METHODS

## Data

Whole genome sequencing (WGS), whole exome sequencing (WXS), and targeted gene panel sequencing data was collected of colorectal adenocarcinoma (COADREAD), lung adenocarcinoma (LUAD), multiple myeloma (MM), pancreatic adenocarcinoma (PAAD), and skin cutaneous melanoma (SKCM). WXS and WGS data were downloaded from cBioportal [23550210, 22588877], which included the relevant TCGA projects and other smaller studies. In addition, data from ICGC for pancreatic cancer [21436628], melanoma and colorectal cancer were also included. Panel data were retrieved from the GENIE consortium [CIT]. GENIE data are an aggregation of several different panels raging from ~30 genes to ~600. *KRAS* was included in all the libraries. A detailed list of all cancer studies can be found in **supplementary table #.**

## Hypermutated samples cutoff

Some of the samples in COADREAD had 5 to 10 times more mutations than the average due to microsatellite instability (MSI). A Gaussian mixed model was used to find the optimal cutoff based on available WGS and WXS data (**supplementary figure #)**. The top 17% and 21% of samples were considered hypermutants in WGS and WXS, respectively. The same 17% cutoff was applied to the gene panel data. Hypermutants were not excluded from the identification of mutational signatures as they Signature 6 is caused by MSI. we filter out samples with less than 40 mutations.

Macintosh HD:Users:giorgio:Dropbox:hms:june_analysis:cooc_mutex_driver:data:cbioportal:exomeonly:figureForThePaper:supplementary:mutationDistr_exome_genomes.pdf

## Tissue gene expression filter

A conservative filter of tissue-specific gene expression was used to confidently remove genes not expressed in the tissues of study. Normal tissue gene expression data was gathered from the GTEx Portal (12/03/2018) and The Human Protein Atlas (HPA, 12/03/2018) (Uhlén et al. 2015), and tumor expression data was collected from MMRF-CoMMpass (01/14/2019), TCGA-COADREAD, TCGA-LUAD, TCGA-PAAD, and TCGA-SKCM (Cancer Genome Atlas Research Network et al. 2013). **Supplementary Table JHCST1** indicates the number of samples per tissue in the GTEx and tumor gene expression data. A gene was considered “expressed” in a tissue if it met at least one of the following criteria: 1) a median expression level of at least 1 TPM across all samples of the tissue in GTEx, 2) indicated as expressed at at least 1 TPM in the HPA data set for the tissue, 3) expressed with a median level of 1 RESM in the corresponding tumor RNA-sequencing data. **Supplementary Figure JHCS1** indicates the number of genes with at least one mutation that were removed from each tumor type according to this filter.

## Identifying mutational signatures

The genome-wide mutations of a sample can be deconvolved into mutational signatures that represent endogenous or exogenous mutagenic processes [﻿23945592]. Single nucleotide variants (SNVs) from exomes or genomes were divided into 96 types, according to the 6 mutations of a pyrimidine (C>A, C>G, C>T and T>A, T>C, T>G) and the 16 possible combinations of 3’ and 5’ adjacent bases. The MATLAB implementation of Non-Negative Matrix Factorization (NMF) algorithm, SigProfiler [23945592], is used to discover the underlying mutational patterns that are common across tumors. Mutational signatures are discovered separately for each tumor type and optimal number of signatures was determined based on silhouette width and frobenius error.

The spectrum of the signatures discovered by NMF were matched to the COSMIC catalog [23945592]. 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 the COSMIC catalog to include these additional signatures. In particular, NMF results yielded multiple subtypes of Signature 7 as reported previously in Refs [bioarxiv 322859v1, 28467829]. In SKCM we also found a signature that is dominantly C>A that is not a subtype of Signature 7 but is due to indirect consequences of UV exposure named Signature 38 [bioarxiv: 322859v1]. Three versions of the signature associated to POLE mutations, Signature 10, were discovered which is in agreement with Ref [bioarxiv: 322859v1]. These 3 types of POLE signatures differed in the dominant part of the mutational spectrum (C>A, C>T or C>G). In LUAD, a signature that has mutations of type C[C>A]N and T[C>A]N that is attributed to 8-oxo-guanine [bioarxiv: 322859v1] was found and also included in our extended catalog. One signature that we discovered COADREAD did not have a good match in any specific signature in literature, although it resembles SBSA in Ref [bioarxiv: 416800] and Signature 34 and 41 in Ref [bioarxiv: 322859v1]. We used this signature as it was discovered by our own NMF rather than matching it to a previously reported version, because the results from different studies was not in good agreement. This new signature that we call Signature N is not contributing to KRAS mutations. We also found 3 signatures that are likely to be artifacts, which also did not contribute to KRAS mutations.

Once the signatures in each sample is determined by NMF we removed signatures that acquire very small exposure in a given sample and recalculated the exposures using Non-Negative Least Squares (NNLS) algorithm. Details of this calculation can be found in the Supplementary Note of [30988514]. The results of this calculation is shown in Supplementary Fig X.

## Probability of *RAS* mutations from mutational signatures

For each sample harboring either a *KRAS* or *NRAS* 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. For example, consider a lung cancer sample harboring a *KRAS* G12C, created by C[C>A]A, and 2 signatures were identified in this sample, 60% Signature 1 and 40% Signature 4. The frequency of the C[C>A]A type in Signature 1 and 4 is 0.0066and 0.0461, respectively (rescaled a range of 0 to 1 as 0.1251686 and 0.8748314). Therefore, the probability of a single *KRAS* G12C mutation is approximately 7 times more likely to originate from Signature 4 than Signature 1. Given the relative contribution of the signatures called, the probability that the G12C mutation was caused by Signature 1 or 4 was:

In conclusion, the chance that the *KRAS* G12C mutation originated from ageing or smoke is 18% and 82%, respectively. The sum across all samples with the same cancer and the same *KRAS* mutation provided an estimate of the global probability of occurrence of a specific *RAS* mutation (**Figure 2B**).

## Co-occurrence Analysis

A Fisher’s exact test was used to check for co-mutational partners between *RAS* alleles other mutated genes with at least two mutated samples. One-tailed tests were performed to check for significant co-occurrence of mutations (right tail) and significant mutual exclusivity (left tail).

## Linear regression of prevalence versus occurrence probability

For each sample harboring a *KRAS* or *NRAS* mutation, the proportion of genome-wide mutations that were identical to the *RAS* mutation was calculated. For example, consider a SKCM tumor with an *NRAS* mutation and a total of 1,500 SNVs. 200 mutations were of the type A[G>T]T (*NRAS* Q61K), 300 of the type G[T>C]T (*NRAS* Q61R), and 400 G[T>A]T (*NRAS* Q61L). The relative probability of randomly generating each of these mutations was therefore:

In Figure 2D, the mean across samples ± X% confidence interval was reported on the y-axis while the frequency of mutation within the tumor type was reported on the x-axis.

Regression lines were calculated by amino acid position with a linear model weighted by the inverse of the standard deviation of each point (thus, more weight was given to those mutations that had a lower variability of their estimate).

## Defining allele-specific synthetic lethality

The most recent data release from the Cancer Dependency Map (DepMap) project [28753430, 29083409] was downloaded on 03/24/2018. For all analyses, the batch-effect corrected, CERES-adjusted dependency scores were used. Thus far, the DepMap project has screened 27 COADREAD, 76 LUAD, 16 MM, 25 PAAD, and 34 SKCM derived cell lines, of which 13 COADREAD, 26 LUAD, and 23 PAAD had a *KRAS* mutation, 8 SKCM had a *NRAS* mutation, and 9 MM had either a *KRAS* or *NRAS* mutation (only including mutations in hotspot codons; **Supplementary Figure JHC1A**). Only alleles found in at least three cell lines of a cancer were used in the following analysis, which restricted the analysis to the most common *KRAS* alleles of COADREAD, LUAD, and PAAD cell lines. The metric for genetic dependency was a measure of the depletion of the CRISPR sgRNA for the given target gene [29083409]; thus a negative score indicated a greater dependency.

The DepMap project identified “pan-essential” genes as those uniformly lethal across cell lines [29083409]; these genes were used to inform the determination of synthetic lethal genes in this analysis. To select genes that showed synthetic lethality in the cell lines of at least one *KRAS* allele, only genes that had an mean dependency score within or less than (i.e. stronger dependency) one standard deviation of the pan-essential genes in cell lines of a *KRAS* allele were tested further. To statistically determine allele-specificity, an ANOVA was used to test for variation in dependency scores in cell lines separated by *KRAS* allele. Due to the relatively low number of cell lines per allele, a *p*-value of 0.10 was used as the threshold.

## The functional enrichment of synthetic lethal genes

WebGestalt was used to identify cellular functions and pathways enriched in genes determined to show allele-specific genetic dependencies [28472511]. The R package ‘WebGestaltR’ was used to interface with the web application. The following databases were used: "geneontology\_Biological\_Process","geneontology\_Cellular\_Component", "geneontology\_Molecular\_Function", "pathway\_KEGG", "pathway\_Panther", "pathway\_Reactome", "pathway\_Wikipathway", "community-contributed\_Hallmark50”. The enriched gene-sets (FDR < 0.10) were manually sorted into the groups presented in the paper.

To specifically investigate the genetic dependencies with key oncogenic signaling pathways, the following KEGG gene-sets from MSigDB C2 curated collection were used: "APOPTOSIS", "CELL\_CYCLE", "JAK\_STAT\_SIGNALING\_PATHWAY", "MAPK\_SIGNALING\_PATHWAY", "MTOR\_SIGNALING\_PATHWAY", "NOTCH\_SIGNALING\_PATHWAY", "P53\_SIGNALING\_PATHWAY", "PHOSPHATIDYLINOSITOL\_SIGNALING\_SYSTEM", "VEGF\_SIGNALING\_PATHWAY", "WNT\_SIGNALING\_PATHWAY" [27899662, 21546393].

## Allele-specific co-mutation subnetwork

Significantly disrupted protein-protein interaction (PPI) subnetworks were identified using HotNet2 (version 1.2.1) [25501392] with samples divided by *RAS* mutation or wild-type *RAS* and cancer type, where mutational frequency of the gene in the subset was used the value for “heat.” The underlying PPI consisted of the binary and complex interactions obtained from the High-quality INTeractomes (HINT) database (11/24/2018) [22846459] and non-expressed genes were removed for each cancer. The simulations were performed over 100 permutated networks with 1,000 heat permutations and a diffusion constant of 0.50. The threshold for the minimum edge weight 𝛿 of the returned subnetwork was determined to be the smallest value with the most number of significant (FDR adjusted *p*-value < 0.10) subnetwork sizes 𝑘. The largest significant subnetwork size was then selected as the threshold.

The allele specificity of a co-mutating subnetwork was determined using a one-sided Fisher’s exact test to compare the frequency of having at least one mutation in a gene in the subnetwork between samples with the *KRAS* allele and all others of the same cancer. In the text, *p­*-values are presented though all had Benjamini-Hochberg FDR adjusted *q*-values below 0.10 [“Controlling the false discovery rate: a practical and powerful approach to multiple testing”].