*KRAS* allele-specific genetic interactions

Joshua Cook1,2,3, 7, Giorgio Melloni3, 7, Doga Gulhan3, Peter J. Park3,4\*, Kevin M. Haigis1,2,5,6\*

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) and Peter J. Park (peter\_park@hms.harvard.edu)

# 

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

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

Previous studies have documented substantial differences in the biochemistry and signaling properties of the common KRas variants (extensively reviewed by3,4). KRas operates as a molecular switch, activating downstream pathways when bound to GTP, but inactive when GDP-bound following the hydrolysis of the -phosphate. This reaction is catalyzed by a GTPase-activating proteins (GAPs) and the exchange of the GDP for a new GTP is facilitated by a guanine nucleotide exchange factors (GEFs).5 Activating KRas mutations result in an increase in downstream pathway activation by increasing the steady-state concentration of GTP-bound KRas. For example, mutations to codons 12, 13, and 61 reduce the rate of intrinsic and/or GAP-mediated hydrolysis, and mutants at 13 and 61, but not 12, also enhance the rate of nucleotide exchange.6,7 Alternatively, A146 mutations do not alter the rate of GTP hydrolysis, but cause hyperactivation through an increased rate of GDP exchange.8–11 Additional biochemical, structural, and signaling distinctions have been identified between different mutant alleles at the same amino acid position.4,6,11–18

Likely as a consequence of their distinct properties, associations have been uncovered between the specific *KRAS* mutation status of a patient’s cancer and its drug-response or clinical outcome.4,19 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.20 It has recently been proposed, via computational and experimental means, that differential interaction kinetics between KRas G13D and the Ras GAP NF-1 explain this effect.21–23 Another example is that the *KRAS* G12D allele is associated with reduced overall survival in advanced PAAD when separately compared to patients with WT *KRAS*, *KRAS* G12R, or *KRAS* G12V.24 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 likely essential to effectively treating *KRAS*-driven cancers. The prominent reductionist strategy of treating all *KRAS* mutants the same has so far proven insufficient. Thus, the current study describes genetic interactions found in tumors with the common *KRAS* alleles in COAD, LUAD, MM, and PAAD. The origins of *KRAS* mutations were studied to assess the extent to which latent mutational processes determined the allelic distribution. Further, comutation networks were constructed for each *KRAS* allele and interrogated to identify properties of the alleles. Finally, allele-specific genetic dependencies were analyzed 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 (1,280 after removing hypermutated samples), 891 LUAD, 1,201 MM, and 1,395 PAAD samples. In addition, there were targeted-sequencing data available for 3,329 COAD (2,865 after removing hypermutated samples), 4,160 LUAD, 61 MM, and 919 PAAD samples. More information on the sequencing data is available in the Methods and Supplementary Tables 1 and 2.

In the data collected for this study, *KRAS* was most frequently mutated at four “hotspots” (Supplementary Table 3). Of these hotspots, codon 12 mutations accounted for 76.8% of all mutations followed by codon 13 (11.4%), 61 (8.1%) and 146 (3.7%), when adjusted for the incidence of the cancer (Fig. [1](#fig:mutational-signatures-main)a). PAAD had the greatest frequency of *KRAS* mutations at 86.3%, followed by COAD (41.4%), LUAD (35.3%), and MM (21.9%) (Fig. [1](#fig:mutational-signatures-main)b). Further, there was substantial variability of the alleles found at these hotspots across the *KRAS*-driven cancers (Fig. [1](#fig:mutational-signatures-main)a). Notably, the most variation in *KRAS* alleles was found in MM, and it was the only cancer where a non-G12 allele was the most frequent. COAD had a unique enrichment of G13D and A146T alleles, while PAAD was distinct in its high frequency of G12R mutations.

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

Most *KRAS* mutations were caused by single nucleotide variants. An exception was the exceedingly rare, though transformative *in vitro*,5 G12F allele (0.4% of all *KRAS* mutations) caused by the dinucleotide substitution c.34\_35GG>TT. Glycine 12 and 13 can be transformed to six different amino acids (A, C, D, R, S, and V) through single nucleotide changes in the first two guanine residues. Glutamine 61 can be mutated to six other amino acids (E, H, K, L, P, and R) and a stop codon via a single nucleotide mutation. Alanine 146 can become one of six other amino acids (E, G, P, S, T, and V) from mutations to a single nucleotide.

One explanation for the distinct allelic frequencies across cancer types is that tissue-specific mutational processes drive the selection for mutations. To explore this hypothesis, the active mutational processes in the tumor samples were elucidated using mutational signatures25 (Supplementary Tables 4 and 5). 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 using non-negative matrix factorization and measured in each sample using non-negative least squares regression (see methods for the complete details; Supplementary Fig. [1](#sfig:mutational-signatures-supp)). The distribution of the levels of each mutational signature were generally uniform within a cancer, regardless of the *KRAS* allele of the tumor (Fig. [5](#sfig:mutational-signatures-supp)c). An exception was for microsatellite instable (MSI) tumors, where the signatures associated with that characteristic dominated (these samples were only included in Supplementary Fig. [1](#sfig:mutational-signatures-supp)a and b, but were removed for the rest of the study). In LUAD, there tended to be reduced levels of signature 4, caused by tobacco smoke,26 in tumors with *KRAS* G12D mutations compared to the others. For each cancer with a *KRAS* mutation, the probability that the allele was caused by each detectable mutational process was calculated. The average of these probabilities for each allele are shown in Fig. [1](#fig:mutational-signatures-main)d. Most of the common *KRAS* mutations in COAD, MM, and PAAD, were likely caused by mutations in “clock-like” signatures 1 and 5, mutations believed to accumulate with age.27 LUAD was the only cancer with *KRAS*-mutant samples enriched for a mutational signature of exogenous cause: the *KRAS* G12A/C/V mutations were primarily attributable to mutations caused by tobacco smoke (signature 4) while *KRAS* G12D mutations were most likely attributable to clock-like mutations (Fig. [1](#fig:mutational-signatures-main)d). Signature 8, of unknown etiology, had a substantial probability of causing some of the *KRAS* alleles across all four cancers.

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

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

Whether mutational signatures represent the mechanism driving *KRAS* allelic diversity between tissues was further analyzed by calculating the predicted frequency of each allele based on the frequency of mutations in the same trinucleotide context throughout the genome (Fig. [1](#fig:mutational-signatures-main)e; Supplementary Table 6). The null hypothesis tested was that, assuming the cancer would acquire a *KRAS* mutation, any of the common alleles (found in greater than 3% of the tumor samples for a given cancer) were sufficient. Thus, the frequency of the *KRAS* alleles would be determined by the mutational processes, alone. The predicted frequencies were compared against the observed allele frequencies. The alleles above the diagonal line were predicted to be more frequent than observed, while those below the line were more frequently observed than predicted by the mutational signatures. In COAD, G13D was predicted to be the most frequent allele, and G12D/V mutations were considerably underestimated (Chi-squared test, p < 0.05). Inversely, the frequencies of G12S and A146T mutations were significantly overestimated in COAD (Chi-squared test, p < 0.05, circles). In LUAD, the frequencies of the G12A/D/V alleles were predicted quite accurately, 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 mutational signature 4 caused by tobacco smoke (Fig. [1](#fig:mutational-signatures-main)c, d), though this calculation suggests there is additional biological pressure promoting this mutation in LUAD. The frequencies of the *KRAS* alleles were best predicted by mutational signatures 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. A linear model was fit for the observed and predicted allele frequencies for each cancer, and the coefficient of determination (R value) was reported as an indicator of goodness-of-fit. These values ranged from around 0.25 to 0.56 and none of the linear models were statistically significant (t-test, p-value < 0.05) suggesting that the observed and predicted allele frequencies were quite dissimilar. 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. This observation suggests that the particular biologic properties of the alleles drive their selection, warranting further investigation into their genetic interactions.

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

We reasoned that if biological selection is driving *KRAS* allele selection in cancer, then distinct functions of each mutant form of KRas 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 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.32–35 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.36,37

To this end, 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 the Row-Column Test from Leiserson *et al.*38 to identify reduced rates of comutation (Supplementary Table 7). The result of the comutation analysis on COAD tumors was a weakly connected network of the *KRAS* alleles with only a few genes linking the alleles together (Fig. [2](#fig:comutation-main)a). 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. [2](#fig:comutation-main)b).

Consistent with the idea that each allele is functionally distinct, there were a substantial number of genes detected to comutate with just one *KRAS* allele. To gain functional insight into the network, genes known to physically interact with KRas,14 signal up- or downstream of KRas,39,40 or are known oncogenes41,42 were extracted (Fig. [2](#fig:comutation-main)b). Several alleles had reduced comutation with *NRAS* and *BRAF* and increased comutation with *APC* and *PIK3CA*, interactions that have been previously documented.32,34,35,43–50 Some novel interactions included increased comutation of *PORCN* with *KRAS* A146T, *MTOR* with G12C, and *SMAD4* with G12V. Further, several of the alleles showed enrichment for cellular functions in their comutation networks (Fig. [2](#fig:comutation-main)c). One of the strongest effects was an enrichment in the G12D comutation network of interactors with *YWHAZ*, a 14-3-3 scaffolding protein implicated in modulating many interactions including the activity of Rho guanine nucleotide exchange factor 7 on RAC1 in phagocytosis and cell adhesion.51 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. [2](#fig:comutation-main)d and e. These include increased comutation with *AMER1*, a negative regulator of Wnt signaling52,53.

The *KRAS* allele-specific comutation network uncovered in LUAD was far larger than that of COAD (Supplementary Fig. [2](#sfig:luad-comutation-network)). 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 KRas 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. [2](#sfig:luad-comutation-network)). There were several intriguing cellular processes enriched in the LUAD networks for each allele (Fig. [2](#fig:comutation-main)c). For example, *KRAS* G12C had increased and reduced comutation interactions with many genes encoding proteins that interact with Myc ("PPI of MYC (TF)"), compared to how reduced comutation interactions drove the enrichment of focal adhesion genes with *KRAS* G12D (Fig. [2](#fig:comutation-main)).

Conducting this analysis in MM was hampered with the fact that this cancer is known to be frequently multi-clonal. 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.42,57 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. [3](#sfig:mm-comutation-heatmap)). 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. [4](#sfig:paad-comutation-network)). There were numerous genes that had opposing comutation interactions with different alleles. Four of these were direct interactors of *KRAS*,14 signal through *KRAS*,39,40 or are known oncogenes41,42 (Supplementary Fig. [4](#sfig:paad-comutation-network)). 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. [2](#fig:comutation-main)c) including the protein-protein interaction networks (PPIN) of SMAD1-3 and TGF- signaling. Inspecting the underlying genes further revealed that the genes that caused the enrichment varied for each protein (Fig. [2](#fig:comutation-main)f). For instance, the comutation events of *ACVR1B* with *KRAS* were primarily with Q61H, whereas those with *FLNA* were mostly with G12R. These subtle differences suggest that specific and nuanced alterations of SMAD signaling best complement a given *KRAS* allele in PAAD.

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

The perturbations necessary to drive cancer expose vulnerabilities not present in the 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.58,59 As the *KRAS* alleles have measurably different signaling behaviors and genetic interactions, they likely have specific genetic vulnerabilities. To this end, data from a genome-wide, CRISPR-Cas9 knock-out screen of cancer cell lines60,61 were used to identify genes with *KRAS* allele-specific genetic dependencies. For the *KRAS* alleles in which there were at least 3 cell lines, allele-specific enrichment 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 pairwise t-tests.

For COAD, there were only 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. [3](#fig:coad-dependency-main)a). For example, many genes comprising Complex I of the electron transport chain had a greater lethal effect when knocked out in cell lines with *KRAS* G12V mutations than in *KRAS* G12D, G13D, or WT cell lines (Fig. [3](#fig:coad-dependency-main)b). Alternatively, the *KRAS* G13D cell lines were less affected when genes involved in the complement immune pathway were targeted (Fig. [3](#fig:coad-dependency-main)c). 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 77 genes were hierarchically clustered into six groups by their dependency scores (Figure [3](#fig:coad-dependency-main)d; Supplementary Table 8). Genes in clusters 1 and 3 tended to have reduced genetic dependency scores in cell lines with *KRAS* G12D and G12V, respectively, compared to the other cell lines. Alternatively, clusters 4, 5, and 6 demonstrated increased dependency in *KRAS* G13D, G12V, and G12D cell lines, respectively, compared to the other cell lines. Genes in cluster 2 tended to have a stronger dependency in *KRAS* WT cell lines. One notable gene with allele-specific associations was the kinetochore-associated protein (*KNTC1*), a regulator of the mitotic checkpoint,62–64 which demonstrated moderate to strong lethal effects when knocked out in almost every cell line except for those with a *KRAS* G12V allele (Figure [3](#fig:coad-dependency-main)e). Also, isocitrate dehydrogenase (*IDH1*), a component of the citric acid cycle in cellular carbon metabolism,65 was found to have a negligible effect on growth reduction when knocked-out in *KRAS* G12D cell lines, an intermediate effect in G12V cell lines, and the strongest effect in G13D and WT cell lines (Figure [3](#fig:coad-dependency-main)e). Interestingly, a negative regulator of the MAPK pathway,66 *WDR26*, was found to be almost essential for the *KRAS* G12D cell lines, though resulted in more moderate growth reduction when knocked out in the other cell lines (Figure [3](#fig:coad-dependency-main)e). Increased expression of *WDR26* has previously been implicated in driving breast cancer by serving as a scaffolding protein in the PI3K-Akt pathway67.

The same analysis was conducted on the LUAD cell lines with WT *KRAS* or G12C or G12V mutations. GSEA of the dependency scores highlighted a reduced dependency on the genes in the p53 hypoxia pathway and an increased dependency for genes in the Bard1 pathway for G12C cell lines (Supplementary Fig. [5](#sfig:luad-dependency-gsea)). The latter enrichment suggests these cell lines are more susceptible to inhibitors of postreplication DNA-damage repair mechanisms because the enrichment was driven by many components of the Fanconi anemia (FA) pathway responsible for resolving DNA interstrand crosslinks.68 A gene-by-gene analysis revealed allele-specific genetic dependencies in 583 genes that were further hierarchically clustered into four groups (Supplementary Fig. [6](#sfig:luad-dependency-heatmap); Supplementary Table 9), suggesting there were large differences between the cell lines of these three *KRAS* alleles.

For the genetic dependency analysis of PAAD, the only *KRAS* alleles with a sufficient number of cell lines were G12D, G12R, and G12V. GSEA revealed substantial differences in the dependencies of critical cellular pathways (Supplementary Fig. [7](#sfig:paad-dependency-gsea)). For instance, the G12D cell lines tended to be more dependent on the pathway involving the G12/G13 subunits of G protein-coupled receptors, which regulate the actin cytoskeleton during proliferation.69–71 Moreover, the G12R cell lines demonstrated a greater dependency on the genes at the G2 and M DNA-damage checkpoint. This enrichment was driven by two of the three components of the MRE11-RAD50-NBN (MRN) complex, *MRE11* and *NBN*. Interestingly, the cell lines with *KRAS* G12V mutations were more sensitive to knock-out of genes in the Hedgehog pathway. In these cell lines, 94 individual genes demonstrated *KRAS* allele-specific genetic dependency (Supplementary Fig. [8](#sfig:paad-dependency-heatmap); Supplementary Table 10). Notably, the deletion of *JUN* had no effect in *KRAS* G12V cell lines while it slowed growth in G12D and G12R cell lines. *JUN* encodes the transcription factor c-JUN, which was previously demonstrated to play a role in the transcriptional repression of the tumor suppressors *CDKN2A* and *CDKN2B* in *KRAS*-driven COAD.72 This is consistent with the reduced dependency of G12V cell lines on c-Jun NH2 terminal kinases (JNK) activation from the gene set analysis (of note, the enriched gene set did not include *JUN*, itself; Supplementary Fig. [7](#sfig:paad-dependency-gsea), Supplementary Fig. [9](#sfig:paad-dependency-JUN)). Leading this enrichment was *MAPK8* (JNK-1), which, when knocked out, led to increased growth in every case, though most strongly in G12V cell lines (Supplementary Fig. [9](#sfig:paad-dependency-JUN)). These data suggest a reduced dependency on the activation of c-Jun via JNK signaling, potentially pointing to a tumor suppression pathway with greater potency in PAAD expressing *KRAS* G12V.

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

Integrating the results from the allele-specific comutation analysis with those from the dependency analyses provided further insight into the distinctions between the *KRAS* alleles. Surprisingly, there was little overlap between the genes found to comutate with an allele and those with differential dependency - the only overlap was found within the genes resulting from analysis of *KRAS* G12C in LUAD (Fig. [4](#fig:results-integration-main)a). One of these genes was *STK11*, the gene encoding STK11 (also known as LKB1), a tumor suppressor that controls the activity of AMP-activated protein kinases (AMPK) to regulate cellular processes including metabolism, apoptosis, and the DNA-damage response.73,74 The high rate of comutation between *KRAS* and *STK11* has been documented previously, though not specifically with *KRAS* G12C. Previous studies have indicated unique biological properties of LUAD tumors with mutations in both *KRAS* and *STK11*, including distinct expression profiles,75 worse clinical prognosis,76,77 and reduced response to immunotherapy.78 The results presented here may suggest a unique synergism between the G12C mutant and *STK11* loss-of-function mutations. Importantly, due to the strong influence of smoking-induced mutations to the prevalence of *KRAS* G12C in LUAD, there was no apparent, nor statistically detectable (Fisher’s exact test, p > 0.05), difference in the types of *STK11* mutations between G12C-mutant samples and the rest of the LUAD tumor samples (Supplementary Fig. [10](#fig:results-integration-main)). Thus, it is unlikely that the genetic associations found here were driven by latent mutational processes, but instead they were determined by the selective advantage that concomitant *KRAS* G12C and *STK11* mutations have in a nascent tumor.

To further mine the genetic interactions found for each *KRAS* allele, functional information was incorporated by annotating the nodes of the PPIN with the associations. For each allele in each cancer, the geodesic distance (the number of nodes on the shortest path) between the genes with allele-specific interactions was on average shorter than between nodes selected randomly on the network (Fig. [4](#fig:results-integration-main)b). This observation indicated that, instead of being randomly distributed throughout the PPIN, there were functional patterns in the identified genes.

To inspect these cellular functions, subnetworks of the PPIN of the proteins with allele-specific interactions were extracted and compared. For COAD, the largest components of each allele’s subnetwork were centered around KRas and the other oncogenes most frequently mutated in the cancer: p53, BRAF, PIK3CA, APC, and NRAS (Fig. [4](#fig:results-integration-main)c highlighted in red). Surrounding these nodes were others that had either comutation or dependency interactions with multiple *KRAS* alleles (in brown). Finally, there were communities of nodes with associations to only a single *KRAS* allele, some representing distinct biological functions. For instance, the G12D allele was again associated with cell adhesion and motility via the module consisting of Myosin-6 and 9 (MYH6/9) two components of lamanin (LAMA1/2), and neural cell adhesion molecule L1 (L1CAM) signaled to via 6 integrin (ITGA6), PI3K (PIK3CA), and PIP5K1- (PIP5KA).

For LUAD, the complexity of the networks for the G12C and G12V alleles necessitated inspecting them separately and limiting each to just their largest connected components (Supplementary Fig. [11](#sfig:luad-integrated-ppin)). This revealed clusters of functionally related proteins that have either a comutation or dependency interaction specifically with one *KRAS* allele. For instance, many proteins involved in Wnt signaling, such as -catenin, LRP-6, Notch 1, and Pygopus homolog 1 were found in the G12C PPI subnetwork. In the G12V PPI subnetwork was a well connected component of vasoconstriction regulators and another of proteins involved in ubiquitination, specifically linked to DSB repair proteins.

The subnetwork surrounding *KRAS* composed of the genes with interactions with *KRAS* alleles in PAAD (Supplementary Fig. [12](#sfig:paad-integrated-ppin)) demonstrated a similar pattern as seen in the COAD. The subnetwork was centered around the well-known drivers of PAAD (red nodes) and there was considerable overlap between the alleles (brown nodes). Beyond that, there were clusters of nodes with interactions with only one *KRAS* allele. Some of these were functional clusters such as a collection of extra-cellular matrix proteins with genetic interactions with the G12R allele. Overall, these subnetworks display cellular processes that have strong allele-specific associations composed of a combination of comutation and dependency interactions.

# Discussion

This study addressing the genetic complexity of cancer through a 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. To investigate these properties, we conducted statistical tests to determine patterns of comutating genes and genetic dependencies for each *KRAS* allele in each cancer. The former identified genes that comutated with specific *KRAS* alleles at an unexpectedly high frequency, suggesting they were alterations that cooperated with the *KRAS* allele to promote positive selection in the tumor. 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 that the various oncogenic *KRAS* mutations are not biologically redundant, but instead have distinct properties relevant to the treatment of *KRAS* mutant tumors.

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, KRas function, and cooperating genetic events.. While the intrinsic biochemical biochemical properties of a KRas mutant are maintained in each cancer, their downstream signaling properties, and ultimately their effects on tumorigenesis, are determined by basal configuration of the tissue-specific signaling network.

More broadly, this study has broad implications for the understanding of oncogene biology and for cancer therapy. Whether a targeted therapy directly inhibits the activated oncoprotein 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,79,80 which consequently determines their response to different inhibitors.81,82 Moreover, the response of HER2 mutant cancers to HER2 inhibition varies depending on the tissue-of-origin of the cancer (PMID 29420467), 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,84,85 which included relevant projects from The Cancer Genome Atlas (TCGA)50,86,87 and other smaller studies. Additional data were acquired from the International Cancer Genome Consortium (ICGC) for pancreatic cancer88 and colorectal cancer. MM WES data were gathered from the Multiple Myeloma Research Foundation (MMRF)-CoMMpass online repository.89 Panel data for multiple cancers were retrieved from AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE v5).90 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 Table.

## 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)91 and The Human Protein Atlas (HPA, 12/03/2018),92,93 and tumor expression data was collected from MMRF-CoMMpass (01/14/2019), TCGA-COAD, TCGA-LUAD, and TCGA-PAAD.50,86,87,89 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 distribution across COAD, LUAD, MM, and PAAD of mutations over the four hotspots on *KRAS* was calculated by accounting for the 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:94 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.94,95 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.

## Predicting effect of mutations on gene or protein function

The effect of a mutation on the function of a gene or encoded protein was predicted using SIFT,96,97 PolyPhen2,98 LRT,99 MutationTaster,100 MutationAssessor,101,102 FATHMM,103 MetaSVM,104 and MetaLR.104 Known clinically significant mutations were acquired from ClinVar.105 Annotations were applied using ANNOVAR.106 A mutation was marked as deleterious if it was predicted as such by at least one of the prediction methods or declared as such by ClinVar.

## Protein-Protein Interaction Network (PPIN)

The PPIN used throughout the study was the combination of interactions from STRING,107,108 HINT,109 and BioPlex.110

## Identifying mutational signatures

The genome-wide mutations of a sample can be deconvolved into mutational signatures that represent endogenous or exogenous mutagenic processes.25 Single nucleotide variants (SNVs) from exomes or genomes were divided into 96 types, according to the 6 mutations of a pyrimidine (C>A, C>G, C>T and T>A, T>C, T>G) and the 16 possible combinations of 3’ and 5’ adjacent bases. The MATLAB111 implementation of NMF algorithm, SigProfiler,25 was used to discover the underlying mutational patterns that are common across tumors. Mutational signatures were discovered separately for each tumor type and the optimal number of signatures was determined based on silhouette width and Frobenius error.112

The spectrum of the signatures discovered by NMF were matched to the COSMIC catalog.113 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.114,115 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.115 Three versions of the signature associated to POLE mutations, signature 10, were discovered (previously reported in115). 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-guanine115 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 SBSA116 and signatures 34 and 41 in reference.115 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 artifacts117 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.118 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 of a tumor sample acquiring the *KRAS* mutation from signature from all signatures can be calculated using Eq. [[eq:kras\_mutation\_from\_signature]](#eq:kras_mutation_from_signature).

## Predicting *KRAS* allele frequency by mutational signatures

The mutational signatures are linear combinations of the 96-dimension spectrum of possible mutations (see "Identifying mutational signatures" above). Thus, assuming the null hypothesis that the prevalence of active mutational processes alone determines the frequency of *KRAS* alleles in a cancer 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 leat 3% of the tumor samples. The 95% confidence intervals were bootstrapped.119 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 comutation partners with at least three comutation events were considered. Further, only genes with a mutation frequency of at least 1% or a comutation frequency with a *KRAS* allele of at least 10% were considered.

The Row-Column Test for Exclusivity (RC-test) was used to identify reduced frequency of comutation between *KRAS* alleles and other mutated genes.38 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. Only genes with a mutational frequency of at least 2% and at least 10 mutually exclusive events were considered.

## Functional enrichment

The R interface to the online *Enrichr* tool was used to identify enriched gene sets in the comutation networks and allele-specific synthetic lethal clusters.120–122 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 only exception to this was the removal of the LUAD cell lines with *KRAS* G13D mutations because this allele is exceedingly rare in LUAD. This is supported by the fact that knocking out *KRAS* in these cell lines had an equivalent effect than when the gene was knocked out in *KRAS* WT cell lines: the average ( std. dev.) dependency score for G13D LUAD cell lines was -0.55 0.26, compared to that of *KRAS* WT cell lines: -0.55 0.28. The rest of the *KRAS* mutant samples demonstrated a far greater dependency on *KRAS*: -1.26 0.33.

The genetic dependency score is often linked to the expression of the gene. Thus, if the RNA expression of the gene could explain the dependency score (linear model, p-value < 0.01), the gene was not tested for *KRAS* allele-specific genetic dependency. Of the remaining genes, an ANOVA was used to measure if the mean dependency scores for the cell lines grouped by *KRAS* allele were different (p-value < 0.01). These genes were declared as deferentially dependent by *KRAS* allele. For these genes, pairwise Student’s t-tests were used to compare the dependency scores of each group. These contrasts were used to decide with which *KRAS* allele(s) a gene shows differential dependency (Benjamini-Hochberg FDR adjusted p-value < 0.05).

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

The GSEA tool (version 3.0) was acquired from the online GSEA portal (https://www.gsea-msigdb.org/gsea/index.jsp). Gene sets were acquired through MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp; 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.

## 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. Python123 and R124 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. These data were generated as part of the Multiple Myeloma Research Foundation Personalized Medicine Initiative.

# Author contributions

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

# Competing interests

The authors declare that they have no competing interests.

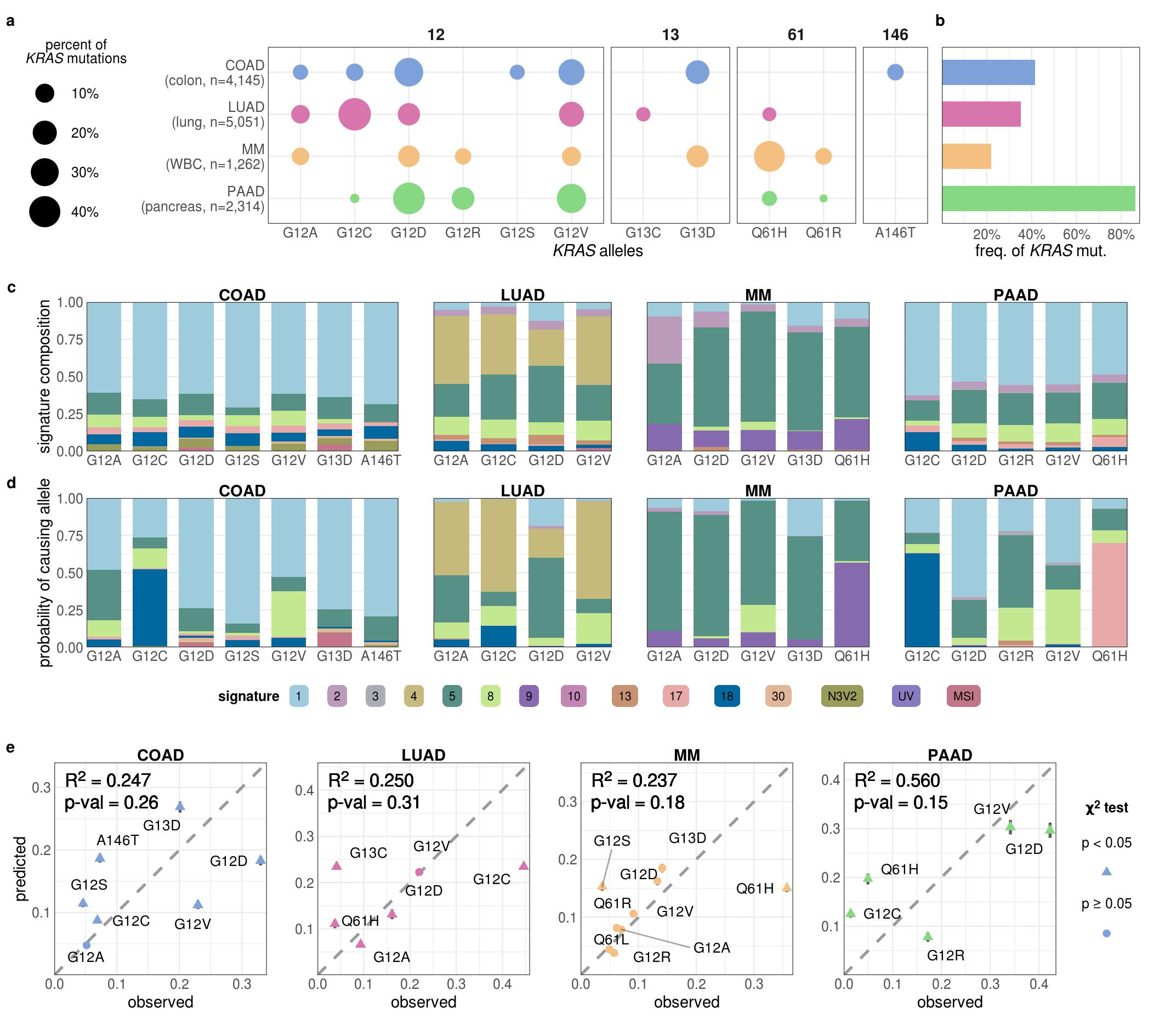
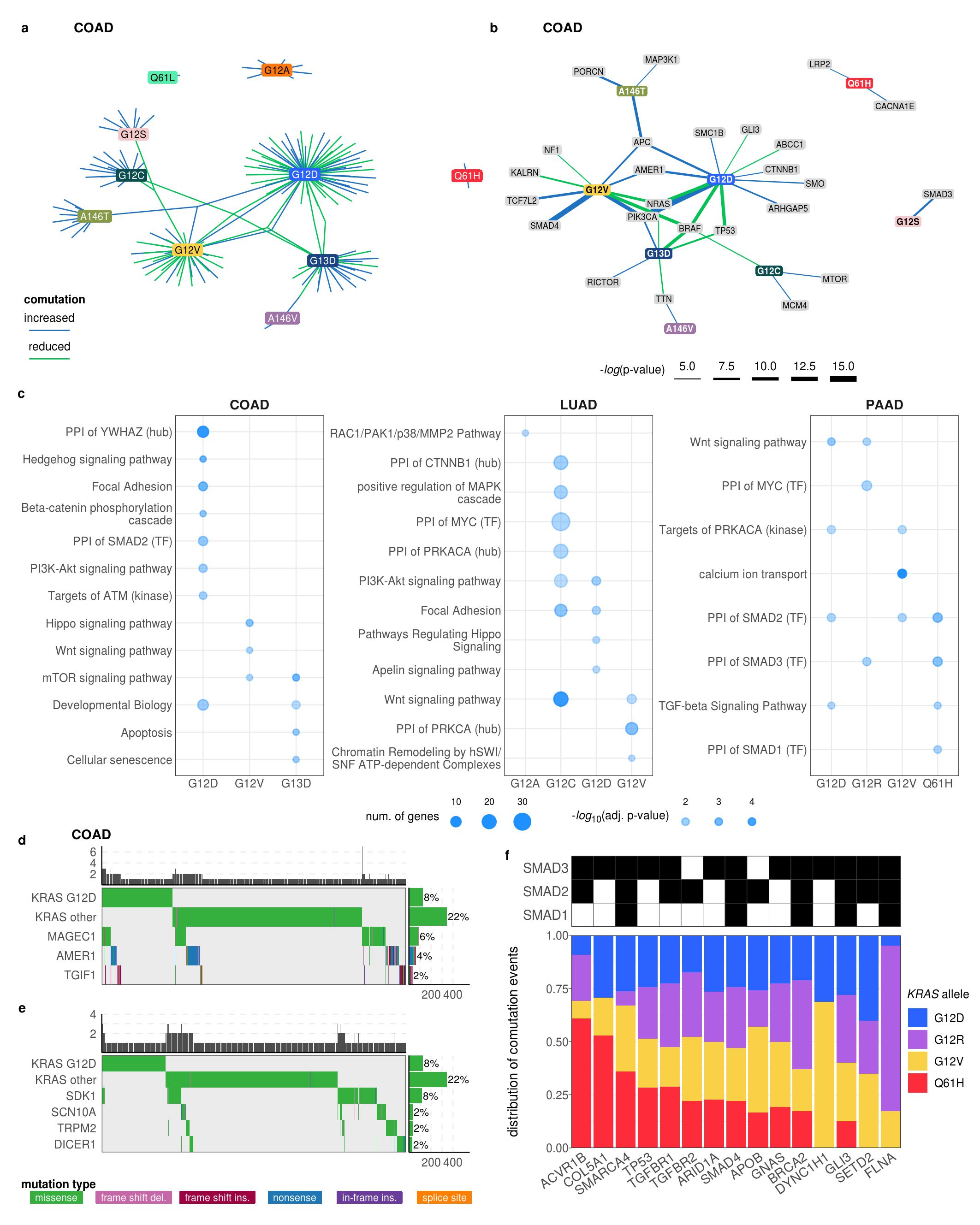


Figure 1. **The contribution of mutational processes to KRAS mutagenesis.** **a.** The distribution of KRAS allele frequencies at the four hotspots, codons 12 (left), 13 (middle-left), 61 (middle-right), and 146 (right) in each cancer. The size of the circle reflects the percent of KRAS mutations that 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. **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. **e.** The predicted vs. observed frequency of KRAS alleles for all 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 0.05). Error bars indicate 95% confidence intervals of the predicted values.



**Figure 2. 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. **b.** A subset of the network shown in panel a of genes encode proteins 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.** 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 rates 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.

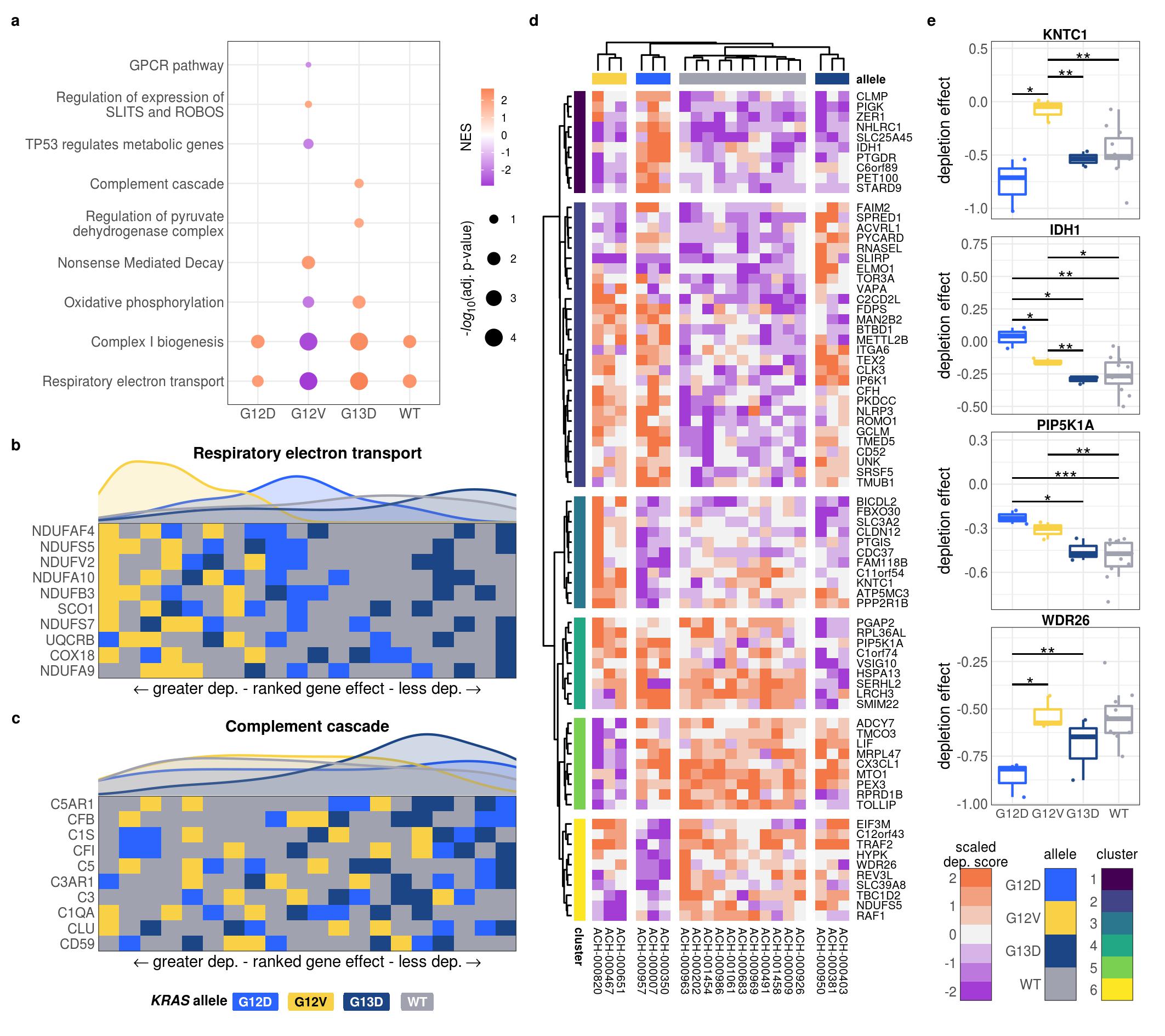
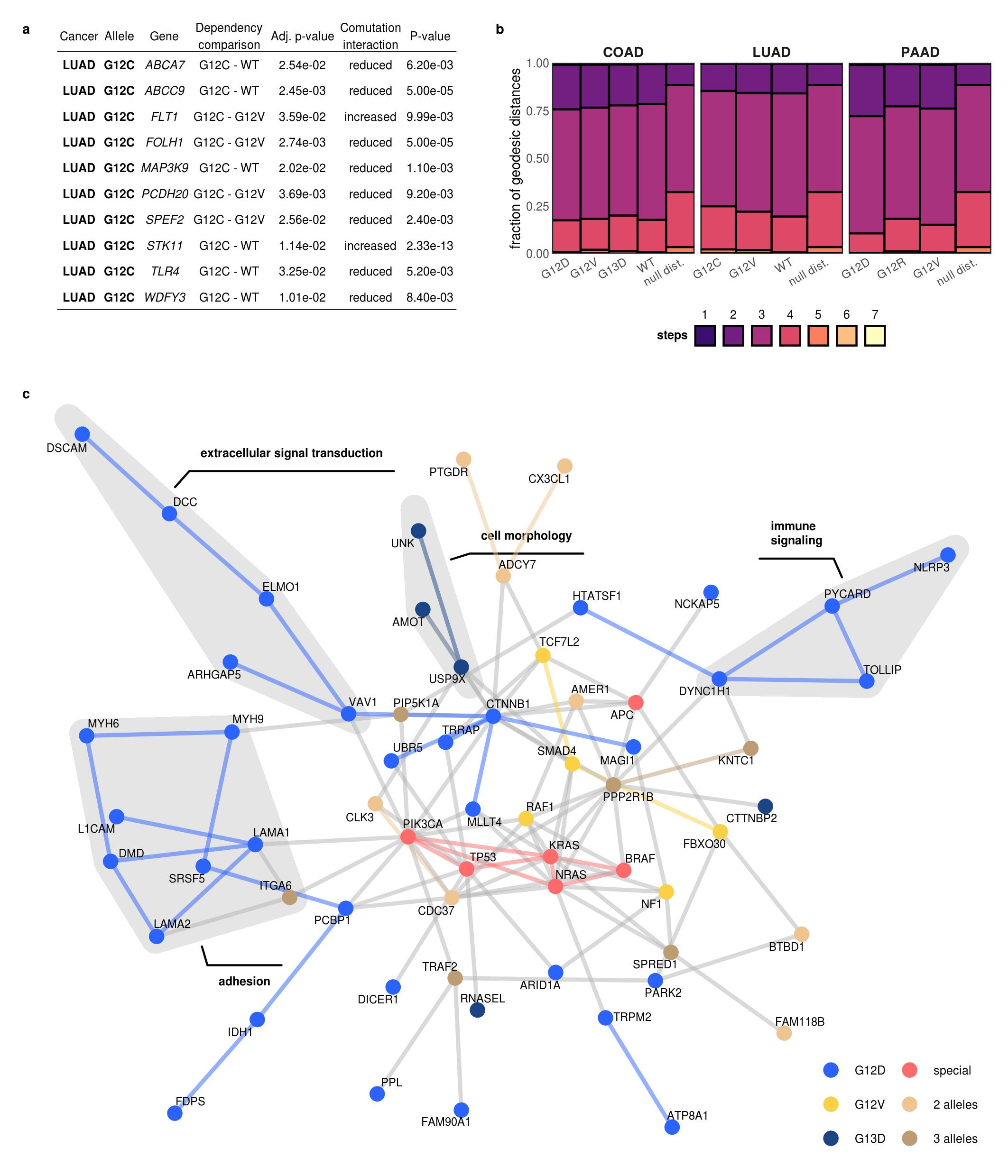
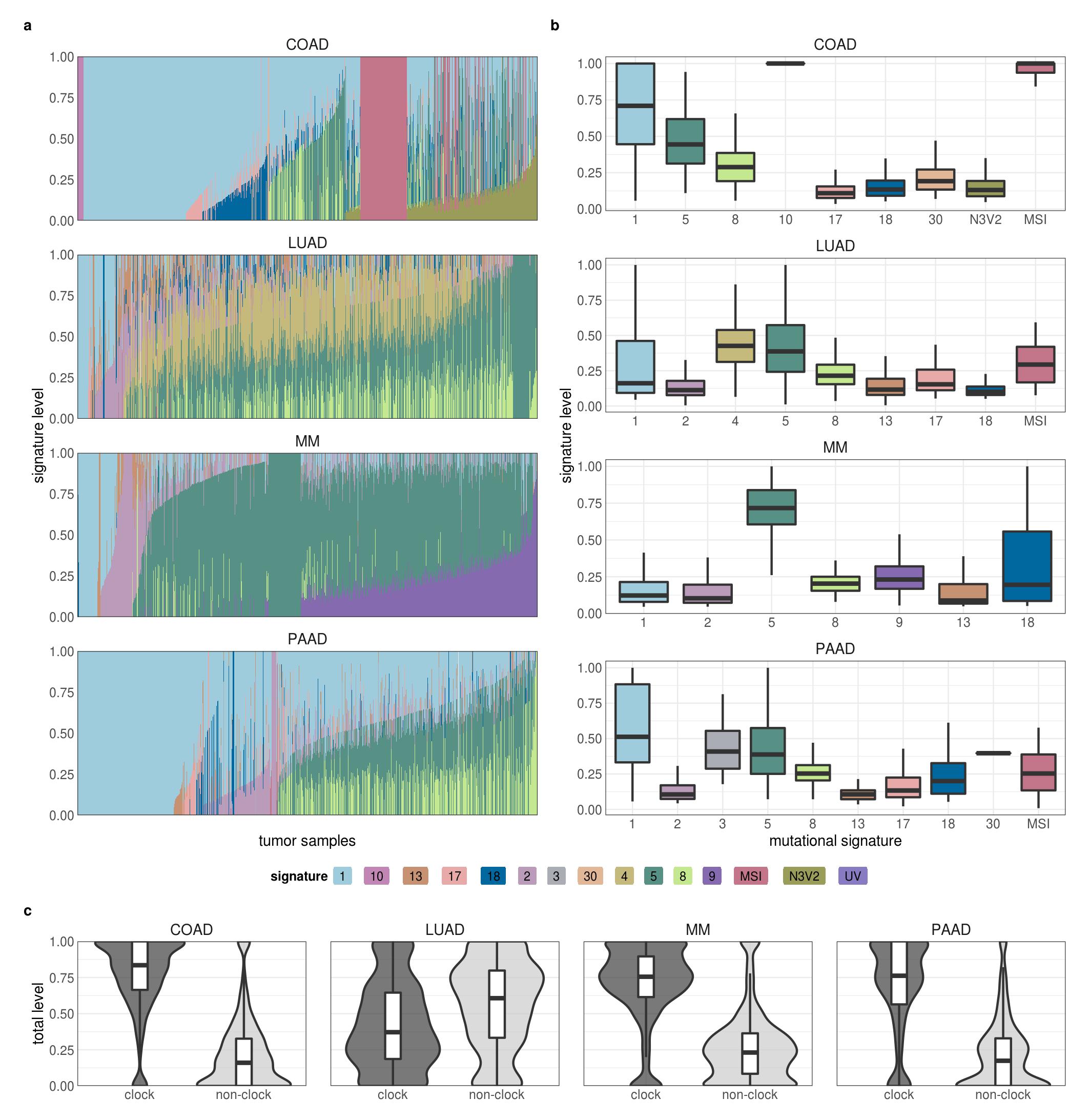


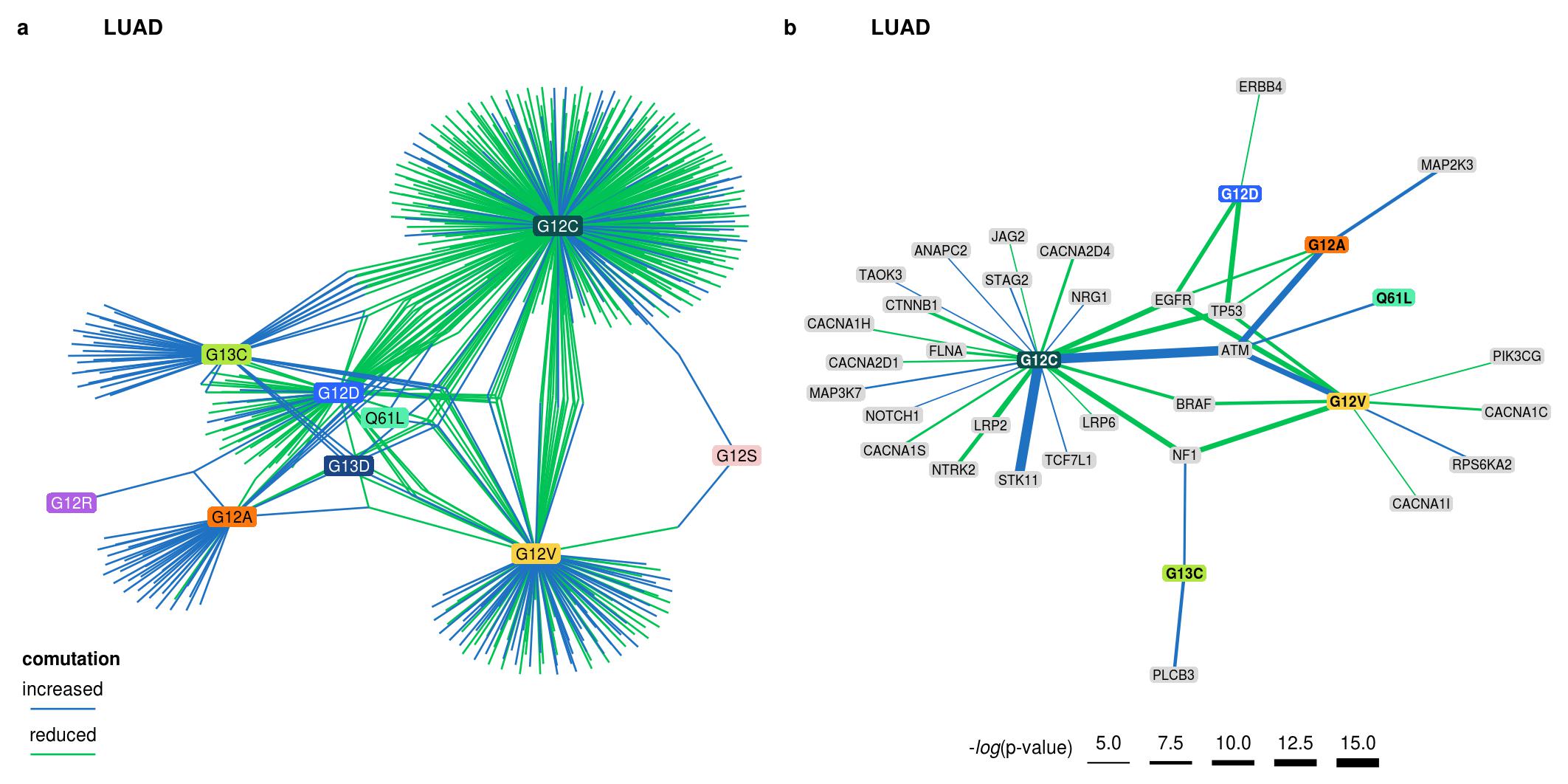
Figure 3. **Allele-specific genetic dependencies in COAD cell lines.** **a.** Gene sets with significant enrichment for increased (lower dependency score; purple) or reduced (higher dependency score; orange) genetic dependency in COAD cell lines. The size of the dot relates the p-value of the association 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 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. **d.** Clustered heatmaps of the genes that demonstrated differential genetic dependency amongst cell lines of different KRAS alleles in COAD cell lines. Each column is a cell line labeled by its DepMap ID and each row is a gene. **e.** Examples of genes that demonstrated differential genetic dependency amongst cell lines of different KRAS alleles in COAD (\*: p<0.05, \*\*: p<0.01, \*\*\*: p<0.001; p-values were adjusted using the Benjamini-Hochberg FDR correction method).



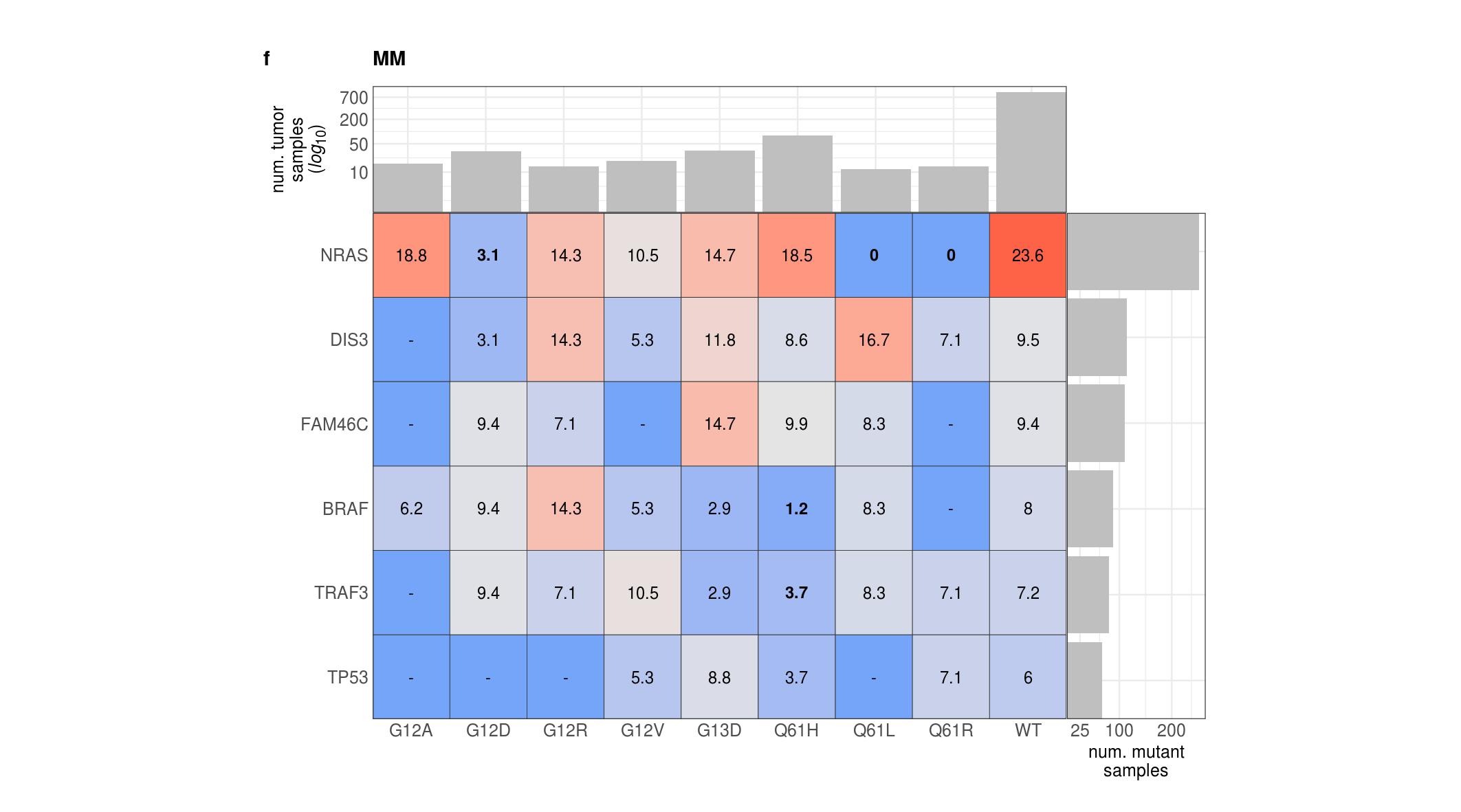
**Figure 4. Integration of comutation and genetic dependency results.** **a.** A table of the genes found to comutate and demonstrate differential dependency with the same *KRAS* allele and cancer. The "Dependency comparison" indicates the comparison of dependency scores found to be statistically significant; the "Adj. p-value" is the Benjamini-Hochberg-adjusted p-value for the comparison. The "Comutation interaction" indicates if the comutation interaction was a reduced or increased rate of comutation; the "P-value" is the p-value for this interaction. **b.** The connectivity of the proteins encoded by the genes found to comutate or demonstrate differential dependence with each *KRAS* allele on the PPIN. The right-most column represents the null distribution of the connectivity of randomly selected pairs of nodes. The bars indicate the distribution of the geodesic distances (shortest path length) between nodes in the PPIN. **c.** The largest connected subnetworks of the PPIN of genes found to comutate or demonstrate differential dependence with a *KRAS* allele in COAD was extracted and merged. Nodes present in multiple subnetworks are colored shades of brown, save for prominent oncogenes and tumor suppressors in COAD, shown in red ("special"). The rest of the nodes are colored by which allele they are associated with. The edges share the color of the nodes they connect if the nodes are the same color, otherwise they are grey. Curated groups of proteins with related cellular functions are highlighted and labeled.



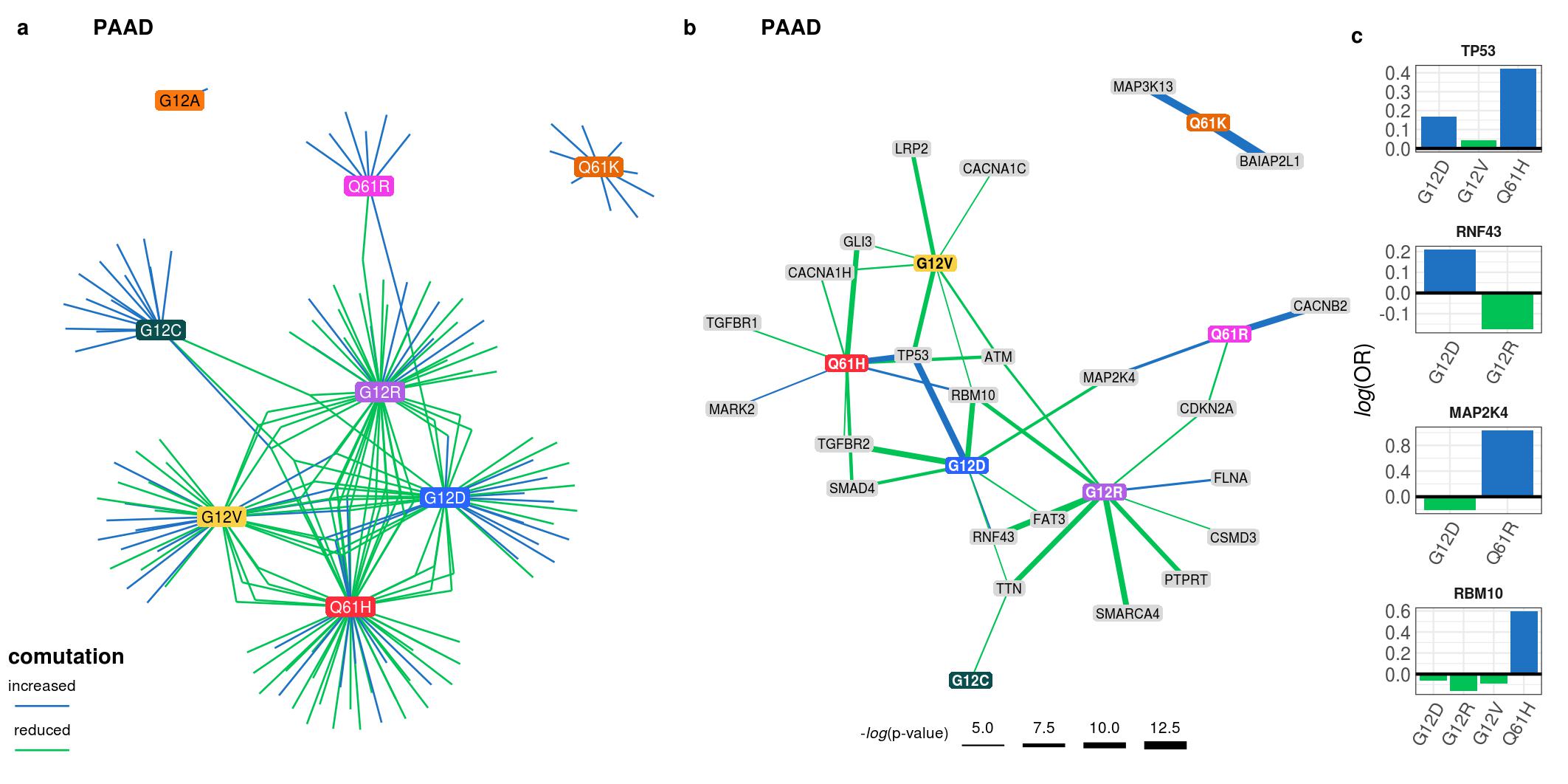
Supplementary Figure 1. **Mutational signatures in tumor samples.** **a, b.** The detected level of the mutational signatures in each tumor sample. In **a**, each tumor sample is a column. **c.** The average levels of clock (signatures 1 and 5) and non-clock (all other signatures) in the tumor samples.



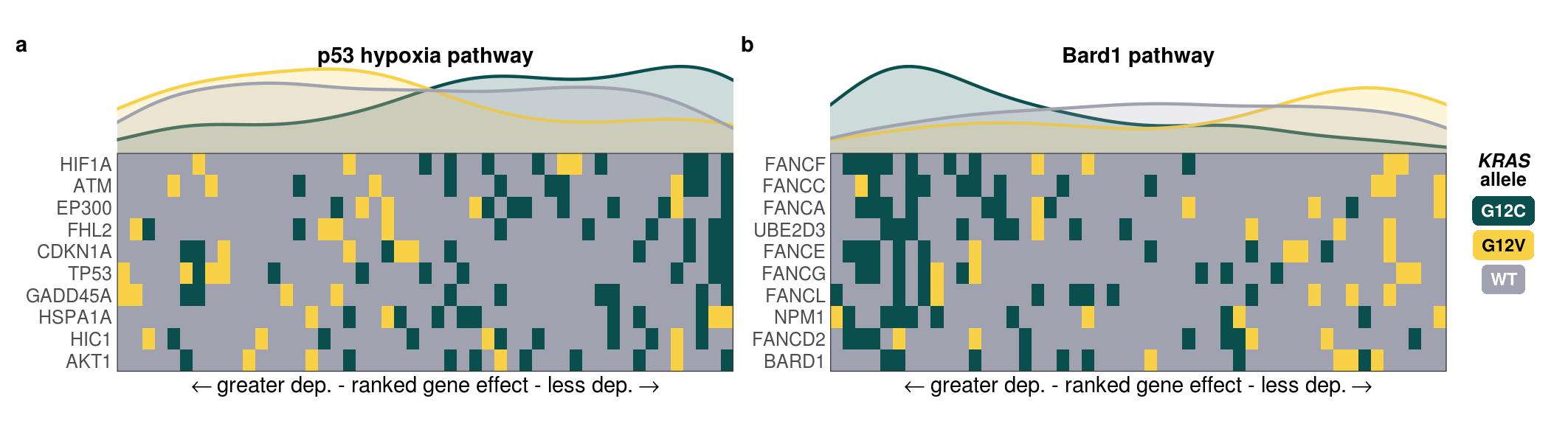
Supplementary Figure 2. **The comutation networks of KRAS alleles in LUAD.** **a.** The comutation network of the KRAS alleles in LUAD where each edge represents a comutation interaction between an allele and another gene. The color of the edge indicates whether the interaction was an increase (blue) or decrease (green) in the frequency of comutation. **b.** A subset of the network shown in **a** of genes in one of its canonical up- or downstream signaling pathways. The width of the edge indicates the strength of the association.



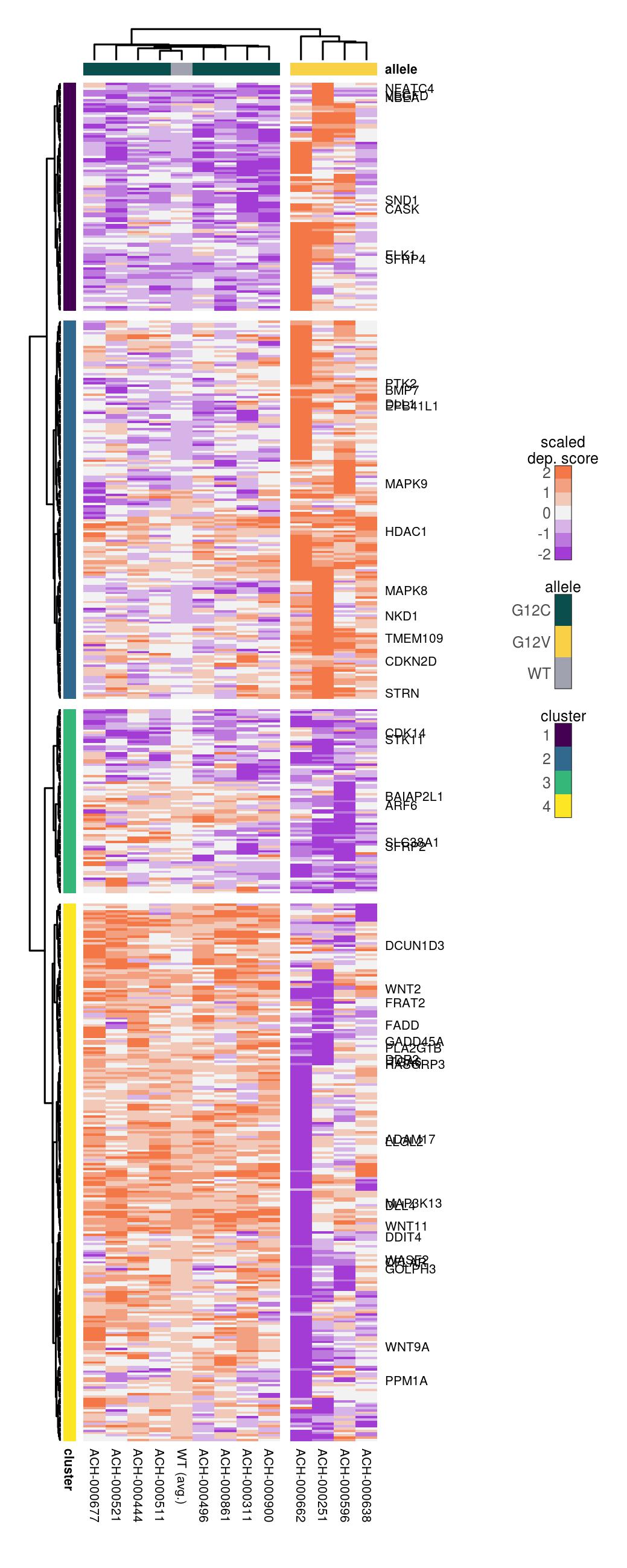
Supplementary Figure 3. **The comutation frequencies between known MM driving genes and KRAS alleles.** The color is correlated with the comutation frequency (the fraction of cells with the KRAS allele that also have a mutation in the other driver gene), indicated in each cell. Bold percent values indicate 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.



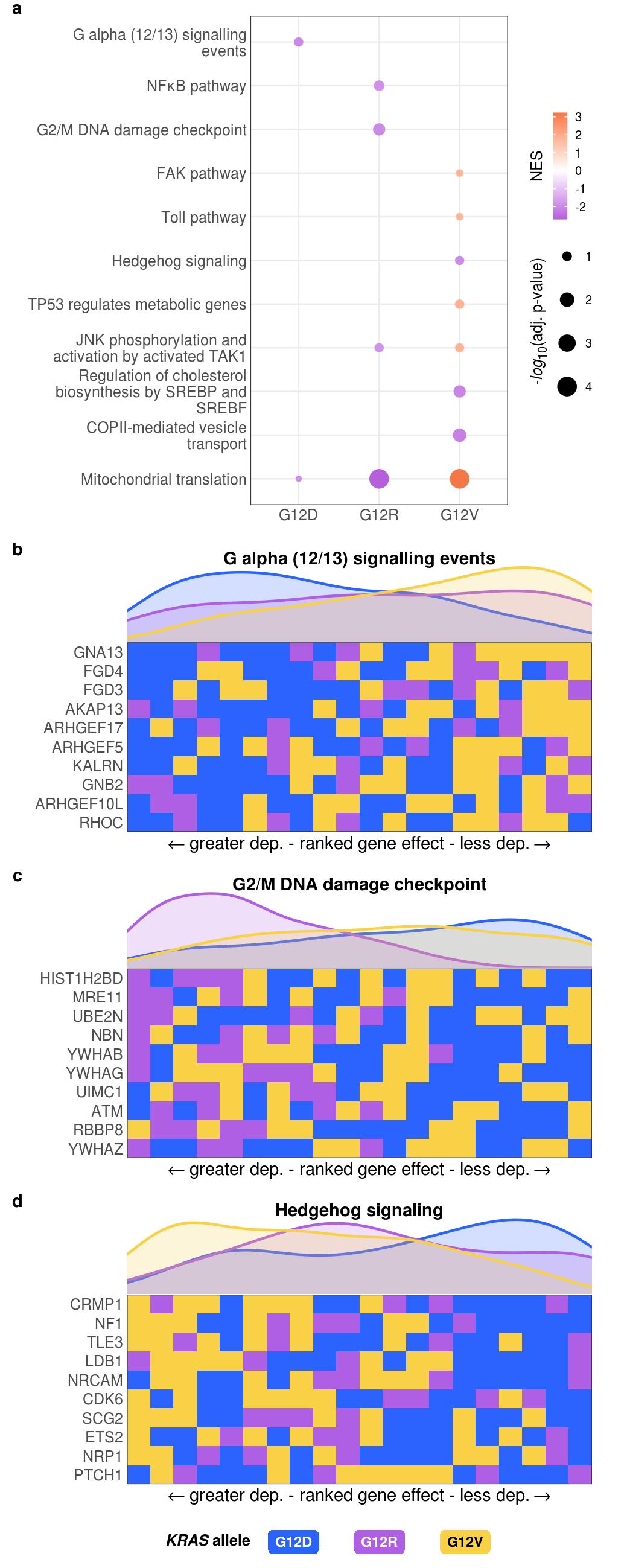
Supplementary Figure 4. **The comutation networks of KRAS alleles in PAAD.** **a.** The comutation network of the KRAS alleles in PAAD where each edge represents a comutation interaction between an allele and another gene. The color of the edge indicates whether the interaction was an increase (blue) or decrease (green) in the frequency of comutation. **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.



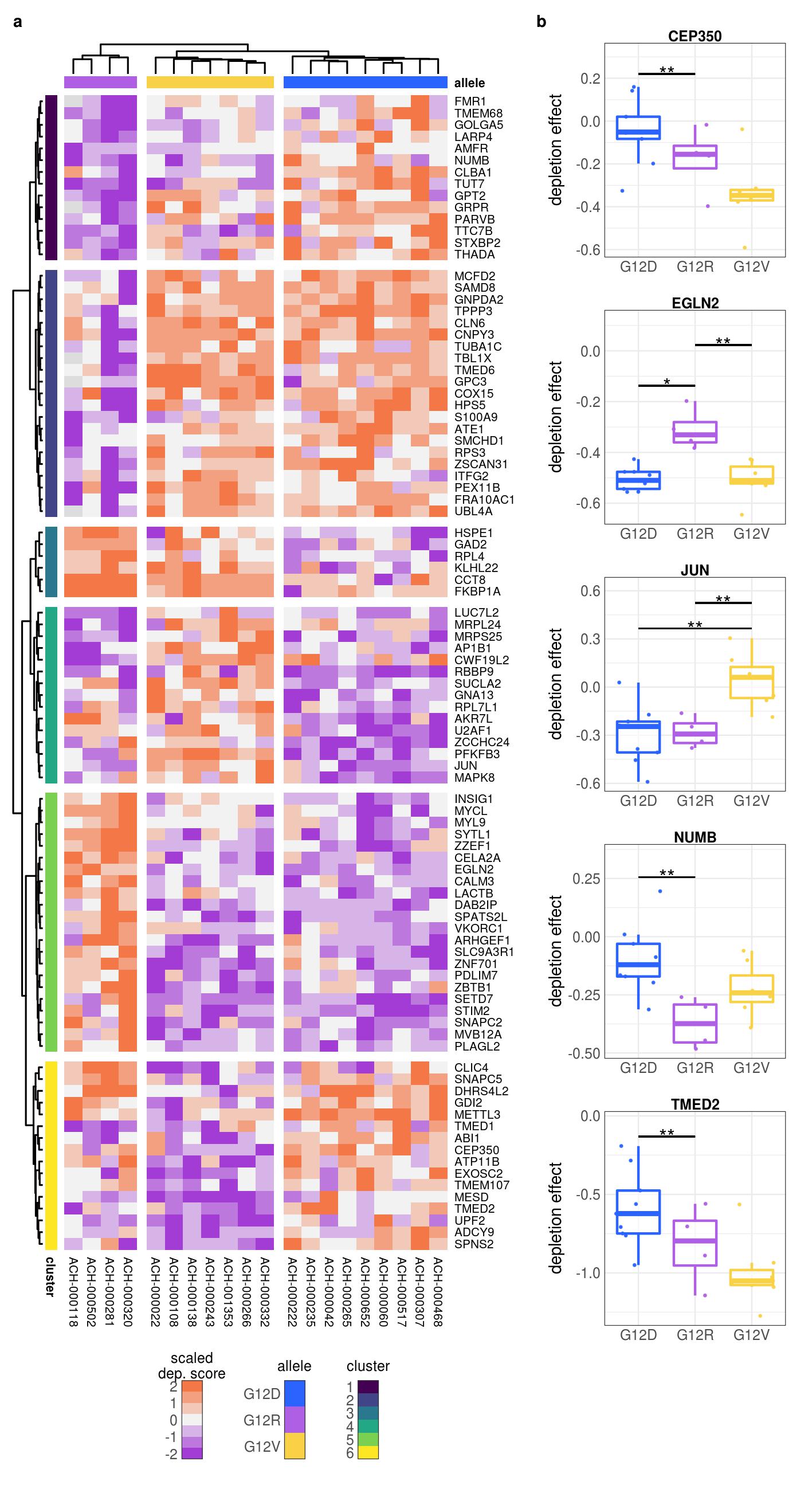
Supplementary Figure 5. **Gene sets with enriched genetic dependency in LUAD cell lines** **a, b.** Heatmaps ranking the cell lines by dependency score of the genes at the leading edge of enrichment for two gene sets in LUAD cell lines. Each row represents a gene and each cell represents a cell line 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.



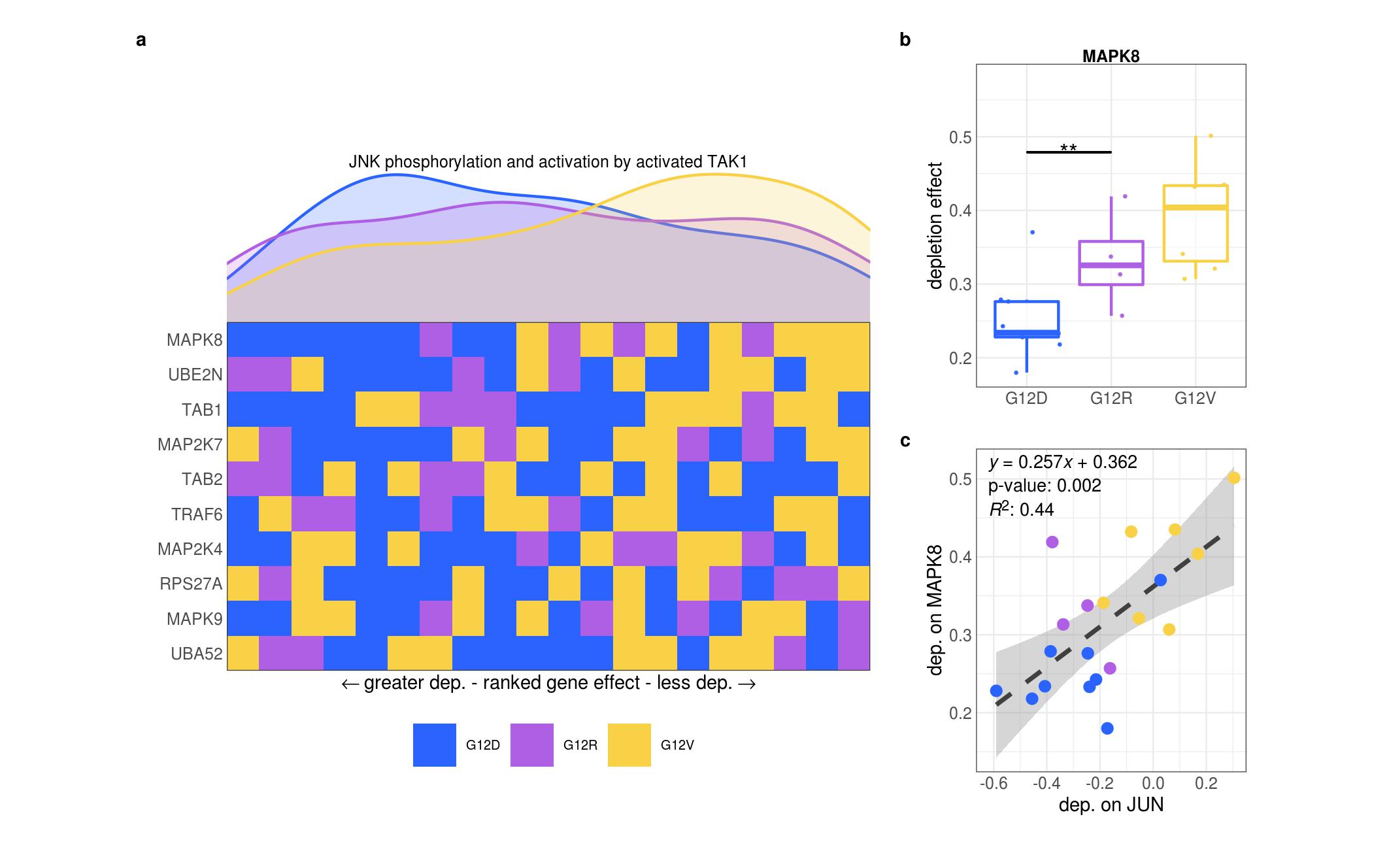
*Supplementary Figure 6.* ***Allele-specific genetic dependencies in LUAD cell lines.*** *Clustered heatmap of the genes that demonstrated differential genetic dependency amongst LUAD cell lines of different KRAS alleles. Each column is a cell line labeled by its DepMap ID and each row is a gene. Only genes known to be involved in KRAS signaling or previously implicated in driving cancer are labeled. Also, for visualization purposes, the scores of the WT cell lines were averaged into a single, representative group, labeled "WT (avg.)."*



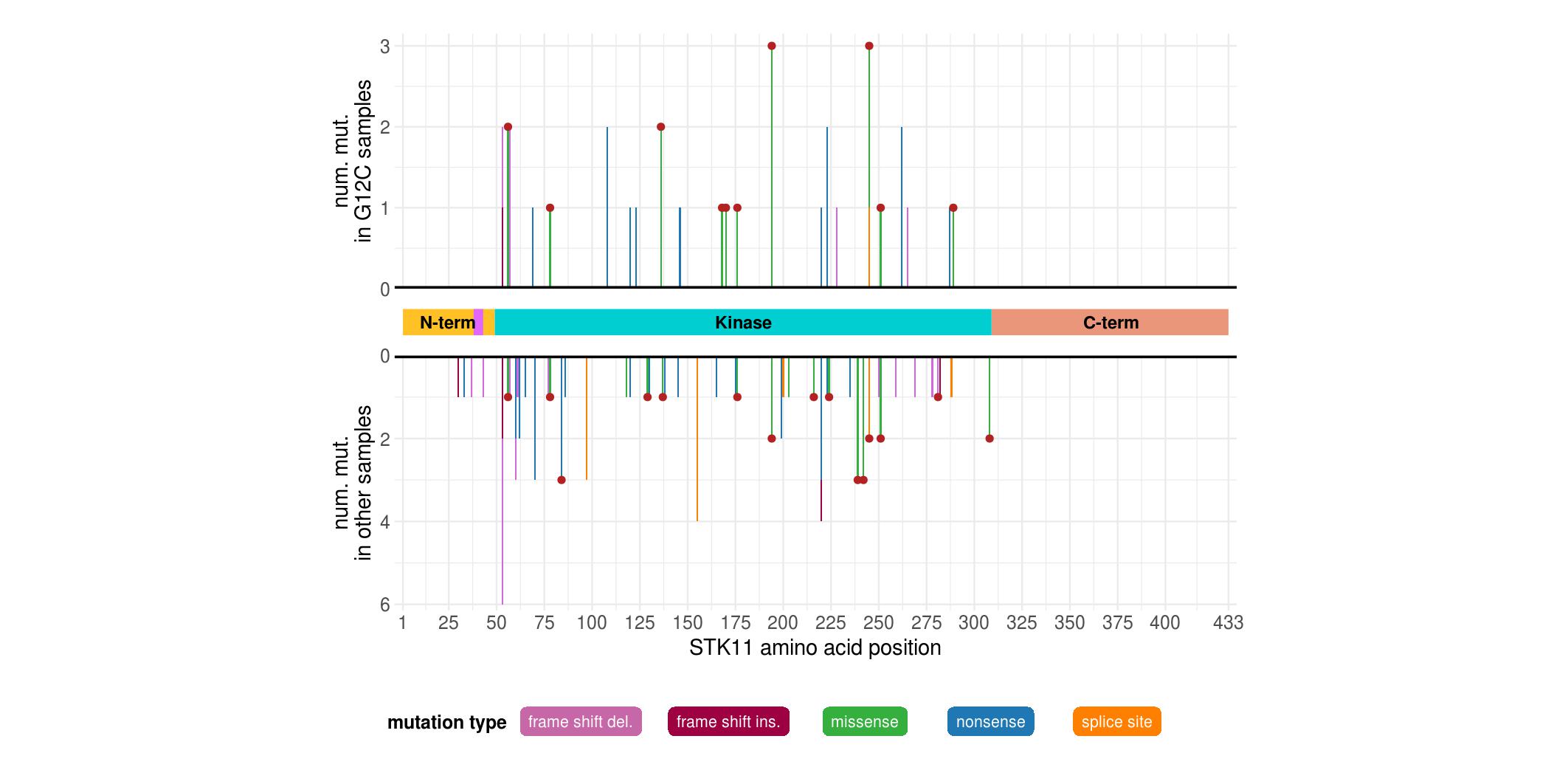
*Supplementary Figure 7.* ***Cellular processes enriched for greater or lesser genetic dependencies in PAAD cell lines separated by KRAS allele.******a.*** *Gene sets with significant enrichment for increased (lower dependency score; purple) or reduced (higher dependency score; orange) genetic dependency in PAAD cell lines. The size of the dot relates the p-value of the association and the color indicated 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.*



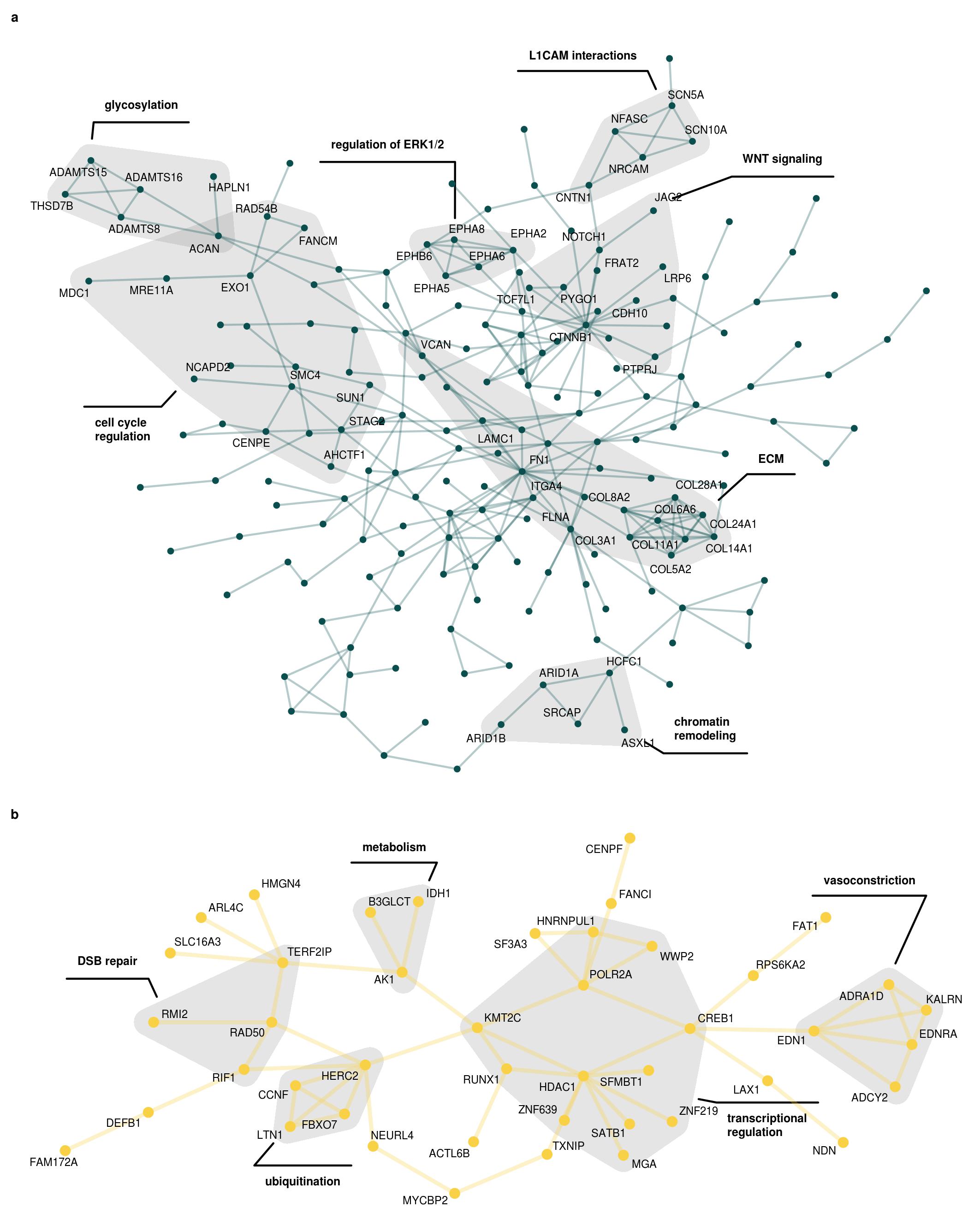
Supplementary Figure 8. **Individual genes with differential genetic dependency by KRAS allele in PAAD cell lines.** **a.** Clustered heatmaps of the genes that demonstrated differential genetic dependency amongst cell lines of different KRAS alleles in PAAD cell lines. Each column is a cell line labeled 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 in PAAD (\*: p < 0.05, \*\*: p < 0.01, \*\*\*: p < 0.001; p-values were adjusted using the Benjamini-Hochberg FDR correction method).



Supplementary Figure 9. **Reduced dependency on JNK signaling in PAAD cell lines with KRAS G12V mutations.** **a.** The "JNK phosphorylation and activation by activated TAK1" gene set was significantly enriched for reduced genetic dependency in PAAD cell lines with KRAS G12V. 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 each gene. Thus, each column indicates a rank. The line plots above the heatmap indicate the representation (density) of each KRAS allele at each rank across the genes. **b.** The genetic dependency on MAPK8 of cell lines of different KRAS alleles in PAAD (\*\*: p < 0.01; p-values were adjusted using the Benjamini-Hochberg FDR correction method). **c.** The genetic dependency on JUN and MAPK8 of cell lines of different KRAS alleles in PAAD. Each point is a cell line colored according to its KRAS allele. The dashed line represents the line-of-best fit. The equation, p-value, and for the line are shown in the top-right.



Supplementary Figure 10. **The protein modifications to STK11 in LUAD.** The mutations along STK11 separated by those in KRAS G12C-mutant samples (top) and the rest of the samples (bottom). The color of the vertical bars indicates the type of mutation, and a red dot at the top indicates if missense mutations at the location are predicted or known to be functionally damaging. The domains of STK11 are shown between the bar-plots. The magenta portion of the N-terminus is the nuclear localization signal.



*Supplementary Figure 11.* ***PPIN subnetworks from the integration of comutation and genetic dependency results for KRAS G12C and G12V in LUAD.*** *The largest connected subnetworks of the PPIN of genes found to comutate or demonstrate differential dependence with only KRAS G12C (****b****) of G12V (****b****) in LUAD were extracted. Due to the order of the G12C subnetwork, only a selection of the nodes are labeled. Curated groups of proteins with shared cellular functions are highlighted and labeled.*



Supplementary Figure 12. **PPIN subnetworks from the integration of comutation and genetic dependency results for PAAD.** The largest connected subnetworks of the PPIN of genes found to comutate or demonstrate differential dependence with a KRAS allele in PAAD was extracted and merged. Nodes present in multiple subnetworks are colored shades of brown, save for prominent oncogenes and tumor suppressors in COAD, shown in red ("special"). The rest of the nodes are colored by which allele they are associated with. Curated groups of proteins with shared cellular functions are highlighted and labeled.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

36. Unni, A. M., Lockwood, W. W., Zejnullahu, K., Lee-Lin, S.-Q. & Varmus, H. Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. *eLife* **4**, e06907 (2015).

37. Ambrogio, C., Barbacid, M. & Santamaría, D. In vivo oncogenic conflict triggered by co-existing KRAS and EGFR activating mutations in lung adenocarcinoma. *Oncogene* **36**, 2309–2318 (2017).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

52. Grohmann, A., Tanneberger, K., Alzner, A., Schneikert, J. & Behrens, J. AMER1 regulates the distribution of the tumor suppressor APC between microtubules and the plasma membrane. *Journal of cell science* **120**, 3738–47 (2007).

53. Tanneberger, K. *et al.* Structural and functional characterization of the Wnt inhibitor APC membrane recruitment 1 (Amer1). *The Journal of biological chemistry* **286**, 19204–14 (2011).

54. Svoronos, A. A., Engelman, D. M. & Slack, F. J. OncomiR or Tumor Suppressor? The Duplicity of MicroRNAs in Cancer. *Cancer research* **76**, 3666–70 (2016).

55. Anastasiadou, E., Jacob, L. S. & Slack, F. J. Non-coding RNA networks in cancer. *Nature reviews. Cancer* **18**, 5–18 (2018).

56. Shankar, S. *et al.* KRAS Engages AGO2 to Enhance Cellular Transformation. *Cell reports* **14**, 1448–1461 (2016).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

79. Yao, Z. *et al.* BRAF Mutants Evade ERK-Dependent Feedback by Different Mechanisms that Determine Their Sensitivity to Pharmacologic Inhibition. *Cancer cell* **28**, 370–83 (2015).

80. Yao, Z. *et al.* Tumours with class 3 BRAF mutants are sensitive to the inhibition of activated RAS. *Nature* **548**, 234–238 (2017).

81. Dagogo-Jack, I. *et al.* 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**, 158–165 (2019).

82. Bracht, J. W. P. *et al.* 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**, (2019).

83. Evers, B., Bernards, R. & Beijersbergen, R. L. The Good, the Bad, and the Ugly: in search of gold standards for assessing functional genetic screen quality. *Molecular systems biology* **10**, 738 (2014).

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

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

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

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

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

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

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

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

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

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

94. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA: a cancer journal for clinicians* **70**, 7–30 (2020).

95. Meza, R., Meernik, C., Jeon, J. & Cote, M. L. Lung cancer incidence trends by gender, race and histology in the United States, 1973-2010. *PloS one* **10**, e0121323 (2015).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

112. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell reports* **3**, 246–59 (2013).

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

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

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

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

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

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

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

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

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

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

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

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