

## **The frequency of Ras mutations in cancer**

Ian A. Prior<sup>1\*</sup>, Fiona E. Hood<sup>1</sup> and James L. Hartley<sup>2</sup>

CAN-19-3682

Running Title: Ras cancer statistics

Keywords: HRAS, KRAS, NRAS, cancer, mutation, genetics

Word Count: 3220 + references and Figure legends

1. Division of Cellular and Molecular Physiology, Institute of Translational Medicine, University of Liverpool, L69 3BX, UK.

2. NCI RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland, 21701, USA.

\* Corresponding Author: Ian A. Prior

Email: [iprior@liv.ac.uk](mailto:iprior@liv.ac.uk)

Tel: +44-151-794-5332

Fax: +44-151-794-4434

### **Conflict of interest statement**

All authors declare that they have no conflicts of interest relevant to this manuscript.

### **Author contributions**

All authors have contributed to the writing and design of this manuscript. IAP is responsible for the conception, analysis, design and writing.

## ABSTRACT

Ras is frequently mutated in cancer; however, there is a lack of consensus in the literature regarding the cancer mutation frequency of Ras with quoted values varying from 10-30%. This variability is at least in part due to the selective aggregation of data from different databases and the dominant influence of particular cancer types and particular Ras isoforms within these datasets. In order to provide a more definitive figure for Ras mutation frequency in cancer, we have cross-referenced the data in all major publicly accessible cancer mutation databases to determine reliable mutation frequency values for each Ras isoform in all major cancer types. These percentages have then been applied to current US cancer incidence statistics to estimate the number of new patients each year that have Ras-mutant cancers. We find that ~19% of cancer patients harbor Ras mutations; equivalent to ~3.4 million new cases per year worldwide. We discuss the Ras isoform and mutation-specific trends evident within the datasets that are relevant to current Ras-targeted therapies.

## INTRODUCTION

Ras proteins activate signaling networks controlling cell proliferation, differentiation and survival (1). They are encoded by three ubiquitously expressed genes, HRAS, KRAS and NRAS, that share significant sequence homology and largely overlapping functions (2). Ras proteins cycle between an inactive GDP-bound conformation and an active GTP-bound conformation. Activation is facilitated by guanine nucleotide exchange factors (GEFs), and inactivating GTP hydrolysis is enhanced by GTPase-activating proteins (GAPs). Ras activation causes a conformational change that allows engagement with more than 20 different proteins from 10 effector families (3). The most intensively studied of these from a cancer and therapeutic perspective have been the RAF and PtdIns-3 kinase families. Activated Ras concentrates effector proteins into plasma membrane signaling nanoclusters where they can interact with necessary proteins and lipids to control downstream pathways (4). Mutations of Ras that render the protein constitutively active are widely observed in cancer; however, there are distinctive patterns in the mutation frequencies associated with each Ras gene and cancer type (5).

## ONCOGENIC RAS

Gain-of-function missense mutations promote oncogenesis with almost all detected in patients clustering in three hotspots at codons 12, 13 and 61 (5). These result in enhanced GTP binding due to fast exchange of nucleotide and/or impairment of GAP binding (6). Although these mutants are all activating, they are not equal in their oncogenic potential and differences in patient survival are associated with different Ras mutants (7-11). Mutation-specific oncogenesis was clearly demonstrated *in vivo* using a library CRISPR gene-editing approach that allowed 12 different activating codon 12 and 13 mutations of KRAS to be simultaneously compared in each mouse and found that only five of these mutants resulted in the development of lung tumors (12). The frequency of individual mutations in patients also varies across tissue types and between isoforms suggesting contextual influences that determine which isoform-mutation combinations have a selective advantage in different cancer types (5).

The interplay between three main factors determines whether conditions are permissive for initiation and progression of Ras-dependent oncogenesis and might explain why specific Ras isoforms and mutations are associated with distinct cancer types (5). The first is Ras dosage that is defined by expression levels and relative activation state (13). The proportion of a Ras population that is GTP-bound varies from

30-90% dependent upon which mutation is present (14). Furthermore, the stability of the active state can vary dependent on whether the mutant is fast-cycling or GAP-insensitive (6). Together with the fact that Ras expression levels vary over 100-fold between isoforms and across tissues (15), this means that significant differences in Ras signaling capacity can result depending on the tissue/isoform/mutation combination. Only a subset of these combinations will be optimal since too much Ras signaling promotes senescence or cell death whilst too little fails to initiate tumorigenesis (16-18). Importantly, the narrow range of permissive signaling capacity is subject to change over time to facilitate tumor progression and resistance to therapeutics (19).

The second factor is signal specificity associated with each Ras isoform and their individual mutations. The extent to which Ras isoforms display differential coupling to effector pathways is not well understood. It is also confounded by differences in expression/dosing; however, *in vivo* experiments expressing Ras isoforms from the same genetic locus to avoid dosage influences still revealed that Ras isoforms cannot fully recapitulate the functions of each other (20,21). Isoform-specific signaling is thought to be mediated by differential intracellular localisation that favors preferential coupling to specific effector pathways (4,22,23), and by distinct biochemical properties imparted by allosteric lobe sequence variations between each isoform (24). Recent *in vitro* analysis revealed distinct binding preferences for Ras-Raf interactions with BRAF binding being highly selective for KRAS whilst CRAF was critical for HRAS-mediated MAP kinase signaling (25). Mutational-specificity is also important for Ras biology (12,26-29), and structural and biochemical features underpinning mutational differences in nucleotide cycling, allosteric regulation and GEF, GAP and effector interactions are now being defined (14,25,28,30-32).

The final factor is cellular and tissue context that contributes the genetic, epigenetic and proteomic landscapes in which Ras networks operate. This heterogeneity can result in different proliferative potential depending on the capacity of the oncogene to engage the subset of drivers in that cell or tissue (28,33,34). Ras dosing and signaling specificity titrated against these backgrounds will favor selection of different Ras variant combinations in each tissue. Cellular context is also important: an example of this was seen in the spatial and cell type-specific variation in MAP kinase activation in KRAS<sup>G12V</sup> mutated mouse colon that was shaped by cell type differences in expression of regulatory proteins of the MAP kinase pathway (35).

Our understanding of cancer is facilitated by observing patterns found in cancer mutation databases. Those data have highlighted the isoform/mutation combinations most frequently seen in each tissue and this has helped to inform the development and testing of potential Ras-targeted therapies in appropriate patient groups. However, it is also important to note that the lack of consensus among these datasets that can result in incorrect estimates of the true disease burden associated with each Ras isoform.

## **CANCER MUTATION DATABASES**

We incorporate four leading cancer mutation databases into our analyses. The largest of these is the Catalogue of Somatic Mutations in Cancer (**COSMIC**), that contains manually curated data from the cancer literature comprising ~9.7 million coding mutations from ~1.4 million samples (including ~34,000 whole genomes; (36)). The most refined database is The Cancer Genome Atlas (**TCGA**) that has molecularly characterised tumor samples from ~11,300 patients representing 33 cancer types. All samples in TCGA have been subject to comprehensive genomic, epigenomic,

transcriptomic and proteomic analysis to better understand the oncogenic systems biology of molecular subtypes of cancer (37). The co-ordinated management within the TCGA Program means that verification and quality control of sample type is also likely to be the most consistent versus larger datasets derived from multiple independent studies. The Memorial Sloan Kettering Cancer Centre (MSKCC) **cBioPortal** facilitates meta-analysis of TCGA and ~130 other datasets comprising ~40,000 samples (38). The International Cancer Genome Consortium (**ICGC**) data portal performs a similar function with multi-omic data for 22 cancer types currently curated from ~24,000 samples derived from patients from around the world (39). There is overlap in the data present within each database, with almost all TCGA data also found within the cBioPortal data portal, ~50% of TCGA data are present within the ICGC data portal and all of the data found in TCGA, cBioPortal and ICGC is collated within COSMIC (Figure 1).

With such rich and integrated datasets available, it might seem surprising that there is still no consensus on the disease burden associated with a major oncogene such as mutant Ras. For example, 11.6% of TCGA samples, 17.5% of cBioPortal samples, 19.3% of ICGC samples and 24.8% of COSMIC samples are Ras mutant (Figure 1). These differences are driven by the different priorities underpinning sample curation. COSMIC collates widely from the cancer literature and includes a high proportion of tissues such as colon and lung that contain high percentages of KRAS mutations (Figure 1B). In contrast, the other datasets consist of studies where Ras mutation status was not a factor in their collection and consequently contain larger proportions of breast, brain, kidney and liver samples where Ras mutations are rarely observed.

In fact, none of the percentages accurately reflect Ras disease burden since none of the datasets accurately recapitulate the relative frequency of each disease in the patient population. An example of this is seen in the TCGA dataset, where the ten rarest cancers representing 2.7% of new cases/year in the US account for ~20% of TCGA samples. Given this, how should the datasets be used best to understand Ras mutation patterns? It seems likely that the sheer volume of Ras mutant samples within the COSMIC dataset (Figure 1), means that for many cancer types broadly accurate conclusions can be drawn regarding the association of particular Ras isoforms and the types of mutations observed. In contrast, the smaller datasets are particularly suited to comparative analysis of genome-wide changes and genetic associations with mutant Ras.

## **RAS PATTERNS ACROSS DATASETS**

Comparison of KRAS mutation data for major KRAS-associated cancers illustrates the challenge of reaching a consensus across the datasets (Supplementary Table 1). For example, colorectal cancer exhibits KRAS mutation frequencies of ~33% in the COSMIC dataset comprising ~75,000 tested samples. The frequencies of 40-45% suggested by the other datasets are based on small sample sizes of fewer than 500. Notably however, the private Foundation Medicine (FM) dataset comprising 13,336 colorectal samples reports a KRAS mutation frequency of ~50% (40). This may be due to higher sensitivity of the recent genetic screening methods employed by FM versus the long-term aggregate data in COSMIC. A second point of difference is that the samples within the FM dataset were from patients that presented with advanced metastatic disease in contrast to the heterogeneity of samples collated in COSMIC that were derived from a wide range of studies. For lung cancer, similar disparities are evident with COSMIC reporting lower percentages of ~21% and the small-scale

datasets giving intermediate values compared to a recent large-scale FM study that found that 31% of 5,749 lung adenocarcinoma samples were KRAS mutant (41).

The collated pancreatic data reveals a consistent anomaly: according to literature sources 90-98% of pancreatic cancers are KRAS mutated (1,3,42); while most datasets contain lower values of 70-75% (Supplementary Table 1). The stroma of pancreatic ductal adenocarcinomas (PDACs) are extensively infiltrated with cancer-associated fibroblasts that are not Ras mutant (43). It seems likely that the lower estimates across the datasets are confounded by stromal sampling resulting in reduced sensitivity for positively identifying KRAS mutations in the small subpopulations of cancer cells present. Consistent with this, a recent large-scale cancer genetics study by Foundation Medicine of 3,594 primary and metastatic PDAC samples where a required threshold of cancer cell content was rigorously verified in each sample reported that 88% of samples contain KRAS mutations (44).

Where publicly available datasets such as the TCGA excel is in multi-gene comparisons and in the integrated analysis of a wide range of types of genetic changes. The Pan-Cancer Atlas used TCGA data to define the molecular subtypes of cancer and to generate deep understanding of the genetic programmes associated with driving different cancer types (45). Analysis of 85 genes within the immediate receptor tyrosine kinase (RTK)-Ras network (37), reveals a wide range of Ras pathway dependencies ranging from 30-96% of samples across the TCGA cancer types (Supplementary Figure 1). Some of these such as pancreatic cancer (PAAD in TCGA terminology) are highly linked to Ras mutation; however, most cancers exhibit activating genetic changes in the Ras network independently of mutant Ras. Targeting Ras and Ras network genes is clearly relevant in nearly all cancer types, even those where Ras mutations are rarely found.

Amplification of non-mutated Ras is a feature of a subset of cancers including esophageal (ESCA), stomach (STAD), ovarian (OV) and testicular (TGCT) where it represents the dominant type of Ras genetic change (Supplementary Figure 1). In each of these cases KRAS is by far the most frequently amplified and the most frequently mutated Ras isoform. We haven't formally checked whether it is the wild type or mutant allele that is amplified; however, in these specific cancer types where amplification is seen more often than mutation, at least a subset of these events will be in the wild type allele. Changes in Ras dosage are associated with progression and response to therapy (19); and the deletions and amplifications observed in some cancer types may reflect this. Copy number analysis reveals that the Ras isoforms display distinct patterns, with KRAS typically amplified whilst HRAS is more often deleted when copy number changes occur (Supplementary Table 2b). Some of these changes appear to be reciprocal; for example, in squamous cell lung cancer (LUSC), HRAS and NRAS exhibit copy number losses whereas KRAS increases (Supplementary Table 2b). Alternatively, in bladder (BLCA), ovarian (OV) and testicular (TCGT) cancer, HRAS is decreased but both NRAS and KRAS show a clear tendency for increasing copy number. Together, these suggest that some interesting isoform-specific biology may be at play that could be worth further investigation.

Thus the lack of consensus at the tissue level remains problematic; however, the major themes within each dataset in terms of preferential coupling of particular Ras isoforms to specific cancer types and the patterns of mutations are consistent across all datasets. In the absence of access to large private cancer genetics datasets that have the benefits of scale, consistently high quality of curation and comprehensive genomic screening, we will use the COSMIC dataset together with selected publicly

available data from the Foundation Medicine database to collate Ras isoform mutations patterns across a wide range of cancer types.

## **RAS MUTATION FREQUENCIES IN CANCER**

In order to estimate Ras disease burden, it is necessary to convert the Ras mutation frequencies found in cancer genetics databases into patient numbers based on current cancer incidence data. We have collated frequencies for a wide range of cancers from all four databases (COSMIC, cBioPortal, ICGC and TCGA; Supplementary Table 2). The data are derived from formally verified cancer types rather than samples with only a generic tissue-based categorization and we have used the TCGA naming system where relevant to facilitate cross-comparison between databases. Data from COSMIC, together with publicly available Foundation Medicine data for all three Ras isoforms in colorectal adenocarcinoma (COAD, READ) and KRAS mutation frequencies in lung (LUAD) and pancreatic (PAAD) adenocarcinomas are presented in Table 1. In 2018 the American Cancer Society estimated that ~1.7 million new cases of cancer were diagnosed in the United States (46). The 29 cancer types presented in Table 1 represent ~80% coverage of United States cancer cases. Note that we have not included non-melanoma skin cancers in Table 1 since they almost always present as benign, they are typically under-reported in cancer statistics and they are not included in global cancer incidence reports.

We estimate that ~19% of cancer patients will harbor a Ras mutation; this is equivalent to ~260,000 new cases per year in the United States. Globally, there are currently ~18 million new cancer diagnoses per year (47). Whilst acknowledging that the incidence of different cancer types varies around the world, a simple extrapolation of our observations suggests that there are ~3.4 million new cancer cases worldwide per year with a Ras mutation. KRAS is the most frequently mutated of the three Ras isoforms in 19 out of the 29 cancer types in Table 1 and is responsible for 75% of Ras mutant cancers. NRAS (17% of patients) and HRAS (7%) show strong coupling to only a small subset of cancer types. Isoform-specific coupling is particularly evident for major KRAS cancer types and for NRAS in melanoma (SKCM; Table 1). In contrast, thyroid cancer sub-types (THCA, THCAF) are notable for displaying high levels of mutation in all three Ras isoforms.

Although KRAS is the major cancer-causing isoform, patient numbers for the other Ras isoforms are still significant and translate into ~230,000 patients globally for HRAS and ~560,000 for NRAS. This highlights the importance of targeting all isoforms and not just the current concentration on developing KRAS-targeted therapies. Farnesyl transferase inhibitors (FTIs) that failed clinical trials more than 20 years ago (48) are enjoying a renaissance due to personalized medicine approaches. KRAS and NRAS do not respond to FTIs; however, HRAS is sensitive (48). Ras mutation profiling means that suitable patients can now be identified and the FTI Tipifarnib is currently progressing through Phase 2 clinical trials for use in HRAS mutant head and neck, leukemia, lymphoma and thyroid cancers (ClinicalTrials.Gov). Whilst most of these cancers represent obvious choices for use of FTIs (Table 1), there are appreciable numbers of potential beneficiaries across a wide range of other cancer types. This includes breast cancer where we estimate that globally there are ~17,000 new cases per year that will harbor mutant HRAS. In addition to targeted use of existing FTIs, an exciting new FTI has recently been developed that overcomes KRAS resistance and potentially opens the way to pan-Ras inhibition (49).

Examining mutation-specific patterns also reveals patterns associated with distinct tissue types (Supplementary Table 3). For example, KRAS codon 13 mutations are

particularly associated with the gastro-intestinal tract and some blood cancers where rare mutations of A146 are also mostly observed. There are 19 different activating codon 12, 13 or 61 mutations that can be created in each isoform by a single base change. Five mutations (G12D, G12V, G12C, G13D and Q61R) account for 70% of all Ras-mutant patients. G12C mutations are frequently found in lung cancer due to G:C>T:A transversions associated with bulky adducts generated by the mutagens in tobacco smoke (50). Chemical inhibitors of G12C-mutated KRAS have been developed (51-53). These compounds preferentially bind to GDP-bound KRAS and prevent exchange for GTP and interaction with effectors. Compounds developed by Mirati, Wellspring, Janssen, and Amgen have now entered Phase I/II clinical trials and Amgen recently reported the initial results of their trials with AMG 510 where it was tolerated by patients and stabilized or partially regressed their non-small cell lung tumors (54). Associated *in vivo* studies revealed a synergy with immunotherapy and excitingly 9 out of 10 mice showed complete and curative tumor regression when AMG 510 was used in combination with anti-PD-1. Whilst lung cancer has been the focus of these trials due to the smoking-associated prevalence of KRAS<sup>G12C</sup>, many other cancer types contain appreciable numbers of potential beneficiaries of treatment (Supplementary Table 3), especially when considered on a global scale. It is also important to be able to target other mutations such as G12V and G12D that each have a 2-3-fold higher disease burden than G12C. Several compounds targeting these other mutations are now in development and entering trials.

## **CONCLUDING REMARKS AND FUTURE DIRECTIONS**

We have determined the global disease burden associated with Ras mutations for different cancer types. ~19% of cancer patients harbor Ras mutations with KRAS responsible for 75% of that number. Our meta-analysis revealed the differences in sampling and interpretation that underlie the lack of consensus that has prevailed to date. Converting the frequencies into patient numbers also helps to refocus attention onto the substantial populations of patients that might benefit from anti-Ras therapeutics in cancers where Ras is not frequently mutated. After a prolonged period where it seemed that Ras was undruggable, we are now entering an era where it seems feasible that we will have Ras-targeted precision therapy options available that can be tailored to individual mutations and cancers. The patient number estimates give an indication of the size of the pool potentially available for clinical trials. Given the isoform-, mutation- and tissue-specific differences in Ras biology that are now evident, it will be important to have access to even larger databases with high quality sample curation and genome-wide profiling to develop deeper understanding that will inform these precision medicine approaches.

## **ACKNOWLEDGEMENTS**

This work was supported by funding from North West Cancer Research (CR1166) and NWCR endowment funding for IAP, and by Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E for JLH. We gratefully acknowledge the contributions of the National Cancer Institute, MSKCC, ICGC, the Sanger Centre and others for generating the cancer genetics databases that made this work possible.

## **REFERENCES**

1. Simanshu DK, Nissley DV, McCormick F. RAS Proteins and Their Regulators in Human Disease. *Cell* **2017**;170:17-33

2. Mo SP, Coulson JM, Prior IA. RAS variant signalling. *Biochem Soc Trans* **2018**;46:1325-32
3. Hobbs GA, Der CJ, Rossman KL. RAS isoforms and mutations in cancer at a glance. *Journal of Cell Science* **2016**;129:1287-92
4. Zhou Y, Hancock JF. Ras nanoclusters: Versatile lipid-based signaling platforms. *Biochim Biophys Acta* **2015**;1853:841-9
5. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res* **2012**;72:2457-67
6. Smith MJ, Neel BG, Ikura M. NMR-based functional profiling of RASopathies and oncogenic RAS mutations. *P Natl Acad Sci USA* **2013**;110:4574-9
7. Seeburg PH, Colby WW, Capon DJ, Goeddel DV, Levinson AD. Biological properties of human c-Ha-ras1 genes mutated at codon 12. *Nature* **1984**;312:71-5
8. Trahey M, Milley RJ, Cole GE, Innis M, Paterson H, Marshall CJ, *et al.* Biochemical and biological properties of the human N-ras p21 protein. *Mol Cell Biol* **1987**;7:541-4
9. Der CJ, Finkel T, Cooper GM. Biological and biochemical properties of human rasH genes mutated at codon 61. *Cell* **1986**;44:167-76
10. Bournet B, Muscari F, Buscail C, Assenat E, Barthelet M, Hammel P, *et al.* KRAS G12D Mutation Subtype Is A Prognostic Factor for Advanced Pancreatic Adenocarcinoma. *Clin Transl Gastroenterol* **2016**;7:e157
11. Blons H, Emile JF, Le Malicot K, Julie C, Zaanani A, Tabernero J, *et al.* Prognostic value of KRAS mutations in stage III colon cancer: post hoc analysis of the PETACC8 phase III trial dataset. *Ann Oncol* **2014**;25:2378-85
12. Winters IP, Chiou SH, Paulk NK, McFarland CD, Lalgudi PV, Ma RK, *et al.* Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. *Nat Commun* **2017**;8:2053
13. Li S, Balmain A, Counter CM. A model for RAS mutation patterns in cancers: finding the sweet spot. *Nat Rev Cancer* **2018**;18:767-77
14. Killoran RC, Smith MJ. Conformational resolution of nucleotide cycling and effector interactions for multiple small GTPases determined in parallel. *J Biol Chem* **2019**;294:9937-48
15. Newlaczyl AU, Coulson JM, Prior IA. Quantification of spatiotemporal patterns of Ras isoform expression during development. *Sci Rep* **2017**;7:41297
16. Pershing NL, Lampson BL, Belsky JA, Kaltenbrun E, MacAlpine DM, Counter CM. Rare codons capacitate Kras-driven de novo tumorigenesis. *J Clin Invest* **2015**;125:222-33
17. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **1997**;88:593-602
18. Sarkisian CJ, Keister BA, Stairs DB, Boxer RB, Moody SE, Chodosh LA. Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis. *Nat Cell Biol* **2007**;9:493-505
19. Burgess MR, Hwang E, Mroue R, Bielski CM, Wandler AM, Huang BJ, *et al.* KRAS Allelic Imbalance Enhances Fitness and Modulates MAP Kinase Dependence in Cancer. *Cell* **2017**;168:817-29 e15
20. Drosten M, Simon-Carrasco L, Hernandez-Porras I, Lechuga CG, Blasco MT, Jacob HK, *et al.* H-Ras and K-Ras Oncoproteins Induce Different Tumor Spectra When Driven by the Same Regulatory Sequences. *Cancer Res* **2017**;77:707-18



21. Potenza N, Vecchione C, Notte A, De Rienzo A, Rosica A, Bauer L, *et al.* Replacement of K-Ras with H-Ras supports normal embryonic development despite inducing cardiovascular pathology in adult mice. *EMBO Rep* **2005**;6:432-7
22. Santra T, Herrero A, Rodriguez J, von Kriegsheim A, Iglesias-Martinez LF, Schwarzl T, *et al.* An Integrated Global Analysis of Compartmentalized HRAS Signaling. *Cell Rep* **2019**;26:3100-15 e7
23. Adhikari H, Counter CM. Interrogating the protein interactomes of RAS isoforms identifies PIP5K1A as a KRAS-specific vulnerability. *Nat Commun* **2018**;9:3646
24. Johnson CW, Reid D, Parker JA, Salter S, Knihtila R, Kuzmic P, *et al.* The small GTPases K-Ras, N-Ras, and H-Ras have distinct biochemical properties determined by allosteric effects. *J Biol Chem* **2017**;292:12981-93
25. Terrell EM, Durrant DE, Ritt DA, Sealover NE, Sheffels E, Spencer-Smith R, *et al.* Distinct Binding Preferences between Ras and Raf Family Members and the Impact on Oncogenic Ras Signaling. *Mol Cell* **2019**;76:1-13
26. Hammond DE, Mageean CJ, Rusilowicz EV, Wickenden JA, Clague MJ, Prior IA. Differential reprogramming of isogenic colorectal cancer cells by distinct activating KRAS mutations. *J Proteome Res* **2015**;14:1535-46
27. Burd CE, Liu W, Huynh MV, Waqas MA, Gillahan JE, Clark KS, *et al.* Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. *Cancer Discov* **2014**;4:1418-29
28. Poulin EJ, Bera AK, Lu J, Lin YJ, Strasser SD, Paulo JA, *et al.* Tissue-Specific Oncogenic Activity of KRAS(A146T). *Cancer Discov* **2019**;9:738-55
29. Stolze B, Reinhart S, Bullinger L, Frohling S, Scholl C. Comparative analysis of KRAS codon 12, 13, 18, 61, and 117 mutations using human MCF10A isogenic cell lines. *Sci Rep* **2015**;5:8535
30. Hunter JC, Manandhar A, Carrasco MA, Gurbani D, Gondi S, Westover KD. Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. *Mol Cancer Res* **2015**;13:1325-35
31. Johnson CW, Lin YJ, Reid D, Parker J, Pavlopoulos S, Dischinger P, *et al.* Isoform-Specific Destabilization of the Active Site Reveals a Molecular Mechanism of Intrinsic Activation of KRas G13D. *Cell Rep* **2019**;28:1538-50 e7
32. Buhrman G, Kumar VS, Cirit M, Haugh JM, Mattos C. Allosteric modulation of Ras-GTP is linked to signal transduction through RAF kinase. *J Biol Chem* **2011**;286:3323-31
33. Sack LM, Davoli T, Li MZ, Li Y, Xu Q, Naxerova K, *et al.* Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell* **2018**;173:499-514 e23
34. Haigis KM, Cichowski K, Elledge SJ. Tissue-specificity in cancer: The rule, not the exception. *Science* **2019**;363:1150-1
35. Brandt R, Sell T, Luthen M, Uhlitz F, Klinger B, Riemer P, *et al.* Cell type-dependent differential activation of ERK by oncogenic KRAS in colon cancer and intestinal epithelium. *Nat Commun* **2019**;10:2919
36. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **2019**;47:D941-D7
37. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **2018**;173:321-37

38. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2012**;2:401-4
39. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, *et al.* The International Cancer Genome Consortium Data Portal. *Nat Biotechnol* **2019**;37:367-9
40. Serebriiskii IG, Connelly C, Frampton G, Newberg J, Cooke M, Miller V, *et al.* Comprehensive characterization of RAS mutations in colon and rectal cancers in old and young patients. *Nat Commun* **2019**;10:3722
41. Suh JH, Johnson A, Albacker L, Wang K, Chmielecki J, Frampton G, *et al.* Comprehensive Genomic Profiling Facilitates Implementation of the National Comprehensive Cancer Network Guidelines for Lung Cancer Biomarker Testing and Identifies Patients Who May Benefit From Enrollment in Mechanism-Driven Clinical Trials. *Oncologist* **2016**;21:684-91
42. Kleeff J, Korc M, Apte M, La Vecchia C, Johnson CD, Biankin AV, *et al.* Pancreatic cancer. *Nat Rev Dis Primers* **2016**;2:16022
43. Ligorio M, Sil S, Malagon-Lopez J, Nieman LT, Misale S, Di Pilato M, *et al.* Stromal Microenvironment Shapes the Intratumoral Architecture of Pancreatic Cancer. *Cell* **2019**;178:160-75 e27
44. Singhi AD, George B, Greenbowe JR, Chung J, Suh J, Maitra A, *et al.* Real-Time Targeted Genome Profile Analysis of Pancreatic Ductal Adenocarcinomas Identifies Genetic Alterations That Might Be Targeted With Existing Drugs or Used as Biomarkers. *Gastroenterology* **2019**;156:2242-53 e4
45. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **2018**;173:291-304 e6
46. American\_Cancer\_Society. Cancer Facts & Figures 2018. Atlanta: American Cancer Society; 2018.
47. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **2018**;68:394-424
48. Whyte DB, Kirschmeier P, Hockenberry TN, Nunez-Oliva I, James L, Catino JJ, *et al.* K- and N-Ras are geranylgeranylated in cells treated with farnesyl protein transferase inhibitors. *J Biol Chem* **1997**;272:14459-64
49. Novotny CJ, Hamilton GL, McCormick F, Shokat KM. Farnesyltransferase-Mediated Delivery of a Covalent Inhibitor Overcomes Alternative Prenylation to Mislocalize K-Ras. *Acs Chem Biol* **2017**;12:1956-62
50. Seo KY, Jelinsky SA, Loechler EL. Factors that influence the mutagenic patterns of DNA adducts from chemical carcinogens. *Mutat Res* **2000**;463:215-46
51. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **2013**;503:548-51
52. Patricelli MP, Janes MR, Li LS, Hansen R, Peters U, Kessler LV, *et al.* Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. *Cancer Discovery* **2016**;6:316-29
53. Xiong Y, Lu J, Hunter J, Li L, Scott D, Choi HG, *et al.* Covalent Guanosine Mimetic Inhibitors of G12C KRAS. *ACS Med Chem Lett* **2017**;8:61-6
54. Canon J, Rex K, Saiki AY, Mohr C, Cooke K, Bagal D, *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **2019**

## FIGURE LEGENDS

**Figure 1.** *Overview of Ras data in cancer genetics databases.* **A.** The sampling relationship, the number of samples tested, and the number of Ras mutant samples identified in each of the publicly accessible databases and data portals. **B.** The sample tissue composition of each database is not equivalent. Data sources: COSMIC v85, cBioPortal v2.2.0, ICGC release 27, TCGA release 12.0 all accessed contemporaneously.

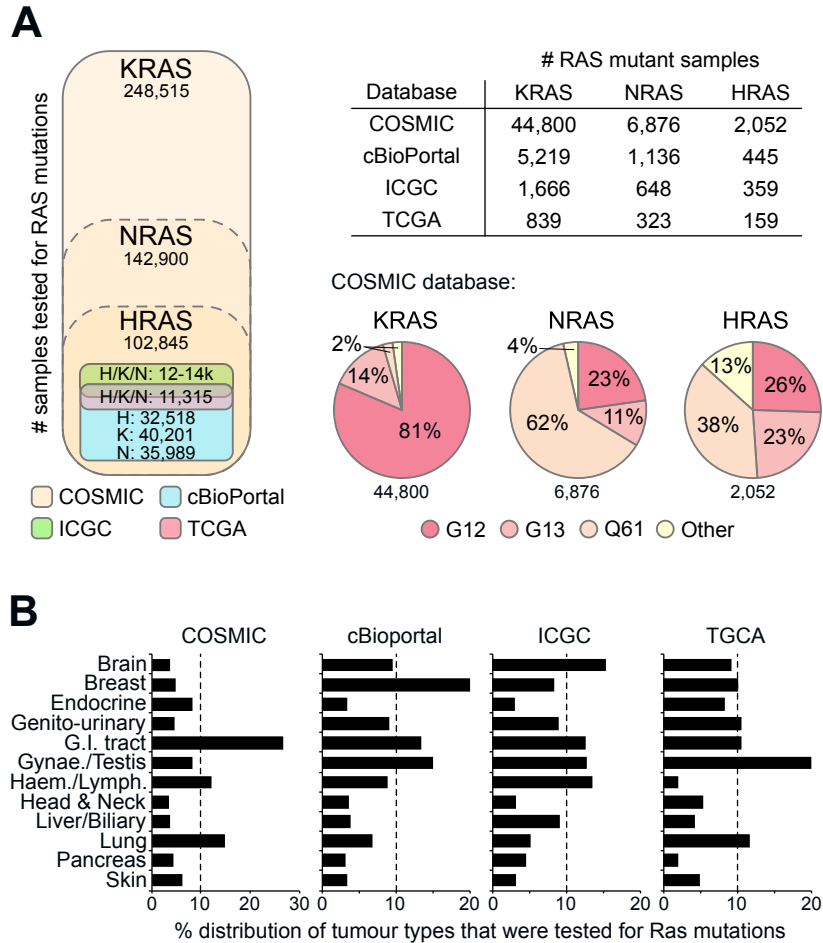
**Table 1.** *Ras mutation frequency and cancer-associated disease burden.* Mutation frequencies are based on COSMIC data together with selected publicly available COAD (H, K, N), READ (H, K, N), LUAD (K) and PAAD (K) data from the Foundation Medicine database (40,41,44). Mutation frequencies are applied to the most recent American Cancer Society data on cancer incidence (total patients) to estimate new cancer cases per year in the USA.

**Supplementary Figure 1.** *Genetic alterations in the RTK/Ras network across cancer types.* TCGA data was collated based on 85 RTK/RAS-relevant genes listed in Supplementary Table 1b. Amplification/deletion represents at least a +2/-2 change in copy number.

**Supplementary Table 1.** *Comparison of KRAS mutation frequencies in each database.* Significant variability in sampling depth and mutation frequency is evident between each database. Summary statistics from databases with tissue type classification are included together with disease-specific values when available.

**Supplementary Table 2.** *Mutation statistics in major cancer genetics databases.* Ras isoform mutation data (Table 1a) and copy number variance (CNV) analysis (Table 1b) are included together with details on how data was collated from COSMIC v85, cBioPortal v2.2.0, ICGC release 27 and TCGA release 12.0.

**Supplementary Table 3.** *Ras mutation-specific patterns in cancer.* Data are all collated from COSMIC v85 and where indicated converted in patient numbers using ACS 2018 cancer incidence statistics.

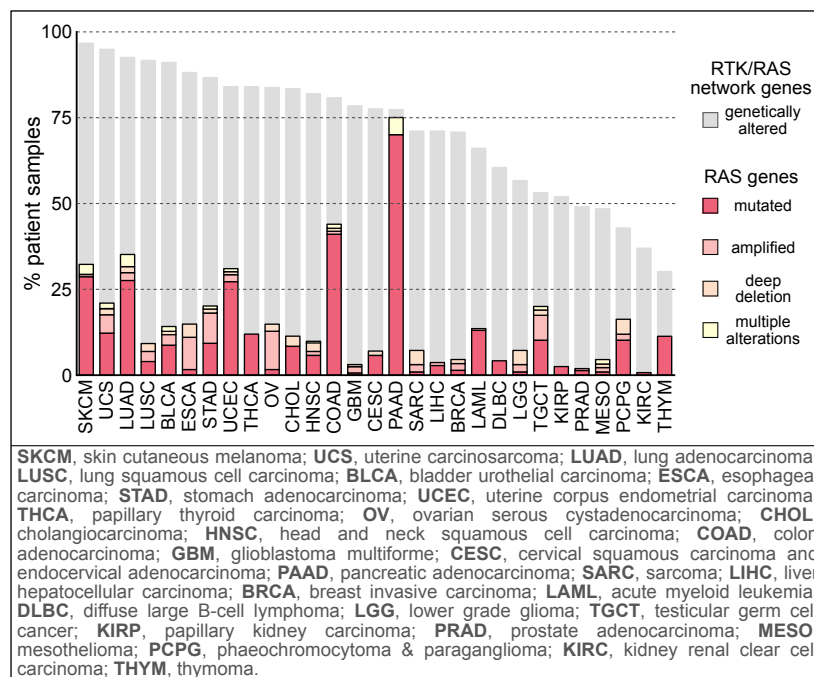


**Figure 1. Overview of Ras data in cancer genetics databases. A.** The sampling relationship, the number of samples tested, and the number of Ras mutant samples identified in each of the publicly accessible databases and data portals. **B.** The sample tissue composition of each database is not equivalent. Data sources: COSMIC v85, cBioPortal v2.2.0, ICGC release 27, TCGA release 12.0 all accessed contemporaneously.

cancer types	HRAS		KRAS		NRAS		% mutated			Est. no. new patients/year (USA)				total patients	
	mut.	tested	mut.	tested	mut.	tested	H	K	N	HRAS	KRAS	NRAS	*RAS		
C.N.S.	GBM	8	1,520	14	1,732	11	1,731	0.5	0.8	0.6	25	39	30	94	4,776
	LGG	1	1,571	19	1,864	11	1,802	0.1	1.0	0.6	5	73	44	121	7,164
Head & Neck	HNSC	180	3,507	95	4,833	52	3,192	5.1	2.0	1.6	3,320	1,272	1,054	5,646	64,690
Endocrine	THCA	85	3,881	78	4,690	279	4,734	2.2	1.7	5.9	927	704	2,494	4,125	42,323
	THCAA	40	794	71	826	144	779	5.0	8.6	19	57	97	209	362	1,129
	THCAF	47	675	31	673	131	694	7.0	4.6	19	393	260	1,065	1,718	5,643
Thoracic	LUAD	7	3,576	2,186	6,838	72	6,992	0.2	32	1.0	183	29,956	964	31,103	93,612
	LUSC	8	1,273	193	4,849	10	1,748	0.6	4.0	0.6	441	2,794	402	3,637	70,209
Breast	BRCA	38	6,329	108	8,109	22	5,902	0.6	1.3	0.4	1,613	3,578	1,001	6,193	268,670
	ESCA	7	2,667	52	3,087	2	2,134	0.3	1.7	0.1	45	291	16	353	17,290
	STAD	11	2,241	158	3,812	7	1,286	0.5	4.1	0.5	129	1,088	143	1,359	26,240
Core G.I.	SIAD	0	122	261	989	2	209	0.0	26	1.0	0	2,763	100	2,863	10,470
	COAD	59	11,400	5,749	11,400	484	11,400	0.5	50	4.2	486	48,999	4,083	53,568	97,220
	READ	4	1,936	978	1,936	79	1,936	0.2	50	4.1	103	26,063	2,116	28,282	51,610
Accessory G.I.	LIHC	2	1,621	27	2,090	9	1,824	0.1	1.3	0.5	47	491	187	725	37,998
	CHOL	0	757	530	2,345	32	882	0.0	23	3.6	0	954	153	1,107	4,222
	GBC	2	241	125	776	5	313	0.8	16	1.6	101	1,964	195	2,259	12,190
	PAAD	0	2,057	3,163	3,594	9	2,093	0.0	88	0.4	0	48,787	238	49,025	55,440
Genito-urinary	KIRC	3	1,433	19	1,811	5	1,584	0.2	1.0	0.3	96	480	144	720	45,738
	BLCA	219	3,136	129	2,565	31	2,296	7.0	5.0	1.4	5,670	4,083	1,096	10,849	81,190
Gynaecological	PRAD	40	2,673	117	3,430	8	2,700	1.5	3.4	0.3	2,464	5,618	488	8,570	164,690
	OV	2	938	259	2,974	17	1,046	0.2	8.6	1.6	47	1,914	361	2,323	22,240
	UCEC	4	1,657	380	2,256	22	799	0.2	17	2.8	147	10,224	1,671	12,042	60,701
	CESC	24	963	85	1,308	5	892	2.5	6.5	0.6	330	860	74	1,265	13,240
Skin	SKCM	37	2,974	56	3,399	1,986	11,749	1.2	1.6	17	1,136	1,504	15,428	18,067	91,270
Haem.	ALL	4	856	265	3,400	358	3,715	0.5	7.8	9.6	28	465	574	1,067	5,960
	AML	1	4,178	345	6,571	1,291	9,024	0.0	5.3	14	5	1,025	2,793	3,822	19,520
	CML	2	639	73	1,144	130	1,341	0.3	6.4	9.7	26	538	817	1,382	8,430
	PCM	3	367	178	1,006	209	1,082	0.8	18	19	252	5,444	5,944	11,639	30,770
estimated total new cancer cases per year (USA)										18,075	202,327	43,884	264,286	1,414,644	
% total cancer patients										1.3	14.3	3.1	18.7		

C.N.S., central nervous system; G.I., gastrointestinal. GBM, glioblastoma multiforme; LGG, lower grade glioma; HNSC, head and neck squamous cell carcinoma; THCA, papillary thyroid carcinoma; THCAA, anaplastic thyroid carcinoma; THCAF, follicular thyroid carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; BRCA, breast invasive carcinoma; ESCA, esophageal carcinoma; STAD, stomach adenocarcinoma; SIAD, small intestine adenocarcinoma; COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; LIHC, liver hepatocellular carcinoma; CHOL, cholangiocarcinoma; GBC, gallbladder carcinoma; PAAD, pancreatic adenocarcinoma; KIRC, kidney renal clear cell carcinoma; BLCA, bladder urothelial carcinoma; PRAD, prostate adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; CESC, cervical squamous carcinoma and endocervical adenocarcinoma; SKCM, skin cutaneous melanoma; ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CML, chronic myeloid leukemia; PCM, plasma cell myeloma.

**Table 1. Ras mutation frequency and cancer-associated disease burden.** Mutation frequencies are based on COSMIC data together with selected publicly available COAD (H, K, N), READ (H, K, N), LUAD (K) and PAAD (K) data from the Foundation Medicine database (40,41,44). Mutation frequencies are applied to the most recent American Cancer Society data on cancer incidence (total patients) to estimate new cancer cases per year in the USA.



**Supplementary Figure 1. Genetic alterations in the RTK/Ras network across cancer types.** TCGA data was collated based on 85 RTK/RAS-relevant genes listed in Supplementary Table 1b. Amplification/deletion represents at least a +2/-2 change in copy number.

KRAS cancer data	COSMIC			cBioportal			ICGC			TCGA		
	mutant	tested	%	mutant	tested	%	mutant	tested	%	mutant	tested	%
LARGE INTESTINE	25,095	75,673	33.2	1,088	3,089	35.2	271	669	40.5	-	-	-
colon adenocarcinoma	2,980	9,619	31.0	190	469	40.5	112	254	44.0	174	400	43.5
rectal adenocarcinoma	1,214	3,453	35.2	92	209	44.0	41	94	43.6	57	137	41.6
LUNG	7,080	40,940	17.3	800	3,771	18.9	19	412	4.6	-	-	-
adenocarcinoma	4,163	19,921	20.9	766	2,708	25.0	15	208	7.2	157	567	27.9
non-small cell carcinoma	2,250	10,961	20.5	14	54	25.9	-	-	-	-	-	-
small cell carcinoma	22	1,037	2.1	-	-	-	-	-	-	-	-	-
squamous cell carcinoma	193	4,849	4.0	16	1,005	1.1	4	204	2.0	9	495	1.8
PANCREAS	6,902	11,863	58.2	676	1,021	66.2	601	748	80.3	-	-	-
ductal adenocarcinoma	5,479	7,685	71.3	641	848	75.6	597	659	90.6	137	182	75.3

**Supplementary Table 1.** Comparison of KRAS mutation frequencies in each database. Significant variability in sampling depth and mutation frequency is evident between each database. Summary statistics from databases with tissue type classification are included together with disease-specific values when available.

cancer	HRAS					KRAS					NRAS					mutation bias			
	G12*	G13*	Q61*	A146*	Others	G12*	G13*	Q61*	A146*	Others	G12*	G13*	Q61*	A146*	Others	H	K	N	
GBM	0	0	1	0	7	8	0	1	1	4	4	0	7	0	0	G12A	11	872	94
LGG	0	0	0	0	1	11	1	5	0	2	4	2	4	0	1	G12C	23	2,139	109
HNSC	90	35	35	0	20	63	14	5	1	12	25	2	17	0	8	G12D	50	5,894	641
THCA	16	6	56	0	7	29	10	21	0	18	5	2	255	0	17	G12R	4	886	32
THCAA	9	0	37	0	1	15	7	9	0	0	1	0	126	0	4	G12S	68	640	198
THCAF	9	9	21	0	1	46	15	9	0	1	6	1	137	0	0	G12V	150	4,097	89
LUAD	0	2	2	0	3	3,573	239	81	7	263	8	0	51	0	13	G13A	0	14	12
LUSC	0	3	4	0	1	159	7	13	0	14	1	0	8	0	1	G13C	6	162	31
BRCA	14	3	9	0	12	73	22	9	0	4	3	1	12	0	6	G13D	22	1,448	336
ESCA	2	0	2	0	3	32	9	1	1	9	0	0	1	0	1	G13R	58	26	123
STAD	8	0	1	0	2	98	44	6	5	5	2	1	3	0	1	G13S	9	34	5
SIAD	0	0	0	0	0	201	42	3	3	12	1	0	1	0	0	G13V	21	23	71
COAD	0	2	0	0	11	2,216	582	46	50	86	25	8	44	0	18	Q61E	1	7	5
READ	0	0	0	0	2	861	254	28	33	38	23	5	36	0	4	Q61H	17	318	208
LIHC	0	0	1	0	1	13	6	3	0	5	0	0	9	0	0	Q61K	84	38	1,046
CHOL	0	0	0	0	0	467	37	18	3	5	7	6	18	0	1	Q61L	87	64	323
GBC	0	0	0	0	2	110	11	1	0	3	2	0	2	0	1	Q61P	2	15	38
PAAD	0	0	0	0	0	5,172	65	130	2	110	1	1	2	0	5	Q61R	186	90	1,439
KIRC	1	0	1	0	1	9	6	3	0	1	1	0	2	0	2	A146G	0	0	0
BLCA	111	24	71	0	13	96	11	9	5	8	4	1	22	0	4	A146V	0	42	1
PRAD	2	3	33	0	2	79	26	9	0	3	7	0	1	0	0	A146E/D	0	0	0
OV	0	0	1	0	1	240	11	3	0	2	0	0	15	0	2	A146P	0	9	1
UCEC	0	0	1	0	3	291	64	6	2	17	6	1	9	0	6	A146S	0	0	0
CESC	21	0	2	0	1	69	9	5	0	2	0	0	4	0	1	A146T	1	111	8
SKCM	2	5	25	1	4	27	3	8	2	16	127	93	1,692	7	67	Others	127	812	272
ALL	0	0	1	0	3	140	75	3	15	32	195	89	61	3	10	Total	927	17,741	5,082
AML	0	0	1	0	0	168	82	34	6	55	589	306	333	0	63	<b>New patients/year (USA)</b>			
CML	0	1	0	0	1	32	7	0	10	24	87	26	4	0	13	cancer	K: G12C	K: G12D	
PCM	0	0	1	0	2	49	26	56	14	33	16	27	155	0	11	LUAD	6,537	3,590	

GBM, glioblastoma multiforme; LGG, lower grade glioma; HNSC, head and neck squamous cell carcinoma; THCA, papillary thyroid carcinoma; THCAA, anaplastic thyroid carcinoma; THCAF, follicular thyroid carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; BRCA, breast invasive carcinoma; ESCA, esophageal carcinoma; STAD, stomach adenocarcinoma; SIAD, small intestine adenocarcinoma; COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; LIHC, liver hepatocellular carcinoma; CHOL, cholangiocarcinoma; GBC, gallbladder carcinoma; PAAD, pancreatic adenocarcinoma; KIRC, kidney renal clear cell carcinoma; BLCA, bladder urothelial carcinoma; PRAD, prostate adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; CESC, cervical squamous carcinoma and endocervical adenocarcinoma; SKCM, skin cutaneous melanoma; ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CML, chronic myeloid leukemia; PCM, plasma cell myeloma.

**Supplementary Table 3. Ras mutation-specific patterns in cancer.** Data are all collated from COSMIC v85 and where indicated converted in patient numbers using ACS 2018 cancer incidence statistics.