Article

RAS oncogenic activity predicts response to chemotherapy and outcome in lung adenocarcinoma

Philip East [1], Gavin P. Kelly [1], Dhruva Biswas [2], Michela Marani[3], David C. Hancock[3], Todd Creasy[4], Kris Sachsenmeier[5], Charles Swanton [2], TRACERx consortium*, Julian Downward [3,6] ✉ & Sophie de Carné Trécesson [3] ✉

Activating mutations in *KRAS* occur in 32% of lung adenocarcinomas (LUAD). Despite leading to aggressive disease and resistance to therapy in preclinical studies, the *KRAS* mutation does not predict patient outcome or response to treatment, presumably due to additional events modulating RAS pathways. To obtain a broader measure of RAS pathway activation, we developed RAS84, a transcriptional signature optimised to capture RAS oncogenic activity in LUAD. We report evidence of RAS pathway oncogenic activation in 84% of LUAD, including 65% *KRAS* wild-type tumours, falling into four groups characterised by coincident alteration of *STK11/LKB1*, *TP53* or *CDKN2A*, suggesting that the classifications developed when considering only *KRAS* mutant tumours have significance in a broader cohort of patients. Critically, high RAS activity patient groups show adverse clinical outcome and reduced response to chemotherapy. Patient stratification using oncogenic RAS transcriptional activity instead of genetic alterations could ultimately assist in clinical decision-making.

The *RAS* oncogenes are mutated in close to 20% of all human cancers, acting as drivers of tumour formation and progression. Point mutations occur mainly in codons 12, 13 and 61 of the three isoforms *HRAS, KRAS* and *NRAS*, decreasing the GTPase activity of their encoded proteins and resulting in the accumulation of the GTP-bound, active conformation. *KRAS* is the most mutated *RAS* isoform, with particularly high prevalence in pancreatic ductal adenocarcinoma (88%), colorectal adenocarcinoma (50%) and lung adenocarcinoma (32%)[1]. The extensive literature describing the role of mutant *KRAS* in proliferation, survival, metabolism and motility supports its significant role in tumour aggressiveness, metastasis and resistance to chemotherapy[2–5]. However, there is a lack of consensus in published studies regarding the predictive value of *KRAS* mutations for patient outcome or response to treatment with chemotherapy[6–8]. *KRAS* mutants can also modulate the tumour microenvironment by regulating the expression of numerous cytokines[9]. Moreover, we have demonstrated that *KRAS* mutation promotes the expression of PD-L1, leading to immune evasion in models of human and mouse lung adenocarcinoma[10]. However, although it is clear that *KRAS* mutation does not preclude response to PD-1 immune checkpoint blockade[11,12], no consistent link between *KRAS* mutation and resistance to immunotherapy or PD-L1 expression has been shown in the clinic[13–16]. Therefore, *KRAS* mutational status cannot be used as a predictive factor to select patients for specific therapy regimens[17], with the exception of EGFR-targeted therapy, where *KRAS* mutations are negatively linked to response to EGFR inhibition in colorectal cancer[18].

[1]Bioinformatics and Biostatistics, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [2]Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [3]Oncogene Biology Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [4]Oncology Data Science, Oncology Research and Development, AstraZeneca, 200 Orchard Ridge Drive, Gaithersburg, MD 20878, USA. [5]Oncology Research and Development, AstraZeneca, 35 Gatehouse Drive, Waltham, MA 02451, USA. [6]Lung Cancer Group, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: julian.downward@crick.ac.uk; sophie.decarne@crick.ac.uk

The stratification of patients uniquely on the mutational status of *KRAS* may have complicated the study of RAS mutants in large cohorts of patients. For instance, in The Cancer Genome Atlas (TCGA), 74% of the lung adenocarcinoma (LUAD) tumours are mutated in one or more genes from the broader RAS pathway, taken as running from receptor tyrosine kinases to ERK MAP kinases and phosphoinositide 3-kinases[19]. Here we propose a method of stratification based on RAS-regulated transcriptional activity which predicts outcome and response to treatment in lung adenocarcinoma and other solid cancers. We derived a gene expression signature, RAS84, and applied machine learning techniques to build a classifier to stratify patients according to the expression of RAS84 in their tumour. Using this method, we discovered that RAS transcriptional activity predicted clinical outcome in lung adenocarcinoma and several other solid cancers, where KRAS mutation alone did not. When applied to a cohort of chemotherapy-treated patients, our classifier predicted poor response in subjects with the highest RAS84 level of expression. We anticipate that the use of RAS84 to stratify patients may validate observations that were made in preclinical models of KRAS-mutant cancers, but not confirmed in clinical studies. Our method will offer the possibility to study the impact of oncogenic RAS activity in large cohorts of patients and may help to predict sensitivity to treatment associated with oncogenic RAS activity, such as chemotherapy.

Results

Other members of the RAS pathway signalling network in addition to KRAS can be altered and affect RAS pathway activity. To overcome this issue, we used RAS pathway transcriptional activity as a fingerprint of RAS oncogenic activity. We started by identifying RAS-target genes established in different models from published data as well as our HCT116 recombined for *KRAS* and sh-*KRAS* data (Supplementary Data 1, GSE199871). We focused on the upregulated genes in the context of activated RAS signalling because most of the signatures only had upregulated genes. Moreover, we reasoned that because more genes were induced than inhibited, and when the magnitude was indicated, the mean fold change was higher in induced genes than inhibited genes, the influence of downregulated genes would be minimal in comparison with those of upregulated genes. We selected studies performed in mouse models or human cell lines, where RAS activity was modulated using different methods, such as RNA interference or inhibition or over-expression of the mutant proteins, and where different isoforms and mutants were represented (KRAS mutants G12D, G12V, G12C, G12A, G13D, Q61H, HRAS mutant G12D). The datasets also represented several organs (lung, pancreas, colon, breast, kidney and prostate). To mitigate the possibility of confounding signal from tumour infiltrating immune cells, we removed all genes present in two immune signatures[20,21]. We also assessed the overlap between the gene signatures. Although all were composed of RAS-target genes, we observed little commonality between the signatures (Supplementary Fig. 1a).

RAS84 construction using lung cell line expression data

Our initial goal was to measure RAS pathway activity in tumour cells. We therefore mapped the established signatures to lung cancer cell line data from the Broad Institute Cancer Cell Line Encyclopedia (CCLE)[22] to determine which ones accurately measured oncogenic RAS activity in samples exempt of stroma and immune cells and where *KRAS* mutation is known to be a prevalent cancer driver. We first cleaned up the signatures by removing genes with low expression or variance across the cell lines (Supplementary Fig. 1b). The 166 remaining cell lines comprised 52 cell lines labelled as small-cell carcinoma, 46 adenocarcinoma, 26 squamous-cell carcinoma, 19 NSCLC with no subtype specified and 23 without label or had another histology type. We removed cell lines with oncogenic RAS pathway mutations other than *KRAS* (*BRAF*, *EGFR*, *ERBB2*, *FGFR1*, *FGFR2*, *FGFR3*, *HRAS*, *JAK2*, *KIT*, *NRAS* and *RET*) from the analysis, since these mutations may drive RAS signalling and confound

the analysis. For each signature, we clustered the filtered CCLE lung cancer cell line signature expression matrix into three groups. We named the clusters RAS-high and RAS-low according to the mean expression of the signature genes within the groups and categorised as "unclassified" the group of samples with intermediate mean expression (Supplementary Fig. 1c). We used the distribution of *KRAS* mutations across the RAS-high and RAS-low clusters to assess the ability of the signature to capture RAS oncogenic activity (Fig. 1a). We reasoned that the signatures measuring RAS oncogenic activity in the lung cell line dataset would show enrichment of *KRAS* mutations in the high group given its role in tumour development in the lung. To highlight the specificity of the signatures to measure RAS oncogenic activity, we also assessed a RAS addiction signature[23] and several other oncogenic signalling pathway signatures[24,25]. The RAS addiction signature was derived from KRAS-mutant cell lines dependent on RAS signalling to maintain cell survival. This signature is therefore expected to capture RAS dependency but not RAS oncogenic activity. We identified "RAS pathway", "KRAS^G13D134" and "HRAS" as the best-performing signatures to enrich KRAS mutants in the RAS-high group (p-value < 1e−5) (Fig. 1a). We refined these three signatures by selecting the genes driving the clustering of the RAS-high and RAS-low groups. We ran a differential gene analysis between these groups to identify signature genes upregulated in the RAS-high group of cells (FDR < 0.05) (Supplementary Fig. 1d). From these genes, we constructed our meta-RAS-activity signature, RAS84, and tested it against the CCLE lung cancer cell line data (Fig. 1b, Supplementary Data 2). RAS84 successfully placed 36 out of 42 KRAS mutant lines into the RAS-high group, with six unclassified and none in RAS-low (Fig. 1c). We observed a good overlap of sample labelling between "RAS pathway" and RAS84 classification and some variability when compared with the other signatures (Supplementary Fig. 1f). However, when compared with other RAS and oncogenic signatures, RAS84 gave the most statistically robust separation of the KRAS mutant cell lines from the RAS-low group (Fig. 1d and Supplementary Fig. 1e). Finally, we ran a differential analysis to identify RAS-high-dependent transcriptional changes when compared to RAS-low (2182 genes, fdr < 0.05, −1 > LFC > 1). We found ERK1 and ERK2 cascade and MAPK cascade GO terms (GO:0070371, p-value 4e−7; GO:0000165 p-value 9e−6) enriched in the RAS-high group (Supplementary Data 3).

Using cell line gene expression and *KRAS* mutation data from lung cancer cell lines, we thus demonstrated the ability of RAS expression signatures to measure oncogenic RAS activity in a lung cancer context. We constructed a meta-signature from the best-performing signatures and demonstrated that it performed better than previous signatures at measuring RAS oncogenic activity by classifying KRAS mutant cell lines as RAS oncogenic signalling activated (RAS-high).

RAS84 expression predicts drug sensitivity and resistance in vitro

To determine whether RAS84 expression was associated with anticancer drug response, we analysed drug sensitivity data obtained from the Genomics of Drug Sensitivity in Cancer project (GDSC) and The Cancer Therapeutics Response Portal (CTRP) in the context of RAS high and low CCLE cell lines. We identified drugs with differential drug responses across the two groups (GDSC fdr < 0.05, −1 > log2(delta IC50) > 1) (Fig. 2a, Supplementary Fig. 2 and Supplementary Data 4, 5). We tested for enriched drug target terms within the drugs showing differential response (hypergeometric fdr < 0.05) (Fig. 2b). We found RAS-high cell lines were sensitive to drugs targeting ERK MAPK and EGFR signalling, and also protein stability and degradation, such as HSP90 inhibitors. We were encouraged to see sensitivity to ERK MAPK and EGFR signalling inhibition, confirming high dependence on RAS signalling in these cell lines. In addition, links between RAS mutation and sensitivity to HSP90 inhibition have been identified previously[26]. Conversely, we found RAS-high cell lines to be resistant to drugs targeting DNA replication, mitosis and chromatin histone acetylation. DNA
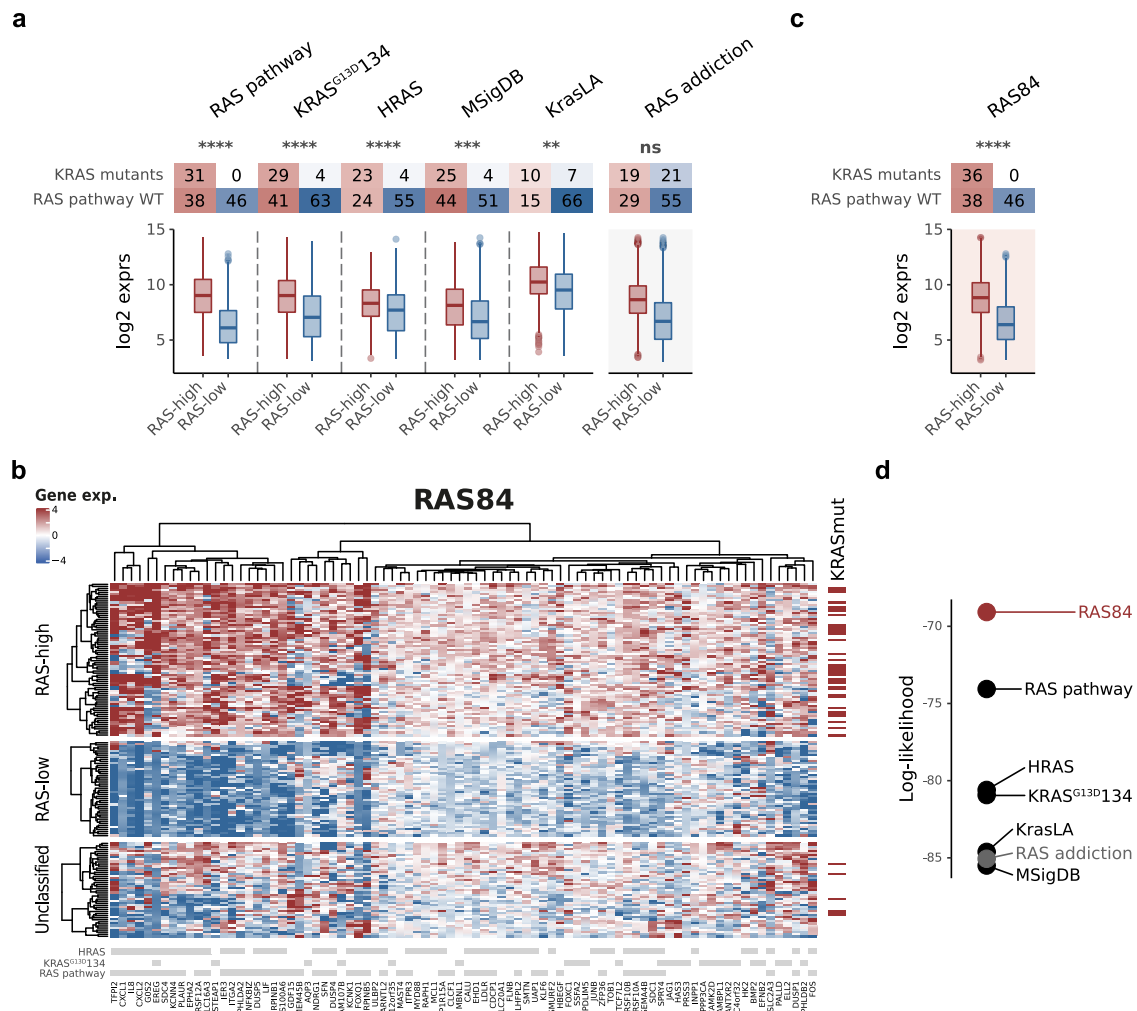
Fig. 1 | In vitro RAS signature derivation. a Contingency tables showing the number of RAS pathway wild-type and KRAS mutant cell lines per RAS-high and RAS-low groups for each signature. RAS pathway wild-type cell lines are those with no oncogenic mutation in any RAS pathway member. RAS-high cell counts are shown in red, RAS-low in blue. The boxplots show the RI distributions for the RAS-high and low groups where the box shows the median and IQR, the whiskers indicate ± 1.5 x IQR, outliers lie outside this range. RAS addiction is presented here as a control signature (RAS pathway $n = 115$, KRAS^{G13D}134 $n = 137$, HRAS $n = 106$, MSigDB $n = 124$, KrasLA $n = 98$, RAS addiction $n = 124$; Chisq test ****$P \leq 0.0001$, ***$P \leq 0.001$, **$P \leq 0.01$, n.s.= $P > 0.05$). **b** Heatmap showing our RAS84 meta-signature genes mapped to filtered (see "Method") CCLE lung cell line data. Cell lines are shown as rows, genes as columns. Groupings of high, medium and low RAS activity are shown as separate clusters, KRAS mutational status is indicated in dark red on the right and parent signature gene membership is indicated in grey at the bottom of the map. **c** Contingency tables showing the number of RAS pathway wild-type and KRAS mutant cell lines per RAS-high and RAS-low groups for RAS84. RAS pathway wild-type cell lines are those with no oncogenic mutation in any RAS pathway member. RAS-high cell counts are shown in red, RAS-low in blue. The boxplots show the RI distributions for the RAS-high and low groups where the box shows the median and IQR, the whiskers indicate ± 1.5 x IQR, outliers lie outside this range. $N = 120$, Chisq test ****$P \leq 0.0001$, ***$P \leq 0.001$, **$P \leq 0.01$, n.s. = $P > 0.05$. **d** Log-likelihood values from a GLM fit (family = binomial) of KRAS mutation status across the three RAS activity groups for each of the signatures. RAS84 is the best-performing signature at segregating KRAS mutants across the RAS activity groups.

replication and mitosis are common chemotherapy targets indicating that RAS84 activity is associated with chemotherapy resistance in vitro. We also tested for KRAS mutation (Fig. 2c) and RAS pathway mutation-dependent drug responses (Fig. 2d). We found both mutant groups were sensitive to just three drugs targeting ERK MAPK signalling. We did not observe resistance to any drugs in these comparisons. This result shows RAS84 better captures RAS-driven drug response than the presence of KRAS mutation alone or wider RAS pathway mutants, highlighting the importance of our transcriptional signature approach.

RAS84 expression is associated with *KRAS* mutation in lung adenocarcinoma

To further validate RAS84 beyond cell lines, we applied it to clinical LUAD expression data from TCGA (512 samples)[27]. Given the increased heterogeneity in signature expression observed in patient tumour

samples when compared to the cell line data, we explored the clustering of the patients beyond the three groups previously used for cell line data. We clustered the patients into five groups (Fig. 3a) and found a low *KRAS* mutation count (6%) in the cluster with the lowest RAS84 expression (chi-square *p*-value 1.05e−08). The other clusters all had high levels of *KRAS* mutation, between 25 and 45% (Fig. 3b). We regrouped the patients using the various signatures described above and found that RAS84 gave optimal segregation of *KRAS* mutant tumours across the groups (Fig. 3c). We labelled each RAS Activity Group (RAG) RAG-0, RAG-1, RAG-2, RAG-3 and RAG-4 ordered low to high by mean RAS84 gene signature expression. To ensure RAS84 expression was predominantly tumour-driven we looked at RAS84 expression in the stroma of five NSCLC samples[28] (Supplementary Fig. 3) and found minimal expression. We assigned a RAS84-Index (RI) value to each patient, defined as the mean expression of the RAS84
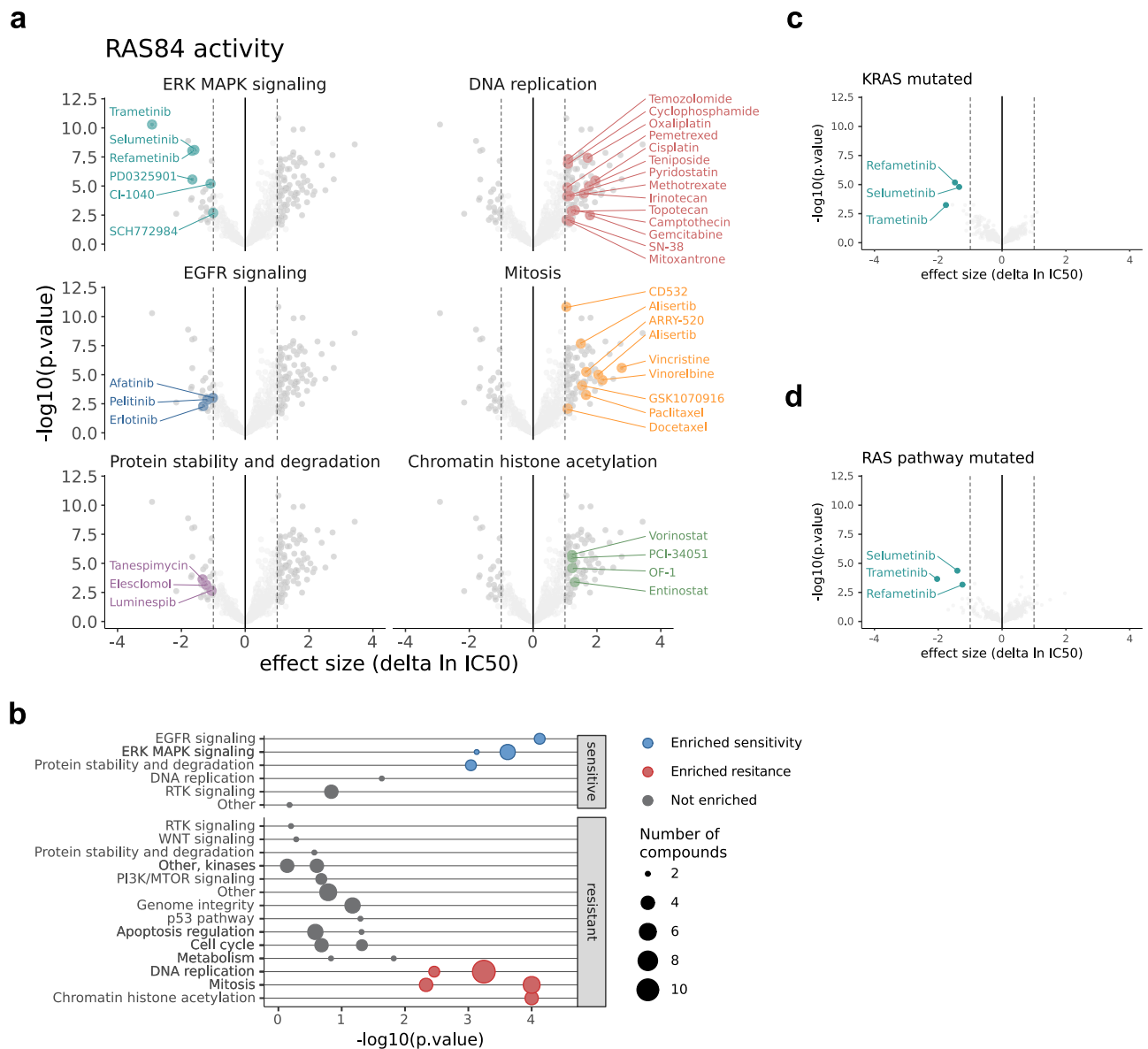
Fig. 2 | In vitro anticancer drug screen. a Volcano plots showing differences in IC50 values between RAS high and low CCLE cell lines. Drugs with enriched target annotations in the significant sensitive and resistant groups are highlighted (hypergeometric test) Drugs with an absolute log2 fold change >1 and fdr < 0.05 (linear model with Benjamini-Hochberg correction) are shown in dark grey. Results from both GDSC1 & 2 are shown. **b** Drug target annotation enrichment in sensitive and resistant drugs from GDSC1 & 2 (fdr < 0.05) in the RAS high CCLE cell lines, determined by hypergeometric test with Benjamini–Hochberg correction. Target terms enriched in the sensitive drugs are shown in blue, the resistant in red. The number of drugs in each group is indicated by the size of the point. All tested targets are shown. **c, d** Volcano plots showing differences in IC50 values between KRAS mutant and wild-type cell lines and RAS pathway mutated and wild-type cell lines (linear model with Benjamini–Hochberg correction). Drugs with enriched target annotations in the significant sensitive and resistant groups are highlighted (hypergeometric test).

genes. To further characterise these groups, we tested if any other reported genomic alteration[19] exhibited a non-random distribution across the five clusters. We identified eight alterations enriched in one or more of the RAGs (chi-square, FDR < 0.05) (Fig. 3d and Supplementary Fig. 4a) and used these cluster-associated alterations to characterise each of the five RAGs (Fig. 3a). *STK11*, *KEAP1*, *RB1*, *TP53*, *ATM* and *CDKN2A* (p-values: <2e−16, 6e−7, 3e−4, 5e−11, 5e−3 and 8e−4) are tumour suppressor genes whereas *EGFR* and *CTNNB1* (p-values: 2e−09 and 0.01) are proto-oncogenes. The under-representation of KRAS mutants characterised RAG-0, but interestingly this group contained a high number of tumour suppressor mutants (*KEAP1*, *RB1* and *TP53*) as well as *CTNNB1* mutants, which could explain how the tumours initiated. Due to the high level of p53 alterations (~70%), we refer to these as P tumours. In addition to frequent *KRAS* mutations, many *EGFR*

mutations also characterised RAG-2 and RAG-3, suggesting the genes driving these clusters reflected RAS pathway activation through upstream receptor tyrosine kinase (RTK) activation. *EGFR* and *KRAS* mutations are mutually exclusive[29,30]. *TP53*, *STK11*/LKB1 and *CDKN2A* were identified as co-mutational partners of *KRAS* in NSCLC where the tumour suppressor gene mutations tend to be mutually exclusive[31]. Along with high *KRAS* mutation rates and RAS84 expression, RAG-1 was characterised by *STK11*/LKB1 mutations (KL tumours) and RAG-4 by *CDKN2A* mutations (KC tumours). *TP53* was frequently altered in several of the clusters, with RAG-3 having the highest rate of p53 alteration (hence KP tumours) after the RAS silent RAG-0. The RAG-2 cluster had modest levels of p53 (*TP53*) alteration (~40%) and more beta catenin (*CTNNB1*) alterations than the other RAS active clusters, but still only about 5%. We refer to these as K tumours, as selective co-occurring
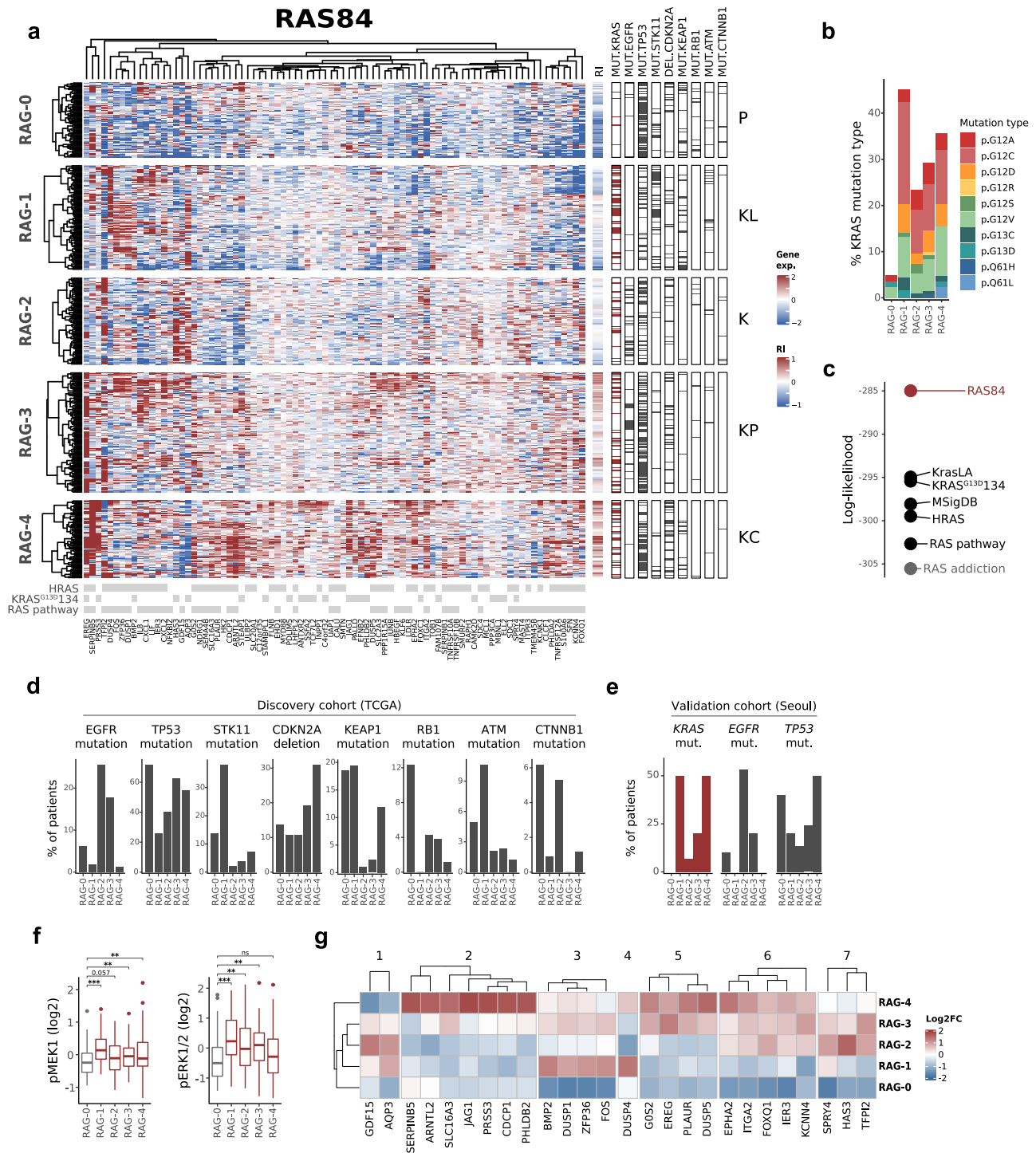
Fig. 3 | LUAD classification by RAS84. a Heatmap showing clustered RAS84 genes and TCGA LUAD cohort patients. Patients are shown as rows, genes as columns. Patients have been clustered into five RAS activity groups (RAGs) by hierarchical clustering using a ward.D2 agglomeration method. Aggregate RAS84-Index (RI) scores are shown to the right of the main heatmap. Genome variants with a significant non-random distribution across the RAGs are shown in the nine columns on the right (chi-square fdr < 0.05). These mutations are used to characterise the five clusters shown by the labels on the right (P, TP53; KL, KRAS/LKB1(STK11); K, KRAS; KP, KRAS/TP53; KC, KRAS/CDKN2A). KRAS mutants are shown in dark red. Parent signature membership is shown in grey at the bottom of the heatmap. **b** The percentage of KRAS mutations per RAG broken down by specific KRAS mutation type. **c** Log-likelihood values from a GLM fit (family = binomial) of KRAS mutation status

across the five RAGs. **d** Bar plots showing the percentage of patients per RAG with EGFR, TP53, STK11 mutations, CDKN2A deletion, KEAP1, RB1, ATM and CTNNB1 mutations found to be significantly associated with any one RAG (fdr < 0.05). **e** EGFR, KRAS and TP53 mutation percentages found to be significantly associated with any one RAG from the Seoul cohort. **f** Boxplots showing The Cancer Protein Atlas (TCPA) RPPA MEK1 and ERK1/2 phosphorylation level distributions across RAGs. Significance levels are shown compared to RAG-0 derived by linear model fit ($n = 349$; RAG-0 $n = 57$, RAG-1 $n = 79$, RAG-2 $n = 61$, RAG-3 $n = 91$, RAG-4 $n = 61$; ****$P \leq 0.0001$, ***$P \leq 0.001$, **$P \leq 0.01$, n.s.= $P > 0.05$) The box shows the median and IQR, the whiskers indicate ± 1.5 x IQR, outliers lie outside this range. **g** Heatmap showing variant mean RAS84 gene expression clusters across the five RAGs.

mutations are not obvious. The clustering in five groups (from C0 with the lowest mean expression level of a given signature, to C4 with the highest) using published signatures only partially overlapped with the RAS84-RAG classification (Supplementary Fig. 4b), although clear similarities were evident.

We validated the patterns of *KRAS*, *EGFR* and *TP53* mutations across our RAGs in an independent lung adenocarcinoma cohort of 87 patients (Seoul cohort, GSE40419) (chisq test *p*-values, *EGFR* 0.0002, *KRAS* 0.0014, *TP53* 0.1) (Fig. 3e). Given the small size of the Seoul cohort it was not possible to cluster the patients in a robust way with all the signatures. We instead analysed the RI distributions for *KRAS* mutant and wild-type patients (Supplementary Fig. 4c, d). RAS84 has highly significant effect size between the two groups. Interestingly, we did not observe any significant association between specific KRAS amino acid mutational changes and RAG, suggesting that specific mutations do not drive RAS activity heterogeneity in distinct ways in the context of lung adenocarcinoma (Fig. 3b).

To determine if RAS84 expression reflected RAS-MAPK signalling activity we looked at ERK1/2 (T202, Y204) and MEK1 (S217, S221) phosphorylation levels within each RAG. We used The Cancer Proteome Atlas (TCPA) reverse-phase protein arrays (RPPA) data[32] for 349 of the TCGA LUAD patients and found an increase in phosphorylation of one or both proteins in all RAGs when compared to RAG-0 (Fig. 3f). At the expression level we found an enrichment of genes associated with the GO term 'ERK1 and ERK2 cascade' when comparing RAG-4 to RAG-0 (GO:0070371, *p*-value 0.0003) (Supplementary Data 6).

To understand what drove MAPK activity in the KRAS wt samples of RAG-3 and -4 (which are not enriched in EGFR mutants), we looked at the expression of RTK ligands and found *EREG* (present in RAS84), *AREG*, *NRG1* and *TGFA* were all enriched in RAG-4 (Supplementary Fig. 5a). *AREG* was also enriched in RAG-3 and *TGFA* in RAG-2, -3 and -4. Interestingly, KRAS mutant and wild-type tumours expressed similar levels of the ligands within each RAG (Supplementary Fig. 5b), suggesting an autocrine signalling loop also exists in KRAS mutant tumours in these groups as previously shown[33,34]. We also looked at the proliferation score distributions across the five groups and did not observe a correlation with RAS84 expression (Supplementary Fig. 6a).

We were also interested in identifying which RAS84 genes were driving the RAGs. We focused on genes that were most variant across the clusters and, via correlation analysis (Supplementary Fig. 6b), found seven small gene clusters (1–7 genes in each) capable of discriminating the five RAGs (Fig. 3g). RAG-4 was characterised by higher expression of genes in cluster 2 when compared to the other RAGs, RAG-3 by the expression of genes in cluster 5 but not 2, RAG-2 by the up-regulation of genes in clusters 6 but not 5 and to an extent the over-expression of cluster 1. RAG-1 patients could be identified by high expression of cluster 3 and low expression of cluster 6, RAG-0 by low expression of cluster 3. We found two clusters whose expression pattern across the five RAGs mirrored that seen in the enriched alterations.

Hence, we have demonstrated that RAS84 performed optimally in classifying KRAS mutant lung adenocarcinoma tumours as active for RAS-driven transcription. We identified five RAGs characterised by distinct associated mutational profiles and we showed RAS84 expression to be reflected by changes at the protein level.

RAS84 expression is mostly clonal

The prognostic value of RAS84, explored below, could make it an attractive potential biomarker. A reliable biomarker should ideally not be affected by the region of sampling and therefore not be perturbed by the intra-tumour heterogeneity observed in most cancers. Recent analyses of signatures derived for prognostication in lung cancer indicate that up to 70% of NSCLC tumours[35] and 40% of LUAD tumours[36] may be subject to sampling bias. To assess the intra-tumour heterogeneity of RAS activity in lung adenocarcinoma, we classified

samples from the multi-region TRACERx cohort into our five RAS Activity Groups (102 samples from 41 patients)[37]. To classify the samples, we trained a support-vector machine (SVM) classifier using the TCGA LUAD classification results (see methods) and used it to assign RAG labels to the TRACERx samples (Fig. 4a). SVM has been used as a classification model with transcriptome data. This classifier will allow the stratification of new patient samples outside of cohort datasets enabling the clinical application of RAS84.

Of the 41 patients, 28 (68%) had multi-region RNA-seq gene expression data available. Of these 28, 16 (57%) patients had all regions falling within the same RAG (Fig. 4a). Twelve (43%) patients had regions that spanned RAGs suggesting a degree of RAS activity heterogeneity in some patients. Of these twelve, nine patients' tumours had regions falling only within neighbouring RAGs (RAG-4/3: four; RAG-3/2: one; RAG-2/1: one; RAG-1/0: three). The remain three of these patients (CRUK0017, CRUK0024 and CRUK0060) span more than two neighbouring RAGs indicating a degree of relatedness in RAS activity across the tumour. CRUK0017 spanned three clusters, RAG-3/2/0. This patient is reported as sub-clonal for *KRAS* mutation by TRACERx (PhyloCCF R1:0; R2:0.84; R4:0.65). Interestingly, the *KRAS* non-mutated region fell into RAG-0. Patients CRUK0024 and CRUK0060 both had regions in two non-adjacent clusters, RAG-3/1. However, we observed that the intra-tumour distances for RAS84 within these group-spanning tumours were still small when compared to the inter-tumour distance distribution (Fig. 4b).

We assessed the intra- and inter-tumour expression variance of RAS84 genes by comparing them to gene sets previously annotated for expression heterogeneity in the TRACERx lung adenocarcinoma cohort. Biswas and colleagues classified all expressed genes into four groups depending on their intra- and inter-tumour expression variance[36]. We found a significant 3.2 fold enrichment of RAS84 genes in the low intra-tumour, high inter-tumour expression variance group (Fisher's exact *p*-value 9.06e−6) (Fig. 4c, Supplementary Data 7). This shows RAS84 is enriched for genes robust to sampling bias.

Altogether, these data suggest that RAS transcriptional activity is comprised of predominantly clonally expressed genes in lung adenocarcinoma tumours, likely reflecting the oncogenic driving capability of the RAS pathway. These data indicate that the RAS84 signature may be relatively robust in the face of sampling bias.

RAS84 predicts survival and response to chemotherapy in lung adenocarcinoma patients

Mutated *KRAS* oncogene promotes tumour progression and metastasis in many preclinical models, but the mutational status of *KRAS* is not reliably associated with outcome in clinical datasets[6,7,38] (Fig. 5a). To determine whether RAS84 had prognostic value in lung adenocarcinoma, we ran a univariate Cox proportional hazards analysis comparing overall survival, and also progression-free survival, across the TCGA LUAD RAGs (*n* = 493, 265 stage I, 117 stage II, 79 stage III and 25 stage IV). We found RAG-4 to be significantly associated with negative outcome in both overall and progression-free survival when compared to RAG-0 (Fig. 5b, c, Supplementary Fig. 7a–f). We also fitted a univariate Cox proportional hazards regression model to the RAS84 RAS Index (RI) values. We found a significant positive association with outcome, showing increased RAS84 expression was a predictor of poor overall survival (coxph HR 2, *p*-value 0.00042). To visualise the ability of RI to predict outcome, we used the model to predict survival time given a two-fold increase or decrease in RI values (Fig. 5d). Since we observed a slight over-representation of stage III tumours and an under-representation of stage I tumours in RAG-4 (Supplementary Fig. 7g, h), we confirmed these findings in a multivariate Cox proportional hazards analysis in an independent lung adenocarcinoma cohort of 103 patients (60 stage I, 19 stage II and 24 stage III) (Uppsala cohort, GSE81089). We first clustered the patients into five groups as previously described and ran the multivariate analysis across the RAGs
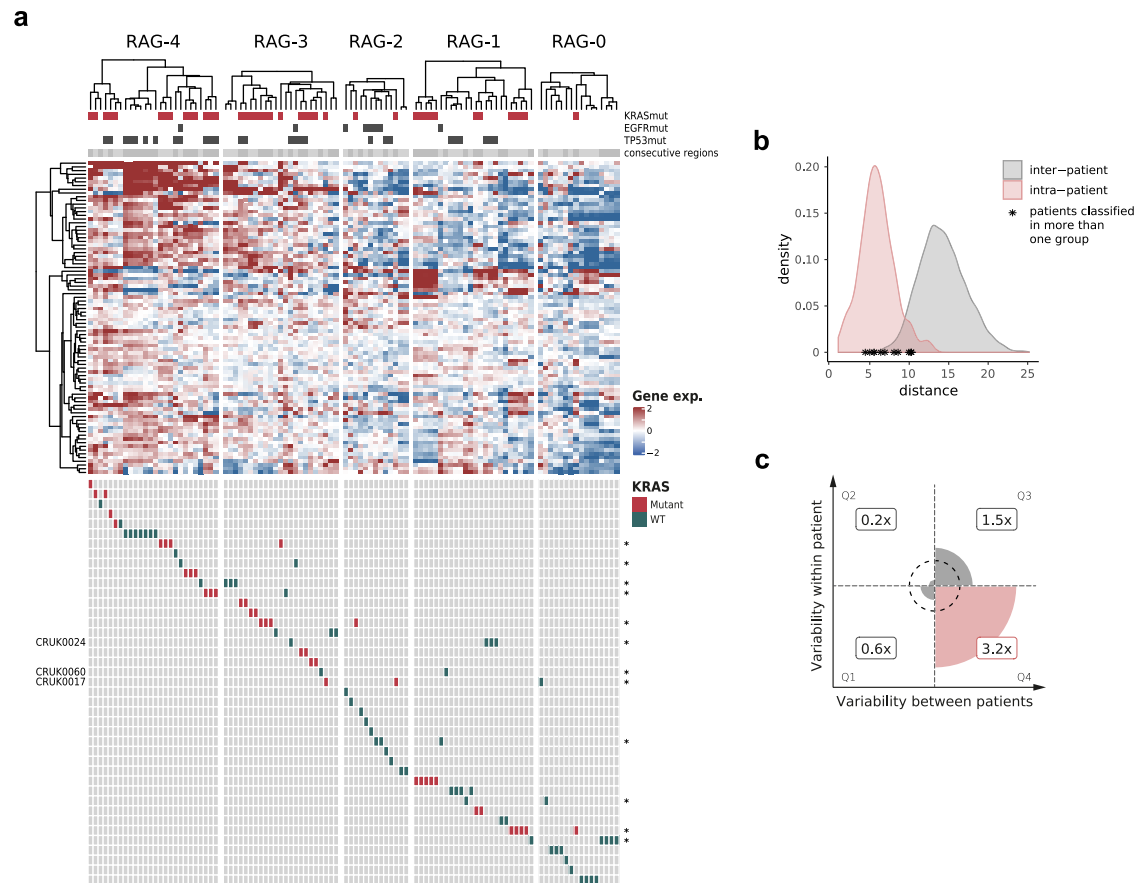
Fig. 4 | Intra-tumour RAS84 heterogeneity. a Heatmap showing RAS84 gene expression across the TRACERx multi-region cohort. The samples have been classified into the five RAGs using an SVM classifier. Contiguous samples from the same patient are indicated by the grey shades just at the top of the heatmap. Gene mutation status is indicated at the top, KRAS in red, EGFR and TP53 in grey. Per-patient regions are indicated by the rows at the bottom of the heatmap. KRAS mutants are shown in red, wild-type in green. Patients with regions spanning multiple RAGs are indicated with an asterisk. Identifiers are given on the left in the three cases where regions did not fall into adjacent RAGs. **b** Intra- (orange) and inter-tumour (grey) sample Euclidian distant distributions. The maximum intra-patient distance for patients with samples spanning different RAGs are indicated with an asterisk. **c** Plot showing the enrichment of RAS84 genes in genes previously classified as low for intra-tumour expression variance and high for inter-tumour expression variance in the TRACERx lung adenocarcinoma cohort. This group is represented by quadrant 4 on the plot showing a 3.2x fold enrichment of RAS84 genes relative to all expressed genes (Fisher's exact enrichment p-value 9.06e−6).

and RI values including TNM stage, World Health Organization (WHO) performance status, smoking history, sex and age in the model. We found RAG to be a significant predictor of outcome (ANOVA LRT p-value 0.032) and specifically RAG-4, RAG-3, and RI to be significantly associated with poor outcome (Fig. 5e–h, Supplementary Fig. 8a–c).

Interestingly, when we combined the KRAS mutational status with RI, the patients with RAS-high/KRAS wt tumours had the same outcome as those with RAS-high/KRAS mut (Supplementary Fig. 8d), confirming that the tumour progression driven by oncogenic RAS activity is independent of KRAS mutational status. We repeated these univariate and multivariate analyses with patient clusters derived using the published signatures. Although we found that three other RAS signatures predicted the highest RAS activity group, C4, to have the worst prognosis, only RAS84 significantly predicted outcome in a multivariate analysis correcting for other predictors of outcome (Fig. 5i, Supplementary Fig. 8e, Supplementary Data 8). This shows RAS84 had prognostic qualities in early-stage lung adenocarcinoma beyond known clinical predictors and suggests RAS pathway activity promotes tumour progression in human lung adenocarcinoma.

Since we observed a high concordance between patients classified as the highest RAS signature group, RAG-4 by RAS84 and C4 RAS pathway signature, we took a closer look at the C4⁺/RAG-4⁺ and C4⁺/RAG-4⁻ patients. Stratified by stage, we found stage III C4⁺/RAG-4⁻ patients to have better OS than C4⁺/RAG-4⁺ patients (Supplementary

Fig. 8f), showing that RAS84 performs well at identifying patients where high RAS activity correlates with poor outcome.

Given that we observed a chemotherapy drug resistance phenotype in vitro, we ran a PFS multivariate Cox proportional hazards analysis using the TEMPUS cohort of adenocarcinoma patients ($n = 94$, 5 Stage I, 17 Stage II, 31 Stage III and 41 Stage IV). We selected patients who had received first-line chemotherapy treatment and constructed PFS intervals using patient records (see "Methods"). We classified the patient tumours using associated RNA-Seq data and our SVM classifier. We modelled PFS with RAG labels along with stage, the administration of radiotherapy, age and sex covariates. We found RAG to be a significant predictor of PFS after chemotherapy (ANOVA LRT p-value 0.043). Specifically, patients in RAG-3 and RAG-4 had a poor response when compared with RAG-2 (p-value 0.006, 0.027; HR 3.04, 2.84) (Fig. 5j and Supplementary Fig. 9a). We also ran the same multivariate analysis testing KRAS mutation as a predictor of PFS. As previously shown[7,8], KRAS mutation did not predict response to chemotherapy (Supplementary Fig. 9b).

To test a possible additive effect of RAG classification and KRAS mutation, we split the RAGs into KRAS mutant and wild-type subgroups and ran the same multivariate analysis as above. Although this subclassification into 10 groups limits the statistical power, we observed that the KRAS wt patients in RAG-4 were the ones responding the worse to chemotherapy (Supplementary Fig. 9c), demonstrating further the independence between KRAS mutational status and
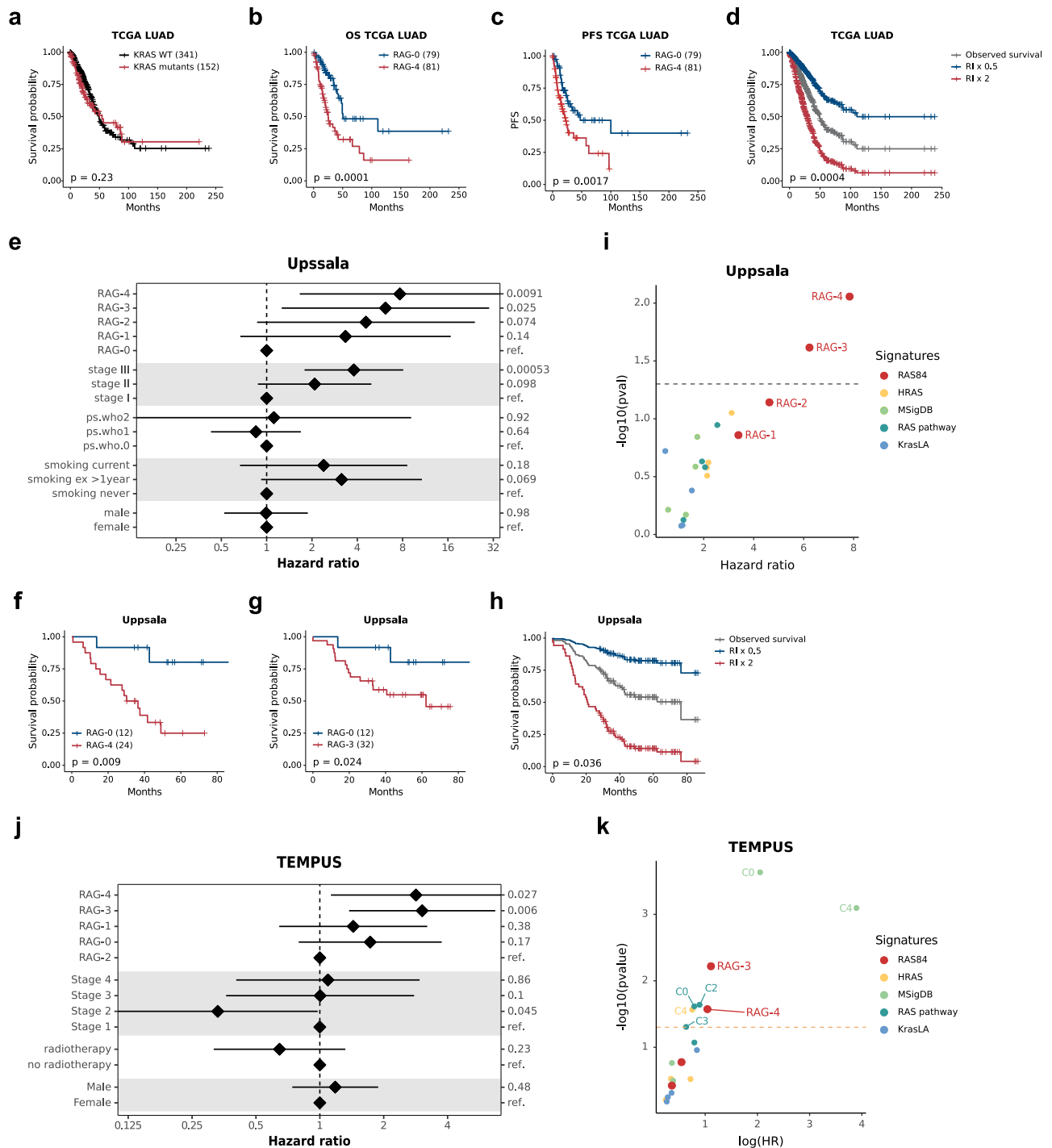
Fig. 5 | RAS84 predicts survival in lung adenocarcinoma. a–d Kaplan–Meier plots showing overall survival data from the TCGA LUAD cohort for patients stratified by KRAS mutation (coxph *p*-value 0.23) (**a**) or from RAG-4 and RAG-0 (coxph *p*-value 0.0001) (**b**), progression-free survival data from the TCGA LUAD cohort for patients from RAG-4 and RAG-0 (coxph *p*-value 0.0017) (**c**) and a Cox proportional hazards regression fit of TCGA LUAD survival data to RI values. The grey survival curve shows the observed data. The blue survival curve is the predicted survival if the RI values decreased 2-fold. The red curve is the predicted survival values if the RI were to increase 2-fold (coxph *p*-value 0.0004) (**d**), Patient number is indicated between brackets. **e** Forest plot showing results from a multivariate Cox proportional hazards analysis of the Uppsala lung adenocarcinoma cohort (*n* = 103 patients). Hazard ratios and 5 and 95% confidence intervals are shown on a natural log scale. **f–h** Kaplan–Meier plots showing overall survival data from the Uppsala cohort for patients from RAG-4 and RAG-0 (multivariate coxph *p*-value 0.0088) (**f**),

RAG-3 and RAG-0 (multivariate coxph *p*-value 0.024) (**g**) and a Cox proportional hazards regression fit of Uppsala survival data to RI values. The grey survival curve shows the observed data. The blue survival curve is the predicted survival if the RI values decreased 2-fold. The red curve is the predicted survival values if the RI were to increase 2-fold (coxph *p*-value 0.036) (**h**). **i** Multivariate *p*-values and hazard ratios plotted for Uppsala RAGs clusters derived from the published signatures. The *p*-values are plotted on a −log10 scale (coxph *p*-value RAS84 RAG-4 0.0088, RAS84 RAG-3 0.024). **j** Forest plot showing results from a multivariate Cox proportional hazards analysis of PFS after chemotherapy in the TEMPUS lung adenocarcinoma cohort (*n* = 100 patients) (coxph *p*-value RAG-high *p*-value 0.0018). Hazard ratios and 5 and 95% confidence intervals are shown on a natural log scale. **k** Multivariate *p*-values and hazard ratios plotted for TEMPUS RAGs and clusters derived from the published signatures. The *p*-values are plotted on a −log10 scale.

phenotypes previously associated with oncogenic RAS activity. We did not observe differences in responses between KRAS mutant and wild-type samples in RAG-3. Altogether, these results show the potential of RAS transcriptional activity to predict response to chemotherapy where KRAS mutation status alone does not.

Finally, we benchmarked RAS84 against published RAS signatures to predict response to chemotherapy. Among the other signatures, only "HRAS" predicted chemoresistance in its highest-signature-expression cluster (C4), however failed to show resistance in C3. "KrasLA" and "RAS pathway" failed to identify C4 as having the worse response to chemotherapy and "MSigDB" also identified C0 as non-responders, therefore failing to show a correlation between RAS activity and resistance to treatment (Fig. 5k). Of note, "MSigDB" failed to identify many patients in C4 (Supplementary Fig. 4c). As shown in the TCGA and Seoul cohorts, the classification varied across the signatures, and RAS84 was still superior to previous RAS signatures at classifying KRAS mutants, although only marginally so with respect to "Ras pathway" (Supplementary Fig. 9d, e).

Because of the clinical potential of RAS84, we sought to identify the genes driving patient stratification to reduce the number of genes in the signature. We started by ensuring there were no highly correlated genes within the signature (Supplementary Fig. 10a, b). We applied Random Forest to our RAG labels and the TCGA RAS84 expression matrix. We used this model to rank the 84 genes by their importance in classifying the patients into the 5 RAGs (Supplementary Data 9). We constructed expression matrices using gene sets iteratively across the gene rank. We built new SVM classifiers using these matrices along with our original RAG labels identified from the RAS84 clustering. We tested these new classifiers against our test data to determine performance (Supplementary Fig. 11a). We achieved maximum accuracy using 55 genes. We determined the sensitivity and specificity to correctly identify each RAG (Supplementary Fig. 11b). We found 43 genes gave a > 0.8 sensitivity in all RAG classes and 11 genes gave >0.9 specificity. We were able to identify a significant association between RAG-4 OS in the TCGA cohort using these 55 genes (HR 2.305, p-value 3.29e−04) (Supplementary Fig. 11c).

We also ran Lasso regression on a TCGA OS Cox Proportional Hazards model including all RAS84 genes as predicators to determine the RAS84 genes specifically associated with outcome. We identified 11 genes to be important (*ARNTL2, CXCL1, EHD1, ITGA2, KCNK1, KLF6, MCL1, RESF1, SDC1, SEMA4B, SERPINB1*). We used this small 11-gene signature mentioned above (hereinafter referred to as "RAS11") in a multivariate Cox Proportional Hazards analysis on the Uppsala cohort and found its mean expression to be significantly associated with outcome (Supplementary Fig. 11d). Importantly, we found that RAS11 expression increased across RAGs defined by RAS84 from RAG-0 to -4 in this cohort, validating that RAS11 is still a good surrogate for oncogenic RAS activity and not only associated with outcome (Supplementary Fig. 11e).

We thus demonstrate the prognostic value of RAG classification and RI quantification in 500+ lung adenocarcinoma patients from two independent cohorts, benchmarked against the failure of KRAS mutational status or previous RAS signatures to predict patient outcomes. We show that 11 of RAS84 genes are sufficient to capture prognosis in an independent cohort of lung adenocarcinoma patients. We also show RAG classification as a predictor of response to chemotherapy, thus demonstrating that RAS84 adds value to current clinical risk factors and response biomarkers.

RAS84 predicts RAS-MAPK pathway activity across cancer types

The degree to which RAS activity is important in tumourigenesis and cancer progression varies across different tissues[19], with some cancers known to be driven largely by RAS mutations (e.g., pancreatic, color-ectal and lung cancers[39,40]) and others not (e.g., uveal melanoma[41], glioblastoma[42], kidney cancer[43]). To determine how RAS84-

determined RAS activity varied across cancer types, we quantified it against all 32 TCGA solid cancers in a pan-cancer analysis. To compare samples across cancers, we calculated an RI value for each sample (Fig. 6a). We identified two distinct cancer populations from the distribution of mean RI values per-cancer (Fig. 6b, Supplementary Fig. 12a). We found four of the top five RAS mutated cancers known to be RAS-driven (RAS mutation frequency: pancreatic adenocarcinoma (PAAD) 71%, colon adenocarcinoma (COAD) 50%, rectum adenocarcinoma (READ) 49% and lung adenocarcinoma (LUAD) 31%) in the highly RAS active group (Supplementary Fig. 12b). We also found *KRAS* mutation to be over-represented within this group (hypergeometric p-value < 2e−16). The other cancers found in the high RI group were stomach adenocarcinoma (STAD) (8.9% RAS mutated), bladder urothelial carcinoma (BLCA) (8.3%), head and neck squamous-cell carcinoma (HNSC) (5.8%), cervical squamous-cell carcinoma and endocervical adenocarcinoma (CESC) (5.2%), lung squamous-cell carcinoma (LUSC) (3.5%) and oesophageal carcinoma (ESCA) (1.2%). RAS is not significantly mutated in these cohorts when compared to those with a lower RAS mutation ratio (Supplementary Fig. 12b). In order to explain the presence of these cancers in the highly RAS active group, we looked at the correlation between RAS pathway alteration status and mean RI across the cohorts. We defined RAS pathway alteration status as the number of patients with at least one alteration in a RAS pathway gene (as defined in TCGA driver pathway analysis[19]) leading to pathway activation. We identified a pan-cancer correlation (spearman coefficient 0.432, p-value 0.0135) (Fig. 6c). Stomach (STAD), bladder (BLCA), head and neck squamous cell (HNSC), oesophageal (ESCA) and lung squamous-cell (LUSC) cancers all had RAS pathway alteration rates above 50% and fell within the 99% confidence interval suggesting RAS pathway alterations other than RAS are driving activity in these cancers. The high RI values in cervical carcinoma (CESC) remain unexplained since the high frequency of *PIK3CA* mutation (28.5%) in this cancer is not significantly associated with RI (Supplementary Fig. 12c).

Skin cutaneous melanoma (SKCM) and thyroid carcinoma (THCA) mean RIs were lower than predicted by their RAS pathway alteration ratio (Fig. 6c, indicated in orange). Interestingly, *NRAS* is the main mutated isoform of RAS in these two cancers. However, RAS mutation does not correlate with RI in SKCM (Supplementary Fig. 12d) and shows a significant inverse correlation in THCA (Fig. 6d) suggesting that mutant RAS is not the main driver of RAS pathway activity in these cancers. Interestingly THCA and SKCM have the highest proportion of BRAF mutation (57% and 50%) compared with the next most common, colon (10%). We found BRAF mutation to be significantly associated with high RI values (Wilcox p-value THCA < 2e−16, SKCM 2e−3) (Fig. 6e) suggesting BRAF to be a key driver of oncogenic RAS pathway activity in these two cancers. Given the lower-than-expected mean RI values it is possible that BRAF activation does not capture the full complexity of RAS pathway activation, possibly due to the use of multiple effector enzyme families by RAS proteins.

We measured a moderate RI in mesothelioma (MESO), kidney renal clear cell carcinoma (KIRC), prostate adenocarcinoma (PRAD), kidney renal papillary cell carcinoma (KIRP), kidney chromophobe (KICH) and liver hepatocellular carcinoma (LIHC) (Fig. 6c, indicated in blue). All fall below the lower 99% CI indicating they have higher than expected mean RI values given their RAS pathway alteration ratio. The absence of RAS pathway alterations suggests that these cancers activate RAS via other mechanisms than the RAS pathway alterations considered here.

To further validate that RAS84 expression could predict high RAS activity in RAS mutants in individual cancers, we looked at pan-RAS mutation (*KRAS, NRAS, HRAS*) distributions across RI values per cohort (Supplementary Fig. 12d). We identified a significant positive association between RAS mutations and RI in pancreatic (PAAD), LUAD, head and neck squamous (HNSC), thymoma (THYM), breast invasive carcinoma (BRCA), uterine carcinosarcoma (UCS), and uterine corpus
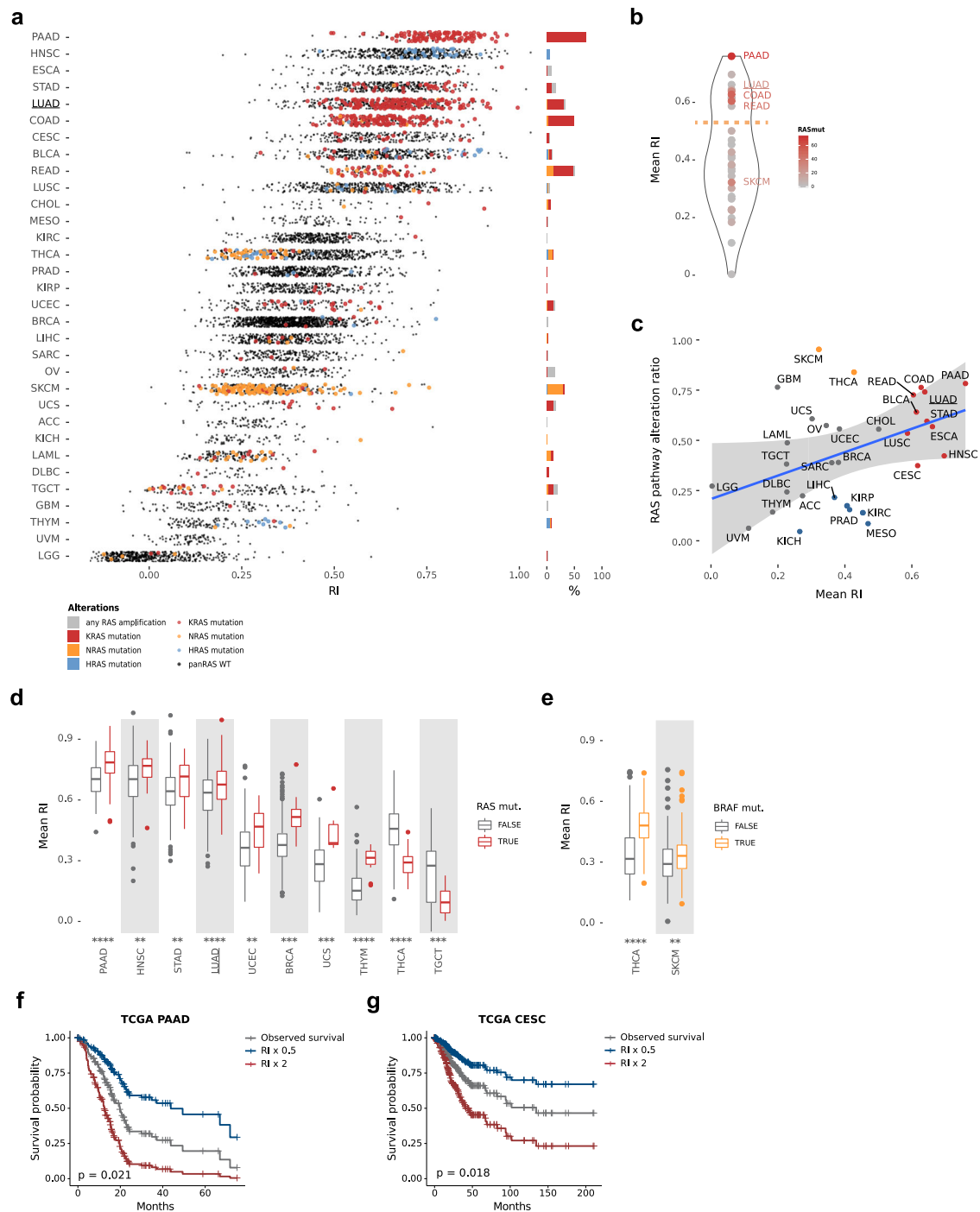
Fig. 6 | RAS Activity Pan-cancer. a RI plotted per patient across the TCGA pan-cancer cohort. RNA-Seq gene counts were VST and z-score normalised per cohort. RAS mutants are highlighted (KRAS in red, HRAS in blue and NRAS in orange). Relative RAS isoform mutation frequencies per cohort are shown in the barchart to the right using the same colours. Frequencies of patients with a RAS gene amplification but no mutation are shown in grey. **b** A violin plot showing the distribution of mean RI across each TCGA cohort. The dotted orange line indicates the distribution minima separating the two cancer populations. **c** The ratio of patients with one or more RAS pathway mutations plotted against the mean RI for each cohort. A linear regression fit is shown in blue with a 99% confidence interval shown by the grey ribbon (spearman coefficient 0.432, *p*-value 0.014). Highly RAS active tumours are shown in red, BRAF-driven tumours in orange and tumours with a RI value below the lower 99% CI are shown in blue. **d** Boxplots showing distributions of

RI values for pan-RAS mutant and wild-type patients. Significant cohorts are shown (linear model, for *n* see Supplementary Data 10; ****$P \le 0.0001$, ***$P \le 0.001$, **$P \le 0.01$, n.s.= $P > 0.05$). The box shows the median and IQR, the whiskers indicate ± 1.5 x IQR, outliers lie outside this range. **e** Boxplots showing distributions of RI values for THCA and SKCM split by BRAF mutation status (linear model fit THCA 281 mut, 192 wt; SKCM *n* = 180 mut, 180 wt; ****$P \le 0.0001$, ***$P \le 0.001$, **$P \le 0.01$, n.s.= $P > 0.05$). The box shows the median and IQR, the whiskers indicate ± 1.5 x IQR, outliers lie outside this range. **f, g** Kaplan–Meier plot showing a Cox proportional hazards regression fit of TCGA PAAD (**f**) or CESC (**g**) cohort survival data to corresponding RI values. The grey survival curve shows the observed data. The blue survival cure is the predicted survival if the RI values decreased 2-fold. The red curve are the predicted survival values if the RI were to increase 2-fold (coxph *p*-value 0.021 and 0.018, respectively).

endometrial carcinoma (UCEC) (Wilcoxon fdr < 0.05) (Fig. 6d, Supplementary Data 10), reinforcing the idea that RAS84 can predict RAS activity within these cancers. For two cancers, thyroid (THCA) and testicular germ cell tumours (TGCT), we observed a negative association of RAS mutation with RI. As mentioned previously, THCA has high *BRAF* mutation levels with some *NRAS* but few *KRAS* mutations. TGCT is characterised by low levels of both *KRAS* (8%) and *NRAS* (3%) mutations. The higher RI activity in the RAS wild-type group suggests that there may be a strong activator of RAS signalling in these samples, such as autocrine growth factor production, or a negative feedback loop in the mutants that is not apparent from the mutational data.

To determine if RAS84 had prognostic qualities in cancers other than LUAD, we ran Cox proportional hazards regression analyses of overall survival against RI values on cohorts within our highly RAS active cancer group. We were able to identify a prognostic association with RAS84 expression for CESC and PAAD (*p*-value < 0.05). Overall survival data with predicted outcome given a two-fold increase and decease in RI are shown in the survival plots (Fig. 6f, g). This result shows RAS activity as defined by RAS84 to be a possible prognostic indicator in other RAS-driven cancers, such as pancreatic (PAAD) and RAS pathway active cancers, such as cervical (CESC).

Discussion

The importance of mutations in *RAS* oncogenes in tumourigenesis, cancer progression and resistance to treatment has been demonstrated in numerous model systems in vitro and in vivo. However, the mutational status of *KRAS* has neither prognostic nor predictive value in human lung cancer, limiting the possibilities to anticipate patient survival and to adapt treatments. From this, one might be tempted to conclude that lung adenocarcinomas lacking a *KRAS* mutation have acquired functionally similar patterns of signalling network activation to those with *KRAS* mutations. However, the development here of a transcriptional measure of oncogenic RAS activity has allowed us to distinguish significant differences in outcome and response to chemotherapy between lung adenocarcinoma patients. We also show that a high proportion of *KRAS* wild-type tumours do exhibit the characteristics of RAS pathway activation. This analysis will facilitate the study of the effect of RAS pathway activity on survival, cancer progression and resistance to treatment in patients and could ultimately inform clinical decisions.

Several groups previously developed "RAS-addiction" or "MEK-sensitivity" signatures using RNA interference against RAS, MEK inhibitor or the RAS processing inhibitor salirasib to investigate resistance to RAS pathway targeted therapy[23,44–46]. Others have also reported approaches to assess RAS activity in tumours from expression data. Bild and colleagues mapped several transformation signatures to expression data from lung tumours and identified a population of patients with deregulation in RAS, Src, β-catenin and Myc activity and poor survival[24]. Sweet-Cordero and colleagues assessed the enrichment of a *Kras*[V12] tumour signature derived from a mouse cancer model in human tumours, observing an enrichment of the signature in human lung adenocarcinoma, but not specifically in KRAS mutants[47]. Nagy and colleagues reported a prognostic value in NSCLC using the mean expression of the top 5 deregulated genes in KRAS mutants versus non-mutants to segregate patients[48]. Way and colleagues used a machine learning approach and trained their classifier to detect *KRAS, HRAS* and *NRAS* mutations and copy number variation across cancers using the TCGA pan-cancer dataset [49]. Of the four published RAS signatures we tested, three predicted outcome in univariate survival analysis[24,43], but only RAS84 conserved its prognostic value when corrected for tumour stage in the multivariate analysis. This finding suggests we were able to extract the RAS-target-genes which capture RAS activity and not only tumour aggressiveness. Precedent RAS signatures likely contained genes associated with tumour progression, but our method based on the classification

of KRAS wt as RAS-low samples worked at extracting RAS-specific genes from these signatures.

We chose to derive our meta-signature in lung cancer because lung adenocarcinoma is known to be RAS-driven with about a 30% *KRAS* mutation rate, but also has 26% of patients that are RAS pathway wild type (no genetic alterations on any of the broader RAS pathway members), thus presenting a potentially wide RAS-activity dynamic range. Our method offers an alternative approach to previously published methods in that it does not rely on the initial segregation of RAS mutant and wild-type patients. We started from genes expressed in RAS active conditions, and we identified those that were good markers of KRAS mutants compared with non-activated RAS pathway tumours. This approach makes our method also sensitive to RAS active tumours driven by non-KRAS mutations. This nuance could explain the greater performance of our approach over others to capture the aggressiveness induced by oncogenic RAS activity. Moreover, using cell line data in the initial derivation of the signature provided a pure tumour cohort, free from the complexities of the tumour microenvironment, which could have introduced noise in identifying the driver genes, thus demonstrated by the low expression of RAS84 in stromal and immune cells.

Our analysis shows that 84% of lung adenocarcinomas exhibit clear evidence of RAS pathway activation independent of *KRAS* mutation status. We show that RAS-pathway-mutation burden is associated with RAS84 activity in our pan-cancer analysis, demonstrating the influence of other RAS-pathway-member mutations. However, there are undoubtedly other indirect mechanisms driving RAS oncogenic activity in KRAS wt tumours, such as epigenetic regulation, inter-exonic variants, influence of the tumour microenvironment and growth factor expression, negative feedback loop regulation, metabolic regulation or others. Using a transcriptional approach presents the advantage of being agnostic to all upstream regulation and negates the need for complete characterisation of all drivers affecting RAS signalling. We identified four groups with different degrees of RAS84 activity. The coincident mutations we observed in RAG-1 to −4 have previously been associated with specific phenotypes by Skoulidis et al. in a cohort of KRAS mutant lung cancer[31]. Our classification includes *KRAS* wt patients (representing 70% of all LUAD), thus broadening the clinical benefits of stratification to all patients.

Many studies over the years described the role of mutant *KRAS* in cancer progression, which might be expected to affect patient survival[5]. Oncogenic RAS promotes cell proliferation[50–55], suppresses apoptosis[56], shifts the metabolic program of cancer cells to sustain hyperproliferation[57,58], promotes angiogenesis[59], increases inflammation[60,61] and remodels the extracellular matrix[62,63]. Moreover, RAS promotes immune evasion by impairing antigen presentation[64], recruiting immunosuppressive cells[65,66] and inducing immune checkpoint ligand expression[10]. Despite extensive literature describing how oncogenic RAS increases tumour aggressiveness, these findings do not reflect patients' survival or response to treatment. An analysis of 227 patients with surgically resected NSCLC showed no association between RAS mutation and relapse[67] and a meta-analysis of 29 studies showed no relation between RAS mutation and survival in lung cancer[68]. The discrepancy between preclinical laboratory experiments and observations made in the clinic could be explained by the fact that some *KRAS* wt tumours can still activate the RAS pathway due to events other than RAS oncogene mutations, such as *BRAF* or *EGFR* mutation. Multiple factors can affect RAS activity from one tumour to another in human cancer, unlike in the controlled isogenic systems typically used in laboratory studies where the only perturbation is the RAS mutation. Based on transcriptional activity, our approach is thus at least partially agnostic to a precise position in the signalling network of genomic alterations, explaining its superiority to predict outcome. Additionally, RAS84 expression appears broadly clonal, showing that RAS activity is generally an early driver event in lung adenocarcinoma. This

observation is an essential consideration for developing clinical bio-markers when assessing tumour RAS activity from a single biopsy.

Adjuvant therapy is currently the first-line treatment for patients with early-stage lung cancers[69]. Although KRAS mutant promotes resistance to chemotherapy in isogenic experiments in vitro and in vivo[2–5], it has no predictive value in patients with lung cancer[6,7]. Using an independent cohort of lung adenocarcinoma (TEMPUS), we show that RAG-3 and -4 patients have a worse progression-free survival in response to first-line chemotherapy. This result is supported by the resistance to 23 chemotherapy drugs we observed in RAS-high lung cell lines in vitro. In both analyses, the classification based solely on the mutational status of KRAS did not reveal increased resistance to che-motherapy in KRAS-mutant cell lines or tumours. We also compared the response to drugs in cell lines mutated on any RAS pathway members versus all RAS pathway wild-type. Interestingly, this classifi-cation was not sufficient to predict resistance to chemotherapy, sug-gesting that other events –perhaps non-genetic– also affect the RAS oncogenic activity that we capture in our approach. Surprisingly, RAG-2 showed the best response to chemotherapy. The absence of coin-cident mutations on tumour suppressor genes may explain the better response observed in this group compared to other RAGs.

We also evaluated the RAS activity signature across different cancer types. In a pan-cancer analysis across all cancer types, we demonstrated high signature expression in tumours known to be RAS-driven, and we showed RAS84 expression to be predictive of RAS pathway mutation burden across cancer types. We predicted nine cancer types to be highly RAS pathway active, almost all of which had a high representation of mutations in RAS pathway genes. Five cancer types (liver hepatocellular carcinoma, kidney renal clear cell and papillary carcinomas, prostate adenocarcinoma and mesothelioma) with low RAS pathway mutation burden classified as moderate RAS pathway activity, indicating that events other than genomic alterations activate RAS signalling in these cancers. In addition, the relationship between RAS84 expression and RAS pathway gene mutations is dif-ferent in cancers with very high BRAF mutation levels, such as thyroid carcinoma and melanoma, which score as only moderate for RAS pathway activity. This suggests that the oncogenic RAS pathway's output varies depending on the mutated gene, with the BRAF mutation not being equivalent to KRAS mutations. This might be expected from our knowledge of the bifurcating nature of the RAS pathway with multiple effector enzyme families directly targeted by RAS proteins. Our results show a great variability of RAS activity amplitude across cancers, highlighting the importance of assessing RAS activity per cancer cohort. Interestingly, the correlation between RAS84 expres-sion and overall survival in pancreatic cancer, where 95% of tumours are mutated on KRAS, shows the direct link between RAS pathway-induced transcriptional activity and tumour aggressiveness.

We validated our observations in small cohorts of patients, which represents a limitation of our study, particularly since our classification stratifies the patients into five groups, limiting the number of patients in each group. Our finding that RAS84 correlates with resistance to chemotherapy needs to be validated in a larger cohort with more early-stage patients. Predicting response to chemotherapy in these patients is crucial to inform clinical decisions regarding the benefit of adjuvant therapies. It will be interesting to evaluate whether RAS84 predicts response to other treatments such as targeted therapy and immu-notherapy. Defining the predictive ability of RAS84 for immunother-apy would be particularly relevant since we and others have shown that mutant KRAS protein can modulate the expression of immunosup-pressive proteins[9,10]. Skoulidis and colleagues showed that KRAS and STK11/LKB1 co-occurring mutations are associated with poor response to PD-1 blockade in NSCLC patients[70]. In our classification, RAG-1 is enriched in STK11/LKB1 mutants, suggesting that this tumour group could be refractory to immune checkpoint blockade (ICB). In the same study, the authors showed that KRAS mutant; TP53 mutant tumours responded better to anti-PD-L1. RAG-3, and to a lesser extend RAG-4, are enriched in TP53 mutants, suggesting that these tumours could respond to ICB. About 20% of patients with NSCLC respond to ICB. The predictive value of RAS84 classification in response to ICB should be investigated in a follow-up study.

The 84 identified genes capture the complexity of RAS pathway-induced gene expression. However, the large number of genes could be a limitation to translating our classifier to a clinical assay. We therefore applied a random forest to our classifier to find the most important genes in driving the stratification in RAGs and found that a reduced number of 55 genes performed well at reproducing the clas-sification observed with RAS84, and predicted RAG-4 to have the worse prognostic in this cohort, but not in a multivariate analysis in an independent cohort. We also took another approach (ran a multi-variate Cox MIC analysis) to identify the RAS-target genes associated with outcome. We obtained a 11-gene signature which conserved its prognostic value in a multivariate analysis in an independent cohort and correlated with RAS activity. However, none of these two smaller signatures showed association with chemoresistance. It could be explained by the fact that different RAS-target genes capture different phenotypes of oncogenic RAS activity. Moreover, the number of genes required to stratify patients would increase with the number of output groups. However, it would be possible to focus on one or two specific groups and define the genes required to identify the patients belong-ing to these groups over the others. We showed RAG-3 and -4 tumours are refractory to chemotherapy. Identifying the driver genes in RAG-3 and -4 patient classification will be valuable in the future if the resis-tance to chemotherapy can be validated in a prospective cohort.

RAS84 captures RAS oncogenic activity in tumour samples better than the mutational status of KRAS when applied to cohorts of lung adenocarcinoma patients and other cancer types. We believe that the stratification of patients based on RAS84 expression will facilitate the study of the effect of RAS on survival, cancer progression and resis-tance to treatment in patients and could ultimately help clinical deci-sion making.

Methods
Selection of the founder gene sets
We selected gene sets from several published data: the RAS addiction signature contained 380 genes upregulated in 5 KRAS-dependent cell lines (4 lung cell lines and 1 pancreatic cell line) compared with 5 KRAS-independent cell lines (4 lung cell lines and 1 pancreatic cell line)[23]; the KrasLA signature contained 89 upregulated genes in mouse lung tumours induced by the spontaneous recombination of KrasLA2 allele compared with normal lung and expressed in human lung tumours[47]; the HRAS transformation signature contained 245 genes correlating with the classification of HMEC samples into oncogene-activated/deregulated versus control[24]; the RAS pathway signature contained 105 genes previously curated from 3 studies including HRAS transforma-tion, KrasLA and a signature of Salirasinib-treated human cancer cell lines[45,46]; the MSigDB signature is the HALLMARK_KRAS_SIGNA-LING_UP meta-signature from MSigDB, which contained a list of 200 genes identified from overlaps between KRAS-related gene sets in other MSigDB collections[25]. We also generated a gene set from in-house data. The data was previously generated using the colon cancer cell line HCT116, which carries a KRASG13D mutation, and its isogenic cell lines Hke3 and Hkh2 where the $KRAS^{G13D}$ allele was deleted by homologous recombination[71]. An Affymetrix analysis was performed on parental and recombined cell lines, and on sh-KRAS and sh-control in the HCT116 cell line. $KRAS^{G13D}$134 was derived by selecting the genes upregulated (L2FC > 1.5) when $KRAS^{G13D}$ was expressed in all three experiments: HCT116 sh-KRAS versus control (6 days), Hke3 versus HCT116 and Hkh2 versus HCT116. We also identified a number of other oncogenic expression signatures to use as controls throughout the analysis (Supplementary Data 11).

Affymetrix analysis of HCT116 sh-*KRAS*

HCT116 cells stably expressing inducible sh-*RAS* vectors[72] were treated with 25 ng/ml doxycycline or vehicle control for 3 days or 6 days. Hke3, Hkh-2 and parental HCT116 were cultured in basal condition. All conditions were done in triplicate. RNA samples were extracted with Qiagen RNeasy kit (Qiagen). RNA samples were then quantified, and 260/280 ratios were measured using NanoDrop spectrophotometers (Thermoscientific) and tested for integrity using a Bioanalyzer system (Agilent). RNA samples were labelled using Affymetrix One Cycle Target labelling Kit (Affymetrix) following the manufacturer's instruction and sequenced using a GeneChip Human Genome U133 Plus 2.0 Assay (Affymetrix).

CCLE

The CCLE microarray expression and mutation data were obtained from the CCLE legacy repository hosted at The Broad (https://data.broadinstitute.org/ccle_legacy_data). We selected lung-derived cell line data. We labelled the cell lines as either *KRAS* mutation positive, RAS pathway mutation positive or RAS pathway mutation negative. Cell lines were labelled as RAS pathway mutation negative if they had no mutation in a RAS pathway gene member defined in Sanchez-Vega and colleagues[19]. We removed cell lines with mutations in RAS pathway members other than *KRAS* from further analysis selecting 166 cell lines (52 small-cell carcinoma, 46 adenocarcinoma, 26 squamous-cell carcinoma, 19 NSCLC with no subtype specified, and 23 without label or had another histology type). We filtered the RAS signatures for genes relevant in the context of lung cancer. We modelled the log COV of expression against the mean log2 expression across all genes using loess regression. Signature genes with positive residuals with respect to the fit and a log2 mean expression value >6 were selected.

We mapped each of the signatures to the cell line expression data and clustered the cell lines using hierarchical clustering with a Ward.D2 agglomeration method. We split the resultant dendrograms into three clusters, labelling each high, low and unclassified for signature expression. The labels were assigned based on ranked mean cluster expression. To assess the performance of each of the signatures we calculated the significance of association of high and low clusters with *KRAS* mutation using a chisq test. To further refine each of the top 3 performing signatures, we ran a differential analysis between the high and low clusters using the Limma package from Bioconductor. We selected signature genes with an fdr <0.05 as those driving the clustering. We merged the differential genes for each signature tested to form our meta-signature, RAS84. We identified RAS-high dependent transcriptional changes between RAS-high and RAS-low groups by limma (3.40.2) analysis on CCLE RMA normalised intensity estimates. Genes with a RAS group mean intensity estimate <6 in both groups were removed from the analysis. Differential genes were selected by FDR < 0.05 and absolute LFC >1. We ran the GO analysis using the clusterProfiler (3.12.0) package from Bioconductor testing all Biological Process terms from org.Hs.eg.db_(3.8.2) (FDR < 0.001).

CCLE drug sensitivity screen

We obtained drug sensitivity data from GDSC (IC50) (v1 367, v2 198 compounds)[73] and CTRP (AUC) (v1 185, v2 481 compounds)[74] for the CCLE cell lines. We clustered the VST normalised RAS84 CCLE RNA-Seq data into two clusters, RAS high and RAS low. We tested for significant differences in drug response values across the two RAS activity clusters by linear model correcting for any KRAS mutation status effect (<0.05 fdr). We analysed each of the two release versions separately for each of the two data repositories. We identified enriched compound target pathways in the GDSC results by hypergeometric test using the TARGET_CATEGORY annotation provided (< 0.05 fdr). We also tested for oncogenic KRAS mutant dependent and oncogenic KRAS pathway dependent drug responses in the GDSC data. We used genotype data

from the CCLE. We called the RAS pathway as mutated if any of the pathway genes contained a mutation[19].

Patient datasets

TCGA pancancer data. All TCGA RNA-Seq gene-level read-counts were downloaded using the TCGAbiolinks (TCGAbiolinks_2.8.4) package from Bioconductor[75] (legacy = TRUE). Raw counts were VST normalised using the varianceStabilizingTransformation function within DESeq2 (DESeq2_1.20.0) from Bioconductor[76]. Normal samples were removed prior to analysis. To compare across the cancer cohorts, we z-score normalised samples and genes. Mutation data were obtained from Sanchez-Vega and colleagues[19] and specific *KRAS* mutation data was downloaded from TCGAbiolinks and integrated with the expression data. Survival and proliferation data were obtained from Thorsson and colleagues[77]. RAS84 gene annotations were mapped to the RNA-Seq feature ids (Supplementary Data 2).

Seoul lung adenocarcinoma cohort, GSE40419. RNA-Seq RPKM values for 87 adenocarcinoma patients were downloaded from GEO using the getGEO function from the GEOQuery Bioconductor package. RPKM values were log2 transformed prior to cluster analysis. Mutation data were obtained from Seo and colleagues[78]. Where multiple features existed per-gene the one with the maximum mean expression value across the cohort was selected.

Uppsala II RNA-Seq. RNA-Seq gene-level read-counts and clinical data were downloaded from the Gene Expression Omnibus (GEO GSE81089). The raw counts were VST normalised using the varianceStabilizingTransformation function within DESeq2 (DESeq2_1.20.0) from Bioconductor[76]. Ensembl gene annotations were obtained using the biomaRt package from Bioconductor. The 103 stage I, II, & III adenocarcinoma samples were selected prior to further analysis (column histology:ch1 == 2, stage.tnm.ch1!= 7). RAS84 gene annotations were mapped to the RNA-Seq feature ids (Supplement data 2). In the case of IER3, which maps to multiple features in this dataset, the feature with the largest mean VST value across all samples was selected (ENSG00000137331).

TRACERx. RNA-Seq gene-level read-counts were obtained for the TRACERx 100 patient cohort. The counts were VST normalised using the varianceStabilizingTransformation function within DESeq2 (DESeq2_1.20.0) from Bioconductor[76]. The 102 adenocarcinoma samples were selected (Histology == "Invasive adenocarcinoma"). The counts were further z-score normalised prior to SVM classification. RAS84 gene annotations were mapped to the RNA-Seq feature ids (Supplement data 2).

Lambrechts scRNA-Seq. Log2 CPM normalised scRNASeq data for B, T, fibroblasts, alveolar, EC and myeloid cells from five lung carcinomas were obtained from ArrayExpress (E-MTAB-6149).

TEMPUS CLINIC-GENOMICS. TEMPUS clinic-genomics is a retrospective lung cancer cohort database containing 1,711 patients. Clinical data were extracted from the TEMPUS real-world oncology database of longitudinal structured and unstructured data from geographically diverse oncology practices, including integrated delivery networks, academic institutions, and community practices. All data were de-identified in accordance with the Health Insurance Portability and Accountability Act (HIPAA). The database extract was retrieved and de-identified in 2018 and contained cohorts with patients' records spanning from 1990–2018. We identified 108 adenocarcinoma patients from the TEMPUS database with first-line chemotherapy treatment and matched RNA-seq molecular data. We calculated a progression-free survival interval from the associated patient clinical histories. We took the start time of treatment as time zero. We used a recorded

recurrence event, a reported progressive disease outcome, a progression in reported tumour stage, death or the administration of an alternative therapy as an endpoint to progression-free interval. In the cases of repeated chemotherapy treatment, we took a gap between treatments of >6 months as a PFS endpoint. In the absence of any endpoint events, we censored on the last follow-up time if no neoplasm was recorded or the last reported outcome if it was one of stable disease, progression-free or partial response. We also integrated stage, the administration of radiotherapy, age and sex data. We applied VST and z-score normalisation to the RNA-Seq gene-level counts across all 633 adenocarcinoma patients in the cohort. We classified patients into RAG groups using our SVM classifier.

RAG Classification

The RAS84 TCGA LUAD VST expression matrix was clustered using hierarchical clustering with a Euclidian distance measure and a ward.D2 agglomeration method (hclust function, R). We split the dendrogram into five clusters and labelled them RAG-0 to RAG-4 based on their mean RAS84 expression value across all samples, lowest to highest. We also calculated an RI value for each sample, defined as the mean VST value across the RAS84 genes. We repeated this analysis for each of the original published signatures. We assessed the performance of each of the signatures by testing for statistically significant differences in the observed *KRAS* mutation frequencies across the five groups using a chi-squared test (chisq.test function, R). We tested all somatic variants reported in Sanchez-Vega and colleagues[19] ($N > 10$) for significant frequency differences across the RAS84 clusters using a chi-square test (fdr < 0.05). Mosaic plots were generated using the vcd package from R (vcd_1.4-4). Specific *KRAS* mutation genotypes were tested individually against a background of all remaining samples using a chi-squared test. The Seoul cohort (GSE40419) was clustered in the same way as the TCGA samples using RAS84KRAS, EGFR and TP53 mutations were tested for significant differences in observed frequencies across the five RAGs using a chi-squared test. To determine KRAS mutation segregation across RAS84 and the published signatures we tested KRAS mutation status against signature mean expression distributions using a linear model. To identify RAG driver genes, we identified genes with the largest deviation in expression from the mean across all samples. We first calculated RAG mean expression values per gene. We then scaled these values to the mean across all samples to calculate a RAG deviation value. Genes with an absolute deviation of >1 were selected. These genes were clustered across RAG mean values by hierarchical clustering using Pearson's correlation and ward.D2 agglomeration (cor and hclust function, R). We identified RAG dependent transcriptional changes by comparing RAGs 1–4 to RAG-0 correcting for tumour purity in the model since we were interested in tumour-specific effects. Tumour purity CPE values were obtained from[79]. Differential genes were identified using DESeq2 (1.24.0) (fdr < 0.05) and shrunken LFC values were generated using the lfcShrink function with type = "ashr"[80]. The genes were further filtered using the shrunken LFC values prior to GO analysis (absolute shrunkenLFC > 1). Go analysis was carried out using goseq (1.36.0) from Bioconductor, accounting for the length bias inherent in RNA-Seq results. Only terms associated with 'Biological Process' were considered, and enriched *p*-values were corrected using Benjamini-Hochberg correction (FDR < 0.05).

RPPA

We obtained level 4 normalised TCPA LUAD RPPA data from https://tcpaportal.org/tcpa/download.html. We identified 349 samples in common between the TCPA LUAD RPPA cohort and our classified TCGA LUAD cohort. We found 216 assayed proteins with values across all samples. We fitted a linear model across the RAGs against RAG-0 as the control for each protein assay. A Benjamini-Hochberg FDR multiple

testing correction was applied across all tests (lm and p.adjust functions from R).

Lung adenocarcinoma OS and PFS analysis

We fitted a univariate Cox proportional-hazard model to RAG labels and TCGA LUAD overall survival data to test RAG as a predictor for outcome using RAS84 and the published signatures. We also fitted a univariate Cox proportional-hazard regression model using RI as a continuous predictor of outcome. We ran OS univariate and multivariate Cox proportional-hazard analysis against RAG and RI in the Uppsala cohort. Overall survival time was calculated by subtracting the surgery date from the vital date (columns vital.date.ch1 - surgery.date.ch1). We compared a reduced coxph model including TNM stage, World Health Organization (WHO) performance status, smoking history, sex and age covariates (columns stage.tnm.ch1, ps.who.ch1, smoking.ch1, gender.ch1 and age.ch1) to a full model including either RAG labels or RI values, using LRT with ANOVA. We accounted for possible non-linear age effects by applying a restricted cubic spline to age using the rcs function from the rms R package (rcs(age, 3)). We carried out this analysis using RAS84 and published signature RAG classifications. We determined RAG classification using the clustering method used in the TCGA cohort classification. We performed a PFS multivariate coxph analysis using the TEMPUS patients. We constructed a reduced model using stage, radiotherapy, gender and age covariates applying a restricted cubic spline as above. We compared this model to a full model including RAG labels using LRT with anova. We assigned RAS84 RAG labels using our SVM classifier. We assigned public signature RAG labels using the clustering approach applied in the TCGA cohort analysis. In our pancancer analysis we fitted univariate Cox proportional-hazard regression models to RI and overall survival data from each TCGA cancer cohort from our high RAS activity cancer group. We identified cancers where RI was a significant predictor of outcome (*p*-value < 0.05). Kaplan–Meier curves were produced as above. These analyses were carried out using the coxph function from the survival R package (survival_3.1–11). Kaplan–Meier curves were produced using the ggsurvplot function from the survminer R package (survminer_0.4.4). In the case of the RI analysis the Kaplan–Meier curves were generated using predicted survival data from the coxph model given a 2-fold increase or decrease in RI. We ran a multivariate Cox MIC analysis[81] on TCGA OS with all RAS84 VST expression vectors as covariates. We used the coxMIC function (method = "BIC") from the coxphMIC R package (coxphMIC_0.1.0). We identified 11 genes with a non-zero beta.MIC value.

Pan cancer RAS84 analysis

To compare RAS84 expression pan-cancer we z-score normalised the previously VST normalised RNA-Seq TCGA data (see section: TCGA Pancancer Data) and merged across the 32 cohorts. To assess RAS activity per sample we calculated an RI value for each sample, defined as the mean expression across the RAS84 genes. To identify high and low RAS active tumours we plotted the distributions of the mean RI values per cohort. From the observed bimodal distribution of cancer RI mean values, we calculated kernel density estimates (density function, R) and split the tumours at the minima between the two population maxima (mean RI value 0.53) (Supplementary Fig. 5a). We identified variants enriched in the high RAS activity tumours by hypergeometric test (fdr < 0.05). To obtain a RAS pathway mutation view we calculated a RAS pathway mutation burden percentage for each sample. We labelled samples as being RAS pathway mutated if they had a mutation in any gene defined in the RAS pathway by Sanchez-Vega et al. We tested for a significant correlation between RAS pathway mutation burden and mean RI using a Pearson's correlation test.

To test for an association between RI values and RAS mutation status, per cancer, we ran Wilcoxon tests across the RAS mutated and

non-mutated groups. We merged the oncogenic mutation calls for KRAS, HRAS and NRAS prior to testing. Cancers with a total RAS mutation count <5 were excluded from the analysis. False discovery rates were calculated to account for multiple testing. We identified significant cancers by FDR < 0.05.

SVM classifier

To facilitate the RAS classification of lung adenocarcinoma RNA-Seq samples we constructed an SVM classifier. We used the RAG labels derived from the LUAD TCGA cluster analysis as class labels and the TCGA LUAD RAS84 expression matrix as training data. The raw gene-level counts were first VST and z-score normalised. We constructed a Radial Sigma SVM using the caret R package[76]. The train function was used to optimise the classifier using a cv resampling strategy with 10 iterations. The classifier was validated against a pre-selected test sub-set of patients (80% training, 20% test). We determined the importance of each of the RAS84 genes in classifying the TCGA patients by fitting a RandomForest model to the data described above. We used the genes ranked by importance to build iterative gene sets which we used to construct new SVM classifiers as described above. We tested these SMVs against our test data to determine model performance, sensitivity and specificity across the importance gene rank.

TRACERx classification

We classified the samples using the z-score VST RAS84 expression matrix and our SVM classifier constructed from the TCGA analysis results and detailed in the classifier methods section above. We called the presence SNVs in each of the groups (PhyloCCF score > 0.05). We calculated a sample distance matrix from the VST expression matrix using Euclidian distance. We plotted the density of distance measures for intra- and inter-tumour distances. We determined the degree of enrichment of RAS84 genes with stable intra-tumour expression, but high inter-tumour variance was assessed relative to the distributions of all expressed genes, in-line with the methods presented in Biswas and colleagues[36].

Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The TRACERx tumour region gene-level RNA sequencing count data of 84 genes used during this study are available through the Cancer Research UK & University College London Cancer Trials Centre (ctc.tracerx@ucl.ac.uk) for non-commercial research purposes, and access will be granted upon review of a project proposal that will be evaluated by a TRACERx data access committee and entering into an appropriate data access agreement subject to any applicable ethical approvals. The TEMPUS data is commercially available from TEMPUS Lab (https://www.tempus.com/life-sciences/data-collaborations/#contact). Affymetrix data that was used to generate the KRAS[G13D]134 signature have been deposited in GEO with the accession code GSE199871. The Uppsala cohort data used in this study are available in the GEO database under accession code GSE81089. The Seoul cohort data used in this study are available in the GEO database under accession code GSE40419. The Lambrechts data used in this study are available in the ArrayExpress database under accession code E-MTAB-6149. All other data associated with this study are present in the paper or supplementary materials, or as cited.

Code availability

All code used in this manuscript is available at https://github.com/FrancisCrickInstitute/RAS84.

References

1. Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J. & Der, C. J. Drugging the undruggable RAS: Mission Possible? *Nat. Rev. Drug Disco.* **13**, 828–851 (2014).
2. Tao, S. et al. Oncogenic KRAS confers chemoresistance by upregulating NRF2. *Cancer Res.* **74**, 7430–7441 (2014).
3. Grabocka, E. & Bar-Sagi, D. Mutant KRAS enhances tumor cell fitness by upregulating stress granules. *Cell* **167**, 1803–1813.e12 (2016).
4. Caiola, E. et al. Base excision repair-mediated resistance to cisplatin in KRAS(G12C) mutant NSCLC cells. *Oncotarget* **6**, 30072–30087 (2015).
5. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774 (2011).
6. Wiesweg, M. et al. Impact of RAS mutation subtype on clinical outcome—a cross-entity comparison of patients with advanced non-small cell lung cancer and colorectal cancer. *Oncogene* **38**, 2953–2966 (2019).
7. Cserepes, M. et al. Subtype-specific KRAS mutations in advanced lung adenocarcinoma: a retrospective study of patients treated with platinum-based chemotherapy. *Eur. J. Cancer* **50**, 1819–1828 (2014).
8. Shepherd, F. A. et al. Pooled analysis of the prognostic and predictive effects of KRAS mutation status and KRAS mutation subtype in early-stage resected non–small-cell lung cancer in four trials of adjuvant chemotherapy. *J. Clin. Oncol.* **31**, 2173–2181 (2013).
9. Carvalho, P. D. et al. KRAS oncogenic signaling extends beyond cancer cells to orchestrate the microenvironment. *Cancer Res.* **78**, 7–14 (2018).
10. Coelho, M. A. et al. Oncogenic RAS signaling promotes tumor immunoresistance by stabilizing PD-L1 mRNA. *Immunity* **47**, 1083–1099.e6 (2017).
11. Herbst, R. S. et al. LBA4 Association of KRAS mutational status with response to pembrolizumab monotherapy given as first-line therapy for PD-L1-positive advanced non-squamous NSCLC in Keynote-042. *Ann. Oncol.* **30**, xi63–xi64 (2019).
12. Gadgeel, S. et al. KRAS mutational status and efficacy in KEYNOTE-189: Pembrolizumab (pembro) plus chemotherapy (chemo) vs placebo plus chemo as first-line therapy for metastatic non-squamous NSCLC. *Ann. Oncol.* **30**, xi64–xi65 (2019).
13. Ji, M. et al. PD-1/PD-L1 expression in non-small-cell lung cancer and its correlation with EGFR/KRAS mutations. *Cancer Biol. Ther.* **17**, 407–413 (2016).
14. Passiglia, F. et al. Efficacy of nivolumab in pre-treated non-small-cell lung cancer patients harbouring KRAS mutations. *Brit J. Cancer* **120**, 57–62 (2019).
15. Adderley, H., Blackhall, F. H. & Lindsay, C. R. KRAS-mutant non-small cell lung cancer: converging small molecules and immune checkpoint inhibition. *Ebiomedicine* **41**, 711–716 (2019).
16. Kartolo, A., Feilotter, H., Hopman, W., Fung, A. S. & Robinson, A. A single institution study evaluating outcomes of PD-L1 high KRAS-mutant advanced non small cell lung cancer (NSCLC) patients treated with first line immune checkpoint inhibitors. *Cancer Treat. Res. Commun.* **27**, 100330 (2021).
17. Román, M. et al. KRAS oncogene in non-small cell lung cancer: clinical perspectives on the treatment of an old target. *Mol. Cancer* **17**, 33 (2018).
18. Lièvre, A. et al. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res.* **66**, 3992–3995 (2006).
19. Sanchez-Vega, F. et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).

20. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

21. Angelova, M. et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol.* **16**, 64 (2015).

22. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

23. Singh, A. et al. A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell* **15**, 489–500 (2009).

24. Bild, A. H. et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2005).

25. Liberzon, A. et al. The Molecular Signatures Database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

26. De Raedt, T. et al. Exploiting cancer cell vulnerabilities to develop a combination therapy for ras-driven tumors. *Cancer Cell* **20**, 400–413 (2011).

27. Collisson, E. A. et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).

28. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).

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

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

31. Skoulidis, F. et al. Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. *Cancer Discov.* **5**, 860–877 (2015).

32. Li, J. et al. TCPA: a resource for cancer functional proteomics data. *Nat. Methods* **10**, 1046–1047 (2013).

33. Kruspig, B. et al. The ERBB network facilitates KRAS-driven lung tumorigenesis. *Sci. Transl. Med.* **10**, eaao2565 (2018).

34. Moll, H. P. et al. Afatinib restrains K-RAS–driven lung tumorigenesis. *Sci. Transl. Med.* **10**, eaao2301 (2018).

35. Lee, W.-C. et al. Multiregion gene expression profiling reveals heterogeneity in molecular subtypes and immunotherapy response signatures in lung cancer. *Mod. Pathol.* **31**, 947–955 (2018).

36. Biswas, D. et al. A clonal expression biomarker associates with lung cancer mortality. *Nat. Med.* **25**, 1540–1548 (2019).

37. Rosenthal, R. et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).

38. Loriot, Y., Mordant, P., Deutsch, E., Olaussen, K. A. & Soria, J.-C. Are RAS mutations predictive markers of resistance to standard chemotherapy? *Nat. Rev. Clin. Oncol.* **6**, 528–534 (2009).

39. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* **3**, 11–22 (2003).

40. Prior, I. A., Lewis, P. D. & Mattos, C. A comprehensive survey of ras mutations in cancer. *Cancer Res.* **72**, 2457–2467 (2012).

41. Helgadottir, H. & Höiom, V. The genetics of uveal melanoma: current insights. *Appl Clin. Genet.* **9**, 147–155 (2016).

42. Zhang, Y., Xia, Q. & Lin, J. Identification of the potential oncogenes in glioblastoma based on bioinformatic analysis and elucidation of the underlying mechanisms. *Oncol. Rep.* **40**, 715–725 (2018).

43. Linehan, W. M. Genetic basis of kidney cancer: Role of genomics for the development of disease-based therapeutics. *Genome Res.* **22**, 2089–2100 (2012).

44. Loboda, A. et al. A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. *BMC Med. Genomics* **3**, 26 (2010).

45. Blum, R. et al. Gene expression signature of human cancer cell lines treated with the ras inhibitor Salirasib (S-Farnesylthiosalicylic acid). *Cancer Res.* **67**, 3320–3328 (2007).

46. Dry, J. R. et al. Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244). *Cancer Res.* **70**, 2264–2273 (2010).

47. Sweet-Cordero, A. et al. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat. Genet.* **37**, 48–55 (2004).

48. Nagy, Á., Pongor, L. S., Szabó, A., Santarpia, M. & Győrffy, B. KRAS driven expression signature has prognostic power superior to mutation status in non-small cell lung cancer: Kras driven expression signature in NSCLC. *Int J. Cancer* **140**, 930–937 (2017).

49. Way, G. P. et al. Machine learning detects pan-cancer ras pathway activation in The Cancer Genome Atlas. *Cell Rep.* **23**, 172–180.e3 (2018).

50. McCarthy, S. A., Samuels, M. L., Pritchard, C. A., Abraham, J. A. & McMahon, M. Rapid induction of heparin-binding epidermal growth factor/diphtheria toxin receptor expression by Raf and Ras oncogenes. *Gene Dev.* **9**, 1953–1964 (1995).

51. Filmus, J. et al. Induction of cyclin D1 overexpression by activated ras. *Oncogene* **9**, 3627–3633 (1994).

52. Albanese, C. et al. Transforming p21 ras mutants and c-Ets-2 activate the cyclin D1 promoter through distinguishable regions. *J. Biol. Chem.* **270**, 23589–23597 (1995).

53. Winston, J. T., Coats, S. R., Wang, Y. Z. & Pledger, W. J. Regulation of the cell cycle machinery by oncogenic ras. *Oncogene* **12**, 127–134 (1996).

54. Schulze, A., Lehmann, K., Jefferies, H. B. J., McMahon, M. & Downward, J. Analysis of the transcriptional program induced by Raf in epithelial cells. *Gene Dev.* **15**, 981–994 (2001).

55. Gangarosa, L. M. et al. A Raf-independent epidermal growth factor receptor autocrine loop is necessary for ras transformation of rat intestinal epithelial cells. *J. Biol. Chem.* **272**, 18926–18931 (1997).

56. Cox, A. D. & Der, C. J. The dark side of Ras: regulation of apoptosis. *Oncogene* **22**, 8999–9006 (2003).

57. Chen, C., Pore, N., Behrooz, A., Ismail-Beigi, F. & Maity, A. Regulation of glut1 mRNA by Hypoxia-inducible Factor-1. *J. Biol. Chem.* **276**, 9519–9525 (2000).

58. Flier, J., Mueckler, M., Usher, P. & Lodish, H. Elevated levels of glucose transport and transporter messenger RNA are induced by ras or src oncogenes. *Science* **235**, 1492–1495 (1987).

59. Rak, J. et al. Mutant ras oncogenes upregulate VEGF/VPF expression: implications for induction and inhibition of tumor angiogenesis. *Cancer Res.* **55**, 4575–4580 (1995).

60. Sparmann, A. & Bar-Sagi, D. Ras-induced interleukin-8 expression plays a critical role in tumor growth and angiogenesis. *Cancer Cell* **6**, 447–458 (2004).

61. Ancrile, B. B., O'Hayer, K. M. & Counter, C. M. Oncogenic ras-induced expression of cytokines: a new target of anti-cancer therapeutics. *Mol. Inter.* **8**, 22–27 (2008).

62. Lengyel, E. et al. Regulation of urokinase-type plasminogen activator expression by an ERK1-dependent signaling pathway in a squamous cell carcinoma cell line. *J. Cell Biochem.* **61**, 430–443 (1996).

63. Sato, H. et al. Expression of genes encoding type IV collagen-degrading metalloproteinases and tissue inhibitors of metalloproteinases in various human tumor cells. *Oncogene* **7**, 77–83 (1992).

64. Lohmann, S., Wollscheid, U., Huber, C. & Seliger, B. Multiple levels of MHC class I down-regulation by ras oncogenes. *Scand. J. Immunol.* **43**, 537–544 (1996).

65. Clark, C. E. et al. Dynamics of the immune reaction to pancreatic cancer from inception to invasion. *Cancer Res.* **67**, 9518–9527 (2007).

66. Thang, N. N. T. et al. Immune infiltration of spontaneous mouse astrocytomas is dominated by immunosuppressive cells from early stages of tumor development. *Cancer Res.* **70**, 4829–4839 (2010).

67. Moldvay, J. et al. Predictive survival markers in patients with surgically resected non-small cell lung carcinoma. *Clin. Cancer Res.* **6**, 1125–1134 (2000).

68. Mascaux, C. et al. The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis. *Brit J. Cancer* **92**, 131–139 (2005).

69. Pignon, J.-P. et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J. Clin. Oncol.* **26**, 3552–3559 (2008).

70. Skoulidis, F. et al. STK11/LKB1 mutations and PD-1 inhibitor resistance in KRAS-mutant lung adenocarcinoma. *Cancer Discov.* **8**, 822–835 (2018).

71. Shirasawa, S., Furuse, M., Yokoyama, N. & Sasazuki, T. Altered growth of human colon cancer cell lines disrupted at activated Ki-ras. *Science* **260**, 85–88 (1993).

72. Ngo, V. N. et al. A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110 (2006).

73. Iorio, F. et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754 (2016).

74. Rees, M. G. et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116 (2016).

75. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2015).

76. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

77. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).

78. Seo, J.-S. et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22**, 2109–2119 (2012).

79. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

80. Stephens, M. False discovery rates: a new deal. *Biostatistics* kxw041. https://doi.org/10.1093/biostatistics/kxw041 (2016).

81. Nabi, R. & Su, X. coxphMIC: an R package for sparse estimation of Cox proportional hazards models. Preprint at https://arxiv.org/abs/1606.07868 (2016).

Acknowledgements

Author contributions

P.E., S.C.T., C.S. and J.D. designed the study, interpreted the results and wrote the manuscript. P.E. and D.B. performed the computational analyses. K.S., T.C., M.M. and D.H. provided datasets for analysis. G.K. carried out statistical analysis. All authors contributed to manuscript revision and review.

Funding

Competing interests

J.D. has acted as a consultant for AstraZeneca, Bayer, Novartis, TheRas, Vividion, Jubilant and has received research funding from AstraZeneca, Bristol Myers Squibb and Revolution Medicines. SCT has acted as a consultant for Revolution Medicines. C.S. is a co-founder of Achilles Therapeutics, receives grant support from Pfizer, AstraZeneca, BMS, Roche–Ventana and Boehringer Ingelheim, has consulted for Pfizer, Novartis, GlaxoSmithKline, MSD, BMS, Celgene, AstraZeneca, Illumina, Genentech, Roche–Ventana, GRAIL, Medicxi and the Sarah Cannon Research Institute, is a shareholder of ApoGen Biotechnologies, Epic Bioscience and GRAIL. K.S. and T.C. were employees of AstraZeneca at the time the study was carried out. None of the other authors of this manuscript have a financial or non-financial interest related to this work.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-33290-0.

Correspondence and requests for materials should be addressed to Julian Downward or Sophie de Carné Trécesson.

Peer review information *Nature Communications* thanks Daniel Murphy the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

TRACERx consortium

Charles Swanton ⓘ² **& Dhruva Biswas** ⓘ²