PNAS
www.pnas.org

Supplementary Information for

Mutation bias within oncogene families is related to proliferation-specific codon usage.

Hannah Benisty[1], Marc Weber[1], Xavier Hernandez-Alias[1], Martin H. Schaefer[1,2,*], Luis Serrano[1,3,4,*]

*Correspondence: martin.schaefer@ieo.it (M.H.S.), luis.serrano@crg.eu (L.S.)

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Datasets S1 to S6

Supplementary Information Text

MATERIAL AND METHODS

Data Sources

Paralogs Ensembl

To define gene families, we retrieved information regarding protein sequence similarity and family membership from Ensembl. Ensembl's family classification often contains outliers that have a much lower sequence similarity compared with the other proteins of the same family. As a larger variation in amino acid sequence implies greater variation of biochemical function, and because in our study we aimed to focus on differences in codon usage , we sought to remove these outliers. We therefore applied another, more stringent filter; for each family we computed the similarity distribution of all members to a consensus member. We then removed all family members that had a lower similarity than the mean similarity minus one standard deviation. We only considered families with at least three members. With this approach, we retrieved 71 gene families that had high amino acid sequence similarity and contained at least three family members. Eight of these gene families contained at least one cancer gene (see below) and 63 families contained no cancer genes (cancer and non-cancer gene families).

TCGA

Mutation data was obtained from The Cancer Genome Atlas (TCGA). We retrieved somatic mutations in coding regions for 20 types of cancer: Bladder Urothelial Carcinoma (BLCA), Breast invasive carcinoma (BRCA), Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), Colon adenocarcinoma (COAD), Glioblastoma multiforme (GBM), Head and Neck squamous cell carcinoma (HNSC), Kidney renal clear cell carcinoma, Kidney renal papillary cell carcinoma (KIRP), Acute Myeloid Leukemia (LAML), Brain Lower Grade Glioma (LGG), Liver hepatocellular carcinoma (LIHC), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), Pancreatic adenocarcinoma (PAAD), Pheochromocytoma and Paraganglioma (PCPG), Prostate adenocarcinoma (PRAD), Skin Cutaneous Melanoma (SKCM), Stomach adenocarcinoma (STAD), Thyroid carcinoma (THCA), and Uterine Corpus Endometrial Carcinoma (UCEC). Together this comprised a set of 5,960 samples.

Cancer gene catalogue

We considered cancer driver genes to be those genes that had a significant ($q<0.01$) number of non-silent mutations in at least 1 out of 21 cancer types in 4,742 patients as defined by Lawrence *et al*. (1).

Coding sequences

The coding sequences of *H. sapiens* were downloaded from the Consensus CDS (CCDS) project (ftp://ftp.ncbi.nlm.nih.gov/pub/CCDS/), release 2016/09/08. In the case of non-

cancer genes, one unique canonical coding sequence was arbitrarily chosen for each protein based on Uniprot mapping to the CCDS. For those genes belonging to the selected cancer gene families, the canonical coding sequence was chosen according to the corresponding canonical protein as defined in Uniprot.

GO gene sets

Gene ontology was downloaded as a MySQL dump of the amiGO database, release 2017/01, and human gene annotations were downloaded from the amiGO database, release 2018/01/04. We defined GO gene sets as follows: i) for each GO term, we retrieved all descendant GO terms (with any kind of relationship type) and assigned all associated genes; ii) we selected all GO terms that had a minimal distance to the root "biological process" term shorter than or equal to 3, and at least 30 associated genes, resulting in a total of 708 gene sets. Note that there is a lot of overlap between these GO gene sets, with a protein appearing on average in 44 sets.

Computational analysis

Codon usage PCA

We applied principal component analysis (PCA) to the relative synonymous codon frequencies (2) of all individual human coding sequences. Note that, contrary to other studies such as Gingold *et al*. (3), we defined our PCA projection based on the codon usage distribution of individual genes, and not of gene sets. By doing so, our projection is independent from the gene ontology annotations. In addition, PCA based on the average codon usage of gene sets may suffer a bias due to the fact that GO gene sets are highly overlapping. Thus, the codon usage of a specific gene may contribute to data points of several gene sets, which may in turn distort the real variation in codon usage. When computing the PCA of individual genes, we first excluded single codon families (AUG and UGG). For coding sequences that lacked codons of a specific family (6.7% of total), we imputed values with the average codon frequency across all genes. We applied the PCA projection to the GO gene sets by computing the mean relative codon frequencies of all genes in the set.
All the source code and data for this analysis is freely available at the Github repository: https://github.com/webermarcolivier/codon_usage_oncogenes.

Low and high proliferation cancer

Low and high proliferating tumors were defined by computing the average Ki67 expression per cancer type. We then divided the log10 (Ki67 expression) space in equally sized ranges ([2.23,2.99] and [3.00,3.76]). Cancer types with low proliferation included PRAD, PCPG, THCA, KIRP, KIRC, and LIHC. Cancer types with high proliferation included PAAD, SKCM, LUAD, BRCA, UCEC, BLCA, LUSC, HNSC, CESC, COAD, and STAD.
To test for a link between RAS mutation status and Ki67 expression on a sample level, we selected those tumors in which multiple patients had mutations in at least two RAS isoforms (SKCM, UCEC, STAD, LUAD, COAD, and THCA). Samples were stratified by the non-silent mutation status of the different RAS isoforms. The binomial test was performed asking how likely it is to see eight times (total number of inter-group comparisons within the same cancer type).

Quantification of tRNA expression

tRNAseq mapping was performed using a specific pipeline for tRNAs (4). The basic pipeline was adapted to paired-end sequencing data. Moreover, given that hydro-tRNA-seq yields short sequences, all reads over 10 nt were included after BBDuk adapter trimming from the BBMap toolkit [v38.22] (https://sourceforge.net/projects/bbmap): k-mer=10 (allowing 8 at the end of the read), Hamming distance=1, length=10–50 bp and Phred>25. Isoacceptors were quantified as reads per million (RPM), summing up all reads mapping to isodecoders that share the same anticodon. Ambiguous reads mapping to genes of different isoacceptors were discarded. In the human reference genome GRCh38 (Genome Reference Consortium Human Reference 38, GCA_000001405.15), a total of 856 nuclear tRNAs and 21 mitochondrial tRNAs were annotated with tRNAscan-SE [v2.0] (5). Trimmed FASTQ files were then mapped using a specific pipeline for tRNAs (4). In short, an artificial genome was first generated by masking all annotated tRNA genes and adding pre-tRNAs (i.e. tRNA genes with 3' and 5' genomic flanking regions) as extra chromosomes. Upon mapping to the artificial genome with Segemehl [v0.3.1] (6), reads that mapped to the tRNA-masked chromosomes and to the tRNA flanking regions were filtered out to remove non-tRNA reads and unmature-tRNA reads, respectively. After this first mapping step, a second library was generated by adding 3' CCA tails and removing introns from tRNA genes. All 100% identical sequences of this so-called mature tRNAs were clustered to avoid redundancy. Next, the subset of filtered reads from the first mapping was aligned against the clustered mature tRNAs using Segemehl [v0.3.1] (6). Mapped reads were then realigned with GATK IndelRealigner [v3.8] (7) to reduce the number of mismatching bases across all reads. For quantification, isoacceptors were quantified as RPM. To increase the coverage of anticodon-level quantification, we considered all reads that map unambiguously to a certain isoacceptor even though they ambiguously map to different isodecoders (i.e. tRNA genes that differ in their sequence but share the same anticodon). Ambiguous reads mapping to genes of different isoacceptors were discarded. The code can be found in the GitHub repository: https://github.com/hexavier/tRNA_mapping.

Relative codon usage

We correlated the relative codon usage of $KRAS_{WT}$ and $KRAS_{HRAS}$, which was calculated by dividing each codon value by the sum of the codon values of a given amino acid. For the purposes of calculating the fold change, we added a pseudo count to all values (+1). For the families, we calculated this fold change in codon usage by comparing the usage of the most mutated gene to the usage of the least mutated gene from the same family (Dataset S5). We performed a sequence alignment using TranslatorX (8) to compare only those codons that align between the two sequences. Finally, we calculated relative codon usage and fold change in the same way as we did for the comparison between $KRAS_{WT}$ and $KRAS_{HRAS}$.

Differential tRNA anticodon abundance

We excluded anticodons for which there are no corresponding tRNA genes (Arg^{GCG}, Gly^{ACC}, His^{ATG}, Leu^{GAG}, Phe^{AAA}, Thr^{GGT} and Val^{GAC}) based on the tRNA gene prediction of the *H. sapiens*

genome GRCh38/hg38 using tRNAscan-SE[63]. Next, we calculated the relative anticodon abundance by dividing the RPM value of each anticodon by the sum of all anticodon RPM values for a given amino acid. Differential relative expression analysis was performed using t-tests, where p-values were FDR-corrected and a cutoff of $q<0.05$ was used.

Differential codon adaptation

We calculated the codon weights (Dataset S4) used in SI Appendix, Fig. S9, S10 and S11 based on wobble base pair interaction rules (9). Similar to tRNA abundances analysis in the section above, we calculated the relative codon adaptation weights by dividing the weight value of each codon by the sum of all codons for a given amino acid. Differential relative codon weight analysis was performed using t-tests, where p-values were FDR-corrected and a cutoff of $q<0.05$ was used.

Statistical analyses

For hypothesis testing, we performed two-sided Student t-tests, two-sided Wilcoxon-Mann-Whitney tests and one-sided binomial tests. In the differential expression analyses, an FDR correction was used to account for multiple testing.

Sample preparation and experimental procedures

Cell lines

In this study we used the HeLa, HEK293 and fibroblast BJ/hTERT (used in Gingold *et al*.(3), kindly provided by Disa Tehler) cell lines. Cells were maintained in a humidified atmosphere at 37°C and 5% CO_2. Cells were grown in DMEM 4.5 g/L Glucose with UltraGlutamine media supplemented with 10% Tet-free FBS (Clontech) and 1% penicillin/streptomycin.

Expression vector design

$KRAS_{HRAS}$ was obtained from the pBABE-Puro-KRas* vector (Addgene#46745). As the amino acid sequence of $KRAS_{HRAS}$ is 100% identical to KRAS, this vector expresses the human KRAS protein from a chimeric cDNA sequence derived primarily from HRAS codons (93% of HRAS synonymous codons) but with a KRAS 3' tail (where the main differences between RAS proteins lie). Details of the nucleotide and amino acid sequences can be found in SI Appendix, Fig. S14 and Dataset S6. For conditional-gene overexpression experiments, $KRAS_{WT}$ and $KRAS_{HRAS}$ were cloned into a modified version of the XLone-GFP vector (10) (Addgene#96930). The modification consisted of replacing the promoter of XLone-GFP with a bidirectional TRE3G promoter (Clontech), which allows the simultaneous expression of both KRAS genes. We used a FLAG tag on $KRAS_{WT}$ and a 3xHA tag on $KRAS_{HRAS}$ to distinguish them by size. We also use two additional expression constructs as controls. One construct had the two tags swapped (i.e. FLAG-$KRAS_{HRAS}$ and 3xHA-$KRAS_{WT}$), and the other construct had the position of the genes with respect to the promoter swapped. Each vector was co-transfected with the pCYL43 (11) plasmid containing the PiggyBac transposase in different cell lines. Cells were selected with blasticidin (HeLa: 5 µg/mL, HEK293: 15 µg/mL, BJ/hTERT: 5 µg/mL). Gene expression

was induced with doxycycline (HeLa: 100 ng/mL, HEK293: 12 ng/mL, BJ/hTERT: 500 ng/mL). We used the same approach to simultaneously express $RAC1_{WT}$ and $RAC1_{RAC3}$. The codon usage of $RAC1_{RAC3}$ exactly corresponds to the codons used by RAC3 except for those codons corresponding to different amino acids (see SI Appendix, Fig. S15 and Dataset S6). The cDNA chimeric sequence for $RAC1_{RAC3}$ was synthesized by Integrated DNA Technologies.

Serum starvation assay

BJ/hTERT, HEK293 and HeLa cells were grown in starvation media (1% Tet-free FBS) or non-starvation media (10% Tet-free FBS) for 48 hours. The expression of both $KRAS_{WT}$ and $KRAS_{HRAS}$ was measured after inducing with doxycycline overnight.

Flow cytometry

BJ/hTERT, HEK293 and HeLa cells were seeded in 6-well plates and maintained in starvation media (1% Tet-free FBS) or non-starvation media (10% Tet-free FBS) for 48 hours. To measure the cell cycle state of the cells, culture medium was supplemented with 1 µg/mL Hoechst 33342 (H3570, ThermoFisher Scientific) and cells were incubated for 1hour at 37ºC. Next, cells were trypsinized and resuspended with 350 µL of media containing Hoechst 33342. At least 10,000 cells were analyzed by flow cytometry (BD LSR II). FlowJo software was used for gating and analysis.

Cell lines assay

Established HeLa, HEK293 and BJ/hTERT cells were induced with doxycycline and the expression was measured after incubating overnight.

Cell growth

The cells were seeded at a density of 25,000 cells per well in a 12-well plate and the counts were performed with Countess cell counting chamber slides and the Countess automated cell counter (ThermoFisher). Counts were carried out every 24 hours.

mRNA quantification

RNA isolation was performed with the RNeasy kit (Qiagen). $KRAS_{WT}$ and $KRAS_{HRAS}$ transcript abundances were quantified by RT-qPCR (Power SYBR Green RNA-to-CT 1-Step Kit, ThermoFisher). Primers for FLAG-$KRAS_{WT}$ amplification: forward 5'-CAAGGACGACGATGACAAG-3' and reverse 5'-GAGAATATCCAAGAGACAGGTT-3'. Primers for 3xHA-$KRAS_{HRAS}$ amplification: forward 5'-CCTGACTATGCGGGCTATC-3' and reverse 5'-GGGTCGTATTCGTCCACAA-3'. For the expression of the constructs in which the tags were swapped, we used the following primers: FLAG-$KRAS_{HRAS}$ forward 5'-CAAGGACGACGATGACAAG-3' and reverse 5'-GGGTCGTATTCGTCCACAA-3', and 3xHA-$KRAS_{WT}$ forward 5'-CCTGACTATGCGGGCTATC-3' and reverse 5'-GAGAATATCCAAGAGACAGGTT-3'. As both genes were in the same expression

cassette, for each sample, the Ct values for $KRAS_{WT}$ were normalized to $KRAS_{HRAS}$, ΔCt = ($CtKRAS_{WT}$-$CtKRAS_{HRAS}$) and represented as $2^{-\Delta Ct}$. Primers for FLAG-$RAC1_{WT}$ amplification: forward 5'-CAAGGACGACGATGACAAG-3' and reverse 5'-GTCCAGCTGTATCCCATAAGC -3'. Primers for 3xHA-$RAC1_{RAC3}$ amplification: forward 5'-CCTGACTATGCGGGCTATC-3' and reverse 5'- GATGTACTCTCCGGGGAAG -3'. As both genes were in the same expression cassette, for each sample, the Ct values for $KRAS_{WT}$ were normalized to the $KRAS_{HRAS}$, ΔCt = ($CtRAC1_{WT}$-$CtRAC1_{RAC3}$) and represented as $2^{-\Delta Ct}$.

Quantitative protein blots

Cells were lysed using the M-PER buffer (ThermoFisher) supplemented with anti-proteases. Protein concentration was measured using the BCA Protein Assay Kit (Pierce). Equal amounts of each sample were mixed with 1x Laemmli buffer and boiled for 5 min. Samples were separated using 12% polyacrylamide gels (BioRad). Transfer was performed using the iBlot system (Invitrogen). Membranes were treated with Li-COR Odyssey blocking buffer for 1 hour at room temperature, then incubated with primary antibody (1:1,000) in 0.2% Tween-20/Li-COR odyssey blocking buffer overnight at 4°C. Following three 5 min washes in TBS-T, the membrane was incubated with secondary antibody (1:10,000) in 0.2% Tween-20/Li-COR Odyssey blocking buffer for 45 min at room temperature. Following three 5 min washes in TBS-T, the membrane was scanned using the Li-COR Odyssey Imaging System. We used the following primary antibodies: anti-pan-RAS (Abcam, ab52939), anti-FLAG (Sigma, F3165), anti-HA (Sigma, 11583816001), and anti--actin (Sigma, A2228). These were respectively detected using goat anti-rabbit (Abcam, ab216773), goat anti-mouse (ab216772) or goat anti-mouse (Abcam, ab216776) IgG antibody conjugated to an IRdye at 800CW and 680CW. Visualization and quantification were performed using ImageJ and Image Studio Lite (LI-COR).

Hydro-tRNA sequencing

Total RNA from HEK293, BJ/hTERT and HeLa cells were extracted using the miRNeasy Mini kit (Qiagen). For each sample, 20 µg of total RNA was treated following the hydro-tRNAseq protocol (12). Briefly, total RNA was resolved on a 15% Novex TBE-urea gel (ThermoFisher) and fragments from 60–100 nt were size-selected and subjected to alkaline hydrolysis (10 mM sodium carbonate and 10 mM sodium bicarbonate) for 10 min at 60ºC. This was followed by dephosphorylation and rephosphorylation. Similarly to conventional small RNA-seq library preparation, samples were adaptor-ligated, reverse-transcribed, and PCR-amplified for 14 cycles. Sequencing of the resulting cDNA was performed using an Illumina HiSeq 2500 platform in 50 bp paired-end format. Raw data for HEK293 and HeLa have been deposited in the ArrayExpress database (13) at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-8144. BJ/hTERT sequencing data have been generated by the same group using the same protocol and have been previously published (14).

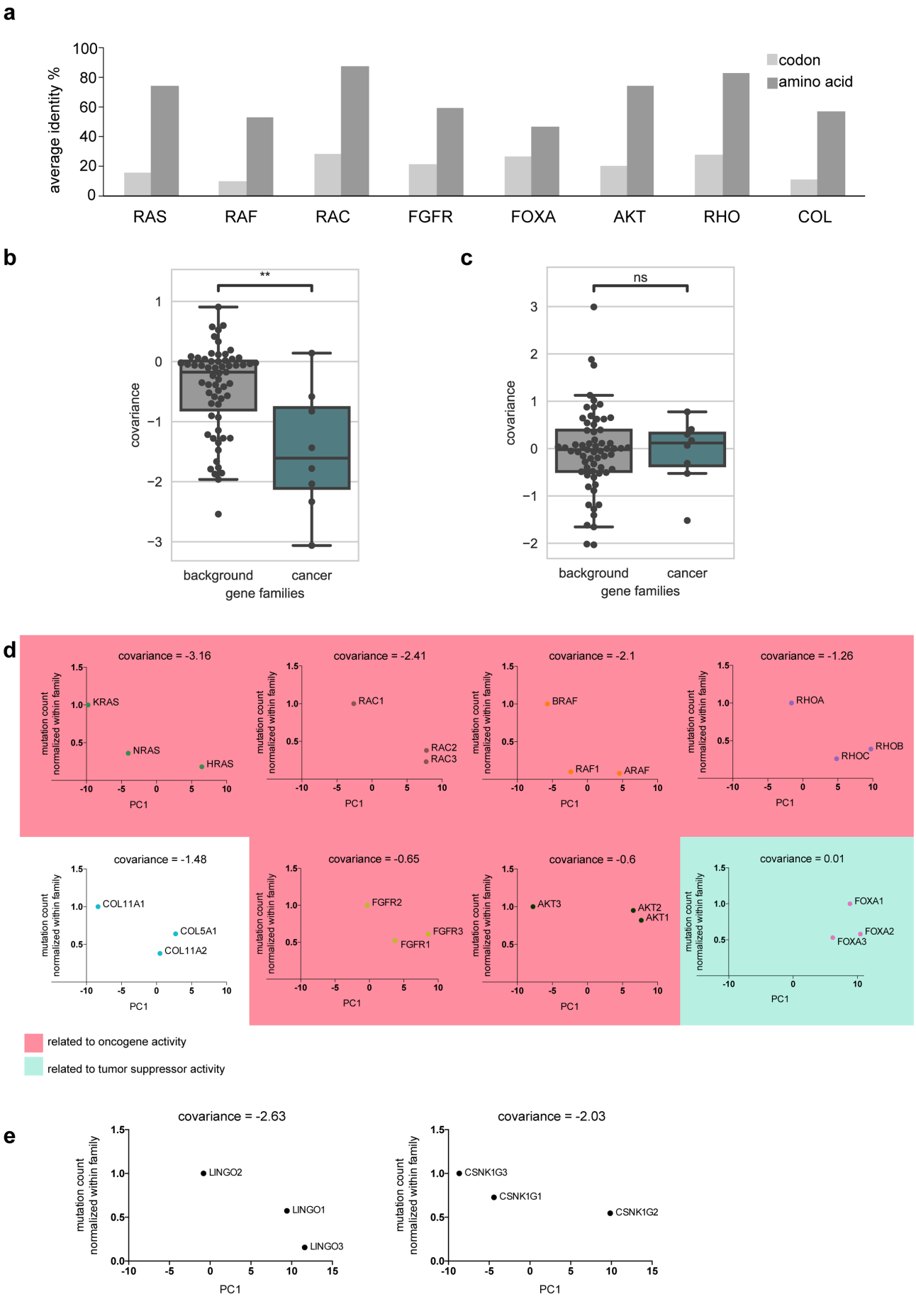

a
average identity %
codon
amino acid
RAS
RAF
RAC
FGFR
FOXA
AKT
RHO
COL
b
covariance
**
background
cancer
gene families
c
covariance
ns
background
cancer
gene families
d
covariance = -3.16
KRAS
NRAS
HRAS
covariance = -2.41
RAC1
RAC2
RAC3
covariance = -2.1
BRAF
RAF1
ARAF
covariance = -1.26
RHOA
RHOB
RHOC
covariance = -1.48
COL11A1
COL5A1
COL11A2
covariance = -0.65
FGFR2
FGFR3
FGFR1
covariance = -0.6
AKT3
AKT2
AKT1
covariance = 0.01
FOXA1
FOXA2
FOXA3
mutation count normalized within family
PC1
related to oncogene activity
related to tumor suppressor activity
e
covariance = -2.63
LINGO2
LINGO1
LINGO3
covariance = -2.03
CSNK1G3
CSNK1G1
CSNK1G2

Fig. S1. Association between codon usage and mutation frequency in genes from eight different families. a Average codon and amino acid identity of gene families. **b** Distribution of covariance of non-synonymous mutation count normalized within family and PC1. Covariance is significantly more negative for cancer gene families than for the background non-cancer related gene families (**p<0.008, W.M.W. test). In particular, the covariance is negative for seven (RAS, RAF, RHO, RAC, FGFR, AKT and COL) out of eight cancer gene families. **c** Distribution of covariance of synonymous mutation count normalized within family and PC1. There is no significant difference in covariance between cancer gene families and the background non-cancer related gene families (ns, not significant; p<0.76, W.M.W. test). **d** Distribution of covariance of non-synonymous mutation count normalized within family and PC1 for the eight cancer gene families. Families are ordered from lowest to highest covariance. Pink corresponds to genes related to oncogenes and cyan corresponds to tumor suppressors (p<0.007, binomial test). The distinction between these two classes of genes was performed using OncodriveROLE (15). The family COL, which has no classification in OncodriveROLE, is not highlighted with any color. **E** Covariance of non-synonymous mutation count normalized within family and PC1 for the two background families LINGO and CSNK1G. The covariance of these families is similar to the covariance of the cancer gene families RAS, RAF and RAC, denoting the possible role of LINGO2 and CSNK1G3 in cancer.

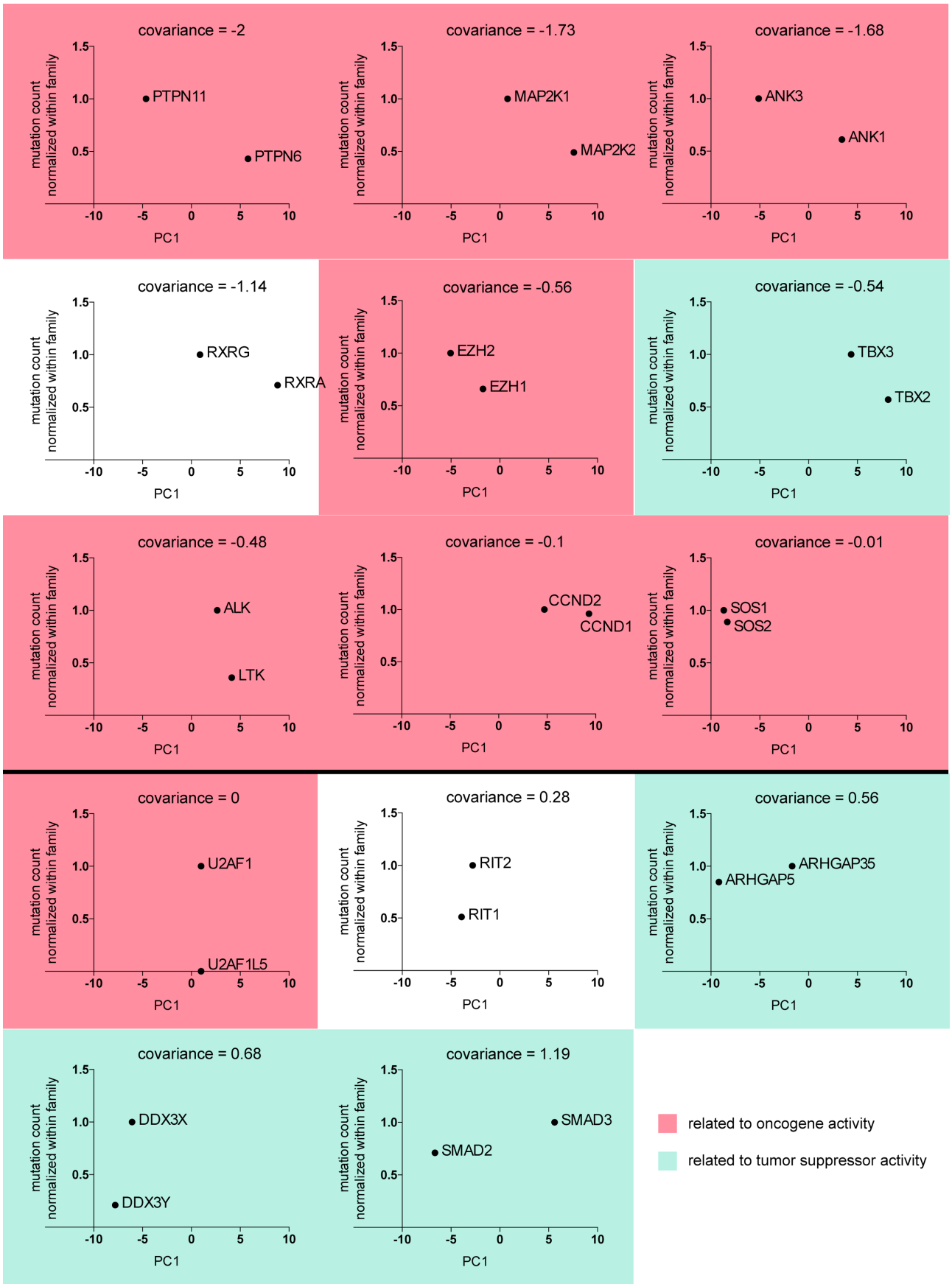

covariance = -2
PTPN11
PTPN6
covariance = -1.73
MAP2K1
MAP2K2
covariance = -1.68
ANK3
ANK1
covariance = -1.14
RXRG
RXRA
covariance = -0.56
EZH2
EZH1
covariance = -0.54
TBX3
TBX2
covariance = -0.48
ALK
LTK
covariance = -0.1
CCND2
CCND1
covariance = -0.01
SOS1
SOS2
covariance = 0
U2AF1
U2AF1L5
covariance = 0.28
RIT2
RIT1
covariance = 0.56
ARHGAP35
ARHGAP5
covariance = 0.68
DDX3X
DDX3Y
covariance = 1.19
SMAD3
SMAD2
mutation count normalized within family
PC1
related to oncogene activity
related to tumor suppressor activity

Fig. S2. Covariance of cancer gene families. Distribution of covariance of non-synonymous mutation count normalized within family and PC1 for 14 cancer gene families composed of two members. Families are ordered from lowest to highest covariance. Pink corresponds to genes related to oncogenes and cyan corresponds to tumor suppressors ($p<0.019$, binomial test). The distinction between these gene classes was performed using OncodriveROLE (15). The families with no classification in OncodriveROLE are not highlighted with any color.

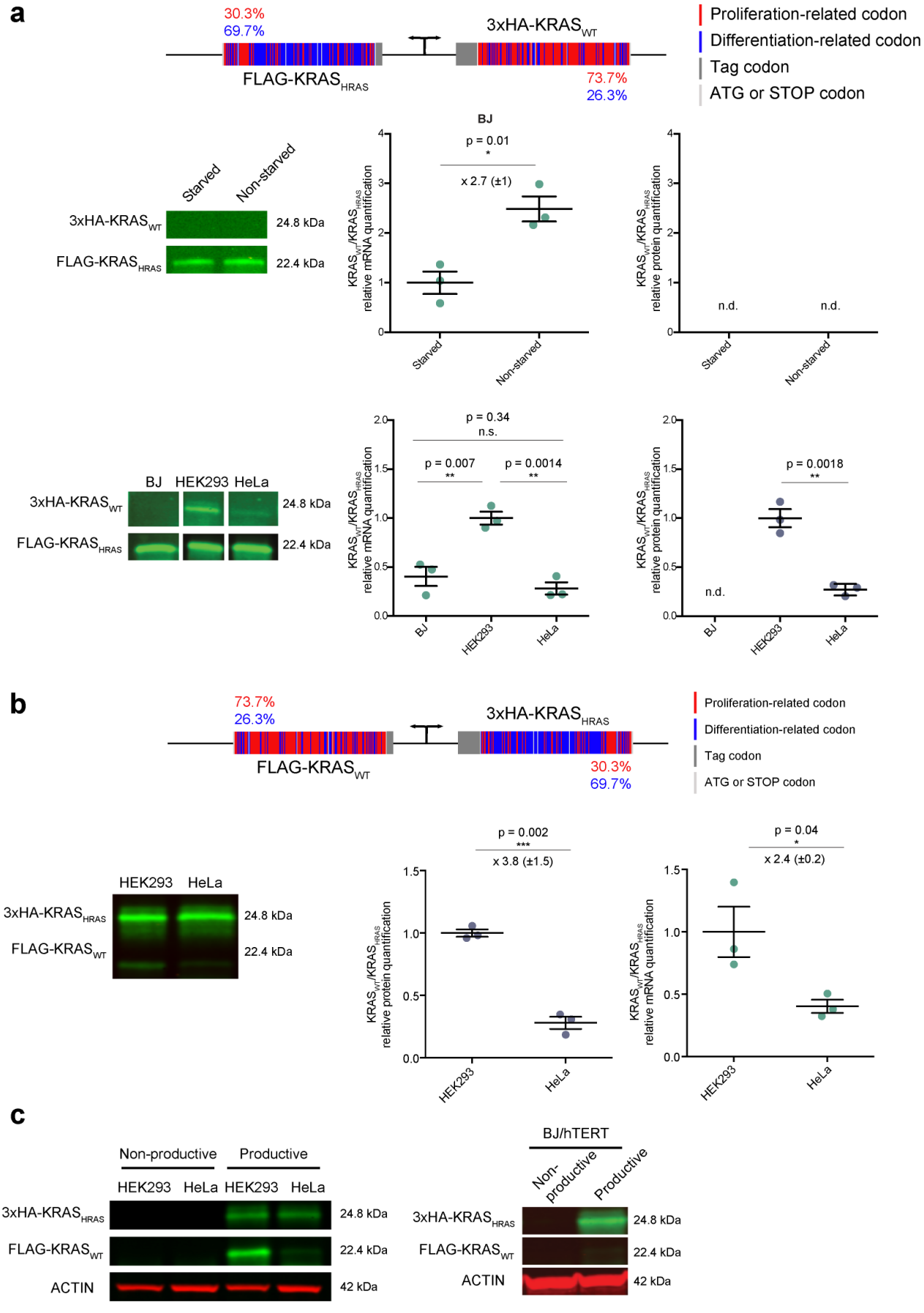

a
30.3%
69.7%
FLAG-KRAS$_{HRAS}$
3xHA-KRAS$_{WT}$
73.7%
26.3%
Proliferation-related codon
Differentiation-related codon
Tag codon
ATG or STOP codon
Starved
Non-starved
3xHA-KRAS$_{WT}$
FLAG-KRAS$_{HRAS}$
24.8 kDa
22.4 kDa
BJ
p = 0.01
*
x 2.7 (±1)
KRAS$_{WT}$/KRAS$_{HRAS}$ relative mRNA quantification
KRAS$_{WT}$/KRAS$_{HRAS}$ relative protein quantification
n.d.
BJ HEK293 HeLa
p = 0.34
n.s.
p = 0.007
**
p = 0.0014
**
p = 0.0018
**
b
73.7%
26.3%
FLAG-KRAS$_{WT}$
3xHA-KRAS$_{HRAS}$
30.3%
69.7%
HEK293 HeLa
p = 0.002
***
x 3.8 (±1.5)
p = 0.04
*
x 2.4 (±0.2)
c
Non-productive
Productive
HEK293 HeLa HEK293 HeLa
ACTIN
42 kDa
BJ/hTERT
Non-productive
Productive

Fig. S3. Controls of the bidirectional expression vector in which the tags are swapped, the position of the genes in relation to the promoter are switched, and the plasmid is non-productive. a The protein and mRNA levels of KRASWT and KRASHRAS with the tags swapped in BJ/hTERT, HEK293 and HeLa cells. Quantification of the proteins in BJ/hTERT cells is not possible due to low fluorescence signal (not detected, n.d.), possibly due to minimal integration of the expression cassette; therefore, we report transcript levels. In starved and non-starved BJ/hTERT cells, the KRASWT/ KRASHRAS transcript ratio has the same fold change irrespective of the tag's position (see main Fig. 2c). The KRASWT/KRASHRAS transcript and protein ratios vary in the different cell lines but, similarly, is not affected by the position of the tag (see main Fig. 3a and b). b The protein and mRNA levels of KRASWT and KRASHRAS with the position of the genes in relation to the promoter switched. Expression in HEK293 and HeLa cells is compared. The KRASWT/KRASHRAS mRNA and protein ratios vary in the different cell lines irrespective of gene position. Error bars represent SEM of three independent experiments (n.s., not significant; *$p<0.05$; **$p<0.01$; ***$p<0.001$; unpaired t test). c Comparison of the expression between the productive and non-productive (without a translation initiation site and ATG) expression cassette. No KRASWT or KRASHRAS expression is observed when translation is suppressed.

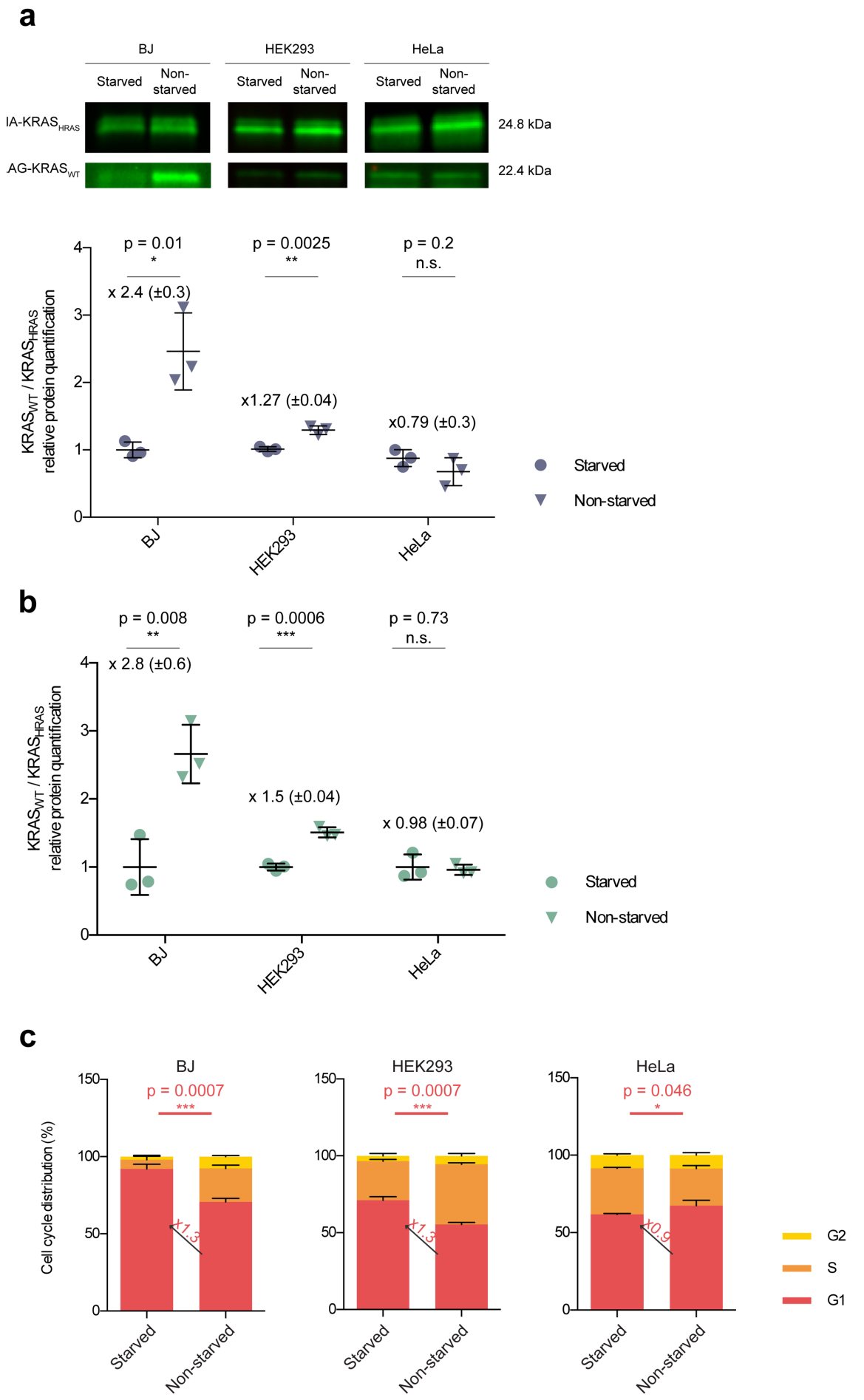

a
BJ
HEK293
HeLa
Starved
Non-starved
IA-KRAS$_{HRAS}$
AG-KRAS$_{WT}$
24.8 kDa
22.4 kDa
p = 0.01
*
p = 0.0025
**
p = 0.2
n.s.
x 2.4 (±0.3)
x1.27 (±0.04)
x0.79 (±0.3)
KRAS$_{WT}$ / KRAS$_{HRAS}$ relative protein quantification
Starved
Non-starved
b
p = 0.008
**
p = 0.0006
***
p = 0.73
n.s.
x 2.8 (±0.6)
x 1.5 (±0.04)
x 0.98 (±0.07)
c
BJ
HEK293
HeLa
p = 0.0007
***
p = 0.0007
***
p = 0.046
*
Cell cycle distribution (%)
x1.3
x1.3
x0.9
G2
S
G1

Fig. S4. Comparison of state-specific changes in the $KRAS_{WT}$/$KRAS_{HRAS}$ ratio between BJ/hTERT, HEK293 and HeLa cells. a Western blot analysis of the levels of $KRAS_{WT}$ and $KRAS_{HRAS}$ in starved and non-starved BJ/hTERT, HEK293 and HeLa cells. The $KRAS_{WT}$/$KRAS_{HRAS}$ protein ratio increases significantly from the quiescent to the proliferative state in BJ/hTERT and HEK293 cells, but not in HeLa cells. **b** The same observation applies to the transcript levels. **c** Cell cycle measurements by flow cytometry with starved and non-starved cells. BJ/hTERT and HEK293, but not in HeLa cells display a significant increase in the number of cells arrested when starved. This suggests that starvation in HeLa cells does not impede cell proliferation and as such, we do not observe protein and transcript level changes between starved and non-starved conditions. Quantifications in **a** and **b** are normalized to the starved condition of each cell line. Error bars represent SEM of three independent experiments (ns, not significant; *p<0.05; **p<0.01; ***p<0.001; unpaired Student t test).

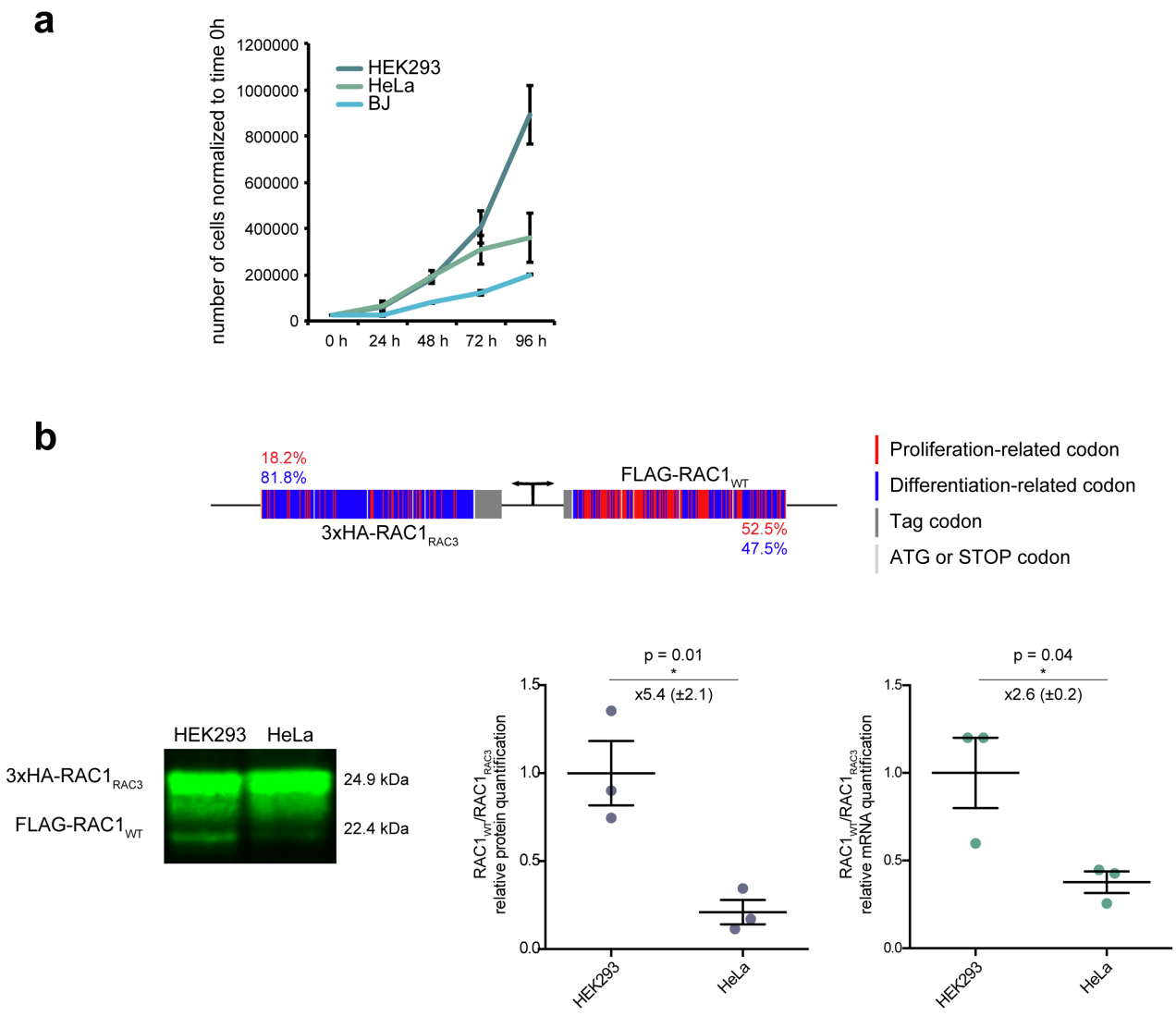


Fig. S5. Growth curves of the three cell lines and codon usage-related expression of RAC1. **a** We measured cell proliferation of HEK293, HeLa and BJ/hTERT cells by counting the cells every 24 hours over 96 hours. **b** Representation of the expression cassette that expresses $RAC1_{WT}$ and $RAC1_{RAC3}$. The $RAC1_{WT}/RAC1_{RAC3}$ protein and mRNA ratios change similarly as for the $KRAS_{WT}/KRAS_{HRAS}$ ratios in both HEK293 and HeLa cells. Error bars represent SEM of three independent experiments (*p<0.05; unpaired Student t test).

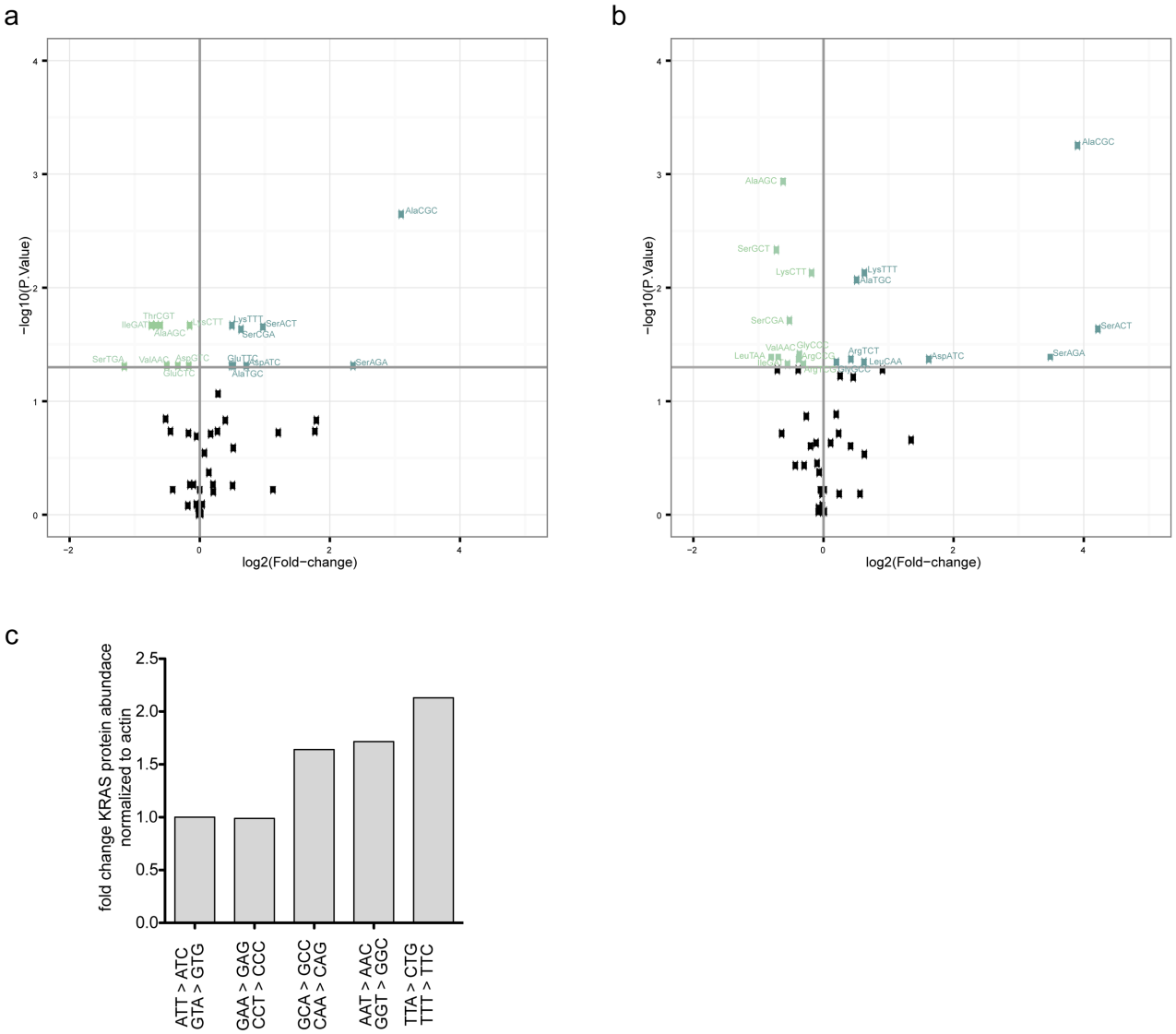
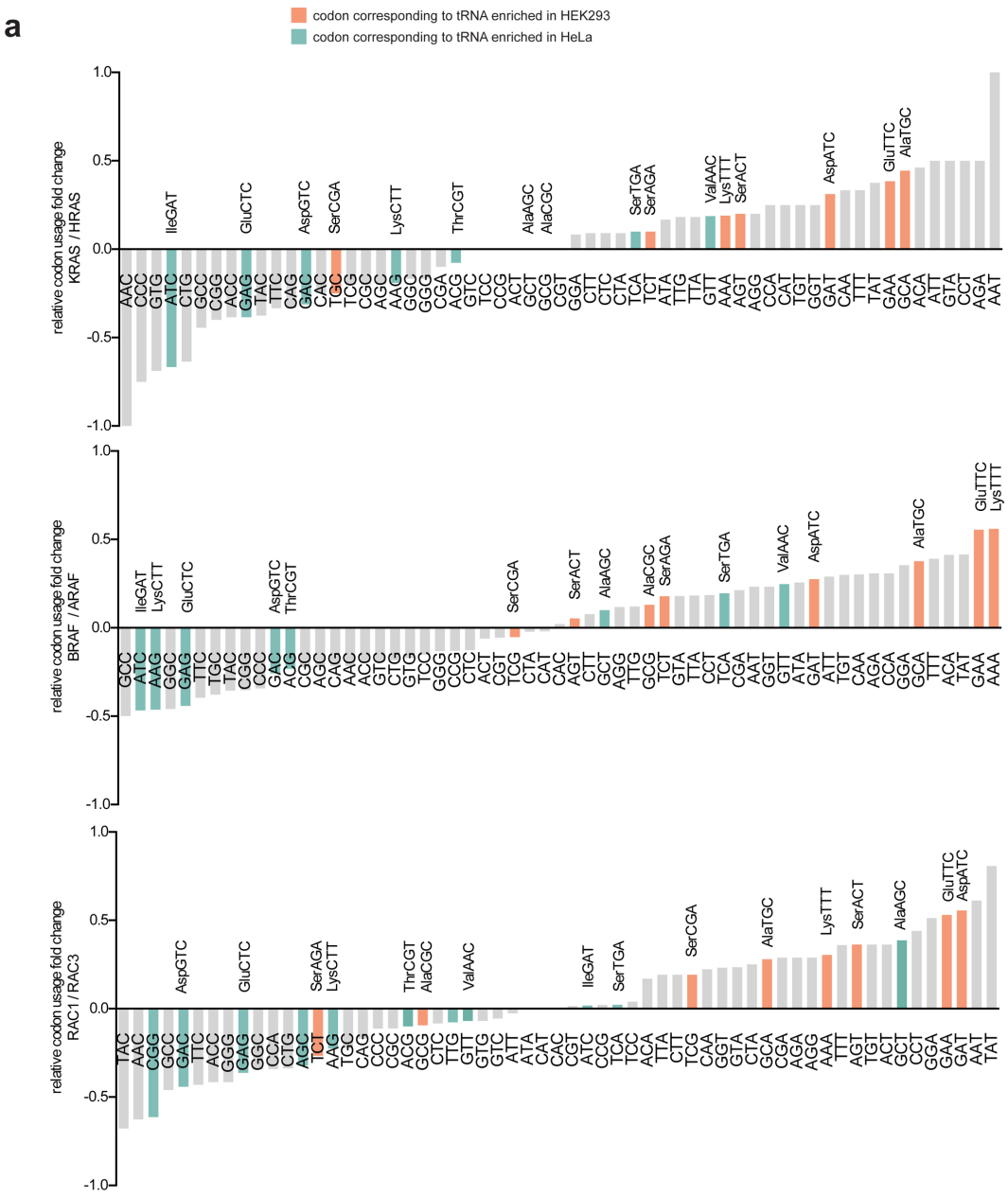


Fig. S6. Differential expression of tRNA in HEK293, HeLa and BJ/hTERT. a Volcano plot showing the log2 fold change of the relative tRNA differential expression between HEK293 and HeLa cells. **b** Volcano plot showing the log2 fold change of the relative tRNA differential expression between HEK293 and BJ/hTERT. Differential relative expression analysis for a and b was performed using the t-test, where p-values were FDR-corrected, with $p<0.05$ as a cutoff. **c** Immunoblot data taken from Lampson *et al*. (16) were quantified where KRAS codons were progressively converted.


a
codon corresponding to tRNA enriched in HEK293
codon corresponding to tRNA enriched in HeLa
relative codon usage fold change KRAS / HRAS
relative codon usage fold change BRAF / ARAF
relative codon usage fold change RAC1 / RAC3
1.0
0.5
0.0
-0.5
-1.0

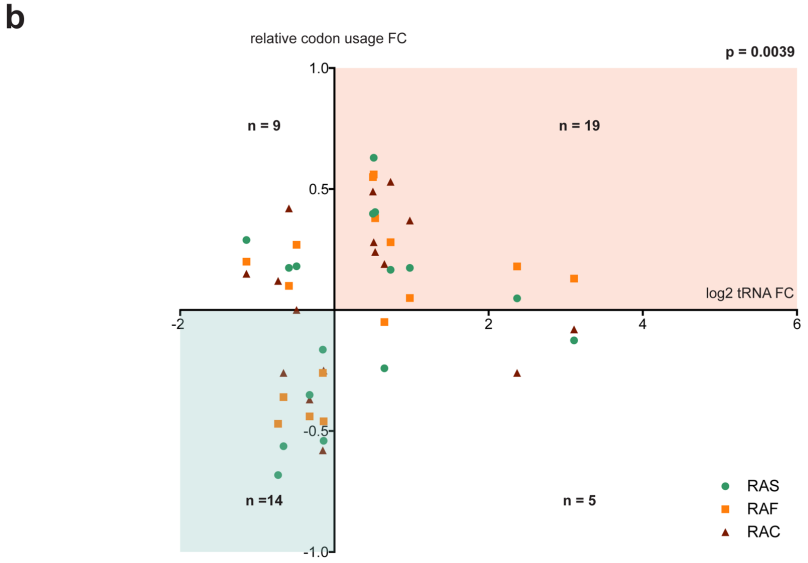

b
relative codon usage FC
p = 0.0039
n = 9
n = 19
n =14
n = 5
log2 tRNA FC
RAS
RAF
RAC

Fig. S7. Relative codon usage between the most frequently and the least frequently mutated gene. a Fold change of the relative codon abundance (pseudocount +1) between the most frequently and the least frequently mutated gene for the three families displaying the highest negative covariance (RAS, RAF, RAC). Codons corresponding to tRNAs differentially expressed between HEK293 and HeLa cells are highlighted. **b** Scatter plot representing relative codon usage change versus log2 fold change of tRNA expression. Generally, the tRNAs enriched in HEK293 cells match the codons enriched in KRAS, BRAF and RAC1, whereas the tRNAs enriched in HeLa cells match the codons enriched in HRAS, ARAF and RAC3 ($p<0.0039$, binomial test).

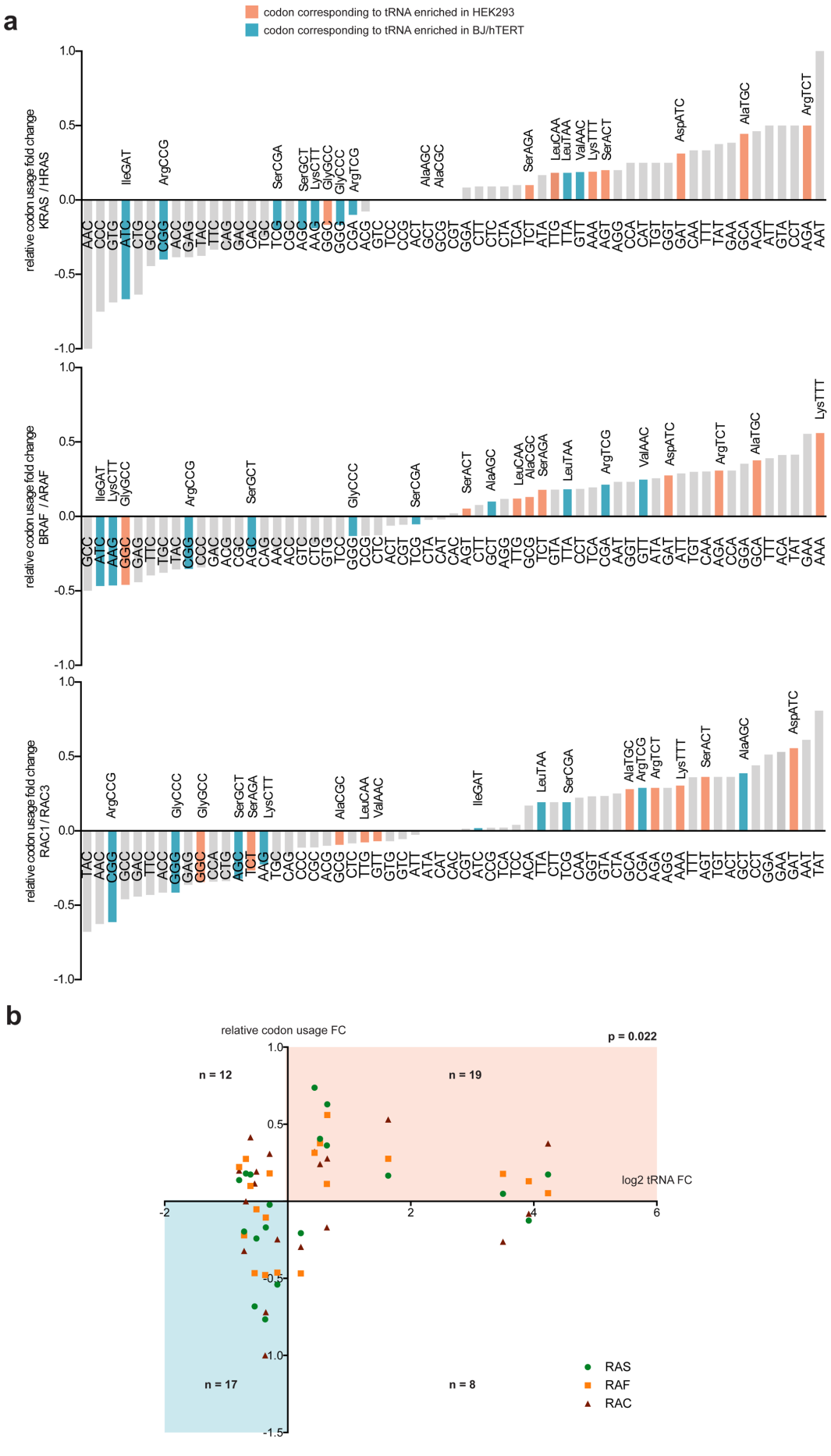

a
codon corresponding to tRNA enriched in HEK293
codon corresponding to tRNA enriched in BJ/hTERT
relative codon usage fold change KRAS / HRAS
relative codon usage fold change BRAF / ARAF
relative codon usage fold change RAC1 / RAC3
b
relative codon usage FC
log2 tRNA FC
p = 0.022
n = 12
n = 19
n = 17
n = 8
RAS
RAF
RAC

Fig. S8. Relative codon usage between the most frequently and the least frequently mutated gene. a Fold change of the relative codon abundance (pseudocount +1) between the most frequently and the least frequently mutated gene for the three families displaying the highest negative covariance (RAS, RAF, RAC). Codons corresponding to tRNAs differentially expressed between HEK293 and BJ/hTERT cells are highlighted. **b** Scatter plot representing the relative codon usage change versus log2 fold change of tRNA expression. Generally, the tRNAs enriched in HEK293 cells match the codons enriched in KRAS, BRAF and RAC1, whereas the tRNAs enriched in BJ/hTERT cells match the codons enriched in HRAS, ARAF and RAC3 ($p<0.022$, binomial test).

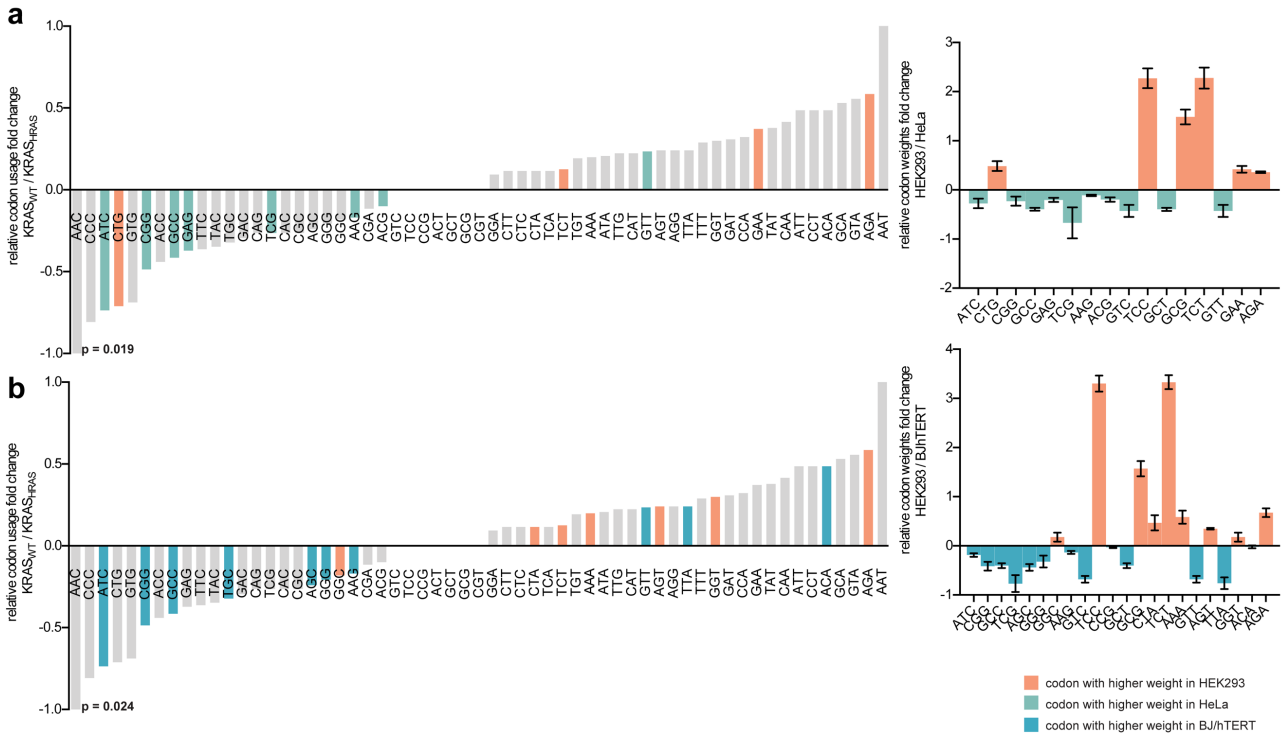
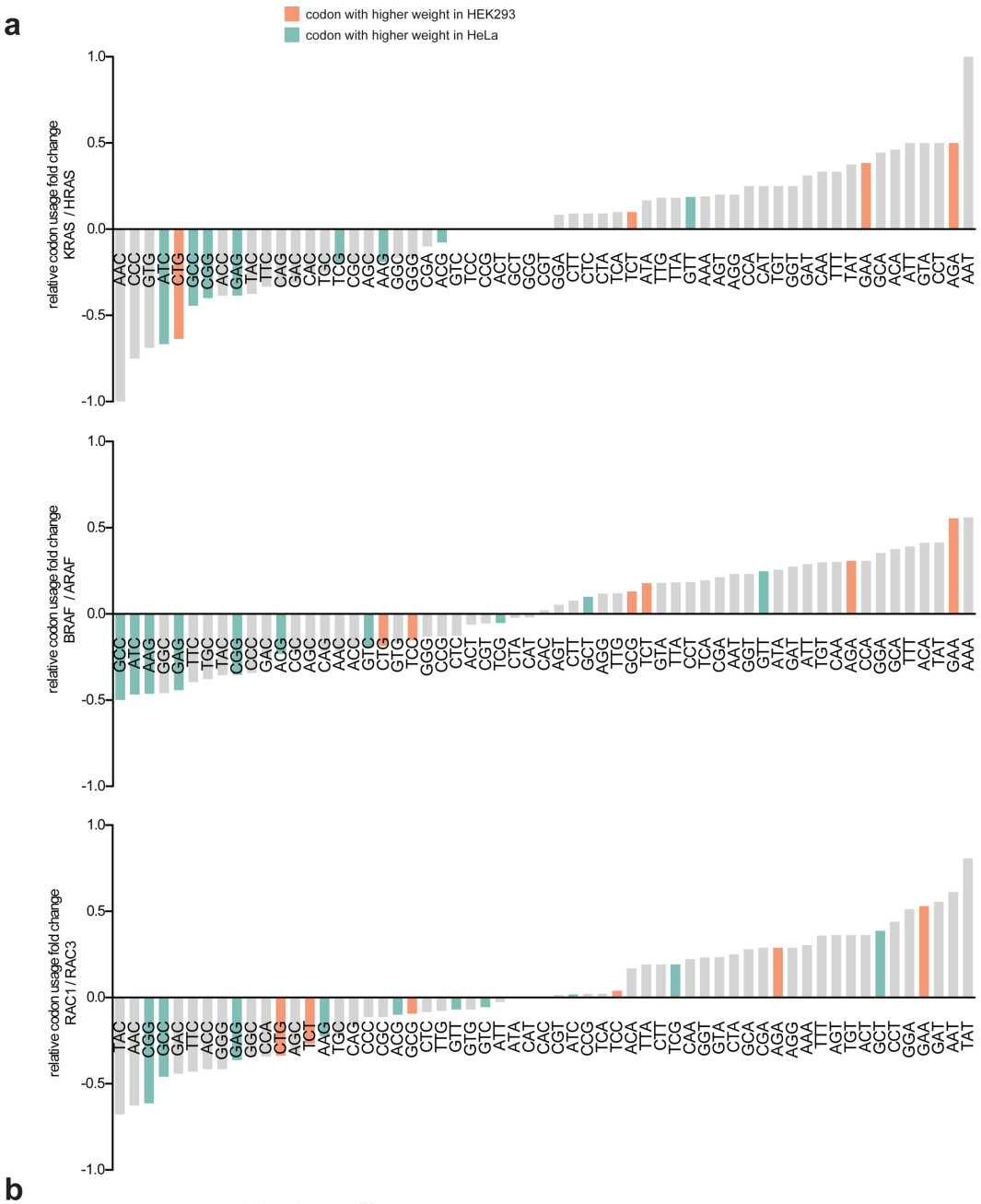


Fig. S9. Association of differentially weighted codons and relative codon usage of $KRAS_{WT}$ and $KRAS_{HRAS}$. a Log2 fold change of the relative codon usage (pseudocount +1) between $KRAS_{WT}$ and $KRAS_{HRAS}$. Codons corresponding to codons that are differentially weighted between HEK293 and HeLa cells are highlighted. The right panel represents the log2 fold change of relative codon weights between HEK293 and HeLa cells. Log2 fold change of the relative codon usage (pseudocount +1) between $KRAS_{WT}$ and $KRAS_{HRAS}$. Codons corresponding to codons that are differentially weighted between HEK293 and BJ/hTERT cells are highlighted. The right panel represents the log2 fold change of relative codon weights between HEK293 and BJ/hTERT cells. Error bars represent SEM of three independent experiments. Binomial tests (one-sided, p-values shown below plots) were performed in a and b by calculating the probability of the correct number of associations between relative codon weight and codon usage.


a
codon with higher weight in HEK293
codon with higher weight in HeLa
relative codon usage fold change
KRAS / HRAS
BRAF / ARAF
RAC1 / RAC3
1.0
0.5
0.0
-0.5
-1.0

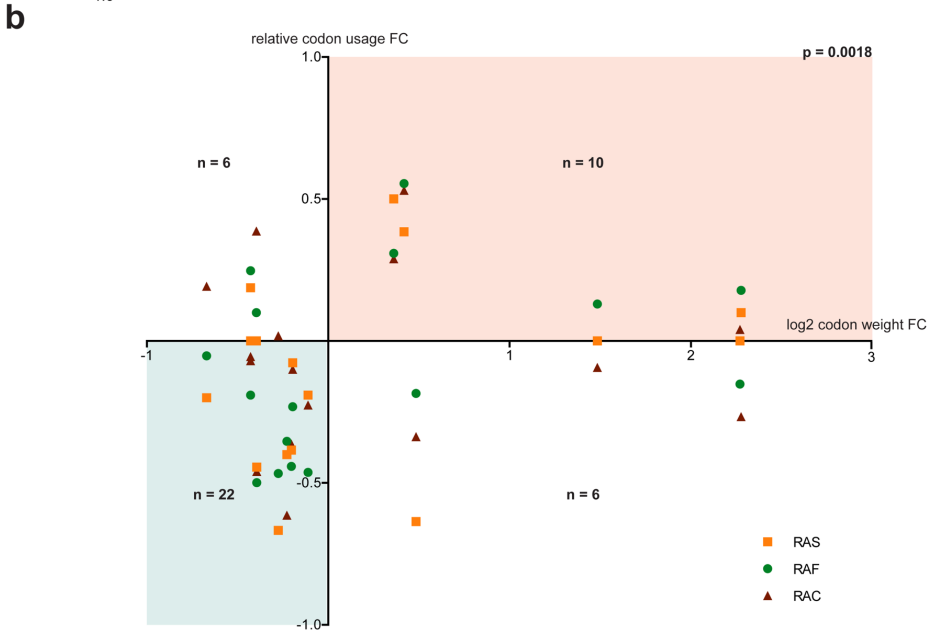

b
relative codon usage FC
log2 codon weight FC
p = 0.0018
n = 6
n = 10
n = 22
n = 6
RAS
RAF
RAC
1.0
0.5
-0.5
-1.0
-1
1
2
3

Fig. S10. Relative codon usage between the most frequently and the least frequently mutated gene. a Fold change of the relative codon abundance (pseudocount +1) between the most frequently and the least frequently mutated gene for the three families displaying the highest negative covariance (RAS, RAF, RAC). Codons corresponding to differentially weighted codons between HEK293 and HeLa cells are highlighted. **b** Scatter plot representing the relative codon usage change versus log2 fold change of codon weights. Generally, the codons enriched in HEK293 cells match the codons enriched in KRAS, BRAF and RAC1, whereas the codons enriched in HeLa cells match the codons enriched in HRAS, ARAF and RAC3 ($p<0.0018$, binomial test).

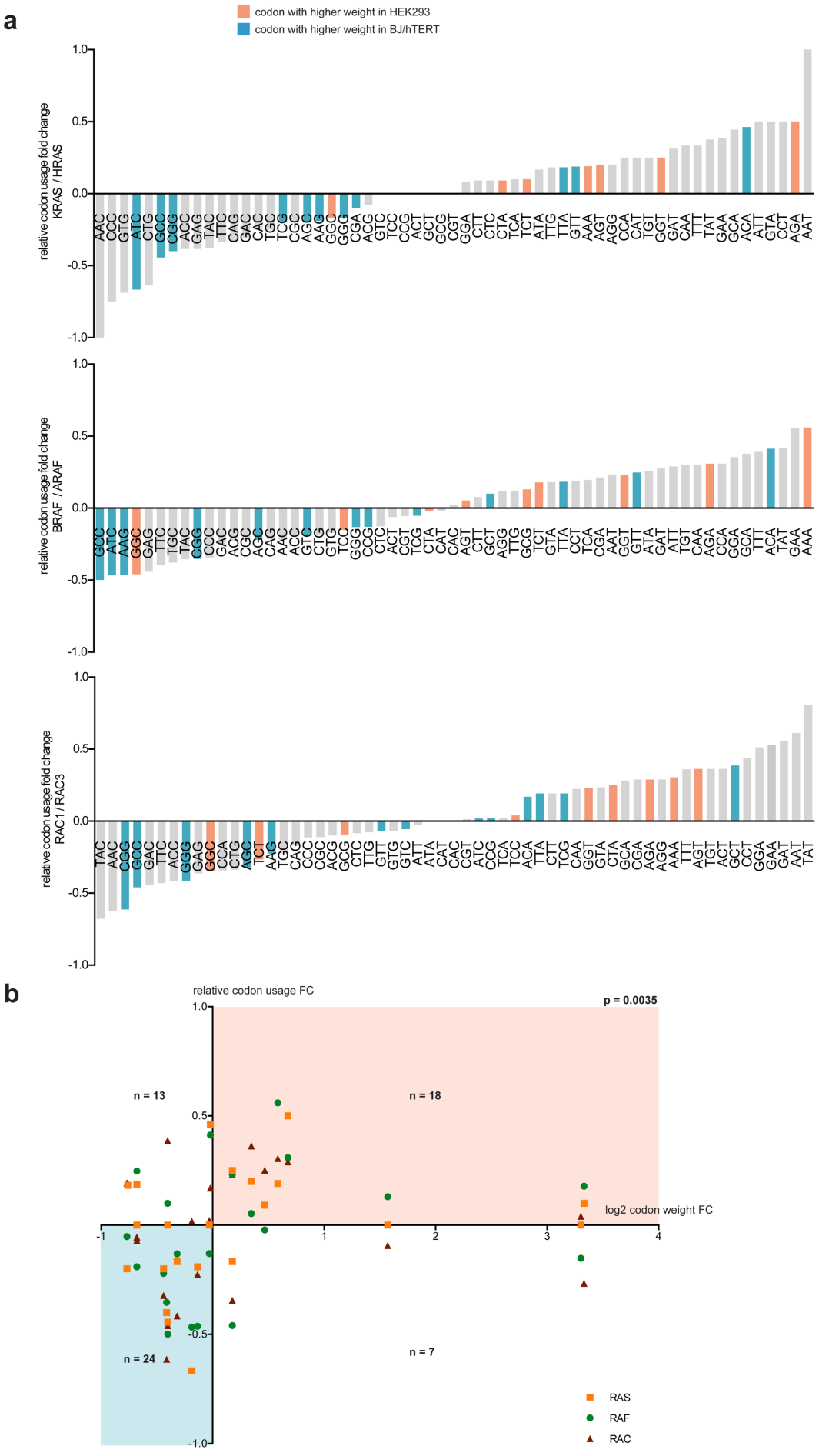

a
codon with higher weight in HEK293
codon with higher weight in BJ/hTERT
relative codon usage fold change
KRAS / HRAS
BRAF / ARAF
RAC1 / RAC3
1.0
0.5
0.0
-0.5
-1.0
b
relative codon usage FC
log2 codon weight FC
p = 0.0035
n = 13
n = 18
n = 24
n = 7
RAS
RAF
RAC

Fig. S11. Relative codon usage between the most frequently and the least frequently mutated gene. a Fold change of the relative codon abundance (pseudocount +1) between the most frequently and the least frequently mutated gene for the three families displaying the highest negative covariance (RAS, RAF, RAC). Codons corresponding to differentially weighted codons between HEK293 and BJ/hTERT cells are highlighted. **b** Scatter plot representing the relative codon usage change versus log2 fold change of codon weights. Generally, the codons enriched in HEK293 cells match the codons enriched in KRAS, BRAF and RAC1, whereas the codons enriched in BJ/hTERT cells match the codons enriched in HRAS, ARAF and RAC3 ($p<0.0035$, binomial test).

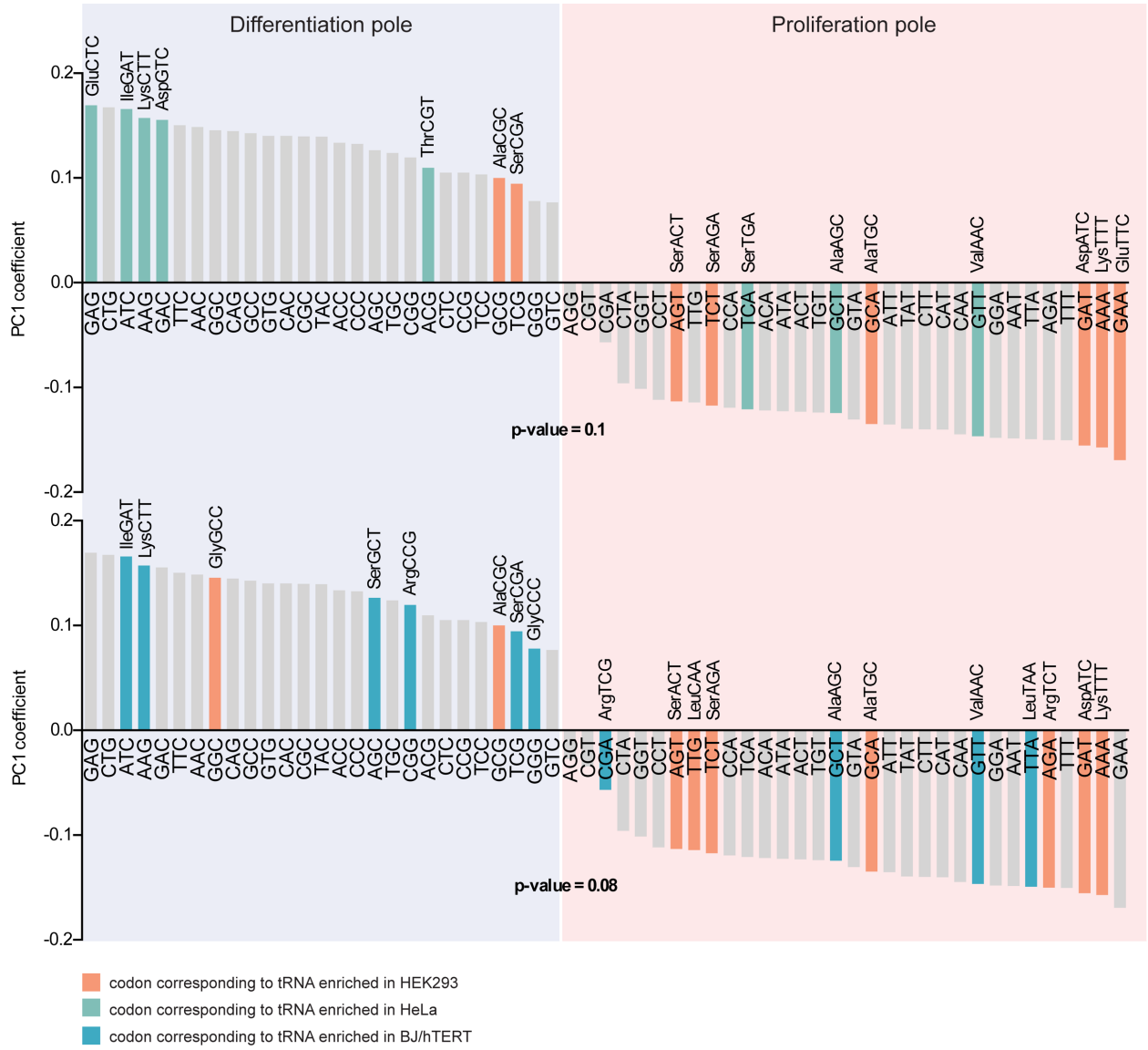


Fig. S12. Proliferation-related versus differentiation-related codons. Codons are ordered according to their first component coefficient (PC1). Note that the scale of the values is arbitrary, as only the relative values are important (direction of the vector in the multidimensional space). Negative values indicate a negative PC1 towards the proliferation pole, whereas positive values move towards the differentiation pole. Cognate codons corresponding to tRNAs differentially expressed between HEK293 and HeLa cells (upper) or between HEK293 and BJ/hTERT cells (lower) are highlighted. We test if the codons from the proliferation pole are preferentially associated to the tRNAs upregulated in HEK293 in comparison the HeLa and BJ/hTERT, and correspondingly, if the codons from the differentiation pole are preferentially associated to the downregulated tRNAs. Binomial tests (one-side, p-values are shown below plots) were performed by calculating the probability of the correct number of associations occurring between relative tRNA expression and expected PC1.

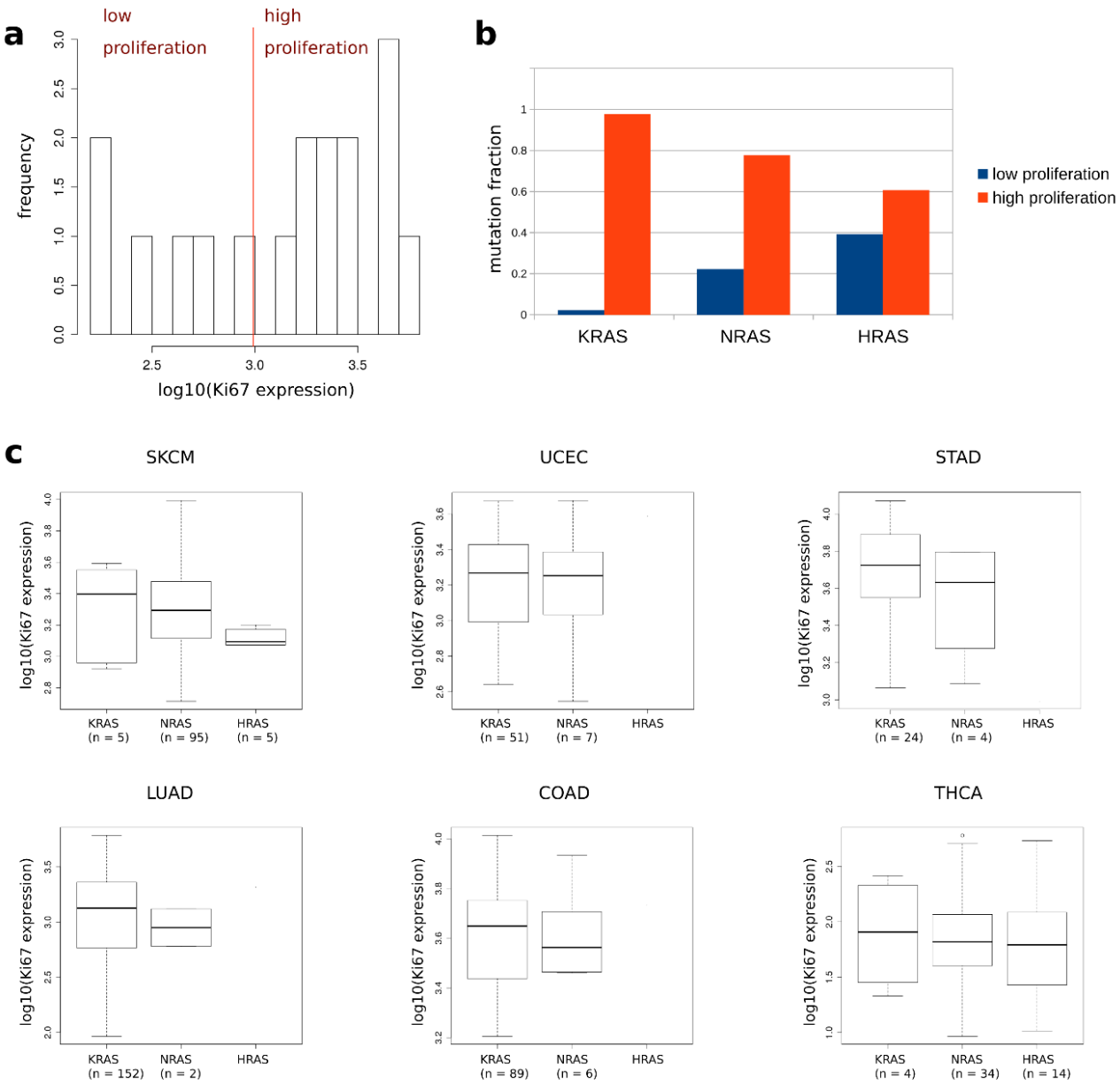


Fig. S13. RAS genes display different mutation frequencies in low and high proliferating cancer types. a Distribution of average Ki67 expression values across cancer types (left) enables splitting cancer types into low and high proliferation types (vertical red line, splits the Ki67 expression interval into equally sized halves). **b** The relative proportion of mutations in KRAS, HRAS and NRAS differs in low and high proliferation ($p<2.2e-16$, Chi-Squared test). **c** KRAS-mutated tumors have a higher median Ki67 expression compared to NRAS- or HRAS-mutated tumors ($p=0.008$, binomial test). Only boxes with more than one sample are shown.

```
CLUSTAL O(1.2.4) multiple sequence alignment


KRAS          ATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAAGAGTGCCTTGACG      60
KRASHRAS      ATGACGGAATATAAGCTTGTGGTGGTGGGCGCTGGAGGCGTGGGAAAGAGTGCCCTGACC      60
HRAS          ATGACGGAATATAAGCTGGTGGTGGTGGGCGCCGGCGGTGTGGGCAAGAGTGCGCTGACC      60
              ***** ******** ** ***** ** ** ** ** ** ** ** ********  ****

KRAS          ATACAGCTAATTCAGAATCATTTTGTGGACGAATATGATCCAACAATAGAGGATTCCTAC      120
KRASHRAS      ATCCAGCTGATCCAGAACCACTTTGTGGACGAATACGACCCCACTATAGAGGATTCCTAC      120
HRAS          ATCCAGCTGATCCAGAACCATTTTGTGGACGAATACGACCCCACTATAGAGGATTCCTAC      120
              ** ***** ** ***** ** ************** ** ** ** ***************

KRAS          AGGAAGCAAGTAGTAATTGATGGAGAAACCTGTCTCTTGGATATTCTCGACACAGCAGGT      180
KRASHRAS      CGGAAGCAGGTGGTCATTGATGGGGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGC      180
HRAS          CGGAAGCAGGTGGTCATTGATGGGGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGC      180
               ******* ** ** ******** ** ** ** ** ***** ** ** ** ** ** **

KRAS          CAAGAGGAGTACAGTGCAATGAGGGACCAGTACATGAGGACTGGGGAGGGCTTTCTTTGT      240
KRASHRAS      CAGGAGGAGTACAGCGCCATGCGGGACCAGTACATGCGCACCGGGGAGGGCTTCCTGTGT      240
HRAS          CAGGAGGAGTACAGCGCCATGCGGGACCAGTACATGCGCACCGGGGAGGGCTTCCTGTGT      240
              ** *********** ** *** ************** * ** *********** ** ***

KRAS          GTATTTGCCATAAATAATACTAAATCATTTGAAGATATTCACCATTATAGAGAACAAATT      300
KRASHRAS      GTGTTTGCCATCAACAACACCAAGTCTTTTGAGGACATCCACCAtTACAGGGAGCAGATC      300
HRAS          GTGTTTGCCATCAACAACACCAAGTCTTTTGAGGACATCCACCAGTACAGGGAGCAGATC      300
              ** ******** ** ** ** ** ** ***** ** ** ***** ** ** ** ** **

KRAS          AAAAGAGTTAAGGACTCTGAAGATGTACCTATGGTCCTAGTAGGAAATAAATGTGATTTG      360
KRASHRAS      AAACGGGTGAAGGACTCGGAgGACGTGCCCATGGTGCTGGTGGGGAACAAGTGTGACCTG      360
HRAS          AAACGGGTGAAGGACTCGGATGACGTGCCCATGGTGCTGGTGGGGAACAAGTGTGACCTG      360
              *** * ** ******** ** ** ** ** ***** ** ** ** ** ** *****  **

KRAS          CCTTCTAGAACAGTAGACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATTCCT      420
KRASHRAS      cCTtCACGCACTGTGGAcaCTaaGCAGGCTCAGGACCTCGCCCGAAGCTACGGCATCCCC      420
HRAS          GCTGCACGCACTGTGGAATCTCGGCAGGCTCAGGACCTCGCCCGAAGCTACGGCATCCCC      420
               ** *  * ** ** **  *    ************ * **  **** ** ** ** **

KRAS          TTTATTGAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATTAGTT      480
KRASHRAS      TtCATCGAGACCTCGGCCAAGACCCGGCAGGGAGTGGAtGATGCCTTCTATACACTAGTT      480
HRAS          TACATCGAGACCTCGGCCAAGACCCGGCAGGGAGTGGAGGATGCCTTCTACACGTTGGTG      480
              *  ** ** ** ** ** *****  * ***** ** ** ********** **  * **

KRAS          CGAGAAATTCGAAAACATAAAGAAAAGATGAGCAAAGATGGTAAAAAGAAGAAAAAGAAG      540
KRASHRAS      CGAGAAATTCGAAAACATAAAGAAAAGATGAGCAAAGACGGTAAAAAGAAGAAAAAGAAG      540
HRAS          CGTGAGATCCGGCAGCACAAGCTGCGGAAGCTGAACCCTCCTGATGAGAGTGGCCCCGGC      540
              ** ** ** **  * ** **      ** *   **      * *  ***

KRAS          TCAAAGAC---AAAGTGTGTAATTATGTAA    567
KRASHRAS      TCAAAG---ACAAAGTGTGTAATTATGTAA    567
HRAS          TGCATGAGCTGCAAGTGTGTGCTCTCCTGA    570
              *  * *      ********  *    * *
```

```
CLUSTAL O(1.2.4) multiple sequence alignment


KRAS          MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG      60
KRASHRAS      MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG      60
HRAS          MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG      60
              ************************************************************

KRAS          QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDL      120
KRASHRAS      QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDL      120
HRAS          QEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDL      120
              ***********************************:***********:**************

KRAS          PSRTVDTKQAQDLARSYGIPFIETSAKTRQGVDDAFYTLVREIRKHKEKMSKDGKKK-KK      179
KRASHRAS      PSRTVDTKQAQDLARSYGIPFIETSAKTRQGVDDAFYTLVREIRKHKEKMSKDGKKK-KK      179
HRAS          AARTVESRQAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQHKLRKLNPPDESGPG      180
               :***:::************:***********:***********:** :  :  .:.

KRAS          KSKTKCVIMX        189
KRASHRAS      KSKTKCVIMX        189
HRAS          CMSCKCVLSX        190
                . ***: *
```

Fig. S14. Nucleotide and amino acid alignment of KRAS, HRAS and $KRAS_{HRAS}$.

```
CLUSTAL O(1.2.4) multiple sequence alignment


RAC1            ATGCAGGCCATCAAGTGTGTGGTGGTGGGAGACGGAGCTGTAGGTAAAACTTGCCTACTG      60
RAC1RAC3        ATGCAGGCCATCAAGTGCGTGGTGGTCGGCGACGGCGCCGTGGGGAAGACATGCTTGCTG      60
RAC3            ATGCAGGCCATCAAGTGCGTGGTGGTCGGCGACGGCGCCGTGGGGAAGACATGCTTGCTG      60
                ***************** ******** ** ***** ** ** ** ** ** *** * ***

RAC1            ATCAGTTACACAACCAATGCATTTCCTGGAGAATATATCCCTACTGTCTTTGACAATTAT      120
RAC1RAC3        ATCAGCTACACGACCAACGCCTTCCCCGGAGAGTACATCCCCACCGTTTTTGACAACTAC      120
RAC3            ATCAGCTACACGACCAACGCCTTCCCCGGAGAGTACATCCCCACCGTTTTTGACAACTAC      120
                ***** ***** ***** ** ** ** ***** ** ***** ** ** ******** **

RAC1            TCTGCCAATGTTATGGTAGATGGAAAACCGGTGAATCTGGGCTTATGGGATACAGCTGGA      180
RAC1RAC3        TCTGCCAACGTGATGGTGGACGGGAAACCAGTCAACTTGGGGCTGTGGGACACAGCGGGT      180
RAC3            TCTGCCAACGTGATGGTGGACGGGAAACCAGTCAACTTGGGGCTGTGGGACACAGCGGGT      180
                ******** ** ***** ** ** ***** ** **  ****  * ***** ***** **

RAC1            CAAGAAGATTATGACAGATTACGCCCCCTATCCTATCCGCAAACAGATGTGTTCTTAATT      240
RAC1RAC3        CAGGAGGACTACGATCGGCTGCGGCCACTCTCCTACCCCCAAACTGACGTCTTTCTGATC      240
RAC3            CAGGAGGACTACGATCGGCTGCGGCCACTCTCCTACCCCCAAACTGACGTCTTTCTGATC      240
                ** ** ** ** **  *  * ** ** ** ***** ** ***** ** ** **  * **

RAC1            TGCTTTTCCCTTGTGAGTCCTGCATCATTTGAAAATGTCCGTGCAAAGTGGTATCCTGAG      300
RAC1RAC3        TGCTTCTCTCTGGTGAGCCCGGCCTCCTTCGAGAATGTTCGTGCCAAGTGGTACCCGGAG      300
RAC3            TGCTTCTCTCTGGTGAGCCCGGCCTCCTTCGAGAATGTTCGTGCCAAGTGGTACCCGGAG      300
                ***** ** ** ***** ** ** ** ** ** ***** ***** ******** ** ***

RAC1            GTGCGGCACCACTGTCCCAACACTCCCATCATCCTAGTGGGAACTAAACTTGATCTTAGG      360
RAC1RAC3        GTGCGGCACCACTGCCCCAACACGCCCATCATCCTGGTGGGCACCAAGCTGGACCTCCGC      360
RAC3            GTGCGGCACCACTGCCCCCACACGCCCATCCTCCTGGTGGGCACCAAGCTGGACCTCCGC      360
                ************** *** **** ****** **** ***** ** ** ** ** **  *

RAC1            GATGATAAAGACACGATCGAGAAACTGAAGGAGAAGAAGCTGATTCCCATCACCTATCCG      420
RAC1RAC3        GACGACAAGGACACCATTGAGAAACTGAAGGAGAAGAAGCTGATTCCCATCACCTACCCA      420
RAC3            GACGACAAGGACACCATTGAGCGGCTGCGGGACAAGAAGCTGGCACCCATCACCTACCCA      420
                ** ** ** ***** ** ***   ***  *** *********   ********** **

RAC1            CAGGGTCTAGCCATGGCTAAGGAGATTGGTGCTGTAAAATACCTGGAGTGCTCGGCGCTC      480
RAC1RAC3        CAGGGCCTGGCCATGGCCAAGGAGATTGGCGCTGTGAAATACCTGGAGTGCTCAGCCCTG      480
RAC3            CAGGGCCTGGCCATGGCCCGGGAGATTGGCTCTGTGAAATACCTGGAGTGCTCAGCCCTG      480
                ***** ** ********   *********  **** ***************** ** **

RAC1            ACACAGCGAGGCCTCAAGACAGTGTTTGACGAAGCGATCCGAGCAGTCCTCTGCCCGCCT      540
RAC1RAC3        ACCCAGCGGGGCCTGAAGACAGTGTTTGACGAGGCGATCCGCGCGGTGCTCTGCCCGCCC      540
RAC3            ACCCAGCGGGGCCTGAAGACAGTGTTTGACGAGGCGATCCGCGCGGTGCTCTGCCCGCCC      540
                ** ***** ***** ***************** ******** ** ** ***********

RAC1            CCCGTGAAGAAGAGGAAGAGAAAATGCCTGCTGTTGTAA   579
RAC1RAC3        CCAGTGAAGAAGAGGAAGAGAAAGTGCCTGCTGTTGTAA   579
RAC3            CCAGTGAAGAAGCCGGGGAAGAAGTGCACCGTCTTCTAG   579
                ** *********  *  **  ** ***    * ** **


CLUSTAL O(1.2.4) multiple sequence alignment


RAC1              MQAIKCVVVGDGAVGKTCLLISYTTNAFPGEYIPTVFDNYSANVMVDGKPVNLGLWDTAG      60
RAC1RAC3          MQAIKCVVVGDGAVGKTCLLISYTTNAFPGEYIPTVFDNYSANVMVDGKPVNLGLWDTAG      60
RAC3              MQAIKCVVVGDGAVGKTCLLISYTTNAFPGEYIPTVFDNYSANVMVDGKPVNLGLWDTAG      60
                  ************************************************************

RAC1              QEDYDRLRPLSYPQTDVFLICFSLVSPASFENVRAKWYPEVRHHCPNTPIILVGTKLDLR      120
RAC1RAC3          QEDYDRLRPLSYPQTDVFLICFSLVSPASFENVRAKWYPEVRHHCPNTPIILVGTKLDLR      120
RAC3              QEDYDRLRPLSYPQTDVFLICFSLVSPASFENVRAKWYPEVRHHCPHTPILLVGTKLDLR      120
                  ************************************************:***:*********

RAC1              DDKDTIEKLKEKKLIPITYPQGLAMAKEIGAVKYLECSALTQRGLKTVFDEAIRAVLCPP      180
RAC1RAC3          DDKDTIEKLKEKKLIPITYPQGLAMAKEIGAVKYLECSALTQRGLKTVFDEAIRAVLCPP      180
RAC3              DDKDTIERLRDKKLAPITYPQGLAMAREIGSVKYLECSALTQRGLKTVFDEAIRAVLCPP      180
                  *******:*::*** ***********:***:******************************

RAC1              PVKKRKRKCLLLX      193
RAC1RAC3          PVKKRKRKCLLLX      193
RAC3              PVKKPGKKCTVFX      193
                  ****  :** ::*
```

Fig. S15. **Nucleotide and amino acid alignment of RAC1, RAC3 and RAC1$_{RAC3}$**

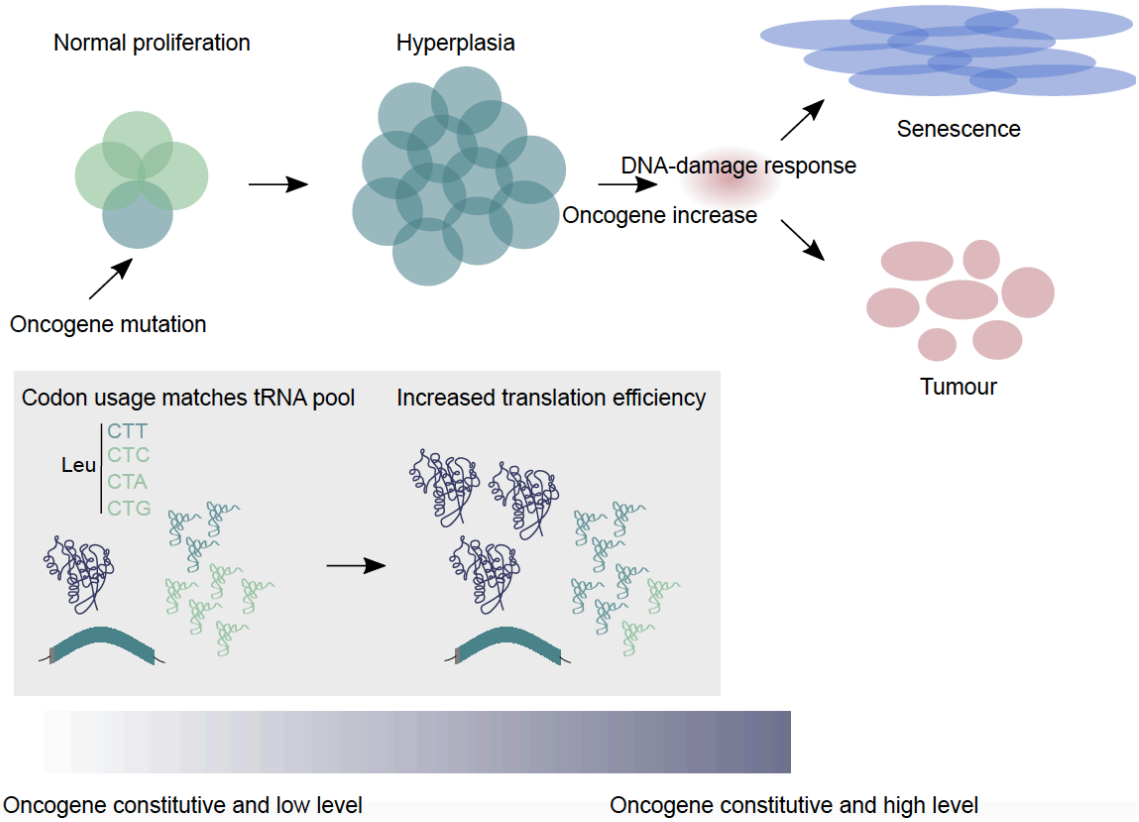


Fig. S16. The carcinogenesis model for RAS-induced tumors proposed by Ferbeyre G in Nat Cell Biol, 2007. Adapted figure from "Barriers to RAS transformation", (17). Oncogenic mutations lead to a constitutive activation of the oncogene that promotes cell proliferation. Low levels of oncogene are not transforming. An increase in oncogene translation efficiency increases oncogene levels. High oncogene levels and activity can transform cells if the cells evade senescence.

CCDS	gene	mutation count TCGA	mutation count normalised within family	protein sequence identity to representative %	similarity cutoff	PC1	covariance
CCDS877.1	NRAS	175	0.36	85.19		-4.03	-3.16
CCDS7698.1	HRAS	86	0.18	100	74.74	6.47	-3.16
CCDS8702.1	KRAS	491	1	84.66		-9.76	-3.16
CCDS5863.1	BRAF	527	1	100		-5.72	-2.1
CCDS35232.1	ARAF	42	0.08	56.98	48.69	4.56	-2.1
CCDS2612.1	RAF1	53	0.1	56.79		-2.38	-2.1
CCDS13945.1	RAC2	15	0.38	92.19		7.77	-2.41
CCDS11798.1	RAC3	9	0.23	92.71	89.08	7.73	-2.41
CCDS5348.1	RAC1	40	1	100		-2.6	-2.41
CCDS2795.1	RHOA	46	1	100		-1.62	-1.26
CCDS1699.1	RHOB	18	0.39	83.16	79.25	9.72	-1.26
CCDS854.1	RHOC	12	0.26	91.71		4.82	-1.26
CCDS9994.1	AKT1	36	1	100		7.68	-0.6
CCDS31076.1	AKT3	44	0.27	82.88	76.94	-7.77	-0.6
CCDS12552.1	AKT2	42	0.55	81.29		6.54	-0.6
CCDS31298.1	FGFR2	122	1	100		-0.35	-0.83
CCDS6107.1	FGFR1	63	0.52	67.06	60.93	3.91	-0.83
CCDS3353.1	FGFR3	75	0.61	64.36		8.27	-0.83
CCDS13147.1	FOXA2	31	0.58	57.02		10.46	0.01
CCDS12677.1	FOXA3	28	0.53	50.86	46.54	6.17	0.01
CCDS9665.1	FOXA1	53	1	100		8.84	0.01
CCDS6982.1	COL5A1	203	0.64	100		2.73	-1.48
CCDS43452.1	COL11A2	121	0.38	71.27	65.88	0.49	-1.48
CCDS53348.1	COL11A1	317	1	75.08		-8.34	-1.48

Table S1. Relevant data for all genes belonging to the eight different gene families studied.

Dataset S1 (separate file): Relevant data pertaining to the background and cancer gene families.

Dataset S2 (separate file): Processed tRNA expression (RPM) data.

Dataset S3 (separate file): Differential tRNA abundance. The relative anticodon abundance is calculated by dividing the RPM value of each anticodon by the sum of all anticodon RPM values for a given amino acid. Differential relative expression analysis between HEK293 and HeLa cells as well as between HEK293 and BJ/hTERT cells were performed using the t-test, where p-values were FDR-corrected and q<0.05 as a cut-off.

Dataset S4 (separate file): Codon weights. The weights of every codon based on the wobble base pairs codon-anticodon interaction rules as described by dos Reis *et al.* (9).

Dataset S5 (separate file): Fold change of the relative codon usage between KRAS/HRAS, BRAF/ARAF, RHOA/RHOC, RAC1/RAC3, COL11A1/COL11A2 and KRASWT/KRASHRAS.

Dataset S6 (separate file): Proliferation- and differentiation-related codon usage for KRASWT, KRASHRAS, RAC1WT and RAC1RAC3. Codon distribution is based on findings from Gingold *et al.* 2014 (3).

SI References

1. M.S. Lawrence, *et al.*, Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
2. P.M. Sharp, W.H. Li, The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
3. H. Gingold, *et al.*, A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**, 1281–1292 (2014).
4. A. Hoffmann, *et al.*, Accurate mapping of tRNA reads. *Bioinformatics* **34**, 2339 (2018).
5. P.P. Chan, T.M. Lowe, GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184-9 (2016).
6. S. Hoffmann, *et al.*, Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.* **5**, e1000502 (2009).
7. A. McKenna, *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
8. F. Abascal, R. Zardoya, M.J. Telford, TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7-13 (2010).
9. M. dos Reis, R. Savva, L. Wernisch, Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044 (2004).
10. L.N. Randolph, X. Bao, C. Zhou, X. Lian, An all-in-one, Tet-On 3G inducible PiggyBac system for human pluripotent stem cells and derivatives. *Sci. Rep.* **7**, 1549 (2017).
11. W. Wang, *et al.*, Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc Natl Acad Sci USA* **105**, 9290–9295 (2008).
12. T. Gogakos, *et al.*, Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep.* **20**, 1463–1475 (2017).

13. N. Kolesnikov, *et al.*, ArrayExpress update--simplifying data submissions. *Nucleic Acids Res.* **43**, D1113-6 (2015).

14. X. Hernandez-Alias, H. Benisty, M.H. Schaefer, L. Serrano, Translational efficiency across healthy and tumor tissues is proliferation-related. *Mol. Syst. Biol.* **16**, e9275 (2020).

15. M.P. Schroeder, C. Rubio-Perez, D. Tamborero, A. Gonzalez-Perez, N. Lopez-Bigas, OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. *Bioinformatics* **30**, i549-55 (2014).

16. B.L. Lampson, *et al.*, Rare codons regulate KRas oncogenesis. *Curr. Biol.* **23**, 70–75 (2013).

17. G. Ferbeyre, Barriers to Ras transformation. *Nat. Cell Biol.* **9**, 483–485 (2007).