins in polarized epithelial cells [61], had a more severe reduction on growth in *KRAS* G13D cell lines compared to the others (Figure 4e). Also, a regulator of apoptosis previously linked to dysregulated expression in cancer, *TFPT*, demonstrated

significantly greater dependency in G12D cell lines. Interestingly, *STARD9*, a gene encoding a kinesin required for mitotic spindle assembly [62], had moderate growth defects when knocked-out in all cell lines except those with a *KRAS* G12D mutation. Lastly, the kinetochore-associated protein (*KNTC1*), a regulator of the mitotic checkpoint [63, 64, 65], which demonstrated moderate to strong lethal effects when knocked out in almost every cell line except for those with a *KRAS* G12V allele (Figure 4e).

For the genetic dependency analysis of PAAD, the *KRAS* alleles with a sufficient number of cell lines were G12D, G12R, and G12V (there were not enough WT *KRAS* cell lines to include in the analysis). GSEA revealed substantial differences in the dependencies of critical cellular pathways (Supplementary Fig. 6). For instance, the G12D cell lines demonstrated a reduced dependency on the genes at the G2 and M DNA-damage checkpoint. Moreover, the G12R cell lines were less dependent on PI3K signalling downstream of FGFR1, driven through a reduced dependency on *FRS2* (fibroblast growth factor receptor substrate 2) and *GRB2*, which encodes a protein linking EGFR to the GEF SOS1. Similarly, the cell lines with *KRAS* G12V mutations were less sensitive to the knock-out of genes implicated in cellular senescence. This enrichment was driven by a significantly reduced dependence upon *JUN*, which encodes the transcription factor c-JUN, and a beneficial impact on growth (a positive dependency score) from knocking-out *MAPK8* (JNK-1), which regulates c-JUN via phosphorylation (Supplementary Fig. 7). In these cell lines, 130 individual genes demonstrated *KRAS* allele-specific genetic dependency (Supplementary Fig. 7; Supplementary Table 10). Several noteworthy interactions include a regulator of cell cycle progression, *KHDRBS1* [66], the oxygen sensor, *EGLN2* [67], and a stabilizer of p53, *BRI3BP* [68]. Overall, the *KRAS* alleles were associated with substantially different genetic dependencies on specific cellular processes, signaling pathways, and individual genes.

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

As emphasized by the weakly connected comutation networks, the *KRAS* alleles are not acting in the same genetic environments, and, therefore, their allele-specific genetic dependency interactions might be mediated by a comutating partner. To address this hypothesis, we constructed linear models for the dependency score of each gene with allele-specific dependency that included a coefficient for the previously linked *KRAS* allele and a coefficient for the mutation of each gene in its comutation network. These models were then fit with elastic net regression to isolate the most informative predictors, adjusting for the RNA expression of the targeted gene [69].

Some of the models indicated that the mutation of a comutation partner could explain the allele-specific dependency interaction. An example of this was how the dependency of COAD cell lines on *STARD9* was greater in *TP53*-mutant lines than in *KRAS* G12D lines (Fig. 5a). If *TP53* mutations induce a stronger dependency on *STARD9*, the reduced frequency of comutation between *TP53* and *KRAS* G12D would cause the opposite effect to be ascribed to the G12D allele. A similar effect was found between *KRAS* G12D and *SMAD4* in PAAD cell lines for the dependency on *EEF1E1*, *ABI1*, and *MYBL2* (Fig. 5b-d). Because of the reduced comutation interaction between *KRAS* G12D and *SMAD4* in PAAD, the effects of knocking out these genes can be ascribed to an allele-specific effect or to the *SMAD4* mutation. These examples highlight how the allele-specific comutation interactions of *KRAS* can influence the interpretation of other interactions.

Discussion

This study addresses the genetic complexity of cancer through a comprehensive genetic interaction analysis of oncogenic *KRAS* alleles in COAD, LUAD, MM, and PAAD. Measuring the levels of mutational signatures revealed that the cancer-specific distributions of *KRAS* mutations were influenced, but 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. Indeed, our lab has previously demonstrated that mutant forms of K-RAS produce distinct molecular and cellular phenotypes that are largely dependent upon tissue context [4, 70, 71]. To investigate allele-specific genetic properties, we conducted statistical tests to identify 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 that they were alterations that cooperated with the *KRAS* allele to promote tumor growth. Alternatively, some genes comutated with a *KRAS* allele less frequently than expected by chance, suggesting they were 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 support a model in which the various oncogenic *KRAS* mutations are not biologically redundant, but instead have distinct properties that are reflected in their genetic interactions.

This analysis of *KRAS* genetic networks in four different tumor types highlights the tissue-specific nature of

genetic interactions. In places, we focused on the results from the analysis of COAD, as it demonstrated a high variability in the types of *KRAS* alleles, had limited exogenous mutational pressure (in contrast to the effects of smoking-induced mutations in LUAD), and we had a large number of WGS and WES data. However, allele-specific genetic interactions were not consistent between tissues, demonstrating the complex relationship between the tissue-of-origin, K-RAS function, and cooperating genetic events. While the intrinsic biochemical properties of a K-RAS mutant are likely maintained in each cancer, their downstream signaling properties, and ultimately their effects on tumorigenesis, are determined by the basal configuration of the tissue-specific signaling network [70]. Thus, the configuration of the tissue signaling network influences the genetic interactions that arise during cancer progression.

In addition to the importance of tissue-specificity, this study provides compelling evidence that the somatic missense mutations that activate oncogenes are not always equivalent. We and others have demonstrated the distinct effects of the *KRAS* alleles, both computationally and experimentally, revealing many instances of substantial variation between different mutations of the same gene. This is likely a more general principle applicable to many oncogenes, especially those with multiple mutational hotspots.

Finally, this study has broad implications for the understanding of oncogene biology and for cancer therapy. Whether a targeted therapy directly inhibits the activated oncprotein or not, it is important to understand how allele-specific signaling properties and genetic interactions influence therapeutic response. For instance, *BRAF* activating mutations have been classified into three groups defined by their functional effects on the protein product [72, 73], which consequently determines their response to different inhibitors [74, 75]. Moreover, the response of HER2 mutant cancers to HER2 inhibition varies depending on the tissue-of-origin of the cancer [76], which could be due to intrinsic signaling differences between the tissues-of-origin or to cooperating mutations unique to a specific cancer type. For cancer therapy to be truly precise, it will be key to appreciate and understand the complexity of the genetic networks in each cancer type.

Methods

Cancer sample data sources and acquisition

Whole genome sequencing (WGS), whole exome sequencing (WES), and targeted gene panel sequencing ("targeted-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 [77, 78], which included relevant projects from The Cancer Genome Atlas (TCGA) [52, 79, 80] and other smaller studies. Additional data were acquired from the International Cancer Genome Consortium (ICGC) for pancreatic cancer [81] and colorectal cancer. MM WES data were gathered from the Multiple Myeloma Research Foundation (MMRF)-CoMMpass online repository [82]. Panel data for multiple cancers were retrieved from AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE v5) [83]. 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 Tables 1 and 2.

Hypermutated sample cutoff

Some of the COAD samples had 5 to 10-times more mutations than the 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 targeted-sequencing 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) [84] and The Human Protein Atlas (HPA, 12/03/2018) [85, 86], and tumor expression data was collected from MMRF-CoMMpass (01/14/2019), TCGA-COAD, TCGA-LUAD, and TCGA-PAAD [82, 52, 79, 80]. 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.

Calculating overall distribution of hotspot mutations

The frequency of mutations at the four hotspots of *KRAS* across COAD, LUAD, MM, and PAAD was calculated by accounting for the different yearly incidence of each cancer type. The incidence of cancers of the "colorectum," "lung and bronchus," "myeloma," and "pancreas" were obtained from the American Cancer Society [87]: 3,870,000 colorectum, 5,930,000 lung and bronchus, 680,000 myeloma, 1,280,000 pancreas. The incidences of COAD, LUAD, and PAAD were estimated by multiplying the number of cases of their respective tissue by the proportion they constitute: 95%, 50%, and 95%, respectively [87, 88]. The distribution of mutations to the hotspots across all cancers was calculated by finding the frequency within each cancer type, then combining those figures, weighting by their yearly incidence.

Protein-Protein Interaction Network (PPIN)

The PPIN used throughout the study was the combination of interactions from STRING [89, 90], HINT [91], and BioPlex [92].

Identifying mutational signatures

The genome-wide mutations of a sample can be deconvolved into mutational signatures that represent endogenous or exogenous mutagenic processes [26]. 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 MATLAB [93] implementation of NMF algorithm, SigProfiler [26], 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 [94].

The spectrum of the signatures discovered by NMF were matched to the COSMIC catalog [95]. 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 [96, 27]. 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 [27]. Three versions of the signature associated to POLE mutations, signature 10, were discovered (previously reported in [27]). 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-guanine [27] was found. One signature that was discovered in COAD did not have a good match with a previously published signature, although it resembled a signature previously reported to be caused by SBSA [97] and signatures 34 and 41 in reference [27]. 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 artifacts [98] and were removed from downstream analysis. Signatures that contributed to less than 5% of the mutations were also removed from downstream analysis. The levels of each signature in each tumor sample were calculated using Non-Negative Least Squares [99]. The final spectra for each mutational signature and mutational signature composition of each tumor samples can be found in the Supplementary Data.

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 p of a tumor sample a acquiring the *KRAS* mutation k from signature s from all signatures S can be calculated using Eq. 1.

$$p_{k,s} = \frac{c_{s,a}w_{k,s}}{\sum_s^S c_{s,a}w_{k,s}} \quad (1)$$

where $\left\{ \begin{array}{l} c_{s,a} \text{ is the contribution of signature } s \text{ in sample } a. \\ w_{k,s} \text{ is the weight of the mutation } k \text{ in signature } s. \end{array} \right.$

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 and the processes are active with the same probability in any part of the genome, the predicted frequency of each *KRAS* allele can be calculated as the frequency of the same mutation across the entire genome. For each cancer, the pool of possible *KRAS* mutations were restricted to those found in at least 3% of the tumor samples. The 95% confidence intervals were bootstrapped [100]. The predicted frequencies of the *KRAS* alleles for each cancer are available in the Supplementary Table.

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 genes with an overall mutation frequency of at least 1% in the given cancer were considered. In addition, only comutation partners with at least three comutation events or a comutation frequency with a *KRAS* allele of at least 10% (i.e. 10% of the tumors with a *KRAS* allele also had a mutation in the given gene) were considered.

The Row-Column Exclusivity Test was used to identify reduced frequency of comutation between *KRAS* alleles and other mutated genes [40]. This is a permutation-based test that finds the probability of observing the actual number of mutually exclusive events given that the number of times the gene 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 (using just the exonic mutations from WGS). Only genes with a mutational frequency of at least 2% and at least 10 mutually exclusive events were considered.

The Fisher's exact test was used to detect increased comutation interactions because, unlike the Row-Column Exclusivity Test, it could utilize the targeted sequencing data. However, the Row-Column Exclusivity Test outperformed the Row Exclusivity Test, a comparable permutation-based approximation of the Fisher's exact test, in the original publication by Leiserson *et al.* [40], suggesting it would be more sensitive for detecting reduced comutation interactions in the current study.

COAD samples identified as hypermutants were excluded from this analysis as they were likely microsatellite

instable. Thus, these samples would be expected to have a high proportion of passenger mutations that would contribute substantial noise to the identification *KRAS* allele-specific comutation interactions.

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 [101, 102, 103]. The online API was last accessed on April 9, 2020. 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).

Modeling of cancer cell line genetic dependencies

Genetic dependency data was downloaded from the online DepMap portal (<https://depmap.org/portal/download/>) (2019Q3) and the CERES scores were used for all analyses. 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 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 < 0.05 and $R^2 \geq 0.4$), the gene was not tested for *KRAS* allele-specific genetic dependency. Further, genes that tended to show differential dependence on the basis of their mutation status (Wilcoxon rank sum test, p < 0.05) were not included in downstream analysis. 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 (p-value < 0.01). For these genes, one-versus-all Student's t-tests were used to compare the dependency scores of each group of cell lines against the others (Benjamini-Hochberg FDR adjusted p-value < 0.05). These genes were declared as differentially dependent by *KRAS* allele. The box-plots in Fig. 4 and Supplementary Fig. 7 show pairwise t-tests with Benjamini-Hochberg FDR adjusted p-value.

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>; downloaded on

October 15, 2019). 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.

Modelling the effect of comutation events on genetic dependency

For each gene found to have a genetic dependency interaction with a *KRAS* allele, an additional linear model was built to estimate the effect of mutations to genes that comutate with the *KRAS* allele. The linear model regressed on the RNA expression level of the gene and contained binary indicator variables for if the cell line had a mutation in the targeted gene, had the specific *KRAS* allele or another allele, or had a mutation in one of the genes that comutates with the specific *KRAS* allele. Only comutation genes that were mutated in at least 3 cell lines and WT in at least 3 cell lines were included as covariates. To avoid including perfectly correlated variables in the model, comutating genes that were perfectly correlated (i.e. they were mutated in exactly the same cell lines) were merged into a single predictor. After these adjustments, the models had 45, 29, and 16 coefficients for genes with dependency interactions with *KRAS* G12D, G12V, and G13D in COAD cell lines, respectively. For PAAD cell lines, the models had 15, 14, and 8 coefficients for genes with dependency interactions with G12D, G12R, and G12V, respectively. Some models had a fewer covariates because either 1) the targeted gene was not mutated in enough of the cell lines to include the coefficient for this variable in the model, or 2) the targeted gene was mutated in the same cell lines as one or more of the comutating genes resulting in the merging of these variables. Due to the imbalance between the number of covariates and data points (i.e. cell lines), the models were fit with elastic-net regularization [69, 104] constraining the mixing parameter $\alpha \in [0.75, 1]$, thus favoring the L1 penalty.

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 analyses. Python v3.7 [105] and R v3.6.1 [106] were used for most of the analyses.

Acknowledgements

This work was supported by a grant from the National Institutes of Health (R01CA232372 to K.M.H.) and an award from the Cancer Research UK Grand Challenge and the Mark Foundation to the SPECIFICANCER team.

The whole exome sequencing data of MM were acquired from the Multiple Myeloma Research Foundation Personalized Medicine Initiative.

Author contributions

J.H.C., G.E.M.M., P.J.P., and K.M.H. devised the research strategy. J.H.C., G.E.M.M., and D.C.G. performed the analyses. J.H.C., G.E.M.M., P.J.P., and K.M.H. wrote the manuscript. J.H.C., G.E.M.M., P.J.P., and K.M.H. helped to interpret results. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

- [1] Dhirendra K Simanshu, Dwight V Nissley, and Frank McCormick. RAS Proteins and Their Regulators in Human Disease. *Cell*, 170(1):17–33, 6 2017.
- [2] Matthew H Bailey, Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico, Michael C Wendl, Jaegil Kim, Brendan Reardon, Patrick Kwok-Shing Ng, Kang Jin Jeong, Song Cao, Zixing Wang, Jianjiong Gao, Qingsong Gao, Fang Wang, Eric Minwei Liu, Loris Mularoni, Carlota Rubio-Perez, Niranjan Nagarajan, Isidro Cortés-Ciriano, Daniel Cui Zhou, Wen-Wei Liang, Julian M Hess, Venkata D Yellapantula, David Tamborero, Abel Gonzalez-Perez, Chayaporn Suphavilai, Jia Yu Ko, Ekta Khurana, Peter J Park, Eliezer M Van Allen, Han Liang, MC3 Working Group, Cancer Genome Atlas Research Network, Michael S Lawrence, Adam Godzik, Nuria Lopez-Bigas, Josh Stuart, David Wheeler, Gad Getz, Ken Chen, Alexander J Lazar, Gordon B Mills, Rachel Karchin, and Li Ding. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, 174(4):1034–1035, 8 2018.
- [3] Kevin M Haigis. KRAS Alleles: The Devil Is in the Detail. *Trends in cancer*, 3(10):686–697, 10 2017.

- [4] Emily J Poulin, Asim K Bera, Jia Lu, Yi-Jang Lin, Samantha Dale Strasser, Joao A Paulo, Tannie Q Huang, Carolina Morales, Wei Yan, Joshua Cook, Jonathan A Nowak, Douglas K Brubaker, Brian A Joughin, Christian W Johnson, Rebecca A DeStefanis, Phaedra C Ghazi, Sudershan Gondi, Thomas E Wales, Roxana E Iacob, Lana Bogdanova, Jessica J Gierut, Yina Li, John R Engen, Pedro A Perez-Mancera, Benjamin S Braun, Steven P Gygi, Douglas A Lauffenburger, Kenneth D Westover, and Kevin M Haigis. Tissue-Specific Oncogenic Activity of KRASA146T. *Cancer discovery*, 9(6):738–755, 6 2019.
- [5] Mark Steven Miller and Lance D. Miller. RAS mutations and oncogenesis: Not all RAS mutations are created equally. *Frontiers in Genetics*, 2(JAN):1–9, 2012.
- [6] Siqi Li, Allan Balmain, and Christopher M Counter. A model for RAS mutation patterns in cancers: finding the sweet spot. *Nature reviews. Cancer*, 18(12):767–777, 12 2018.
- [7] Mariano Barbacid. ras genes. *Annual review of biochemistry*, 56:779–827, 1987.
- [8] John C Hunter, Anuj Manandhar, Martin A Carrasco, Deepak Gurbani, Sudershan Gondi, and Kenneth D Westover. Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. *Molecular cancer research : MCR*, 13(9):1325–35, 9 2015.
- [9] M. J. Smith, B. G. Neel, and M. Ikura. NMR-based functional profiling of RASopathies and oncogenic RAS mutations. *Proceedings of the National Academy of Sciences*, 110(12):4574–4579, 2013.
- [10] L A Feig and G M Cooper. Relationship among guanine nucleotide exchange, GTP hydrolysis, and transforming potential of mutated ras proteins. *Molecular and cellular biology*, 8(6):2472–8, 6 1988.
- [11] Sarah Edkins, Sarah O'Meara, Adrian Parker, Claire Stevens, Marcelo Reis, Siân Jones, Chris Greenman, Helen Davies, Gillian Dalgliesh, Simon Forbes, Chris Hunter, Raffaella Smith, Philip Stephens, Peter Goldstraw, Andrew Nicholson, Tsun Leung Chan, Victor E Velculescu, Siu Tsan Yuen, Suet Yi Leung, Michael R Stratton, and P Andrew Futreal. Recurrent KRAS codon 146 mutations in human colorectal cancer. *Cancer biology & therapy*, 5(8):928–32, 8 2006.
- [12] Manickam Janakiraman, Efsevia Vakiani, Zhaoshi Zeng, Christine A. Pratilas, Barry S. Taylor, Dhananjay Chitale, Ensar Halilovic, Manda Wilson, Kety Huberman, Julio Cesar Ricarte Filho, Yogindra Persaud, Douglas A. Levine, James A. Fagin, Suresh C. Jhanwar, John M. Mariadason, Alex Lash, Marc Ladanyi,

- Leonard B. Saltz, Adriana Heguy, Philip B. Paty, and David B. Solit. Genomic and biological characterization of exon 4 KRAS mutations in human cancer. *Cancer research*, 70(14):5901–11, 7 2010.
- [13] Nicole L K Pershing, Benjamin L Lampson, Jason A Belsky, Erin Kaltenbrun, David M MacAlpine, and Christopher M Counter. Rare codons capacitate Kras-driven de novo tumorigenesis. *The Journal of clinical investigation*, 125(1):222–33, 1 2015.
- [14] G Aaron Hobbs, Nicole M Baker, Anne M Miermont, Ryan D Thurman, Mariaelena Pierobon, Timothy H Tran, Andrew O Anderson, Andrew M Waters, J Nathaniel Diehl, Bjoern Papke, Richard G Hodge, Jennifer E Klomp, Craig M Goodwin, Jonathan M DeLiberty, Junning Wang, Raymond W S Ng, Prson Gautam, Kirsten L Bryant, Dominic Esposito, Sharon L Campbell, Emanuel F Petricoin, Dhirendra K Simanshu, Andrew J Aguirre, Brian M Wolpin, Krister Wennerberg, Udo Rudloff, Adrienne D Cox, and Channing J Der. Atypical KRASG12R Mutant Is Impaired in PI3K Signaling and Macropinocytosis in Pancreatic Cancer. *Cancer discovery*, 10 2019.
- [15] Tina L. Yuan, Arnaud Amzallag, Rachel Bagni, Ming Yi, Shervin Afghani, William Burgan, Nicole Fer, Leslie A. Strathern, Katie Powell, Brian Smith, Andrew M. Waters, David Drubin, Ty Thomson, Rosy Liao, Patricia Greninger, Giovanna T. Stein, Ellen Murchie, Eliane Cortez, Regina K. Egan, Lauren Procter, Matthew Bess, Kwong Tai Cheng, Chih-Shia Lee, Liam Changwoo Lee, Christof Fellmann, Robert Stephens, Ji Luo, Scott W. Lowe, Cyril H. Benes, and Frank McCormick. Differential Effector Engagement by Oncogenic KRAS. *Cell reports*, 22(7):1889–1902, 2018.
- [16] Joanna R Kovalski, Aparna Bhaduri, Ashley M Zehnder, Poornima H Neela, Yonglu Che, Glenn G Wozniak, and Paul A. Khavari. The Functional Proximal Proteome of Oncogenic Ras Includes mTORC2. *Molecular cell*, 73(4):830–844, 2 2019.
- [17] Nathan T. Ihle, Lauren A. Byers, Edward S. Kim, Pierre Saintigny, J. Jack Lee, George R. Blumenschein, Anne Tsao, Suyu Liu, Jill E. Larsen, Jing Wang, Lixia Diao, Kevin R. Coombes, Lu Chen, Shuxing Zhang, Mena F. Abdelmelek, Ximing Tang, Vassiliki Papadimitrakopoulou, John D. Minna, Scott M. Lippman, Waun K. Hong, Roy S. Herbst, Ignacio I. Wistuba, John V. Heymach, and Garth Powis. Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. *Journal of the National Cancer Institute*, 104(3):228–39, 2 2012.

- [18] Michael Spoerner, Alfred Wittinghofer, and Hans Robert Kalbitzer. Perturbation of the conformational equilibria in Ras by selective mutations as studied by ³¹P NMR spectroscopy. *FEBS letters*, 578(3):305–10, 12 2004.
- [19] Matthew J Smith and Mitsuhiro Ikura. Integrated RAS signaling defined by parallel NMR detection of effectors and regulators. *Nature chemical biology*, 10(3):223–30, 3 2014.
- [20] Tuu Pantzar, Sami Rissanen, Daniel Dauch, Tuomo Laitinen, Ilpo Vattulainen, and Antti Poso. 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(9):e1006458, 9 2018.
- [21] Wendy De Roock, Derek J Jonker, Federica Di Nicolantonio, Andrea Sartore-Bianchi, Dongsheng Tu, Salvatore Siena, Simona Lamba, Sabrina Arena, Milo Frattini, Hubert Piessevaux, Eric Van Cutsem, Chris J O’Callaghan, Shirin Khambata-Ford, John R Zalcberg, John Simes, Christos S Karapetis, Alberto Bardelli, and Sabine Tejpar. Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. *JAMA*, 304(16):1812–20, 10 2010.
- [22] Thomas McFall, Jolene K Diedrich, Meron Mengistu, Stacy L Littlechild, Kendra V Paskvan, Laura Sisk-Hackworth, James J Moresco, Andrey S Shaw, and Edward C Stites. A systems mechanism for KRAS mutant allele-specific responses to targeted therapy. *Science signaling*, 12(600):8288, 9 2019.
- [23] Dana Rabara, Timothy H Tran, Srisathiyaranarayanan Dharmia, Robert M Stephens, Frank McCormick, Dhirendra K Simanshu, and Matthew Holderfield. KRAS G13D sensitivity to neurofibromin-mediated GTP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America*, 116(44):22122–22131, 10 2019.
- [24] Maria Paz Zafra, Direna Alonso-Curbelo, Sukanya Goswami, Emma M Schatoff, Teng Han, John E Wilkinson, and Lukas E Dow. An in vivo KRAS allelic series reveals distinct phenotypes of common oncogenic variants. *bioRxiv*, 14:847509, 2019.
- [25] Barbara Bournet, Fabrice Muscari, Camille Buscail, Eric Assenat, Marc Barthet, Pascal Hammel, Janick Selves, Rosine Guimbaud, Pierre Cordelier, and Louis Buscail. KRAS G12D Mutation Subtype Is A Prognostic Factor for Advanced Pancreatic Adenocarcinoma. *Clinical and translational gastroenterology*, 7(3):e157, 3 2016.

- [26] Ludmil B. Alexandrov, Serena Nik-Zainal, David C. Wedge, Samuel A J R Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, Niccolò Bolli, Ake Borg, Anne-Lise Børresen-Dale, Sandrine Boyault, Birgit Burkhardt, Adam P. Butler, Carlos Caldas, Helen R. Davies, Christine Desmedt, Roland Eils, Jórunn Erla Eyfjörd, John A. Foekens, Mel Greaves, Fumie Hosoda, Barbara Hutter, Tomislav Ilicic, Sandrine Imbeaud, Marcin Imielinski, Marcin Imielinsk, Natalie Jäger, David T W Jones, David Jones, Stian Knappskog, Marcel Kool, Sunil R. Lakhani, Carlos López-Otín, Sancha Martin, Nikhil C. Munshi, Hiromi Nakamura, Paul A. Northcott, Marina Pajic, Elli Papaemmanuil, Angelo Paradiso, John V. Pearson, Xose S. Puente, Keiran Raine, Manasa Ramakrishna, Andrea L. Richardson, Julia Richter, Philip Rosenstiel, Matthias Schlesner, Ton N. Schumacher, Paul N. Span, Jon W. Teague, Yasushi Totoki, Andrew N J Tutt, Rafael Valdés-Mas, Marit M van Buuren, Laura van 't Veer, Anne Vincent-Salomon, Nicola Waddell, Lucy R. Yates, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Jessica Zucman-Rossi, P Andrew Futreal, Ultan McDermott, Peter Lichter, Matthew Meyer-son, Sean M. Grimmond, Reiner Siebert, Elías Campo, Tatsuhiro Shibata, Stefan M. Pfister, Peter J. Campbell, and Michael R. Stratton. Signatures of mutational processes in human cancer. *Nature*, 500(7463):415–21, 8 2013.
- [27] Ludmil B. Alexandrov, Jaegil Kim, Nicholas J. Haradhvala, Mi Ni Huang, Alvin Wei Tian Ng, Yang Wu, Arnoud Boot, Kyle R. Covington, Dmitry A. Gordenin, Erik N. Bergstrom, S. M. Ashiqul Islam, Nuria Lopez-Bigas, Leszek J. Klimczak, John R. McPherson, Sandro Morganella, Radhakrishnan Sabarinathan, David A. Wheeler, Ville Mustonen, PCAWG Mutational Signatures Working Group, Gad Getz, Steven G. Rozen, Michael R. Stratton, and PCAWG Consortium. The repertoire of mutational signatures in human cancer. *Nature*, 578(7793):94–101, 2 2020.
- [28] Ludmil B. Alexandrov, Philip H. Jones, David C. Wedge, Julian E. Sale, Peter J. Campbell, Serena Nik-Zainal, and Michael R. Stratton. Clock-like mutational processes in human somatic cells. *Nature genetics*, 47(12):1402–7, 12 2015.
- [29] Alessandra Viel, Alessandro Bruselles, Ettore Meccia, Mara Fornasarig, Michele Quaia, Vincenzo Canzonieri, Eleonora Policicchio, Emanuele Damiano Urso, Marco Agostini, Maurizio Genuardi, Emanuela Lucci-Cordisco, Tiziana Venesio, Aline Martayan, Maria Grazia Diodoro, Lupe Sanchez-Mete, Vittoria Stigliano, Filomena Mazzei, Francesca Grasso, Alessandro Giuliani, Marta Baiocchi, Roberta Maestro,

Giuseppe Giannini, Marco Tartaglia, Ludmil B. Alexandrov, and Margherita Bignami. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine*, 20:39–49, 6 2017.

- [30] Camilla Pilati, Jayendra Shinde, Ludmil B. Alexandrov, Guillaume Assié, Thierry André, Zofia Hélias-Rodzewicz, Romain Ducoudray, Delphine Le Corre, Jessica Zucman-Rossi, Jean-François Emile, Jérôme Bertherat, Eric Letouzé, and Pierre Laurent-Puig. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *The Journal of pathology*, 242(1):10–15, 2017.
- [31] Igor B Rogozin, Alexander Goncearenco, Artem G Lada, Subhajyoti De, Vyacheslav Yurchenko, German Nudelman, Anna R Panchenko, David N Cooper, and Youri I Pavlov. DNA polymerase η mutational signatures are found in a variety of different types of cancer. *Cell cycle (Georgetown, Tex.)*, 17(3):348–355, 2018.
- [32] Mia Petljak and Ludmil B Alexandrov. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis*, 37(6):531–40, 2016.
- [33] Marketa Tomkova, Jakub Tomek, Skirmantas Kriaucionis, and Benjamin Schuster-Böckler. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biology*, 19(1):129, 12 2018.
- [34] Klaus-Peter Janssen, Paola Alberici, Hafida Fsihi, Claudia Gaspar, Cor Breukel, Patrick Franken, Christophe Rosty, Miguel Abal, Fatima El Marjou, Ron Smits, Daniel Louvard, Riccardo Fodde, and Sylvie Robine. APC and oncogenic KRAS are synergistic in enhancing Wnt signaling in intestinal tumor formation and progression. *Gastroenterology*, 131(4):1096–109, 10 2006.
- [35] Eric R Fearon and Max S Wicha. KRAS and cancer stem cells in APC-mutant colorectal cancer. *Journal of the National Cancer Institute*, 106(2):djt444, 2 2014.
- [36] Eri Sakai, Mizuho Nakayama, Hiroko Oshima, Yuta Kouyama, Atsushi Niida, Satoshi Fujii, Atsushi Ochiai, Keiichi Nakayama, Koshi Mimori, Yutaka Suzuki, Chang Pyo Hong, Chan-Young Ock, Seong-Jin Kim, and Masanobu Oshima. Combined Mutation of Apc, Kras, and Tgfbr2 Effectively Drives Metastasis of Intestinal Cancer. *Cancer research*, 78(5):1334–1346, 2018.

- [37] Mayank Jauhri, Akanksha Bhatnagar, Satish Gupta, Manasa Bp, Sachin Minhas, Yogender Shokeen, and Shyam Aggarwal. 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(2):1010428317692265, 2 2017.
- [38] Arun M. Unni, William W. Lockwood, Kreshnik Zejnullahu, Shih-Queen Lee-Lin, and Harold Varmus. Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. *eLife*, 4(JUNE):e06907, 6 2015.
- [39] C Ambrogio, M Barbacid, and D Santamaría. In vivo oncogenic conflict triggered by co-existing KRAS and EGFR activating mutations in lung adenocarcinoma. *Oncogene*, 36(16):2309–2318, 2017.
- [40] Mark D M Leiserson, Matthew A Reyna, and Benjamin J Raphael. A weighted exact test for mutually exclusive mutations in cancer. *Bioinformatics (Oxford, England)*, 32(17):i736–i745, 2016.
- [41] Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, 45(D1):D353–D361, 2017.
- [42] Minoru Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*, 44(D1):457–62, 1 2016.
- [43] S Bamford, E Dawson, S Forbes, J Clements, R Pettett, A Dogan, A Flanagan, J Teague, P A Futreal, M R Stratton, and R Wooster. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *British journal of cancer*, 91(2):355–8, 7 2004.
- [44] Zbyslaw Sondka, Sally Bamford, Charlotte G. Cole, Sari A. Ward, Ian Dunham, and Simon A. Forbes. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature reviews. Cancer*, 18(11):696–705, 11 2018.
- [45] M Sensi, G Nicolini, C Petti, I Bersani, F Lozupone, A Molla, C Vegetti, D Nonaka, R Mortarini, G Parmiani, S Fais, and A Anichini. Mutually exclusive NRASQ61R and BRAFV600E mutations at the single-cell level in the same human melanoma. *Oncogene*, 25(24):3357–64, 6 2006.

- [46] R Seth, S Crook, S Ibrahem, W Fadhil, D Jackson, and M Ilyas. Concomitant mutations and splice variants in KRAS and BRAF demonstrate complex perturbation of the Ras/Raf signalling pathway in advanced colorectal cancer. *Gut*, 58(9):1234–41, 9 2009.
- [47] J Cisowski, V I Sayin, M Liu, C Karlsson, and M O Bergo. Oncogene-induced senescence underlies the mutual exclusive nature of oncogenic KRAS and BRAF. *Oncogene*, 35(10):1328–33, 3 2016.
- [48] Alyssa L Kennedy, Jennifer P Morton, Indrani Manoharan, David M Nelson, Nigel B Jamieson, Jeff S Pawlikowski, Tony McBryan, Brendan Doyle, Colin McKay, Karin A Oien, Greg H Enders, Rugang Zhang, Owen J Sansom, and Peter D Adams. Activation of the PIK3CA/AKT pathway suppresses senescence induced by an activated RAS oncogene to promote tumorigenesis. *Molecular cell*, 42(1):36–49, 4 2011.
- [49] Grace M. Wang, Hong Yuen Wong, Hiroyuki Konishi, Brian G. Blair, Abde M. Abukhdeir, John P. Gustin, D. Marc Rosen, Samuel Ray Denmeade, Zeshaan Rasheed, William Matsui, Joseph P. Garay, Morassa Mohseni, Michaela J. Higgins, Justin Cidado, Danijela Jelovac, Sarah Croessmann, Rory L. Cochran, Sivasundaram Karnan, Yuko Konishi, Akinobu Ota, Yoshitaka Hosokawa, Pedram Argani, Josh Lauring, and Ben Ho Park. Single copies of mutant KRAS and mutant PIK3CA Cooperate in immortalized human epithelial cells to induce tumor formation. *Cancer Research*, 73(11):3248–3261, 2013.
- [50] Shon Green, Christy L Trejo, and Martin McMahon. PIK3CA(H1047R) Accelerates and Enhances KRAS(G12D)-Driven Lung Tumorigenesis. *Cancer research*, 75(24):5378–91, 12 2015.
- [51] Chen-Hsiang Yeang, Frank McCormick, and Arnold Levine. Combinatorial patterns of somatic gene mutations in cancer. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 22(8):2605–22, 8 2008.
- [52] Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407):330–7, 7 2012.
- [53] Pierre-Olivier Angrand, Inmaculada Segura, Pamela Völkel, Sonja Ghidelli, Rebecca Terry, Miro Brajenovic, Kristina Vintersten, Rüdiger Klein, Giulio Superti-Furga, Gerard Drewes, Bernhard Kuster, Tewis Bouwmeester, and Amparo Acker-Palmer. Transgenic mouse proteomics identifies new 14-3-3-associated proteins involved in cytoskeletal rearrangements and cell signaling. *Molecular & cellular proteomics : MCP*, 5(12):2211–27, 12 2006.

- [54] Annette Grohmann, Kristina Tanneberger, Astrid Alzner, Jean Schneikert, and Jürgen Behrens. AMER1 regulates the distribution of the tumor suppressor APC between microtubules and the plasma membrane. *Journal of cell science*, 120(Pt 21):3738–47, 11 2007.
- [55] Kristina Tanneberger, Astrid S. Pfister, Vitezslav Kriz, Vitezslav Bryja, Alexandra Schambony, and Jürgen Behrens. Structural and functional characterization of the Wnt inhibitor APC membrane recruitment 1 (Amer1). *The Journal of biological chemistry*, 286(22):19204–14, 6 2011.
- [56] Jens G. Lohr, Petar Stojanov, Scott L. Carter, Peter Cruz-Gordillo, Michael S. Lawrence, Daniel Auclair, Carrie Sougnez, Birgit Knoechel, Joshua Gould, Gordon Saksena, Kristian Cibulskis, Aaron McKenna, Michael A. Chapman, Ravid Straussman, Joan Levy, Louise M. Perkins, Jonathan J. Keats, Steven E. Schumacher, Mara Rosenberg, Multiple Myeloma Research Consortium, Gad Getz, and Todd R. Golub. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer cell*, 25(1):91–101, 1 2014.
- [57] Fiona M Behan, Francesco Iorio, Gabriele Picco, Emanuel Gonçalves, charlotte M Beaver, Giorgia Migliardi, Rita Santos, Yanhua Rao, Francesco Sassi, Marika Pinnelli, Rizwan Ansari, Sarah Harper, David Adam Jackson, Rebecca McRae, Rachel Pooley, Piers Wilkinson, Dieudonne van der Meer, David Dow, Carolyn Buser-Doepner, Andrea Bertotti, Livio Trusolino, euan A Stronach, Julio Saez-Rodriguez, Kosuke Yusa, and Mathew J Garnett. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature*, 568(7753):511–516, 2019.
- [58] edmond M Chan, Tsukasa Shibue, James M McFarland, Benjamin Gaeta, Mahmoud Ghandi, Nancy Dumont, Alfredo Gonzalez, Justine S McPartlan, Tianxia Li, Yanxi Zhang, Jie Bin Liu, Jean-Bernard Lazaro, Peili Gu, Cortt G Piett, Annie Apffel, Syed O Ali, Rebecca Deasy, Paula Keskula, Raymond W S Ng, emma A Roberts, Elizaveta Reznichenko, Lisa Leung, Maria Alimova, Monica Schenone, Mirazul Islam, Yosef E Maruvka, Yang Liu, Jatin Roper, Srivatsan Raghavan, Marios Giannakis, Yuen-Yi Tseng, Zachary D Nagel, Alan D'Andrea, David E Root, Jesse S Boehm, Gad Getz, Sandy Chang, Todd R Golub, Aviad Tsherniak, Francisca Vazquez, and Adam J Bass. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature*, 568(7753):551–556, 2019.

- [59] Aviad Tsherniak, Francisca Vazquez, Phil G Montgomery, Barbara A Weir, Gregory Kryukov, Glenn S Cowley, Stanley Gill, William F Harrington, Sasha Pantel, John M Krill-Burger, Robin M Meyers, Levi Ali, Amy Goodale, Yenarae Lee, Guozhi Jiang, Jessica Hsiao, William F J Gerath, Sara Howell, Erin Merkel, Mahmoud Ghandi, Levi A Garraway, David E Root, Todd R Golub, Jesse S Boehm, and William C Hahn. Defining a Cancer Dependency Map. *Cell*, 170(3):564–576, 7 2017.
- [60] Robin M. Meyers, Jordan G. Bryan, James M. McFarland, Barbara A. Weir, Ann E. Sizemore, Han Xu, Neekesh V. Dharia, Phillip G. Montgomery, Glenn S. Cowley, Sasha Pantel, Amy Goodale, Yenarae Lee, Levi D. Ali, Guozhi Jiang, Rakela Lubonja, William F. Harrington, Matthew Strickland, Ting Wu, Derek C. Hawes, Victor A. Zhivich, Meghan R. Wyatt, Zohra Kalani, Jaime J. Chang, Michael Okamoto, Kimberly Stegmaier, Todd R. Golub, Jesse S. Boehm, Francisca Vazquez, David E. Root, William C. Hahn, and Aviad Tsherniak. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature genetics*, 49(12):1779–1784, 12 2017.
- [61] Katia Monastyrskaya, Verónica Sánchez-Freire, Ali Hashemi Gheinani, David J. Klumpp, Eduard B. Babiychuk, Annette Draeger, and Fiona C. Burkhard. miR-199a-5p Regulates Urothelial Permeability and May Play a Role in Bladder Pain Syndrome. *American Journal of Pathology*, 182(2):431–448, 2 2013.
- [62] Jorge Z. Torres, Matthew K. Summers, David Peterson, Matthew J. Brauer, James Lee, Silvia Senese, Ankur A. Ghokar, Yu Chen Lo, Xingye Lei, Kenneth Jung, David C. Anderson, David P. Davis, Lisa Belmont, and Peter K. Jackson. The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. *Cell*, 147(6):1309–1323, 12 2011.
- [63] G. K.T. Chan, S A Jablonski, D A Starr, M L Goldberg, and T J Yen. Human Zw10 and ROD are mitotic checkpoint proteins that bind to kinetochores. *Nature cell biology*, 2(12):944–7, 12 2000.
- [64] F Scaërou, D A Starr, F Piano, O Papoulas, R E Karess, and M L Goldberg. The ZW10 and Rough Deal checkpoint proteins function together in a large, evolutionarily conserved complex targeted to the kinetochore. *Journal of cell science*, 114(Pt 17):3103–14, 9 2001.
- [65] Geert J P L Kops, Yumi Kim, Beth A A Weaver, Yinghui Mao, Ian McLeod, John R Yates, Mitsuo Tagaya, and Don W Cleveland. ZW10 links mitotic checkpoint signaling to the structural kinetochore. *The Journal of cell biology*, 169(1):49–60, 4 2005.

- [66] Isabelle Barlat, Florence Maurier, Marc Duchesne, Estelle Guitard, Bruno Tocque, and Fabien Schweighoffer. A Role for Sam68 in Cell Cycle Progression Antagonized by a Spliced Variant within the KH Domain. *Journal of Biological Chemistry*, 272(6):3129–3132, 2 1997.
- [67] Mircea Ivan and William G Kaelin. The EGLN-HIF O₂-Sensing System: Multiple Inputs and Feedbacks. *Molecular Cell*, 66(6):772–779, 2017.
- [68] Seon Ah Ha, Min Shin Seung, Jin Lee Yong, Sanghee Kim, Kee Kim Hyun, Hong Namkoong, Heejeong Lee, Soo Lee Youn, Young Seok Cho, Gyu Park Yong, Myung Jeon Hae, Changkyu Oh, and Woo Kim Jin. HCCRBP-1 Directly Interacting With HCCR-1 Induces Tumorigenesis Through P53 Stabilization. *International Journal of Cancer*, 122(3):501–508, 2008.
- [69] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 4 2005.
- [70] Douglas K Brubaker, Joao A Paulo, Shikha Sheth, Emily J. Poulin, Olesja Popow, Brian A. Joughin, Samantha Dale Strasser, Alina Starchenko, Steven P Gygi, Douglas A Lauffenburger, and Kevin M Haigis. Proteogenomic Network Analysis of Context-Specific KRAS Signaling in Mouse-to-Human Cross-Species Translation. *Cell systems*, 9(3):258–270, 9 2019.
- [71] Christian W Johnson, Yi-Jang Lin, Derion Reid, Jillian Parker, Spiro Pavlopoulos, Patrick Dischinger, Carrie Graveel, Andrew J Aguirre, Matthew Steensma, Kevin M Haigis, and Carla Mattos. Isoform-Specific Destabilization of the Active Site Reveals a Molecular Mechanism of Intrinsic Activation of KRas G13D. *Cell reports*, 28(6):1538–1550, 8 2019.
- [72] Zhan Yao, Neilawattie M. Torres, Anthony Tao, Yijun Gao, Lusong Luo, Qi Li, Elisa de Stanchina, Omar Abdel-Wahab, David B. Solit, Poulikos I. Poulikakos, and Neal Rosen. BRAF Mutants Evade ERK-Dependent Feedback by Different Mechanisms that Determine Their Sensitivity to Pharmacologic Inhibition. *Cancer cell*, 28(3):370–83, 9 2015.
- [73] Zhan Yao, Rona Yaeger, Vanessa S Rodrik-Outmezguine, Anthony Tao, Neilawattie M Torres, Matthew T Chang, Matthias Drosten, Huiyong Zhao, Fabiola Cecchi, Todd Hembrough, Judith Michels, Hervé Baumert, Linde Miles, Naomi M Campbell, Elisa de Stanchina, David B Solit, Mariano Barbacid, Barry S Taylor, and

- Neal Rosen. Tumours with class 3 BRAF mutants are sensitive to the inhibition of activated RAS. *Nature*, 548(7666):234–238, 2017.
- [74] Ibiayi Dagogo-Jack, Pablo Martinez, Beow Y Yeap, Chiara Ambrogio, Lorin A Ferris, Christine Lydon, Tom Nguyen, Nicholas A Jessop, A John Iafrate, Bruce E Johnson, Jochen K Lennerz, Alice T Shaw, and Mark M Awad. Impact of BRAF Mutation Class on Disease Characteristics and Clinical Outcomes in BRAF-mutant Lung Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 25(1):158–165, 1 2019.
- [75] Jillian Wilhelmina Paulina Bracht, Niki Karachaliou, Trever Bivona, Richard B. Lanman, Iris Faull, Rebecca J. Nagy, Ana Drozdowskyj, Jordi Berenguer, Manuel Fernandez-Bruno, Miguel Angel Molina-Vila, and Rafael Rosell. BRAF Mutations Classes I, II, and III in NSCLC Patients Included in the SLLIP Trial: The Need for a New Pre-Clinical Treatment Rationale. *Cancers*, 11(9), 9 2019.
- [76] David M. Hyman, Sarina A. Piha-Paul, Helen Won, Jordi Rodon, Cristina Saura, Geoffrey I. Shapiro, Dejan Juric, David I. Quinn, Victor Moreno, Bernard Doger, Ingrid A. Mayer, Valentina Boni, Emiliano Calvo, Sherene Loi, Albert C. Lockhart, Joseph P. Erinjeri, Maurizio Scaltriti, Gary A. Ulaner, Juber Patel, Jiabin Tang, Hannah Beer, S. Duygu Selcuklu, Aphrothiti J. Hanrahan, Nancy Bouvier, Myra Melcer, Rajmohan Murali, Alison M. Schram, Lillian M. Smyth, Komal Jhaveri, Bob T. Li, Alexander Drilon, James J. Hardling, Gopa Iyer, Barry S. Taylor, Michael F. Berger, Richard E. Cutler, Feng Xu, Anna Butturini, Lisa D. Eli, Grace Mann, Cynthia Farrell, Alshad S. Lalani, Richard P. Bryce, Carlos L. Arteaga, Funda Meric-Bernstam, José Baselga, and David B. Solit. HER kinase inhibition in patients with HER2-and HER3-mutant cancers. *Nature*, 554(7691):189–194, 2 2018.
- [77] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, Ethan Cerami, Chris Sander, and Nikolaus Schultz. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*, 6(269):pl1, 4 2013.
- [78] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J Byrne, Michael L Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P

- Goldberg, Chris Sander, and Nikolaus Schultz. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*, 2(5):401–4, 5 2012.
- [79] Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, 511(7511):543–50, 7 2014.
- [80] Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer cell*, 32(2):185–203, 2017.
- [81] Christopher J. Scarlett, Elizabeth L. Salisbury, Andrew V. Biankin, and James Kench. Precursor lesions in pancreatic cancer: morphological and molecular pathology. *Pathology*, 43(3):183–200, 4 2011.
- [82] Brian A Walker, Konstantinos Mavrommatis, Christopher P Wardell, T. Cody Ashby, Michael Bauer, Faith Davies, Adam Rosenthal, Hongwei Wang, Pingping Qu, Antje Hoering, Mehmet Samur, Fadi Towfic, Maria Ortiz, Erin Flynt, Zhinuan Yu, Zhihong Yang, Dan Rozelle, John Obenauer, Matthew Trotter, Daniel Auclair, Jonathan Keats, Niccolo Bolli, Mariateresa Fulciniti, Raphael Szalat, Phillippe Moreau, Brian Durie, A. Keith Stewart, Hartmut Goldschmidt, Marc S Raab, Hermann Einsele, Pieter Sonneveld, Jesus San Miguel, Sagar Lonial, Graham H Jackson, Kenneth C Anderson, Herve Avet-Loiseau, Nikhil Munshi, Anjan Thakurta, and Gareth Morgan. A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia*, 33(1):159–170, 2019.
- [83] AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer discovery*, 7(8):818–831, 2017.
- [84] GTEx Consortium, Data Analysis Laboratory, Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration and Visualization—EBI, University of California Santa Cruz Genome Browser Data Integration & Visualization—UCSC Genomics Institute, Lead analysts:, Data Analysis Laboratory, Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript

working group:, Alexis Battle, Christopher D Brown, Barbara E. Engelhardt, and Stephen B. Montgomery. Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204–213, 2017.

- [85] Mathias Uhlen, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, IngMarie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al-Khalili Szigyarto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M Schwenk, Marica Hamsten, Kalle von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar von Heijne, Jens Nielsen, and Fredrik Pontén. Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)*, 347(6220):1260419, 1 2015.
- [86] Mathias Uhlen, Björn M Hallström, Cecilia Lindskog, Adil Mardinoglu, Fredrik Pontén, and Jens Nielsen. Transcriptomics resources of human tissues and organs. *Molecular systems biology*, 12(4):862, 4 2016.
- [87] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2020. *CA: a cancer journal for clinicians*, 70(1):7–30, 1 2020.
- [88] Rafael Meza, Clare Meernik, Jihyoun Jeon, and Michele L Cote. Lung cancer incidence trends by gender, race and histology in the United States, 1973-2010. *PloS one*, 10(3):e0121323, 2015.
- [89] Christian von Mering, Lars J Jensen, Berend Snel, Sean D Hooper, Markus Krupp, Mathilde Foglierini, Nelly Jouffre, Martijn A Huynen, and Peer Bork. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic acids research*, 33(Database issue):433–7, 1 2005.
- [90] Damian Szkłarczyk, Annika L Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T Doncheva, John H Morris, Peer Bork, Lars J Jensen, and Christian von Mering. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1):D607–D613, 1 2019.
- [91] Jishnu Das and Haiyuan Yu. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC systems biology*, 6:92, 7 2012.
- [92] Edward L. Huttlin, Lily Ting, Raphael J. Bruckner, Fana Gebreab, Melanie P. Gygi, John Szpyt, Stanley Tam, Gabriela Zarraga, Greg Colby, Kurt Baltier, Rui Dong, Virginia Guarani, Laura Pontano Vaites, Alban

- Ordureau, Ramin Rad, Brian K. Erickson, Martin Wühr, Joel Chick, Bo Zhai, Deepak Kolippakkam, Julian Mintseris, Robert A. Obar, Tim Harris, Spyros Artavanis-Tsakonas, Mathew E. Sowa, Pietro De Camilli, Joao A. Paulo, J. Wade Harper, and Steven P. Gygi. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*, 162(2):425–440, 7 2015.
- [93] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.
- [94] Ludmil B. Alexandrov, Serena Nik-Zainal, David C. Wedge, Peter J. Campbell, and Michael R. Stratton. Deciphering signatures of mutational processes operative in human cancer. *Cell reports*, 3(1):246–59, 1 2013.
- [95] John G Tate, Sally Bamford, Harry C Jubb, Zbyslaw Sondka, David M Beare, Nidhi Bindal, Harry Boutselakis, Charlotte G Cole, Celestino Creatore, Elisabeth Dawson, Peter Fish, Bhavana Harsha, Charlie Hathaway, Steve C Jupe, Chai Yin Kok, Kate Noble, Laura Ponting, Christopher C Ramshaw, Claire E Rye, Helen E Speedy, Ray Stefancsik, Sam L Thompson, Shicai Wang, Sari Ward, Peter J Campbell, and Simon A Forbes. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research*, 47(D1):D941–D947, 1 2019.
- [96] Nicholas K. Hayward, James S. Wilmott, Nicola Waddell, Peter A. Johansson, Matthew A. Field, Katia Nones, Ann-Marie Patch, Hojabr Kakavand, Ludmil B. Alexandrov, Hazel Burke, Valerie Jakrot, Stephen Kazakoff, Oliver Holmes, Conrad Leonard, Radhakrishnan Sabarinathan, Loris Mularoni, Scott Wood, Qinying Xu, Nick Waddell, Varsha Tembe, Gulietta M. Pupo, Ricardo De Paoli-Iseppi, Ricardo E. Vilain, Ping Shang, Loretta M S Lau, Rebecca A. Dagg, Sarah-Jane Schramm, Antonia Pritchard, Ken Dutton-Regester, Felicity Newell, Anna Fitzgerald, Catherine A. Shang, Sean M. Grimmond, Hilda A. Pickett, Jean Y. Yang, Jonathan R. Stretch, Andreas Behren, Richard F. Kefford, Peter Hersey, Georgina V. Long, Jonathan Cebon, Mark Shackleton, Andrew J. Spillane, Robyn P M Saw, Núria López-Bigas, John V. Pearson, John F. Thompson, Richard A. Scolyer, and Graham J. Mann. Whole-genome landscapes of major melanoma subtypes. *Nature*, 545(7653):175–180, 2017.
- [97] Henry Lee-Six, Sigurgeir Olafsson, Peter Ellis, Robert J Osborne, Mathijs A Sanders, Luiza Moore, Nikitas Georgakopoulos, Franco Torrente, Ayesha Noorani, Martin Goddard, Philip Robinson, Tim H H Coorens, Laura O'Neill, Christopher Alder, Jingwei Wang, Rebecca C Fitzgerald, Matthias Zilbauer, Nicholas Cole-

- man, Kourosh Saeb-Parsy, Inigo Martincorena, Peter J Campbell, and Michael R Stratton. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*, 574(7779):532–537, 10 2019.
- [98] Maura Costello, Trevor J. Pugh, Timothy J. Fennell, Chip Stewart, Lee Lichtenstein, James C. Meldrim, Jennifer L. Fostel, Dennis C. Friedrich, Danielle Perrin, Danielle Dionne, Sharon Kim, Stacey B. Gabriel, Eric S. Lander, Sheila Fisher, and Gad Getz. 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(6):e67, 4 2013.
- [99] Doga C. Gulhan, Jake June-Koo Lee, Giorgio E M Melloni, Isidro Cortés-Ciriano, and Peter J. Park. Detecting the mutational signature of homologous recombination deficiency in clinical samples. *Nature genetics*, 51(5):912–919, 2019.
- [100] Angelo Carty and Brian Ripley. *boot: Bootstrap Functions (Originally by Angelo Carty for S)*, 2019. R package version 1.3-23.
- [101] Edward Y Chen, Christopher M Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R Clark, and Avi Ma'ayan. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14(4):128, 4 2013.
- [102] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, Michael G McDermott, Caroline D Monteiro, Gregory W Gundersen, and Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):90–7, 2016.
- [103] Wajid Jawaid. *enrichR: Provides an R Interface to 'Enrichr'*, 2019. R package version 2.1.
- [104] Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan, and Tyler Hunt. *caret: Classification and Regression Training*, 2019. R package version 6.0-84.
- [105] Guido Van Rossum and Fred L Drake Jr. *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.

[106] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

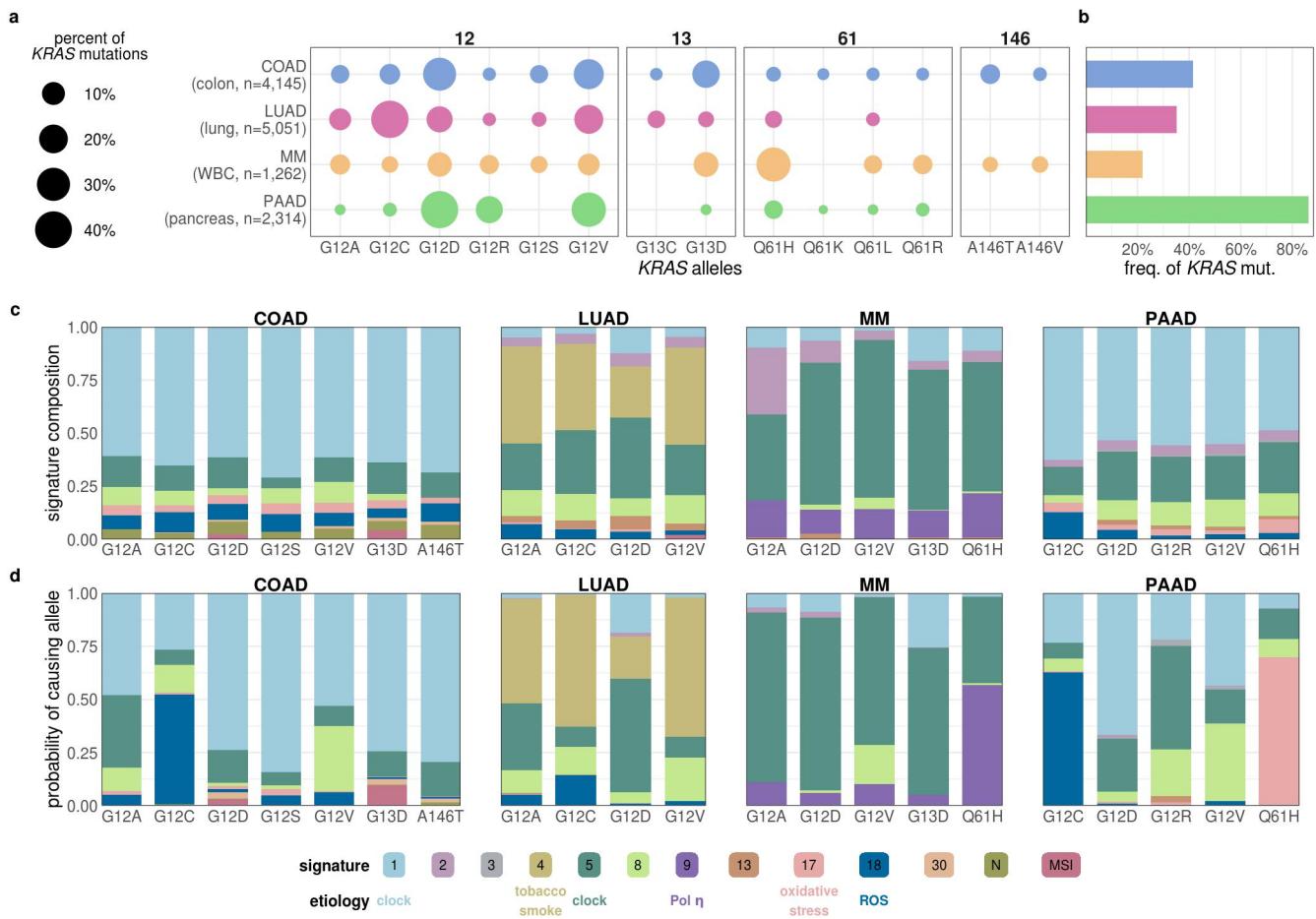


Figure 1: The contribution of mutational processes to KRAS mutagenesis. **a.** The distribution of KRAS allele frequencies at the four hotspots, codons 12 (left), 13 (middle-left), 61 (middle-right), and 146 (right) in each cancer. The size of the circle reflects the percent of KRAS mutations that are the indicated allele in each cancer. Each cancer is assigned a different color. The number of tumor samples whose sequencing data was collected for this study is indicated along the y-axis. **b.** The frequency of KRAS mutations in each cancer. **c.** The average levels of mutational signatures in tumor samples separated by KRAS allele. Each color represents a different mutational signature. Mutational signatures of known etiology are annotated. **d.** The average probability of each mutational signature to have caused the KRAS mutation in a tumor sample. This value accounts for the level of each mutational signature in the tumor sample and the ability of the mutational signature to cause the indicated KRAS allele.

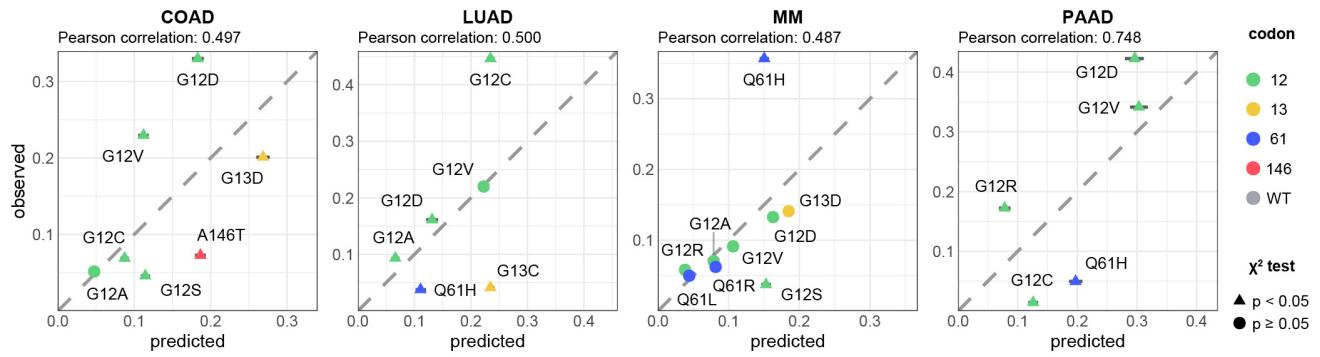


Figure 2: **The predicted frequencies of cancer-specific KRAS alleles.** The predicted vs. observed frequency of KRAS alleles for the common alleles of each cancer. ▲ indicates rejection of the null hypothesis that the observed and predicted frequencies are the same (Chi-squared test, $p < 0.05$). ● indicates the failure to reject the null hypothesis (Chi-squared test, $p \geq 0.05$). Error bars indicate bootstrapped 95% confidence intervals of the predicted values.

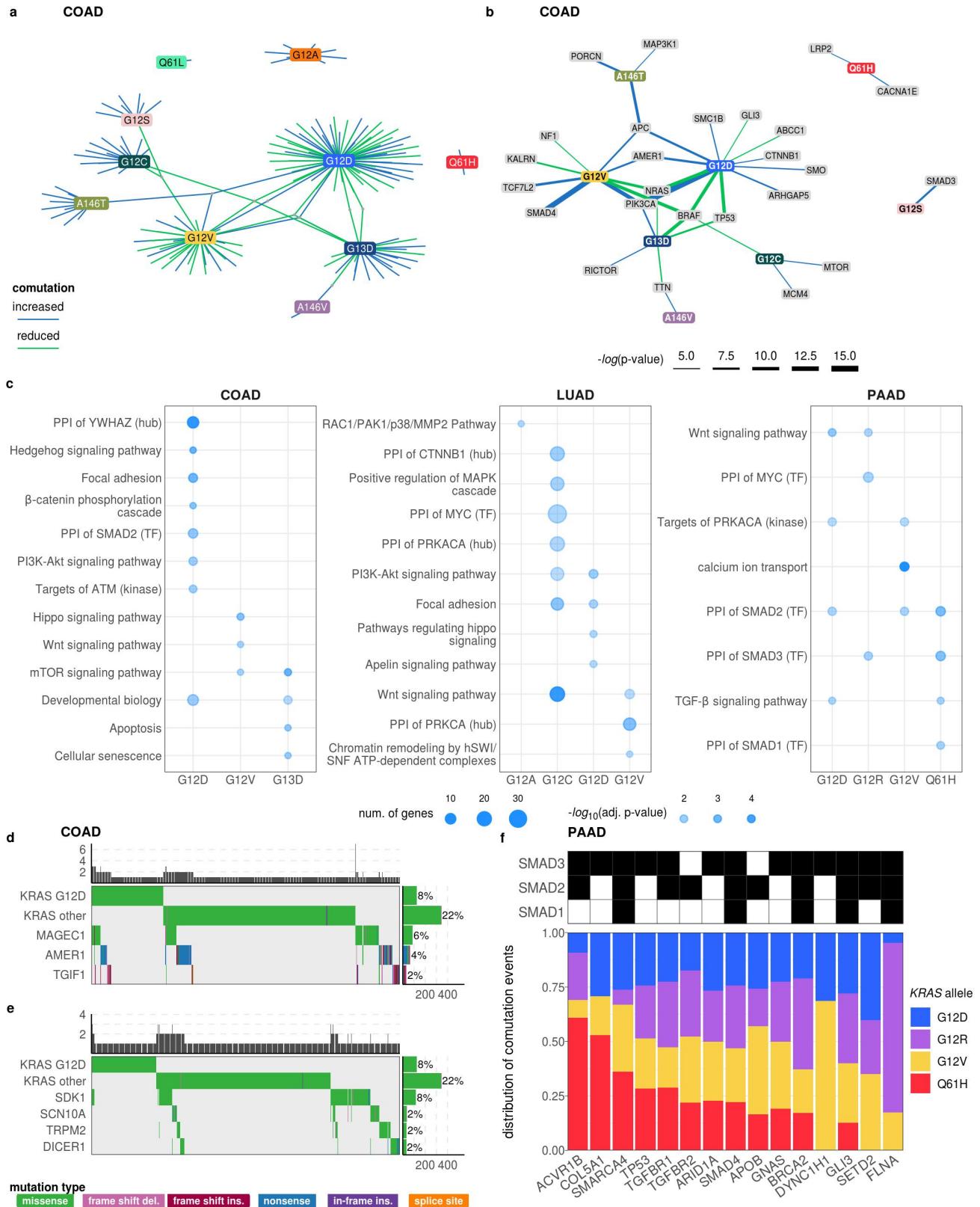


Figure 3: The comutation networks of oncogenic *KRAS* alleles.

Figure 3. The comutation networks of oncogenic KRAS alleles. **a.** The comutation network of the *KRAS* alleles in COAD with each edge representing a comutation interaction between an allele and another gene. The color of the edge indicates whether the interaction was an increase (blue) or decrease (green) in the frequency of comutation. Genes with multiple interactions are represented by a grey dot to disambiguate them from where edges intersect. **b.** A subset of the network shown in panel a of genes that encode proteins known to physically interact with K-RAS, are in one of its canonical up- or downstream pathways, or are validated oncogenes. The width of the edge indicates the strength of the association. **c.** Cellular functions enriched in the comutation networks of the *KRAS* alleles in COAD (left), LUAD (center), and PAAD (right). The size of the dot indicates the number of genes in both the function and the comutation network, and the transparency indicates the p-value of the enrichment. **d,** **e.** A visualization of the increased (**d**) or decreased (**e**) comutation of select genes with *KRAS* G12D in COAD. Rows of the central plot represent genes. Each column of the central plot is a different tumor sample. A filled space denotes a mutation of the gene in the sample, the color describing the type of variant. The bar plots above and to the right indicate the marginal values of the central plot. **f.** A comparison of the comutation frequencies in PAAD of the genes producing proteins in the PPIN of SMAD1-3. Each column is a gene with a comutation interaction with a *KRAS* allele and in at least one of the gene sets. The black tiles on top indicate that the gene was in the PPIN of the indicated SMAD protein. The bar plot shows the distribution of the comutation events of each gene across tumor samples with the various *KRAS* mutations.

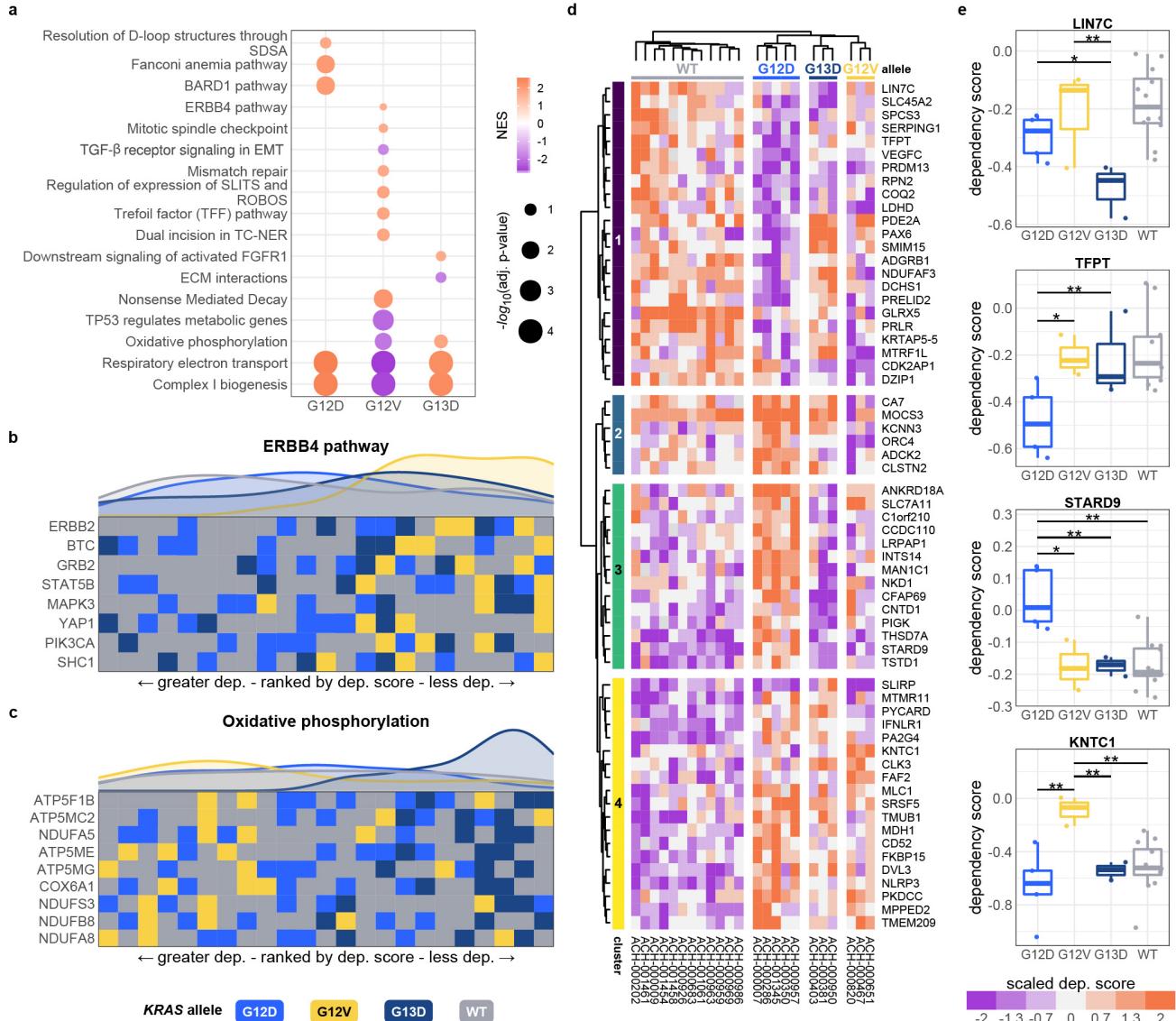


Figure 4: Allele-specific genetic dependencies in COAD cell lines. **a.** Gene sets with significant enrichment for increased (lower dependency score; purple) or reduced (higher dependency score; orange) genetic dependency in COAD cell lines. The size of the dot relates the p-value of the association and the color indicated the strength of the enrichment ("normalized enrichment score"). **b, c.** Heatmaps ranking the cell lines by dependency score of the genes at the leading edge of enrichment for two gene sets. Each row represents a gene and each cell represents a cell line colored by its KRAS allele. The cell lines are arranged in ranking order by their dependency score for the gene. Thus, each column indicates a rank. The line plots above the heatmaps indicate the representation (density) of each KRAS allele at each rank across the genes. **d.** Hierarchically clustered heatmaps of the genes that demonstrated differential genetic dependency amongst cell lines of different KRAS alleles. Each column is a cell line labeled by its DepMap identifier and each row is a gene. **e.** Examples of genes that demonstrated differential genetic dependency amongst cell lines of different KRAS alleles (pairwise t-tests; *: p<0.05, **: p<0.01, ***: p<0.001; p-values were adjusted using the Benjamini-Hochberg FDR correction method).

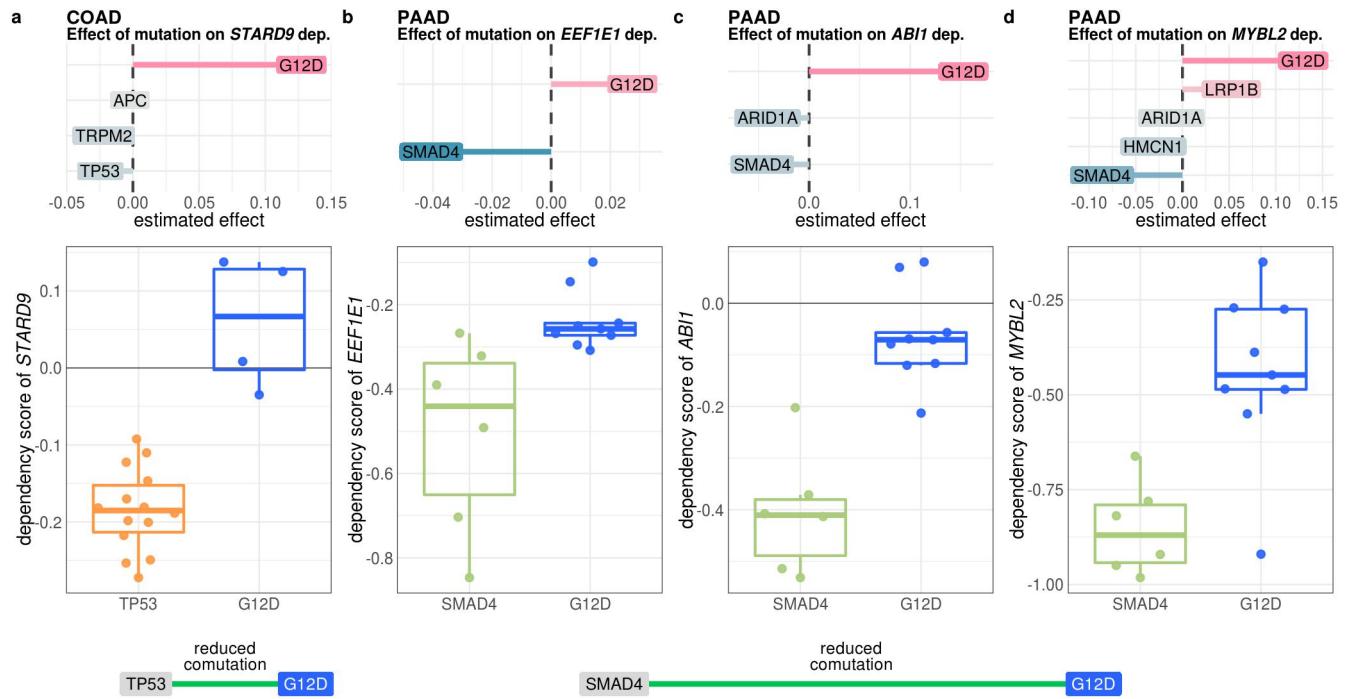
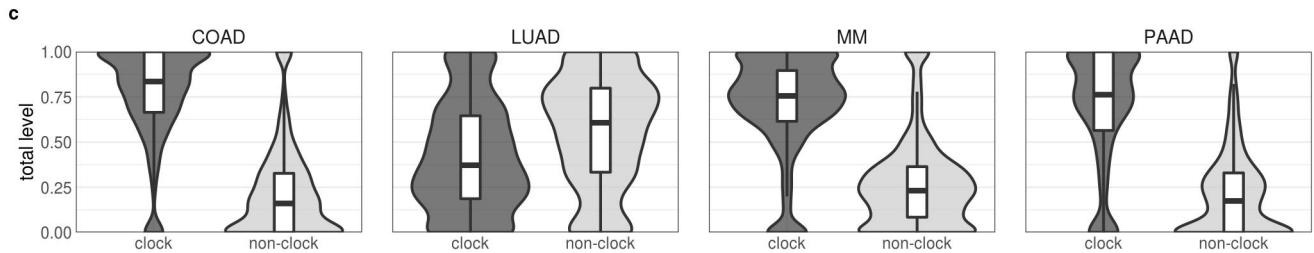
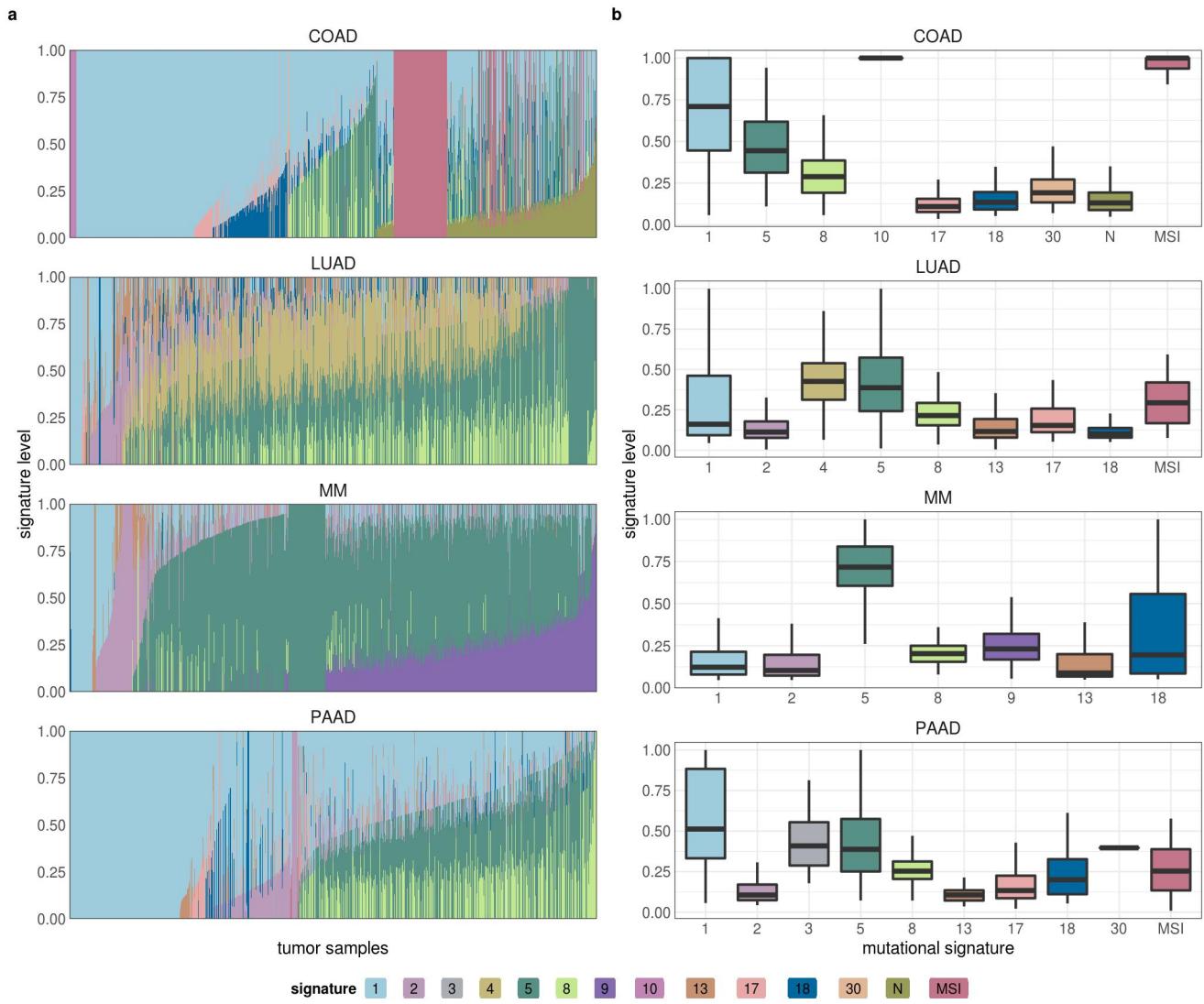
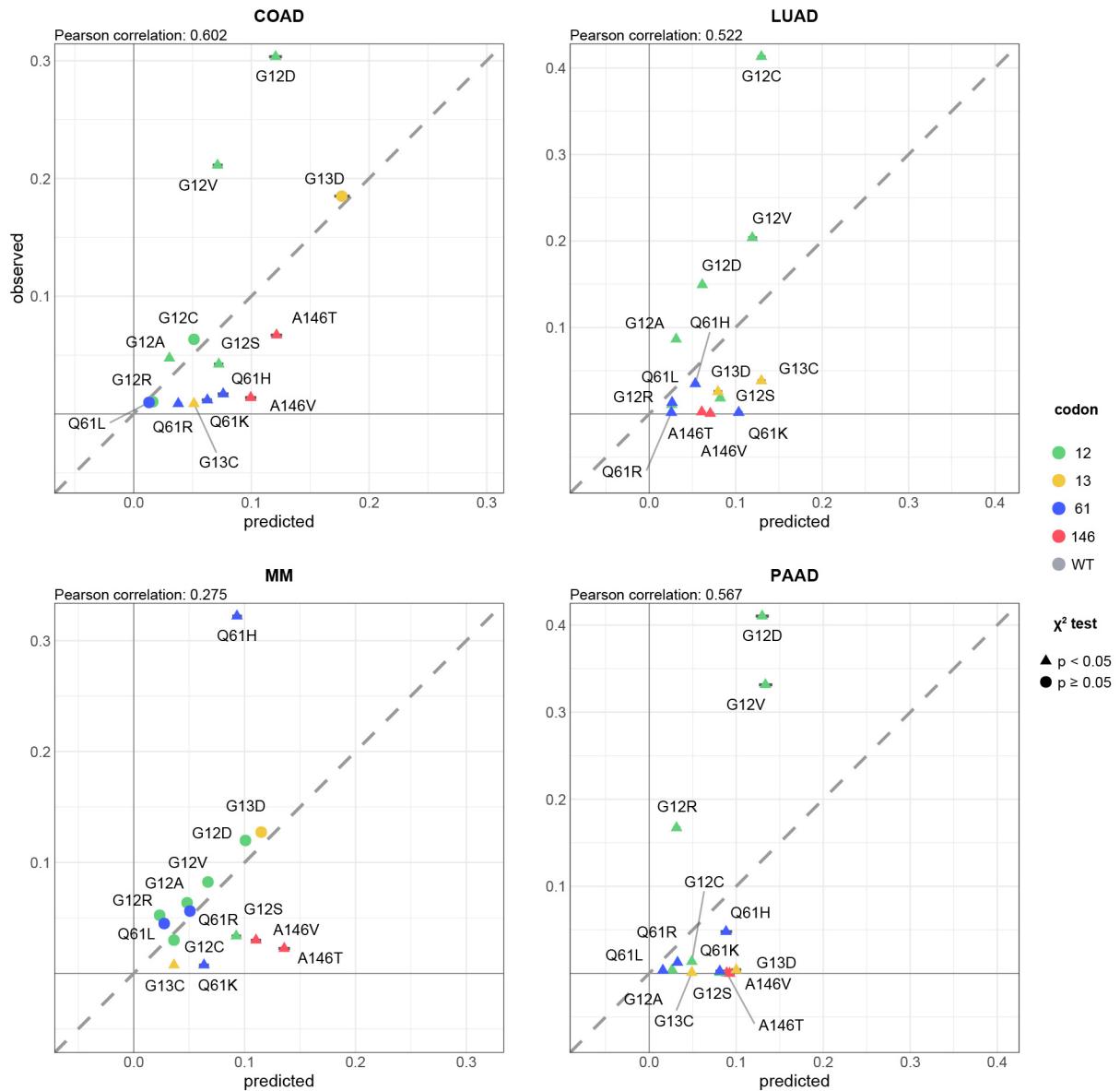


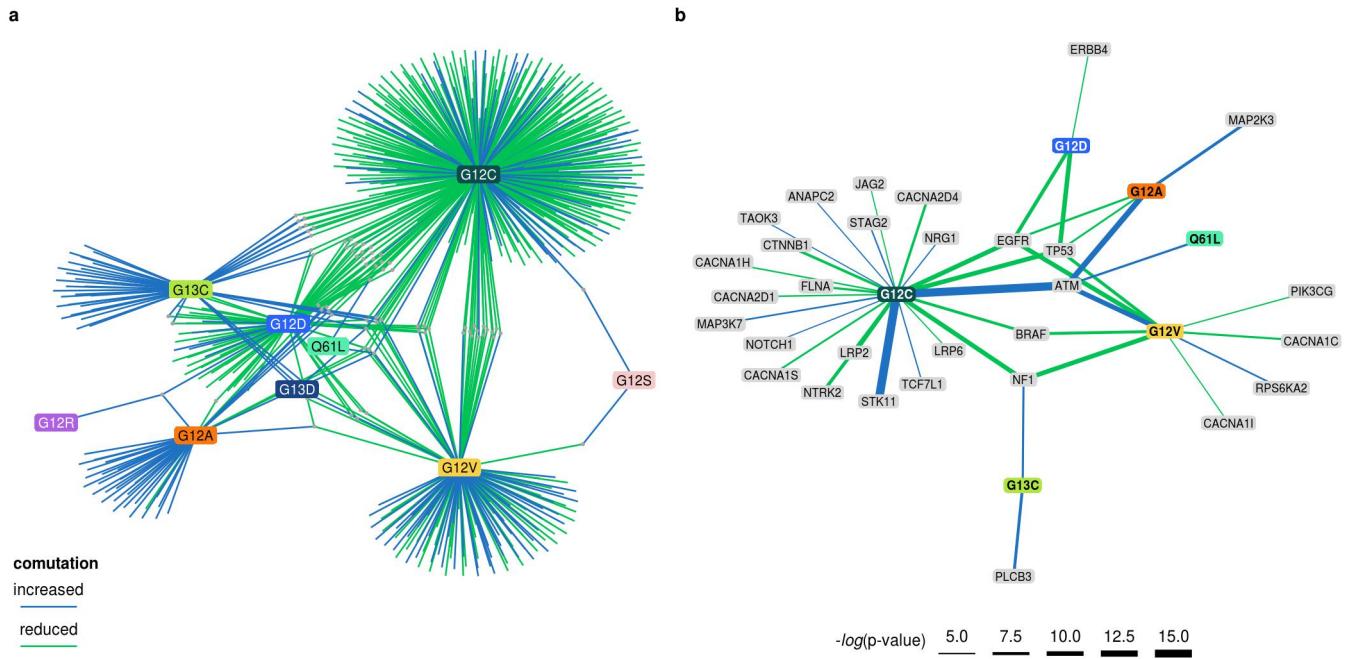
Figure 5: Some dependency interactions can be explained by comutation events. **a.** The non-zero coefficients for the model of *STARD9* dependency in COAD cell lines regressed on *KRAS* G12D (versus all other *KRAS* alleles) and its comutation interactors (top), and the actual dependency scores for *KRAS* G12D mutant and *TP53* mutant cell lines (bottom). Cell lines without either mutation or with both are not shown. *KRAS* G12D has reduced comutation with *TP53* in COAD. **b, c, d.** The non-zero coefficients for the models of (b) *EEF1E1*, (c) *ABL1*, and (d) *MYBL2* dependency in PAAD cell lines regressed on *KRAS* G12D (versus all other *KRAS* alleles) and its comutation interactors (top), and the actual dependency scores for *KRAS* G12D mutant and *SMAD4* mutant cell lines (bottom). Cell lines without either mutation or with both are not shown. *KRAS* G12D has reduced comutation with *SMAD4* in PAAD.



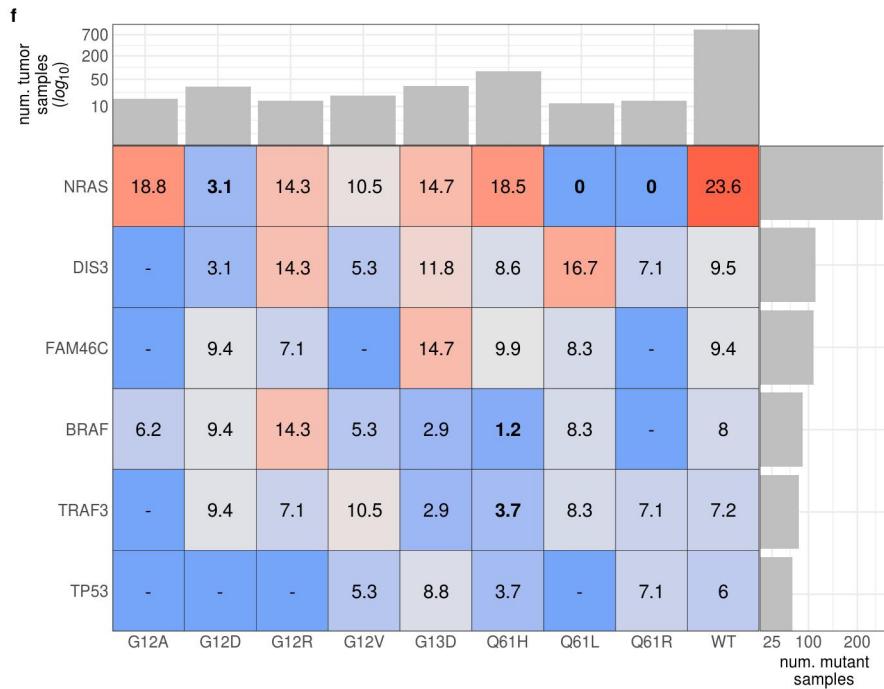
Supplementary Figure 1: **Mutational signatures in tumor samples.** **a, b.** The detected level of the mutational signatures in each tumor sample. In **a**, each tumor sample is a column. **c.** The average levels of clock (signatures 1 and 5) and non-clock (all other signatures) in the tumor samples.



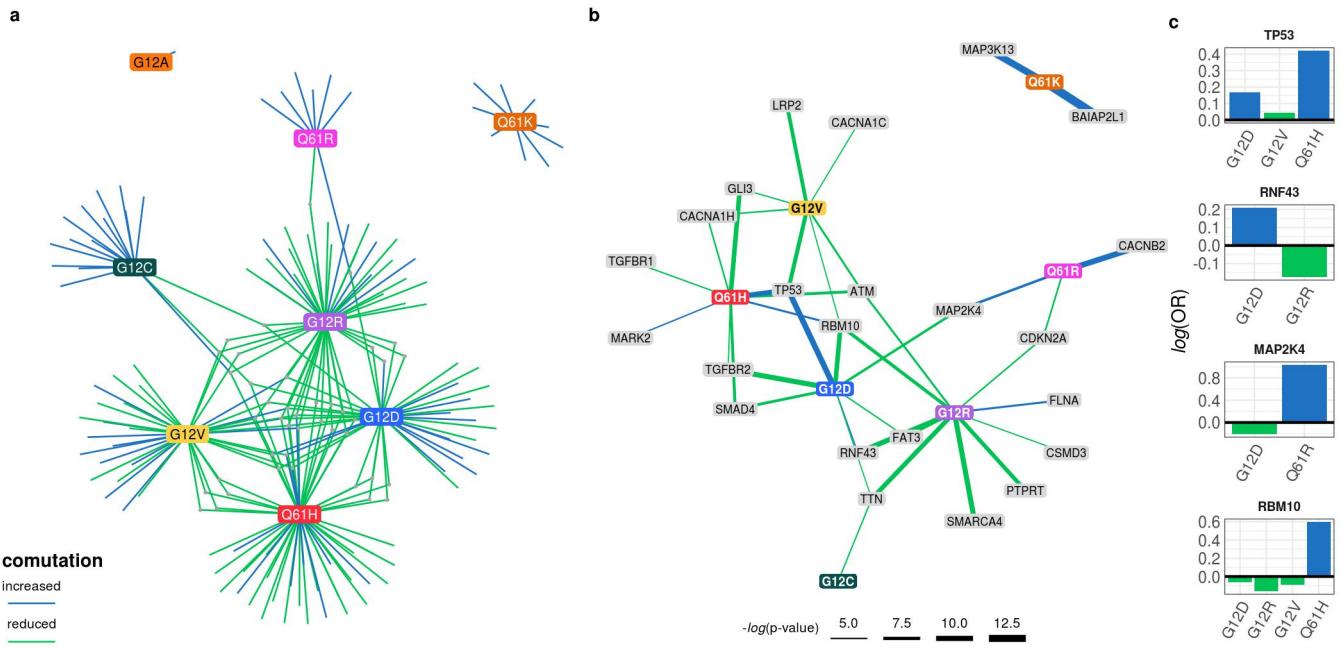
Supplementary Figure 2: **The predicted frequencies of all oncogenic KRAS alleles in each cancer.** The predicted vs. observed frequency of KRAS alleles in each cancer for all KRAS alleles demonstrated to drive any of the four cancers. The KRAS alleles included in the calculation were found mutated frequently in at least one of the four cancer types. ▲ indicates rejection of the null hypothesis that the observed and predicted frequencies are the same (Chi-squared test, $p < 0.05$). ● indicates the failure to reject the null hypothesis (Chi-squared test, $p \geq 0.05$). Error bars indicate bootstrapped 95% confidence intervals of the predicted values.



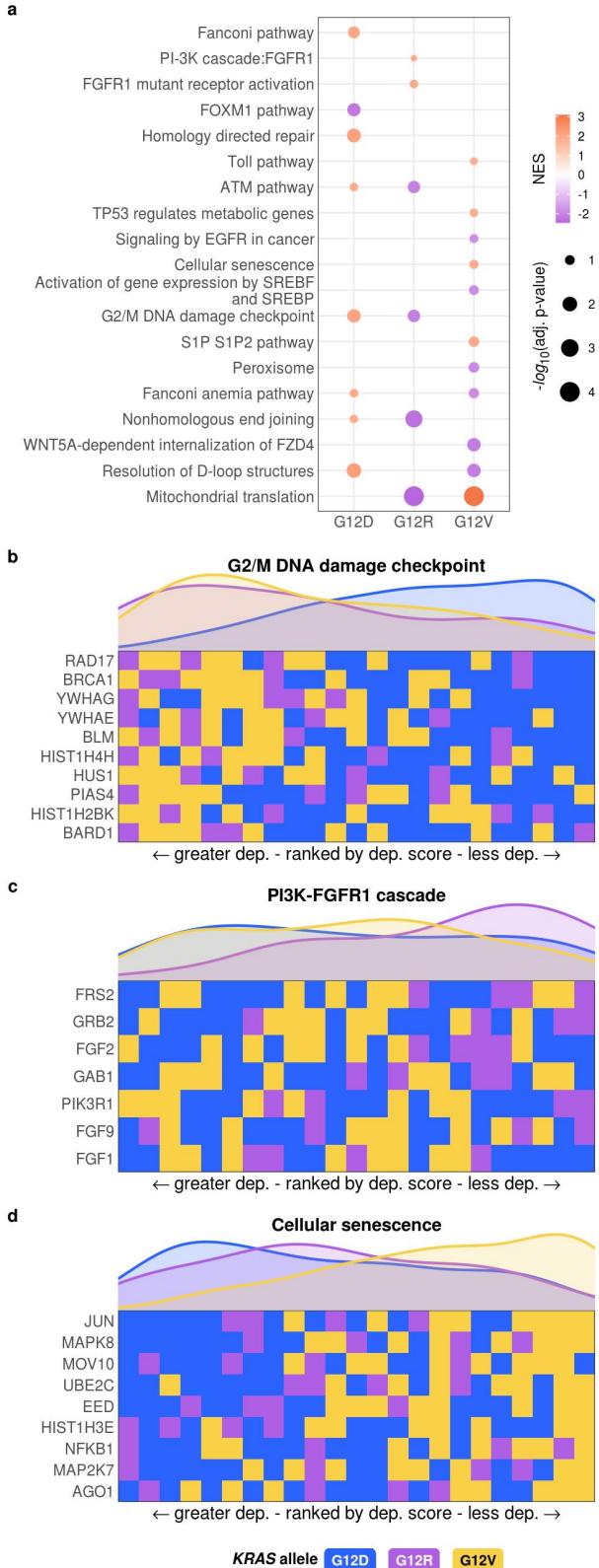
Supplementary Figure 3: **The comutation networks of KRAS alleles in LUAD.** **a.** The comutation network of the KRAS alleles in LUAD where each edge represents a comutation interaction between an allele and another gene. The color of the edge indicates whether the interaction was an increase (blue) or decrease (green) in the frequency of comutation. Genes with multiple interactions are represented by a grey dot to disambiguate them from where edges intersect. **b.** A subset of the network shown in **a** of genes in one of the canonical up- or downstream signaling pathways of K-RAS. The width of the edge indicates the strength of the association.



Supplementary Figure 4: **The comutation frequencies between known MM driver genes and KRAS alleles.** The color is correlated with the comutation frequency (the fraction of cells with the KRAS allele that also have a mutation in the other driver gene), indicated in each cell. Bold percent values indicate statistical significance of a comutation interaction (see Methods). The bar plot along the top indicates the number of samples with the KRAS allele, and the bar plot on the right indicates the number of samples with a mutation in the gene.

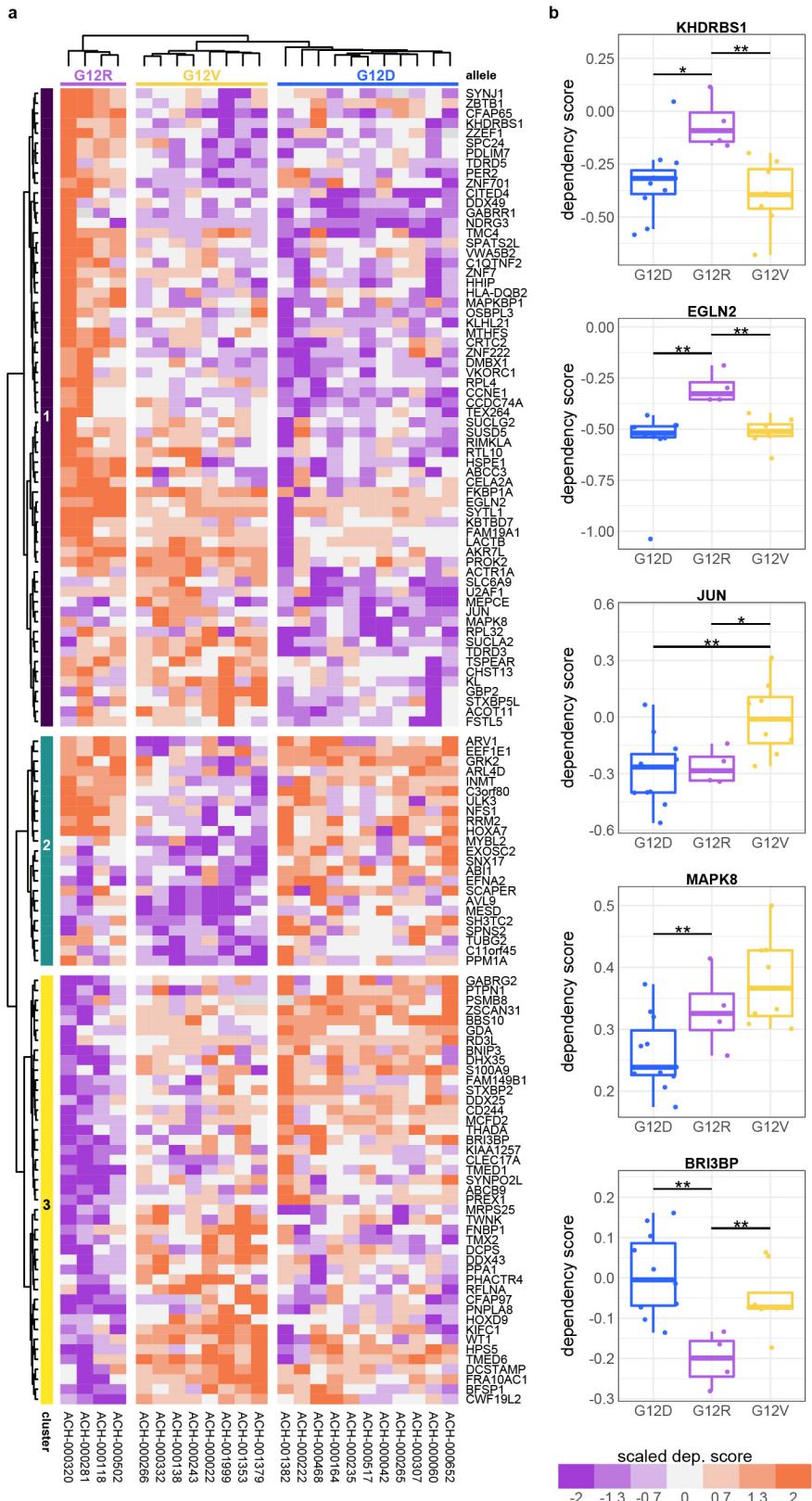


Supplementary Figure 5: **The comutation networks of KRAS alleles in PAAD.** **a.** The comutation network of the KRAS alleles in PAAD where each edge represents a comutation interaction between an allele and another gene. The color of the edge indicates whether the interaction was an increase (blue) or decrease (green) in the frequency of comutation. Genes with multiple interactions are represented by a grey dot to disambiguate them from where edges intersect. **b.** A subset of the network shown in **a** of genes known to physically interact with KRAS, are in one of its canonical up- or downstream pathways, or are validated oncogenes. The width of the edge indicates the strength of the association. **c.** The log-odds of comutation between KRAS alleles and other genes that had detectable opposing comutation interactions with multiple alleles.



Supplementary Figure 6: Allele-specific genetic dependencies on cellular processes and pathways in PAAD cell lines.

Supplementary Figure 6. Cellular processes enriched for greater or lesser genetic dependencies in PAAD cell lines separated by KRAS allele. **a.** Gene sets with significant enrichment for increased (lower dependency score; purple) or reduced (higher dependency score; orange) genetic dependency in PAAD cell lines. The size of the dot relates the p-value of the association, and the color indicates the strength of the enrichment. **b, c, d.** Heatmaps ranking the cell lines by dependency score of genes at the leading edge of enrichment for three gene sets in PAAD. Each row represents a gene and each cell represents a cell line colored by its KRAS allele. The cell lines were arranged in ranking order by their dependency score for the gene. Thus, each column indicates a rank. The line plots above the heatmaps indicate the representation (density) of each KRAS allele at each rank across the genes.



Supplementary Figure 7: Individual genes with differential genetic dependency by KRAS allele in PAAD cell lines.

Supplementary Figure 7. Individual genes with differential genetic dependency by KRAS allele in PAAD cell lines. **a.** Clustered heatmaps of the genes that demonstrated differential genetic dependency amongst PAAD cell lines of different KRAS alleles. Each column is a cell line labeled by its DepMap ID and each row is a gene. **b.** Examples of genes that demonstrated differential genetic dependency amongst cell lines of different KRAS alleles (pairwise *t*-tests; *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$; p-values were adjusted using the Benjamini-Hochberg FDR correction method).