IJC
International Journal of Cancer

KRAS driven expression signature has prognostic power superior to mutation status in non-small cell lung cancer

Ádám Nagy[1,2], Lőrinc Sándor Pongor[1,2], András Szabó[2], Mariacarmela Santarpia[3] and Balázs Győrffy[1,2]

[1] MTA TTK Lendület Cancer Biomarker Research Group, Budapest, Magyar, Hungary
[2] Semmelweis University 2nd Department of Pediatrics, Budapest, Hungary
[3] Medical Oncology Unit, Department of Human Pathology 'G. Barresi', University of Messina, Italy

KRAS is the most frequently mutated oncogene in non-small cell lung cancer (NSCLC). However, the prognostic role of KRAS mutation status in NSCLC still remains controversial. We hypothesize that the expression changes of genes affected by KRAS mutation status will have the most prominent effect and could be used as a prognostic signature in lung cancer.

We divided NSCLC patients with mutation and RNA-seq data into KRAS mutated and wild type groups. Mann-Whitney test was used to identify genes showing altered expression between these cohorts. Mean expression of the top five genes was designated as a "transcriptomic fingerprint" of the mutation. We evaluated the effect of this signature on clinical outcome in 2,437 NSCLC patients using univariate and multivariate Cox regression analysis.

Mutation of KRAS was most common in adenocarcinoma. Mutation status and KRAS expression were not correlated to prognosis. The transcriptomic fingerprint of KRAS include FOXRED2, KRAS, TOP1, PEX3 and ABL2. The KRAS signature had a high prognostic power. Similar results were achieved when using the second and third set of strongest genes. Moreover, all cutoff values delivered significant prognostic power ($p < 0.01$). The KRAS signature also remained significant ($p < 0.01$) in a multivariate analysis including age, gender, smoking history and tumor stage.

We generated a "surrogate signature" of KRAS mutation status in NSCLC patients by computationally linking genotype and gene expression. We show that secondary effects of a mutation can have a higher prognostic relevance than the primary genetic alteration itself.

Tumor Markers and Signatures

Despite advances in the last decade, lung cancer remains the most lethal tumor in women and men still exceeding the combined mortality of breast, prostate, colorectal and pancreatic cancers.[1] Non-small cell lung cancer (NSCLC) accounts for nearly 85% of all lung cancer cases and is further classified into different subtypes including adenocarcinoma (AC), squamous cell carcinoma and large cell carcinoma.[2] Oncogenic mutations not only enable tumor development but also delineate new anticancer therapy targets.[3] To date, five molecularly targeted agents have been approved for the treatment of NSCLC. Targeting the epidermal growth factor receptor (EGFR) was the first pioneer of personalized therapy in lung cancer.[4,5] EGFR kinase domain mutations are present in about 20% of AC patients. Among others, exon 19 deletion and point mutations in exon 18 and 21 deliver high sensitivity to EGFR tyrosine kinase inhibitors (TKIs) erlotinib,[6] gefitinib[7] and afatinib.[8] The EML4-ALK fusion gene is found in 2–7% of AC patients[9] and generally occurs independently of other oncogenic drivers, including EGFR, KRAS or ERBB2 mutations.[10] NSCLC patients carrying ALK rearrangements have shown resistance to anti-EGFR TKIs but sensitivity to ALK inhibitors comprising crizotinib and ceritinib.[11] Upcoming agents explored in preclinical and clinical studies target KRAS, BRAF, ERBB2, PI3KC and translocations involving RET, ROS and amplification of c-MET.[12]

The RAS gene family includes KRAS, NRAS and HRAS and encodes for plasma membrane-localized proteins with intrinsic GTPase activity. RAS proteins serve as molecular switches, regulating intracellular signal transduction pathways in response to stimulation of cell surface receptors (targets reviewed in Ref. 13). Mutations in KRAS, NRAS and HRAS are commonly observed in various tumor types, including NSCLC.[14] The most frequently affected isoform is KRAS, which is mutated in 25% of lung ACs.[15] Interestingly, KRAS mutations are more frequent in smokers and in Caucasians.

Key words: lung cancer, survival, biomarker, microarrays, mutation, SNP, Cox regression, Mann-Whitney analysis, TCGA

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: OTKA; **Grant number:** K108655

DOI: 10.1002/ijc.30509

History: Received 11 Sep 2016; Accepted 28 Oct 2016; Online 8 Nov 2016

Correspondence to: Balázs Győrffy, MD PhD DSc, MTA TTK Lendület Cancer Biomarker Research Group, Magyar Tudósok körútja 2., 1117, Budapest, Hungary, Tel.: +[36305142822], E-mail: gyorffy.balazs@ttk.mta.hu

What's new?

As many as one-quarter of patients with lung adenocarcinoma (AC), a form of non-small cell lung cancer (NSCLC), exhibit tumor-associated mutations in *KRAS*. Whether *KRAS* mutation status and expression are correlated to prognosis, however, remains unclear. In this study, a surrogate signature of *KRAS* mutation status was generated for NSCLC by relating genotype to gene-expression signature. The approach led to the identification of a significant correlation between overall survival in lung AC and the transcriptomic fingerprint of somatic *KRAS* mutations. Three genes strongly influenced by *KRAS* mutation may be relevant to the search for novel NSCLC drug targets.

In NSCLC, the vast majority of KRAS mutations involve codons 12 or 13, with G12C, G12D and G12V mutations being predominant, and are responsible for triggering the constitutively active state of the enzyme.[16]

Mutations in KRAS are associated with an intrinsic EGFR-TKI resistance.[17] Considering the localization of KRAS downstream of the EGF receptors, activating mutations of KRAS render the entire signaling network triggered regardless of upstream inhibition. Therefore, KRAS is a negative predictor for targeted therapy, and mutant patients will not respond to administration of EGFR-TKIs.[18] The correlation between response and KRAS mutations status and resistance against EGFR-TKIs in NSCLC has also been validated in meta-analysis studies[19] and is now listed in the NCCN guidelines (https://www.nccn.org/). KRAS is an idyllic predictive biomarker—the predictive value is over 99%, the mutations are in a confined segment of the gene and we have a reasonable biological hypothesis elucidating its role.

Targeted therapy of KRAS mutant lung cancer could be based on the inhibition of main signaling pathways downstream of the active KRAS, including the RAF-MEK-ERK and the PI3K-AKT-mTOR pathways. Preclinical results show that the MEK1/2 inhibitor selumetinib significantly suppressed tumor growth in KRAS mutant NSCLC xenografts.[20] *In vivo* combination of selumetinib with docetaxel improved median PFS of NSCLC patients (5.3 *vs*. 2.1 months), nevertheless with more adverse events than docetaxel alone.[21] Optimal treatment of KRAS mutant NSCLC patients remains an open issue.

At the same time, it is unclear whether somatic mutations of KRAS *per se* are related to poor survival and treatment resistance. Although at first a prognostic effect was suggested in colon cancer,[22] it was not possible to validate these results in later studies for colon[23] or lung cancer.[24] Differences in effect of KRAS mutations between Asian and non-Asian populations was uncovered in a recent meta-analysis, indicating that in Asian patients the KRAS mutations are a poor prognostic factor.[25]

As a first step of this study, we performed a literature survey to uncover the proportion of papers describing a correlation between survival and KRAS status in NSCLC. The absolute majority of studies investigating this issue utilize the raw mutation status. Because of the crucial role of KRAS in different cell signal transduction pathways, KRAS mutations cannot solely affect the gene itself and the expression of corresponding protein, but can also influence the expression of other downstream genes. Here, we hypothesize that the expression changes of these genes could be used as a surrogate marker of the KRAS mutation status. To evaluate this hypothesis, we divided NSCLC patients into two cohorts—those with a genetic alteration in KRAS gene and wild type—identified the signature of genes showing altered expression between these cohorts. Then, we evaluated the correlation of this "transcriptomic fingerprint" with clinical outcome.

Methods

Identifying studies evaluating the prognostic power of KRAS

We performed a literature search in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) to identify studies assessing the prognostic effect of somatic KRAS mutations in NSCLC patients. In this, the keywords "lung cancer" and "KRAS mutation" were used. We filtered to include articles containing human samples and published in English between 2011 and 2015. We reduced our search results to those articles where the somatic KRAS mutation status was determined. The complete workflow of the literature survey is presented in Figure 1*a*.

Somatic mutation data of NSCLC patients

The statistical analysis used three types of data: genotype data containing somatic mutations and copy number variations (CNVs), RNA-seq gene expression and microarray gene chip data.

Next generation sequencing data (exome sequencing and RNA-seq data) for NSCLC patients including AC and squamous cell lung carcinoma was downloaded from The Cancer Genome Atlas of the National Cancer Institute (TCGA, http://cancergenome.nih.gov)[26,27] (Fig. 1*b*). We used the CGHub repository (https://cghub.ucsc.edu/) to download the aligned TCGA samples using the download client software GeneTorrent (version 3.8.5) for both tumor and normal samples.

Identification of mutations from exome sequencing data was performed by MuTect algorithm using default parameters.[28,29] Hard filters were applied to filter out somatic mutations that had $<20\times$ coverage, and <4 altered reads. We used the human reference genomes GRCh37, GRCh37-lite and HG19 for mutation calling. To annotate the identified mutations we used the dbSNP (bulid 139) and COSMIC (version 68) databases.[30] We applied the SNPeff v3.5 program to functionally annotate VCF files generated by MuTect.[31]

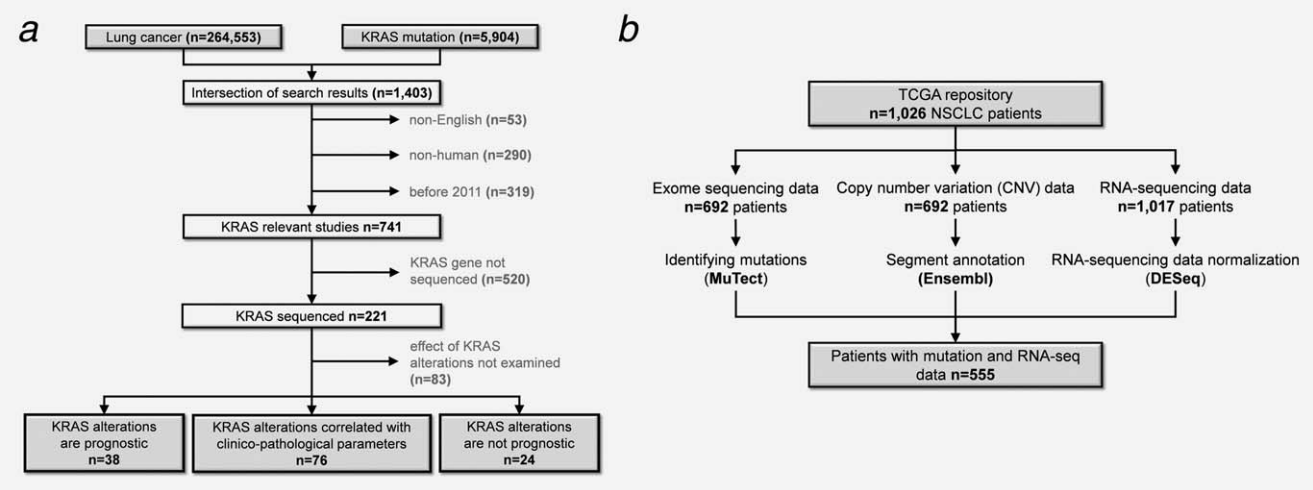

Figure 1. Analysis workflow for the literature survey (*a*) and for the database setup (*b*).

We filtered the CNV data obtained from the TCGA repository according to two parameters: at least 10 probes had to be present at a position with a segment mean above 0.2 for amplification, and under −0.2 for deletions. For the annotation of the filtered segments, we used the Human Gene Sets GTF annotation file downloaded from the Ensemble database of the Human Genome version GRCh37.

Processing of RNA-seq data

We used the preprocessed (level 3) RNA-seq data generated by the Illumina HiSeq 2000 RNA Sequencing Version 2 platform. RNA-seq data was normalized in R v3.2.3 statistical environment (http://www.r-project.org) using DESeq Bioconductor library using a negative binomial distribution.[32] We decided to use DESeq normalization because of its capability to maintain a realistic false-positive rate when compared to other normalization methods for RNA-seq data.

Gene chip database

We used a previously established lung cancer microarray database[33] that contains 2,437 samples measured using Affymetrix HGU133 microarrays (1,800 samples from GEO, http://www.ncbi.nlm.nih.gov/geo/), 504 samples from CaArray project, http://cabig.cancer.gov/ and 133 samples from the TCGA repository). Gene chips were normalized with the MAS5 algorithm by the Affy Bioconductor library. Array quality control was performed as described previously.[34] For each gene, we selected the most reliable probe set using JetSet.[35]

To enable comparison of RNA-seq and gene-chip data, both databases were filtered to include only those genes, which were present on both transcriptomic platforms. This matched gene list contained all together 11,500 genes. To avoid bias due to background noise, genes with a mean expression below 100 (MAS5 normalized expression value) were excluded from further analyses.

Statistical analysis pipeline

Statistical computations were performed in the R v3.2.3 statistical environment. The first step of the analysis was splitting of RNA-seq samples into two cohorts based on the alteration (somatic mutations or CNVs) of KRAS gene. Then we applied Mann-Whitney-Wilcoxon test to identify genes differentially expressed between the mutated and wild type cohorts. The analysis was restricted to patients with a coding mutation in KRAS. We also calculated the fold change value for each gene. In a nonparametric analysis, the mean expression of the top five genes whose expression was significantly associated with genotype alteration of KRAS was designated as the "KRAS-signature."

The following step of the analysis was executed using the gene chip data. The gene chip data was used because there the sample number was a magnitude higher than those with RNA-seq data. In addition, more clinical data was available for these patients. We examined the correlation between the KRAS surrogate signature and overall survival (OS) using Cox proportional hazards regression and by plotting Kaplan-Meier survival plots. Cox regression analysis was performed using the "survival" R package v2.38 downloaded from CRAN (http://CRAN.project.org/package = survival). Kaplan-Meier plots were generated applying the "survplot" R package v0.0.7 (http://www.cbs.dtu.dk/~eklund/survplot). Fulfilment of the proportional hazards assumption necessary for the Cox regression was validated by employing the "coxph" R function. Prevalence of mutations in further cancer genes among KRAS wild type and mutant patients was compared using a χ^2 test. Statistical significance was set at $p < 0.05$.

Results

Literature survey of KRAS mutations in lung cancer

The search in PubMed resulted in 1,403 hits, of which 741 were published between 2011 and 2016, were English and utilized human samples (search was performed on the 31th

Table 1. Studies evaluating the impact of KRAS mutations in lung cancer ($n = 138$)

	All studies *n* (%)	Do not mention any NSCLC subtype *n* (%)	AC *n* (%)	SCC *n* (%)	LCC *n* (%)	LCN *n* (%)	SC *n* (%)	ASCC *n* (%)
KRAS mutations correlated with survival	38 (28%)	17 (12%)	17 (12%)	2 (1%)	2 (1%)	–	–	–
KRAS mutations correlated with other clinicopathological features	76 (55%)	27 (20%)	37 (27%)	4 (3%)	3 (2%)	2 (1%)	1 (1%)	2 (1%)
KRAS mutations did not have a significant effect	24 (17%)	13 (9%)	9 (7%)	1 (1%)	1 (1%)	–	1 (1%)	1 (1%)

NSCLC: non-small cell lung cancer; **AC**: adenocarcinoma; **SCC**: squamous cell carcinoma; **LCC**: large cell carcinoma; **LCN**: large cell neuroendocrine carcinoma; **SC**: sarcomatoid carcinoma; **ASCC**: adenosquamous carcinoma.
For detailed information about each study see Supporting Information **Table S1**.

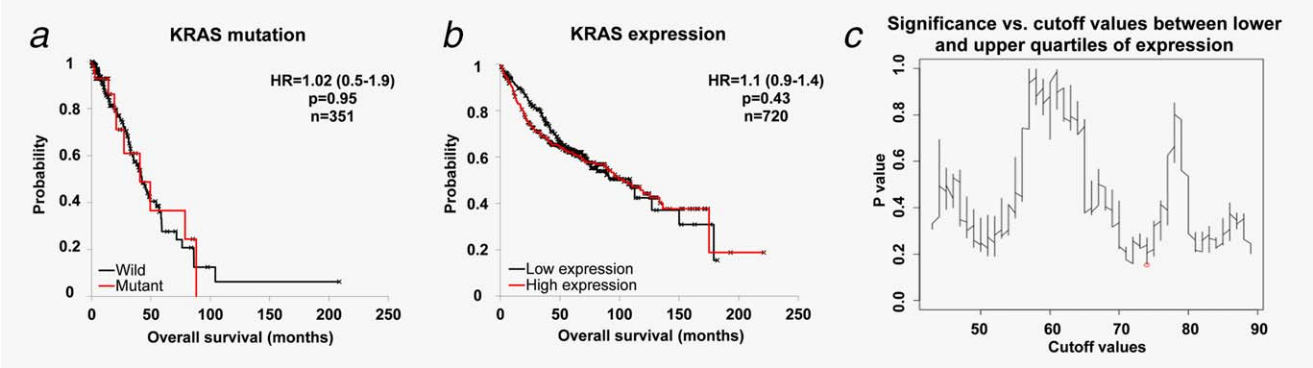

Figure 2. KRAS gene *per se* has no correlation to survival in NSCLC. Analysis of the effect of KRAS mutation (***a***) and expression (***b***) on survival in NSCLC AC patients. When investigating different cutoff values across all patients (***c***), none of the threshold values between the lower and upper quartile of expression reached statistical significance. The strongest achieved *p* values is marked by a red circle in (***c***). [Color figure can be viewed at wileyonlinelibrary.com]

January 2016). Of these, KRAS was sequenced in 221 studies (Supporting Information Table S1). Effect of KRAS mutation status on survival was examined in 138 articles. In these, gene mutations correlated with survival in 38 studies, with clinicopathological parameters, including smoking history, resistance to treatment, tumor growth in 76 studies. Mutation status was not prognostic at all in 24 studies (Fig. 1*a*). Correlation of KRAS mutations with prognosis according to different NSCLC subtypes is shown in Table 1. We have to note that some studies did not mention any lung cancer subtype or contained more than one.

Database construction

The TCGA repository holds 1,026 NSCLC patients—including 522 AC and 504 SCC samples. The distribution of AC and SCC was similar in the TCGA and the microarray databases. OS was available for 967 patients with a median follow up of 8.68 months. When performing survival analysis to inspect the correlation between clinico-pathological parameters and clinical outcome, there were significant correlations between the OS and tumor size ($p = 0.0001$), lymph node status ($p = 2.03E{-}06$) and metastasis ($p = 0.03$).

Microarray data was available for 2,437 patients. Gene expression was determined using three different gene chip platforms (GPL96, GPL570 and GPL3921). OS data was accessible for 2,002 patients and the median follow up was 41.6 months. Survival differences were significant in patients stratified by tumor size ($p < 1E{-}16$), lymph node status ($p < 1E{-}16$) and metastasis ($p = 0.003$).

Neither KRAS mutations nor KRAS gene expression have a prognostic power

First, we examined the impact of KRAS mutations *per se* on OS. In the AC subgroup, including 60 patients harboring any type of KRAS somatic mutation, KRAS mutation status was not significantly correlated with OS (HR = 1.02; 95% CI = 0.5–1.9; $p = 0.95$; Fig. 2*a*). In SCC cohort there was only one patient harboring a KRAS mutation. For this reason, all the subsequent analyses were performed for AC patients only.

We obtained similar results when investigating the effect of KRAS gene expression on OS using the gene chip database (Fig. 2*b*). Lack of any correlation between survival and

Table 2. Nonparametric transcriptomic fingerprint of top five genes correlated to KRAS mutation status (A), KRAS amplification (B) and KRAS deletion (C)

	p–Values	Fold change	Mean expression in mutants	Mean expression in wild
A)				
FOXRED2	1.14E–06	0.62	944	1530
KRAS	6.31E–06	1.29	4472	3462
TOP1	7.06E–06	1.29	10272	7979
PEX3	1.51E–05	0.77	499	647
ABL2	1.81E–05	1.22	2534	2077
B)				
KRAS	1.55E–09	1.94	6553	3379
ETNK1	5.29E–06	1.57	5865	3740
FAM60A	8.00E–06	1.77	7491	4243
TMEM185B	8.33E–06	1.29	1623	1258
CLEC11A	1.73E–05	0.55	298	544
C)				
HSDL2	5.91E–05	1.64	3923	2398
GAK	1.06E–04	0.63	3132	4982
YARS2	1.44E-04	0.61	517	846
MCOLN1	1.50E-04	0.63	733	1155
CMAS	1.65E-04	0.59	1095	1862

prognostic power can be observed when plotting a graph of the achieved *p* values *vs.* all the potential cutoff values (Fig. 2*c*).

Similar non-significant results were obtained when investigating the correlation between survival and KRAS expression using the RNA-seq database (Supporting Information Figure S1).

The KRAS mutation surrogate signature has a significant effect on survival

Each of the 60 patients had a coding mutation in KRAS. Seven of these also had a noncoding mutation. The five strongest genes correlation to KRAS status in the Mann-Whitney test include the FOXRED2, PEX3, KRAS, TOP1 and ABL2 genes (Table 2). The mean expression of these genes was used as a surrogate signature of KRAS mutation status in the gene chip database. When using the median expression as a cutoff, we achieved high association with OS (HR = 2.4; 95% CI = 1.9–3.2; $p = 1.24E{-}12$; Fig. 3*a*). When computing the *p* values achieved in the Cox regression across all possible cutoff values between the lower and upper quartiles of expression, all the *p* values remained highly significant (Fig. 3*b*).

Quality control

To validate the approach of running a nonparametric analysis using the five strongest genes, we have also run the analysis using the second and the third set of five best genes associated with KRAS mutation (Supporting Information Table 2). Both these additional signatures delivered very similar results to the set of the first five genes: the second set reached HR = 1.9; 95% CI = 1.5–2.5 (at $p = 4.7E{-}08$) and the third set delivered a HR = 2.5; 95% CI = 1.9–3.2 (at $p = 4.4E{-}14$; Figs. 3*c* and 3*d*).

The main strength of our analysis was the use of an RNA-seq dataset in the training cohort and then the use of a gene chip dataset for the validation cohort. Since different technology platforms can measure expression of the same gene with different sensitivity, specificity and dynamic range, a strong correlation between these platforms is an important issue for the reliability of our analysis. There were 130 NSCLC samples published in TCGA with simultaneous RNA-seq and gene chip measurement. When comparing the mean expression of surrogate signature of KRAS mutations in these, the two platforms delivered a highly significant correlation (p = 6.4E-11, Spearman rank corr. coeff. = 0.53; Supporting Information Figure S2).

When comparing the prevalence of mutations in KRAS wild and in KRAS mutated patients for all Cosmic Cancer Consensus Genes, only one gene was identified at a false discovery rate below 10%, namely KRAS itself. The following strongest genes (listed in Supporting Information Table S3.) had a very low prevalence (for example FOXO1 was mutated in only one, and FUBP1 in only two patients in the KRAS wild type cohort—the exclusion of patients with these mutations did not changed the KRAS surrogate signature).

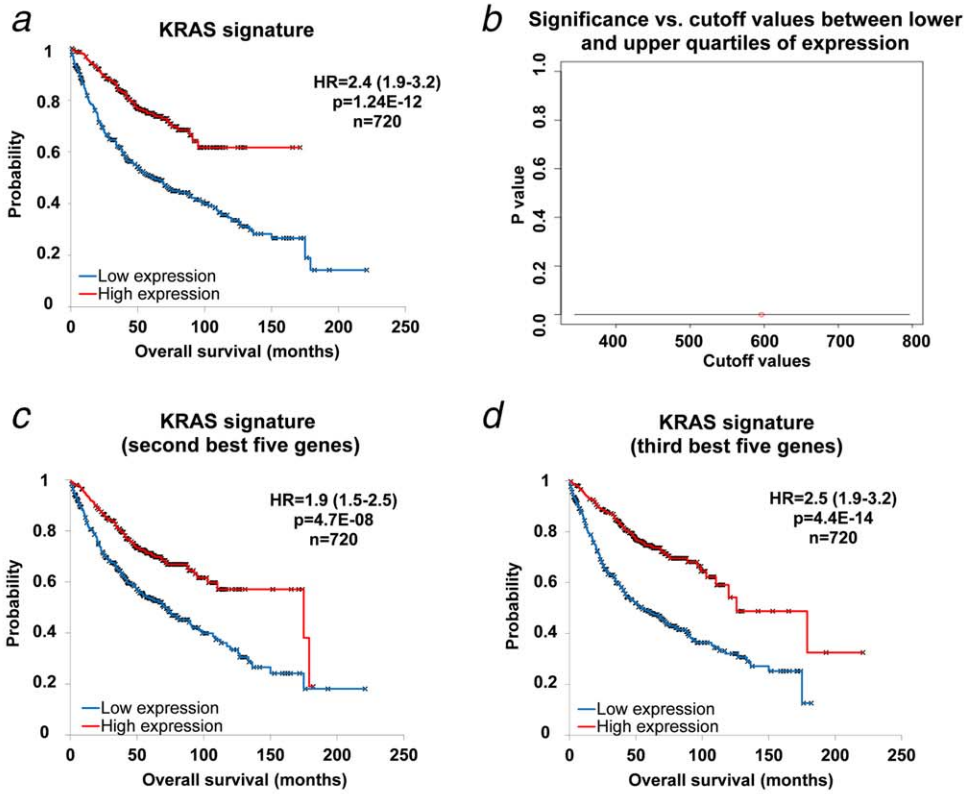

Figure 3. The surrogate signature of KRAS mutation status has a high prognostic power. Signature comprising the mean expression of the top five genes (*a*). When investigating different cutoff values between the lower and upper quartiles of expression for the surrogate signature, every cutoff value achieved high significance (*b*). Similar results were achieved when using the second (*c*) and the third (*d*) set of five strongest genes. [Color figure can be viewed at wileyonlinelibrary.com]

Impact of KRAS amplification and deletion driven signature on survival in AC

We further analyzed the prognostic roles of surrogate signatures of KRAS amplification and deletion. KRAS amplifications were found in 30 patients and deletions found in 12 patients. The KRAS amplifications related signature includes KRAS, ETNK1, FAM60A, TMEM185B and CLEC11A genes (Table 2). Nevertheless, the KRAS amplification signature was not correlated with OS as presented on Supporting Information Figure S3A (HR = 0.91; 95% CI = 0.7–1.2; $p = 0.45$). KRAS deletions associated with the expression of HSDL2, GAK, YARS2, MCOLN1 and CMAS (Table 2). The average expression of these genes was significantly associated with OS (HR = 2.3; 95% CI = 1.8–2.9; $p = 1.8E{-}11$, Supporting Information Figure S3B).

Multivariate analysis

In a multivariate analysis of OS including surrogate signature for KRAS mutation, age, gender, smoking history and tumor stage, only the KRAS signature ($p = 0.01$), age ($p = 0.01$) and stage ($p = 5E{-}07$) emerged as significant factors. When performing the same analysis for KRAS deletion, only age ($p = 0.01$) and stage ($p = 3E{-}07$) remained significant.

Discussion

We have assessed the correlation between KRAS and OS in lung cancer. Different types of data including mutation status and gene expression were assessed at first, but these did not have a significant prognostic power. These results are in line with the outcome of the literature survey, as no correlation to survival was observed in 72% of all studies with KRAS sequencing performed in NSCLC.

KRAS mutations cannot solely affect the gene itself and the expression of the corresponding protein, but can also influence the expression of other downstream genes involved in crucial pathways regulating cell growth, differentiation and apoptosis. Thus, the different expression of these genes in KRAS-mutant tumors might have a more prominent role in affecting patient's clinical outcomes. Having multiple levels of data for the same patients opens a window of opportunity for a two-step analysis: first, a genome wide transcriptomic analysis across all genes was performed to identify genes affected by a KRAS mutation. In the second step, the strongest genes were combined into a "surrogate signature" of KRAS mutation status, and the correlation between this expression based signature and clinical outcome was computed. Interestingly, the surrogate signature had a dramatic effect on patient's outcome.

In our analysis, we set up a signature encompassing the top five genes only. To validate this approach, we also evaluated additional signatures comprising the second and third sets of strongest genes and both these settings delivered similar significance. One might wonder how a limited signature of five genes can provide such robust information in the case of mutant KRAS. The most probable reason for this is the extreme high level of interconnections in the signaling network resulting in a concordant expression change across multiple downstream genes.[13] We have to note that the surrogate signature was highly significant regardless of the used cutoff value and remained significant in a multivariate analysis including clinicopathological parameters.

This approach has an additional advantage: to link a sequence variation to clinical outcome we would need thousands of patients with sufficient follow up data. However, in the TCGA dataset the median follow up of NSCLC patients is merely 8.68 months, therefore preventing a reliable direct testing of the prognostic impact. At the same time, available gene chip datasets comprising NSCLC samples have five times longer follow up times and extensive clinical annotations.

The top five genes affected by a somatic KRAS mutation include FOXRED2, KRAS, TOP1, PEX3 and ABL2. Of these, three genes were already associated with tumor development besides KRAS. TOP1 (topoisomerase 1) plays a crucial role in DNA replication and maintaining genome stability by regulating the supercoiling state of DNA. TOP1 is also the cellular target for irinotecan, an agent approved for advanced or metastatic NSCLC.[36] Higher expression of TOP1 in KRAS mutant patients and correlation to survival suggests that TOP1 inhibitors might have increased benefit when administered to treat patients with a KRAS mutant tumor.

The ABL2 gene is a member of the Abelson family of non-receptor tyrosine protein kinases.[37] These kinases are involved in controlling cell growth, survival, invasion, adhesion and migration.[38] Oncogenic activation of the ABL kinases is most known in Philadelphia-positive (Ph+) human leukemias where the generated BCR-ABL1 fusion protein has a constitutive tyrosine kinase activity. Currently, ABL kinases are targeted by several FDA-approved agents including imatinib and dasatinib to treat patients with BCR-ABL1-positive leukemias.[39] ABL2 amplification is prevalent among ABL alterations, which has been detected in invasive lung and breast carcinomas.[40,41] Similar to TOP1, the expression of ABL2 was also increased in KRAS mutant samples. Thus, agents targeting ABL kinases might have a beneficial effect in patients harboring a KRAS genetic alteration.

The FOXRED2 (FAD-dependent oxidoreductase domain containing 2) gene is also highly expressed in human tumors including NSCLC, colorectal cancer and breast cancer.[42] Hypoxia has recently been associated with oncogenesis in multiple tumors and it has been suggested that administration of HAPs (hypoxia-activated prodrugs) may bring potential therapeutic benefit in some tumors.[43] HAPs, including the nitro aromatic compounds TH-302[44] and PR-104[45] are currently in clinical development. HAPs are metabolized to cytotoxic DNA damaging agents *via* enzyme-catalyzed reductive reactions. FOXRED2, as a HAP reductase, is enzymatically capable to activate HAPs; therefore it contributes to prodrug activation in some human tumors.[42] Our results suggest KRAS mutation as a negative biomarker for HAPs–response might be lower in tumors with a KRAS mutation.

The signature includes five genes only and these could be easily measured by RT-PCR in a tumor sample. The genes are linked to a set of agents (TOP1 inhibitors, ABL kinase inhibitors and HAPs), which could deliver different efficiency in KRAS mutated and wild type patients. A successful validation of these drugs in a future clinical trial would open the possibility to identify patients eligible for a new set of agents depending on the interconnection between gene expression and mutation status in other key genes as well.

In a similar setup, we also determined the effect of surrogate signatures related to KRAS amplifications and deletions on OS in lung cancer. As expected, loss of KRAS was a good prognostic marker.

We have to note some limitations of our analysis. First, the literature survey and the results of mutation calling both show that KRAS mutations are overwhelmingly more frequent in AC compared with other NSCLC subtypes. For this reason, we had to limit our analysis and consequently or conclusions to AC of the lung. A second limitation is the exclusion of additional effects on transcription like epigenetic regulation—the same mutation with or without a methylation event might result in different transcriptional outcome. Unfortunately, methylation data was not available for the investigated patients.

In summary, here we generated a "surrogate signature" of KRAS mutation status in lung AC patients by computationally connecting genotype to an extended gene expression signature. We show that three out of the top five genes influenced by a KRAS mutation could also have a direct pharmacological implication. We used the surrogate signature in a large and well annotated gene chip database to test its prognostic significance and found that the transcriptomic fingerprint of somatic KRAS mutations had a highly significant correlation with OS in lung AC. Our results emphasize the prominence of KRAS and prove that secondary effects can have a superior prognostic relevance compared to the primary genetic alteration.

Author Contribution Statements

BG contributed to the conception and design and writing of the manuscript. ÁN contributed to the data analysis, data interpretation and drafting the manuscript. LP contributed to the data acquisition. AS and MS were involved in data analysis. All authors read and approved the final manuscript.

Competing Interests

The authors declare that they have no conflict of interest.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin* 2015;65:5–29.
2. Ettinger DS, Akerley W, Borghaei H, et al. Non-small cell lung cancer. *J Natl Compr Canc Netw* 2012;10:1236–71.
3. Alamgeer M, Ganju V, Watkins DN. Novel therapeutic targets in non-small cell lung cancer. *Curr Opin Pharmacol* 2013;13:394–401.
4. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–500.
5. Cappuzzo F, Hirsch FR, Rossi E, et al. Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *J Natl Cancer Inst* 2005;97:643–55.
6. Zhou C, Wu YL, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011;12:735–42.
7. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;361:947–57.
8. Belani CP. The role of irreversible EGFR inhibitors in the treatment of non-small cell lung cancer: overcoming resistance to reversible EGFR inhibitors. *Cancer Invest* 2010;28:413–23.
9. Kwak EL, Bang YJ, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010;363:1693–703.
10. Takahashi T, Sonobe M, Kobayashi M, et al. Clinicopathologic features of non-small-cell lung cancer with EML4-ALK fusion gene. *Ann Surg Oncol* 2010;17:889–97.
11. Shaw AT, Yeap BY, Mino-Kenudson M, et al. Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. *J Clin Oncol* 2009;27:4247–53.
12. Boolell V, Alamgeer M, Watkins DN, et al. The Evolution of Therapies in Non-Small Cell Lung Cancer. *Cancers* 2015;7:1815–46.
13. Gyorffy B, Schafer R. Biomarkers downstream of RAS: a search for robust transcriptional targets. *Curr Cancer Drug Targets* 2010;10:858–68.
14. Karnoub AE, Weinberg RA. Ras oncogenes: split personalities. *Nat Rev Mol Cell Biol* 2008;9:517–31.
15. Vasan N, Boyer JL, Herbst RS. A RAS renaissance: emerging targeted therapies for KRAS-mutated non-small cell lung cancer. *Clin Cancer Res* 2014;20:3921–30.
16. Riely GJ, Marks J, Pao W. KRAS mutations in non-small cell lung cancer. *Proc Am Thorac Soc* 2009;6:201–5.
17. Shigematsu H, Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer* 2006;118:257–62.
18. Massarelli E, Varella-Garcia M, Tang X, et al. KRAS mutation is an important predictor of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. *Clin Cancer Res* 2007;13:2890–6.
19. Mao C, Qiu LX, Liao RY, et al. KRAS mutations and resistance to EGFR-TKIs treatment in patients with non-small cell lung cancer: a meta-analysis of 22 studies. *Lung Cancer* 2010;69:272–8.
20. Davies BR, Logie A, McKay JS, et al. AZD6244 (ARRY-142886), a potent inhibitor of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase 1/2 kinases: mechanism of action in vivo, pharmacokinetic/pharmacodynamic relationship, and potential for combination in preclinical models. *Mol Cancer Ther* 2007;6:2209–19.
21. Janne PA, Shaw AT, Pereira JR, et al. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. *Lancet Oncol* 2013;14:38–47.
22. Andreyev HJ, Norman AR, Cunningham D, et al. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J Natl Cancer Inst* 1998;90:675–84.
23. Ogino S, Nosho K, Kirkner GJ, et al. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 2009;58:90–6.
24. Scoccianti C, Vesin A, Martel G, et al. Prognostic value of TP53, KRAS and EGFR mutations in nonsmall cell lung cancer: the EUELC cohort. *Eur Respir J* 2012;40:177–84.
25. Meng D, Yuan M, Li X, et al. Prognostic value of K-RAS mutations in patients with non-small cell lung cancer: a systematic review with meta-analysis. *Lung Cancer* 2013;81:1–10.
26. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543–50.
27. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–25.
28. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–9.
29. Pongor L, Kormos M, Hatzis C, et al. A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. *Genome Med* 2015;7:104
30. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015;43:D805–11.
31. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 2012;6:80–92.
32. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11:R106
33. Gyorffy B, Surowiak P, Budczies J, et al. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PloS One* 2013;8:e82241
34. Gyorffy B, Benke Z, Lanczky A, et al. RecurrenceOnline: an online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data. *Breast Cancer Res Treat* 2012;132:1025–34.
35. Li Q, Birkbak NJ, Gyorffy B, et al. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics* 2011;12:474
36. Georgoulias V, Kouroussis C, Agelidou A, et al. Irinotecan plus gemcitabine vs irinotecan for the second-line treatment of patients with advanced non-small-cell lung cancer pretreated with docetaxel and cisplatin: a multicentre, randomised, phase II study. *Br J Cancer* 2004;91:482–8.
37. Ben-Neriah Y, Daley GQ, Mes-Masson AM, et al. The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. *Science* 1986;233:212–4.
38. Colicelli J. ABL tyrosine kinases: evolution of function, regulation, and specificity. *Sci Signaling* 2010;3:re6
39. Greuber EK, Smith-Pearson P, Wang J, et al. Role of ABL family kinases in cancer: from leukaemia to solid tumours. *Nat Rev Cancer* 2013;13:559–71.
40. Sos ML, Michel K, Zander T, et al. Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. *J Clin Invest* 2009;119:1727–40.
41. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.
42. Hunter FW, Jaiswal JK, Hurley DG, et al. The flavoprotein FOXRED2 reductively activates nitrochloromethylbenzindolines and other hypoxia-targeting prodrugs. *Biochem Pharmacol* 2014;89:224–35.
43. Zardavas D, Baselga J, Piccart M. Emerging targeted agents in metastatic breast cancer. *Nat Rev Clin Oncol* 2013;10:191–210.
44. Meng F, Evans JW, Bhupathi D, et al. Molecular and cellular pharmacology of the hypoxia-activated prodrug TH-302. *Mol Cancer Ther* 2012;11:740–51.
45. Patterson AV, Ferry DM, Edmunds SJ, et al. Mechanism of action and preclinical antitumor activity of the novel hypoxia-activated DNA cross-linking agent PR-104. *Clin Cancer Res* 2007;13:3922–32.