

Identification of Subsets of Genetic Alterations in KRAS-mutant Lung Cancer using Association Rule Mining

Junior Tayou

Abstract

Background: Lung cancer is the leading cause of all cancer death accounting for 1 out of 4 cancer-related death in both men and women. KRAS mutations occur in ~ 25% of patients with lung cancer, and the presence of these mutations is associated with poor prognosis. Efforts to directly target KRAS or associated downstream MAPK or the PI3K/AKT/mTOR pathways have seen little or no benefits. One probable reason for the lack of progress in targeting KRAS-mutant tumors is the co-occurrence of other cell survival pathways and mechanisms.

Method and results: To identify other potential cell survival pathways in subsets of KRAS-mutant tumors, I performed unsupervised machine learning on somatic mutations in metastatic lung cancer from 725 patient samples. I identified 67 other genes that were mutated in at least 10% of the samples with KRAS alterations. This gene list was enriched with genes involved in the MAPK, AKT and STAT3 pathways, cell-cell adhesion, DNA repair, chromatin remodeling, and the Wnt/beta-catenin pathway. I also identified 160 overlapping subsets of 3 or more genes that code for oncogenic or oncosuppressive proteins that were mutated in at least 10% of KRAS-mutant tumors.

Conclusions: In this study, I identified genes that are co-mutated in KRAS-mutant lung cancer. I also identify subpopulations of KRAS-mutant lung cancer based on the set of genes that were also altered in the tumor samples. The design of research models that captures these subsets of KRAS-mutant tumors would enhance our understanding of the disease and facilitate personalized treatment for lung cancer patients with KRAS alterations.

Keywords: Metastatic lung cancer, KRAS, genetic mutations, unsupervised machine learning

Background

Advanced lung cancer is the most commonly diagnosed, and the leading cause of all cancer death worldwide [1, 2]. Morphologically, lung cancer can be broken down into two major categories: non-small cell lung carcinoma (NSCLC) which makes up 85% of all the cases and small cell lung carcinoma (SCLC) [2]. Data accumulated over the last 15 years shows that lung cancer is not a single disease but a collection of genetically defined neoplasms with distinct molecular features and clinical outcomes. The known oncogenic drivers of lung cancer include the receptor tyrosine kinases (RTKs): EGFR, ALK, RET, MET, DDR2, ROS1 and FGFR1 and their downstream facilitators KRAS, PIK3CA, and BRAF [3]. Tumors with EGFR alterations are targetable with EGFR tyrosine kinase inhibitors (TKIs) including erlotinib, gefitinib, and afatinib [4-6]. Rearrangement of ALK and ROS1 genes are targeted with crizotinib and ceritinib [7]. Progress has also been made in targeting mutations in the other RTKs [3, 8]. However, for tumors containing alterations in KRAS, there are no FDA approved drugs. KRAS mutations are present in between 20 and 25% of lung cancer and are associated with poor prognosis [9, 10]. KRAS signal primarily through the MAPK and the PI3K/AKT/mTOR pathways [11-13]. Efforts to directly target KRAS or its downstream effectors have seen little or no benefit [14]. The reason for the lack of success in targeting KRAS effectors might be due to the development of adaptive response mechanisms or the co-occurrence of mutations in other genes that activate different cell survival pathways in KRAS-mutant tumors.

Here, I employed unsupervised machine learning using association rules to identify genes whose mutations co-occurs with KRAS alterations in advanced lung cancer. The building blocks of association rules are the items that may appear in any given transaction [15]. This approach is largely applied in supermarkets to learn about consumer purchasing patterns. Retailers use it to identify products that are usually purchased together in order to tailor their advertisement outreach. Transactions are specified in terms of itemsets. For example, in a grocery store, {Bread, peanut butter, jelly} represent a typical itemset. A rule from this transaction might be expressed in the form {peanut butter, jelly} \rightarrow {bread}, in other words, if peanut butter and jelly (left-hand side or LHS) are purchased together, then bread (right-hand side or RHS) is also likely to be bought. Supermarkets apply association rules to find “interesting” combinations of items that are not as apparent as the peanut butter, jelly and bread example.

I used the Apriori association rule algorithm [16], which states that all subsets of a frequent itemset must also be frequent. For example, for the rule {mutation X} → {mutation Y}, then the items {mutation X}, {mutation Y} and {mutation X and mutation Y} must all be frequent. The strength of a rule is determined by three statistical measures: the support, confidence and lift statistics. The support of a mutation or mutation set is a measure of how frequently it occurs in the tumor samples. The support for one or more mutations in a gene X in N tumor samples is given by

$$\text{Support}(\text{mutation } X) = \frac{\text{the number of samples with mutations in gene } X}{N}$$

The confidence is the measure of the accuracy of a rule. For a given rule {mutation X} → {mutation Y}, the confidence is defined as

$$\text{Confidene}(\text{mutation } X \rightarrow \text{mutation } Y) = \frac{\text{Support}(\text{mutation } X, \text{mutation } Y)}{\text{Support}(\text{mutation } X)}$$

The confidence tells us the proportion of the samples where a mutation in gene X results in a mutation in gene Y. The lift statistic measures how much more likely a mutation or a set of mutations is present in a sample relative to its typical rate of occurrence, given that another mutation or set of mutations are present in the sample. The lift for a given rule {mutation X} → {mutation Y} is define by

$$\text{Lift}(\text{mutation } X \rightarrow \text{mutation } Y) = \frac{\text{Confidence}(\text{mutation } X, \text{mutation } Y)}{\text{Support}(\text{mutation } Y)}$$

For a hypothetical example, {TP53, BRCA1} → {CD43}, support = 0.2, confidence = 0.80, lift = 3, count = 125. This rule can be interpreted as – if a tumor has mutations in both TP53 and BRCA1, it will also contain a mutation in CD43. This rule covers 20% (support = 0.2) of all the tumor samples, and the rule is correct in 80% (confidence = 0.80) of tumors having TP53 and BRCA1 mutations. A lift of 3 implies that samples with mutations in both TP53 and BRCA1 are three times more likely to have a mutation in CD43 than a typical cancer sample.

I applied the Apriori machine learning algorithm to a dataset containing somatic mutations from 725 advanced lung cancer patient samples to identify the subsets of genes that were also altered in tumors with KRAS mutations.

Methods

Data acquisition: The data used in this study was obtained from the CBioportal website. The data represents the findings from a study on the mutational landscape of advanced cancer from more than 10,000 patients by Zehir A et al. [17]. I downloaded the data on October 6th, 2017. After extraction of all the information on mutations including deletion, insertion, substitution and gene rearrangements from all metastatic lung cancer samples, I merged the columns of interest by sample ID. The resulting file containing the genes with alterations was converted into a sparse matrix (supplementary table 1). I converted the genes coding for receptor tyrosine kinases (PTPRT, PTPRD, and PTPRS) to PTPRs and the B melanoma antigen precursors (BAGE 2, 3, 4, and 5) to BAGE.

Apriori algorithm: Association rule was performed using the Apriori algorithm that was developed by Agrawal R and Srikant R [16]. I used the “arules” and “aruleviz” packages in the R programming language to execute the apriori algorithm and for visualization of the resulting rules respectively. For a typical analysis, I used the codes:

```
cancer_rules <- apriori(data, parameter = list(support = 0.01, confidence =  $\alpha$ , minlen =  $\beta$ , maxlen= $\mu$ ))
```

I used a support of 0.01 in all the analyses; this threshold will include only the genes that were mutated in at least 1% of the samples. More than 600 genes were above this threshold. The values for the confidence, the minimum (minlen) and the maximum (maxlen) number of the members in a rule were selected based on the question that was addressed.

Graph analysis: Rules or subrules generated using the Apriori algorithm in R were saved as Gephi file and manipulated in Gephi 0.9.2 to generate directed graphs as shown in figures 2D and 4A. For example, the rule

```
cancer_rules <- apriori(data, parameter = list(support = 0.01, confidence = 0.5, minlen = 2, maxlen=10)),
```

can be save as a Gephi file using the code:

```
saveAsGraph((cancer_rules), file="cancer_rules.graphml").
```

In Gephi, graphs were manipulated to represent nodes with different colors or labels with different font sizes based on the directionality and magnitude of the degree of connectivity of the nodes.

Bioinformatics: I used the REACTOME pathway analysis site to determine the pathways that were enriched in the gene sets obtained from association rule analysis.

Results and discussion

Distribution of mutations of known oncogenic drivers in metastatic lung cancer: The dataset used in this study contained 725 metastatic lung cancer patient samples (93% NSCLC and 7% SCLC)[17]. Consistent with previous studies [18], TP53 was mutated in 63% of all metastatic lung cancer samples. ROS1 and ALK1 mutations were present in 17% and 15% of the tumors respectively. EGFR alterations were present in 29% of the samples, while DDR2, RET, MET and FGFR1 were mutated in 8%, 12%, 9% and 6% respectively. The RTK facilitators PIK3CA, BRAF and KRAS, were altered in 8%, 16% and 23% of the samples respectively. The samples that had KRAS mutations were almost exclusively NSCLC; only 0.6% (1 tumor sample) was SCLC. All the tumors that had EGFR mutations were NSCLC.

Analyses of rules statistics: Using a support of 1% and confidence of 10%, I identified 187,168 rules for the genes that were mutated in all the samples. KRAS mutations were associated with 5,724 of these rules. The supports for the 187,168 rules was in the range of 1 and 25% with most of the rules covering between 1 and 10% of the data (Figure 1A). Rules containing KRAS had a narrower support range. The supports for the KRAS-mutant subrules were predominantly between 1 and 5%. The maximum support for a rule that had a KRAS alteration was 12% (Figure 1B). The confidence for all the rules and the KRAS-mutant subrules were evenly spread between 10 and 100% (Figure 1A, B).

KRAS-mutant cancers were less likely to have EGFR, MET or PIK3CA mutations: I examined the co-occurrence of mutations in KRAS and alteration in other oncogenic drivers in metastatic lung cancer. Consistent with previous reports [19], EGFR and KRAS mutations were almost mutually exclusive; only 3% (support = 0.03) of the tumors had mutations in both oncogenes (Figure 1C). The likelihood for a sample with a KRAS alteration to also have an EGFR mutation was two-fold (lift = 0.46) less than one would expect in a typical metastatic lung cancer sample (Figure 1C). Similarly, the likelihood for mutations in MET (found in 1% of KRAS-mutant samples) or PIK3CA (1% of KRAS-mutant samples) to occur in KRAS-mutant samples were less than one would expect by chance. The likelihoods that a sample with alterations in KRAS will also have mutations in ALK, ROS1, RET, FGFR1, DDR2 or BRAF were all close to 1 (Figure 1C).

Mutations in KRAS were associated with alterations in 67 other genes: To identify potential cell survival pathways that might also be activated in KRAS-mutant tumors. I queried the rules that I generated from the Apriori algorithm for genes associated with KRAS mutations. I identified 67 genes that were mutated in at least 10% of the KRAS-mutant tumors (Figure 1D and Table 1). These genes are involved in diverse biological pathways that are important for cancer cell survival (Figure 1E). Genes involved in RTK signaling and the downstream MAPK, PI3K/AKT/mTOR and the JAK/STAT pathways were predominantly altered (Figure 1E, 2A). This list was also enriched with genes coding for proteins involved in chromatin remodeling and transcriptional regulation by TP53 (Figure 1E, 2C).

Mutation of RTK genes in samples with KRAS mutations: In addition to the known oncogenic RTK drivers discussed earlier, I identified several other RTKs that were altered in KRAS-mutant tumors including members of the EGFR family, the EPH, NTRK, and VEGF receptor families. ERBB4, a member of the EGFR family, was mutated in 17% of the samples with KRAS alterations (Figure 2A). Unlike EGFR (Figure 1C), the likelihood of having a tumor with mutations in ERBB4 given that the tumor has an alteration in KRAS was higher than one would expect by chance in an advanced lung cancer sample (Table 1, lift = 1.35). KRAS-mutant samples also had alterations in several members of the VEGF receptor family including FLT1 (16%), FLT3 (10%), FLT4 (19%) and KDR (15%) receptors (Figure 2A, Table1). The NTRK1 and NTRK3 that codes for TrkA and TrkC receptors were mutated in 10% and 25% of KRAS-mutant samples respectively (Figure 2A, Table

1). Four members of ephrin receptor family: EPHA3 (22%), EPHA5 (22%), EPHA7 (12%) and EPHB1 (11%) were also altered in KRAS-mutant tumors (Table 1). The platelet-derived growth factor receptors, PDGFRA and PDGFRB, were altered in 14% and 11% of the KRAS-mutant samples respectively (Figure 2A, Table 1). Mutations in MET and KRAS were almost mutually exclusive (Figure 1C). However, MET's ligand HGF was mutated in 20% of KRAS-mutant samples (Table 1).

Mutation of genes involved in the MAPK and PI3K/AKT/mTOR pathways: KRAS is an effector of RTK signaling and mutation in KRAS can lead to constitutive activation of the downstream cell survival MAPK pathway irrespective of receptor activation [20, 21] (Figure 2A). Consequently, RTK inhibitors have been shown to be futile in the presence of activating KRAS mutations [20, 21]. BRAF is a serine-threonine kinase directly downstream of KRAS in the MAPK pathway (Figure 2A). Constitutively active BRAF mutations can also drive tumor progression by activating MEK irrespective of the status of RTK and KRAS [22]. In colorectal cancer, it has been shown that mutations in KRAS and BRAF are almost mutually exclusive [23, 24]. However, in the metastatic lung cancer dataset used in this study, BRAF was mutated in 16% of samples with KRAS mutations (Figure 1C). Also, the likelihood of having a KRAS-mutant tumor with a BRAF alteration was 0.95 (Table 1). Activating mutations in the Gas subunit of heterotrimeric G-proteins can also promote cancer cell proliferation via the MAPK pathway [25]. I observed that GNAS which codes for the Gas subunit was mutated in 13% of the KRAS-mutant samples (Figure 2A). The tumor suppressor NF1 was altered in 11% of KRAS-mutant tumors (Figure 2A). NF1 activates the GTPase activity of KRAS [26]. However, loss of NF1 is not redundant in KRAS-induced myeloproliferative disorders as alterations in both genes resulted in a disease with a shorter latency [27]. These observations suggest that NF1 might have additional oncosuppressive functions and the presence of a concurrent mutation in KRAS might lead to a more aggressive disease. In summary, I identified four genes downstream of receptor activation with genetic alterations in the MAPK pathway.

RTKs and KRAS signaling can also activate the PI3K/AKT/mTOR pathway [13]. This pathway is commonly activated and genetically altered in lung cancer [28, 29]. I observed that alterations of several components of the PI3K/AKT/mTOR pathway were present in KRAS-mutant tumors. PIK3CG, PIK3C2G, TSC2, and mTOR were mutated in 12%, 10%, 13% and 13% of the tumors with KRAS alterations respectively (Figure 2A and Table 1). Consistent with previous studies [29], I

observe that mutations of individual components of this pathway were almost mutually exclusive (Data not shown). The tumor suppressor, INPP4B was mutated in 19% of KRAS-mutant tumors (Figure 2A and Table 1). PI(3,4)P2 is essential for AKT activation. INPP4B inactivates AKT by dephosphorylating PI(3,4)P2 [30]. Also, alterations in INPP4B can drive tumorigenesis via activation of serum and glucocorticoid-regulated kinase 3 (SGK3) [31]. STK11 is a serine/threonine kinase that functions as a tumor suppressor by regulating TP53-dependent apoptosis pathways [32] and also by activating TSC1/2 in the PI3K/AKT/mTOR pathway [33, 34] (Figure 3A). STK11 is commonly mutated in lung cancer [19], and it was mutated in 17% of all metastatic lung cancer samples in our dataset (Data not shown). Consistent with prior studies [19, 35], I observed that STK11 mutations were two times more likely to occur in KRAS-mutant cancer as opposed to KRAS wild-type tumors (Table 1). STK11 alterations were found in 35% of tumors with KRAS-mutant tumors. On the contrary, STK11 mutations were three times less likely to occur in EGFR-mutant tumors as opposed to EGFR wild-type cancer samples. Only 6% of the EGFR-mutant tumor had alterations in STK11 (Data not shown). These observations are consistent with the fact that STK11 and KRAS mutation occurs predominantly in smokers while EGFR mutations occur mostly in tumors from patients without a smoking history [19]. STK11 also seems to have additional effects in KRAS-mutant tumors. Mutations in STK11 was shown to suppress the immune surveillance in KRAS-mutant lung cancer, and they might serve as a predictor of de novo resistance in immune checkpoint blockade [36, 37]. In summary, I identified at least six genes downstream of RTK activation that were mutated in the PI3K/AKT/mTOR pathway.

Mutations in protein receptor tyrosine phosphatases and STAT3 pathway: The receptor-type protein tyrosine phosphatases (PTPRT, PTPRD, and PTPRS) were mutated in more than 50% of KRAS-mutant samples. These tumor suppressors were two times more likely to occur in samples with KRAS mutations as opposed to tumors with wild-type KRAS (Table 1). On the other hand, PTPRs mutations were less likely to be present in samples with EGFR alteration (data not shown). Consistent with previous studies [38, 39], PTPRT was mutated in 17% of all metastatic lung cancer samples, while PTPRD and PTPRS were mutated in 18% and 7% of the samples respectively. At least one of these PTPRs was mutated in 33% of all samples (Table 1). The role of PTPRs in lung

cancer is yet to be unraveled. However, phospho-STAT3 is a substrate for these phosphatases [40, 41] and frequent loss of function mutations in PTPRT and PTPRD have been linked to hyperphosphorylation of STAT3 in head and neck squamous carcinoma and in gliomagenesis [42-44]. STAT3 is hyperactive in some lung cancers [45]. This hyperactivity might be due to enhanced signaling from RTKs, cytokines or JAK kinases (altered in 11% of KRAS-mutant tumors) or by inactivating mutations in PTPRs [42, 46, 47].

Alterations in the antioxidant stress response pathway: Another tumor suppressor gene that had a high rate of genetic alterations in KRAS-mutant samples was KEAP1. Concurrent mutations in KRAS and KEAP1 in lung cancer is associated with a more aggressive disease [48]. The KEAP1 gene encodes for Kelch-like ECH associated protein 1 which is a negative modulator of NFE2L2, a master transcriptional regulator of the antioxidant response. Loss of KEAP1 was shown to alter the response of KRAS-mutant lung cancer cells to MAPK inhibitors. Krall E et al. showed that loss of KEAP1 could abolish the increase in reactive oxygen species that are associated with MAPK inhibition. They went further to show that deletion of KEAP1 can change the cell metabolism and enhance the survival and proliferation of the lung cancer cells in the absence of MAPK signaling [49]. In another study by Romero R et al., lung cancer cells with concurrent mutations in KRAS and KEAP1 were shown to be dependent on increase glutaminolysis and were more sensitive to glutaminase inhibition [50]. I observed that KEAP1 was 1.7 times more likely to be mutated in KRAS-mutant as oppose to KRAS-wildtype tumors. I observed that 30% of KRAS-mutant samples had mutations in KEAP1 (Figure 2B and Table 1). On the contrary, EGFR-mutant tumors were more than two-fold less likely to have a mutation in KEAP1 (data not shown).

Mutations in DNA replication stress response genes: Replication stress is an attribute of many malignant cells [51]. ATR mediate response to stress by arresting the progression of the cell cycle while stabilizing and repairing the replication fork. ATM and TP53 maintain the integrity of the genome when ATR is defective. ATM and TP53 arrest cell cycle progression allowing for DNA repair [52, 53]. I observed that TP53, ATM, and ATR were altered in 51%, 22% and 10% of the samples with KRAS mutations respectively (Table 1 and Figure 2C). ATM and ATR mutations were almost mutually exclusive in KRAS-mutant cells, and none of the samples with concurrent mutations in {KRAS, ATM, ATR, TP53}. However, {ATM, KRAS, TP53} and {ATR, KRAS, TP53}

subsets were present in 9.5% and 6% of KRAS-mutant samples (Data not shown). KRAS-mutant samples also had alterations in SETD2 (16%), FANCA (10%) and MRE11 (11%) which codes for proteins that regulate repair of DNA mismatch and double-strand breaks. In summary, six genes that are involved in DNA repair and the replication stress response mechanisms were mutated in at least 10% KRAS-mutant samples.

Mutations in transcription factors and chromatin remodeling genes: KRAS-mutant tumors had alterations SMARCA (19%), ARIDIA (13%), PBRM1 (10%), PRDM9 (10%), CREBBP (13%), TBX3 (11%), ATRX (14%), MLL2 (14%), MLL3 (14%), TERT (13%), ZFX3 (13%) and MED12 (16%) genes that codes for transcription factors and proteins involved in chromatin remodeling (Table 1). MED12 promotes tumorigenesis in part by regulating the Wnt/beta-catenin pathway [54]. I also identified mutations in other genes that are important for the formation of the beta-catenin/TCF transactivating complex including MLL2, MLL3, CREBBP, TERT, and SMARCA. SMARCA, ARIDIA, and PBRM1 are part of the chromatin remodeling complex SWI/SNF. The genes that encode for the subunits of the SWI/SNF complex are altered in more than 20% of all human cancers. Deficiency in one or more of these subunits including ARIDIA is associated with enhanced tumor growth and invasiveness [55]. Taken together, 12 genes coding for transcription factors and chromatin modifying proteins that were altered in at least 10% KRAS-mutant tumors.

Mutations in E3 ubiquitin ligases: Another group of genes that were altered in KRAS-mutant samples were genes coding for E3 ubiquitin ligases. PARK2, TRAF7, and RFW2 were mutated in 12%, 11% and 10% of the samples with KRAS alterations (Table 1). E3 ubiquitin ligases mediate the degradation of many proteins that are important for cancer development [56]. PARK2 deletion or loss of function was shown to promote inflammation, genomic instability, and progression of lung cancer [57, 58]. TRAF and RFW2 can ubiquitinate and promote degradation of TP53 leading to decrease TP53-dependent transcription and apoptosis [59]. In KRAS-mutant samples, mutations of the individual E3 ubiquitin ligases were mutually exclusive.

Mutations in Cadherins and β -catenins: Cadherins and catenins control cell-cell adhesion and dysregulation of their expressions are associated with increased metastasis and poor prognosis of lung and other cancers [60, 61]. I observed that CDH12 and the atypical cadherin FAT1 were

genetically altered in 13% and 15% of KRAS-mutant tumors respectively (Table 1). Mutations in FAT1 and CDH12 were mutually exclusive in KRAS-mutant tumors. CTNNA2 which codes for α -catenin was mutated in 34% of KRAS-mutant samples (Table 1). KRAS-mutant tumors were two-fold more likely to have a mutation in CTNNA2 than KRAS wild-type tumors. Mutations in CTNNA2 were less likely to occur in EGFR samples; only 10% of EGFR samples had alterations in CTNNA2, and half of these samples had concurrent {CTNNA2, EGFR, KRAS} mutations (Data not shown). The effect of CTNNA2 dysfunction have not been addressed in lung cancer, but frequent mutations in CTNNA2 were shown to promote migration and invasiveness of head and neck squamous cell carcinoma cells [62]. The mutation subsets {KRAS,CTNNA2,FAT1} and {KRAS,CTNNA2,CDH12} were present in 5% and 8% of KRAS-mutant tumors respectively. In summary, cadherins and catenins were collectively altered in at least 40% of KRAS-mutant tumors.

Analyses of rules involving three or more genes: To understand the complexity of the sub-populations of KRAS-mutant tumors, I queried the rules that were generated for the subtypes of cancers with mutations in three or more genes. As shown in figure 1B, mutations in KRAS were associated with 5724 rules. When I limited the query to rules with at least three genes and that were present in at least 10% of KRAS-mutant tumors, 160 subrules were generated (Figure 3A and Table S2). Most (79%) of these rules had mutations in three genes. Many of these rules were associated with mutations in PTPRs, TP53, CTNNA2 and the pseudogene LOC1750 (Figure 3A). There was a great degree of “overlap” between the rules. For example, all the 17 rules with CTNNA2 on the RHS were similar to the 95% of the rules with LOC1750 on the RHS.

Mutations in PTPRs can lead to constitutive activation of STAT3 which is associated with many lung cancer [42-44]. PTPRs alterations were collectively altered in at least 50% of KRAS-mutant tumors (Table 1). A concurrent mutation in KRAS might lead to a tumor with hyperactive MAPK, AKT and STAT3 signaling [20, 21]. I explored the subsets of tumors with KRAS and PTPRs mutations to identify other potential pathways that might be defective (Figure 3B). I identified {KRAS, PTPRT, SMARCA4}, {KRAS, PTPRT, MED12}, and {KRAS, PTPRT, MLL2} subsets in 11.9%, 11.3% and 10.7% of KRAS-mutant tumors respectively. In addition to hyperactivity of MAPK, AKT and STAT3 pathways, additional mutations in SMARCA4, MED12 or MLL2 could also implicate

Wnt/beta-catenin pathway and transcriptional regulation by the SWI/SNF complex [54, 55]. Another subset of tumors with KRAS and PTPRT mutations were {KRAS, KEAP1, PTPRT} and {CTNNA2, KRAS, PTPRT} in 14% and 21.4% of the KRAS-mutant tumors respectively. Concurrent mutations in KRAS, KEAP1, and PTPRT can lead to a tumor with hyperactive MAPK, AKT and STAT3 pathways and defective antioxidant response mechanisms [49, 50]. While mutations in CTNNA2, KRAS, and PTPRT could lead to a tumor with defective cell-cell adherence in addition to constitutive activation in the MAPK, AKT and STAT3 pathways [62].

STK11 alterations are associated with poor prognosis in lung cancer [35]. However, there are conflicting reports on the effects of alterations in STK11 in KRAS-mutant tumors. Some researchers reported that concurrent STK11 and KRAS mutations did not affect the aggressiveness of the disease in patients [48, 63]. Findings from other studies show that STK11 can potentiate tumor progression in KRAS-mutant tumors in mouse models [64] and was associated with a more aggressive disease [36]. It is possible that this discrepancy might be due to the presence of other alterations in subpopulations of KRAS^{mut}/STK11^{mut} lung cancer. I observed that about 66% of KRAS^{mut}/STK11^{mut} tumors also had alterations in KEAP1, which regulates oxidative stress response [49, 50]. I also observed that STK11^{mut}/PTPRs^{mut} and STK11^{mut}/CTNNA2^{mut} were present in about 35%, and 33% of the tumors with KRAS mutations (Figure 3C). Such combinations of mutations might affect the biology of the disease leading to different clinical outcomes. The design of research models that can capture such combinations of mutation could enhance our understanding of the disease.

Conclusion

In this study, I have implemented a smart approach to finding subpopulations of metastatic KRAS-mutant lung cancers based on the combination of somatic mutations in genes coding for oncogenic and oncosuppressive proteins. Designing research models that captures these overlapping subpopulations could lead to the development of combination therapies that are effective and can circumvent potential resistance mechanisms in KRAS-mutant metastatic lung cancer.

Figure legend

Figure 1: KRAS-mutant lung cancer is associated with alterations in genes that regulate diverse cellular processes: Scatter plot showing the distribution of (A) all the rules generated for the genes that were mutated in all the metastatic lung cancer samples. (B) The subset of the rules containing KRAS. (C) The association between mutations in KRAS and other tumor drivers. Note that the lift (likelihood) for KRAS-mutant tumors to also have mutations in EGFR, MET or PI3K3CA was less than one would expect by chance. (D). A directed graph showing the association between KRAS and 67 genes. I limited our query to a one to one association and with KRAS on the LHS to identify the genes that were also mutated in KRAS-mutant samples. I used a confidence of 0.1 to include only the genes that were altered in at least 10% of KRAS samples. The subrules generated in R was imported into Gephi and manipulated to represent the nodes with distinct colors based on their degree of connectivity. (E) I used the 67 genes that were changing as input into the REACTOME pathway analysis tool. Shown in E, is a subset of the pathways that were enriched in the 67-member gene list.

Figure 2: Schematics of genes in the MAPK, mTOR and STAT3 pathways, the oxidative stress response and the DNA repair mechanisms that were altered in KRAS-mutant lung cancer: Schematics showing the relationship between the genes that were mutated in KRAS-mutant tumors. Red denotes oncogenic proteins, and blue is for oncosuppressive proteins. (A) The members of the MAPK, PI3K/AKT/mTOR, and STAT3 pathways, (B) KEAP1 regulation of oxidative stress response, (C) TP53/ATM/ATR mediated cell cycle arrests.

Figure 3: Analyses of rules with 3 or more members: (A) I generated all the rules that contained KRAS as in figure 1B. The rules were then subsetted to include only rules with at least three genes, with KRAS on the LHS and that was present in at least 10% of KRAS-mutant samples. The resulting 160 rules were imported as a graph file into Gephi. Files were manipulated to show the nodes of different colors based on the edges coming into the nodes (in-degree). Note that KRAS had an in-degree of zero since it was restricted to the LHS. Mutations in PTPRs were the most connected with 32 in-degrees. (B) and (C) shows the tabular representation of the rules that converge on PTPRs and STK11 nodes.

Table Legend

Table 1: Genes associated with KRAS in metastatic lung cancer: The support is the proportion of metastatic lung cancer covered by the rule. Confidence (accuracy) is the proportion of the KRAS-mutant tumors covered by the rule. Lift is the likelihood of having a mutation in the gene on the right of the rule given that the sample has an alteration in KRAS.

R codes

#Load data and required packages

```
library(arules)
library(arulesViz)
data <- read.transactions("data.csv", sep = ",")
```

#Model training

```
## Support = 0.01, confidence = 0.1 and minimum length of rule = 2
cancer_rules <- apriori(data, parameter = list(support = 0.01, confidence = 0.1, minlen = 2))
```

#Figure 1A

```
plot(cancer_rules)
```

#Figure 1B

```
KRAS_rules <- subset(cancer_rules, subset = items %ain% c("KRAS"))
plot(KRAS_rules)
```

#Figure 1D

```
cancer_rules <- apriori(data, parameter = list(support =
0.01, confidence = 0.1, minlen = 2, maxlen = 2))

KRAS_rules <- subset(cancer_rules, subset = lhs %ain% c("KRAS"))
saveAsGraph((KRAS_rules), file="KRAS_lhs_2-member_rules.graphml")
```

```
## Import and manipulate the saved file in Gephi
```

Figure 3A

```
cancer_rules <- apriori(data, parameter = list(support =  
0.01, confidence = 0.1, minlen = 3))
```

```
KRAS_rules <- subset(cancer_rules, subset = lhs %ain% c("KRAS") & count >= 17)  
saveAsGraph((KRAS_rules), file="KRAS_lhs_3-member_rules_count_17.graphml")
```

```
## Import and manipulate the saved file in Gephi
```

#Figure 3B

```
cancer_rules <- apriori(data, parameter = list(support =  
0.01, confidence = 0.1, minlen = 3))
```

```
KRAS_rules <- subset(cancer_rules, subset = lhs %ain% c("KRAS") & count >= 17)  
## Count >= 17, to include only rules that appear in at least 10% of KRAS-mutant samples  
KRAS_PTPRs_rules <- subset(KRAS_rules, subset = rhs %ain% c("PTPRs"))
```

```
inspect(KRAS_PTPRs_rules) ## Or write to a csv file
```

#Figure 3C

```
cancer_rules <- apriori(data, parameter = list(support =  
0.01, confidence = 0.1, minlen = 3))
```

```
KRAS_rules <- subset(cancer_rules, subset = lhs %ain% c("KRAS") & count >= 17)
```

```
KRAS_STK11_rules <- subset(KRAS_rules, subset = rhs %ain% c("STK11"))  
inspect(KRAS_STK11_rules) ## Or write to a csv file
```

Declarations

Ethics approval and consent of participants

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data supporting the findings from in this study were obtained from the CBioportal site[17]. I have included supplementary files for all the analyses that I performed.

Competing interests

I declare that I have no competing interests

Acknowledgement

Not applicable

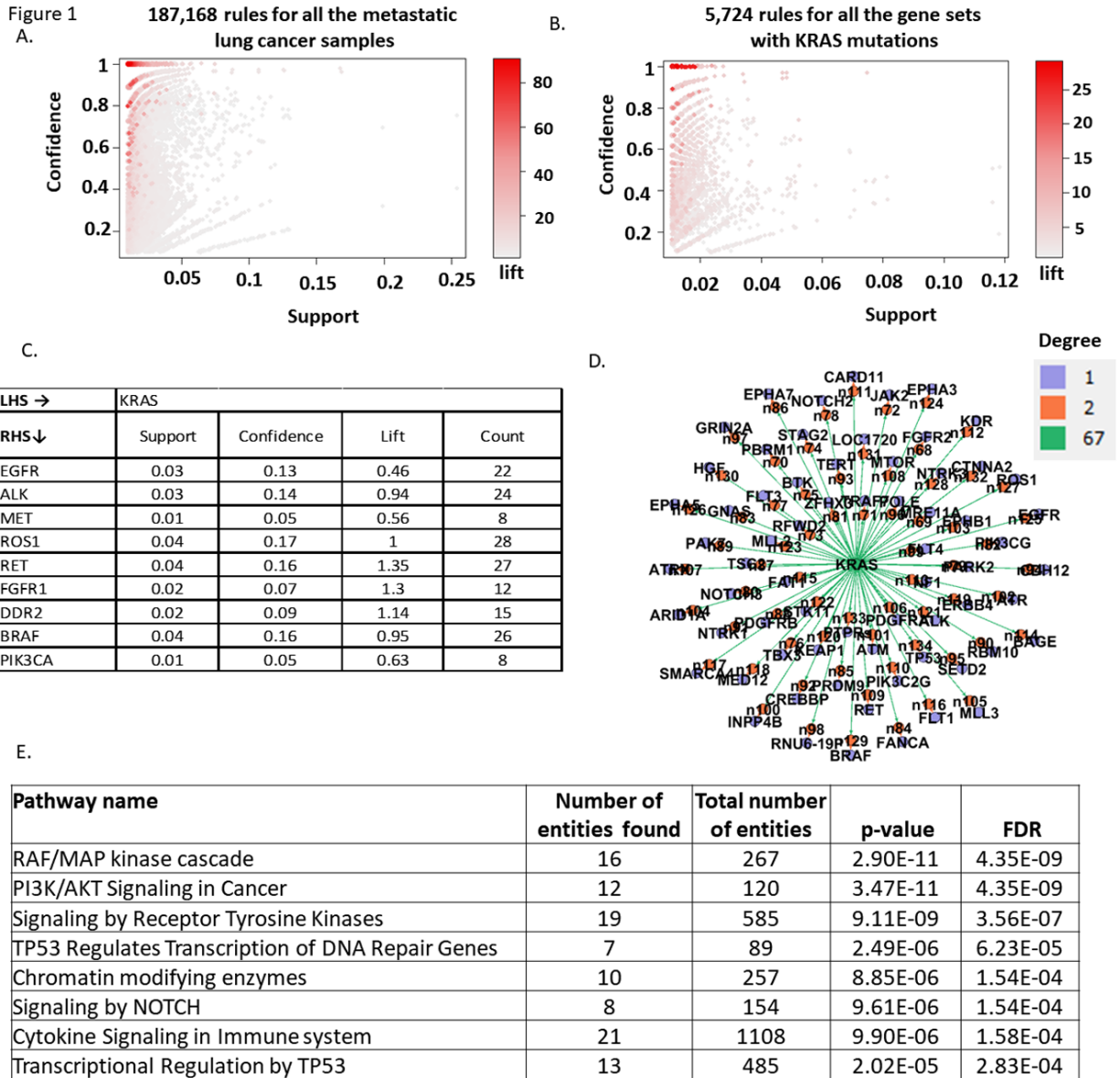
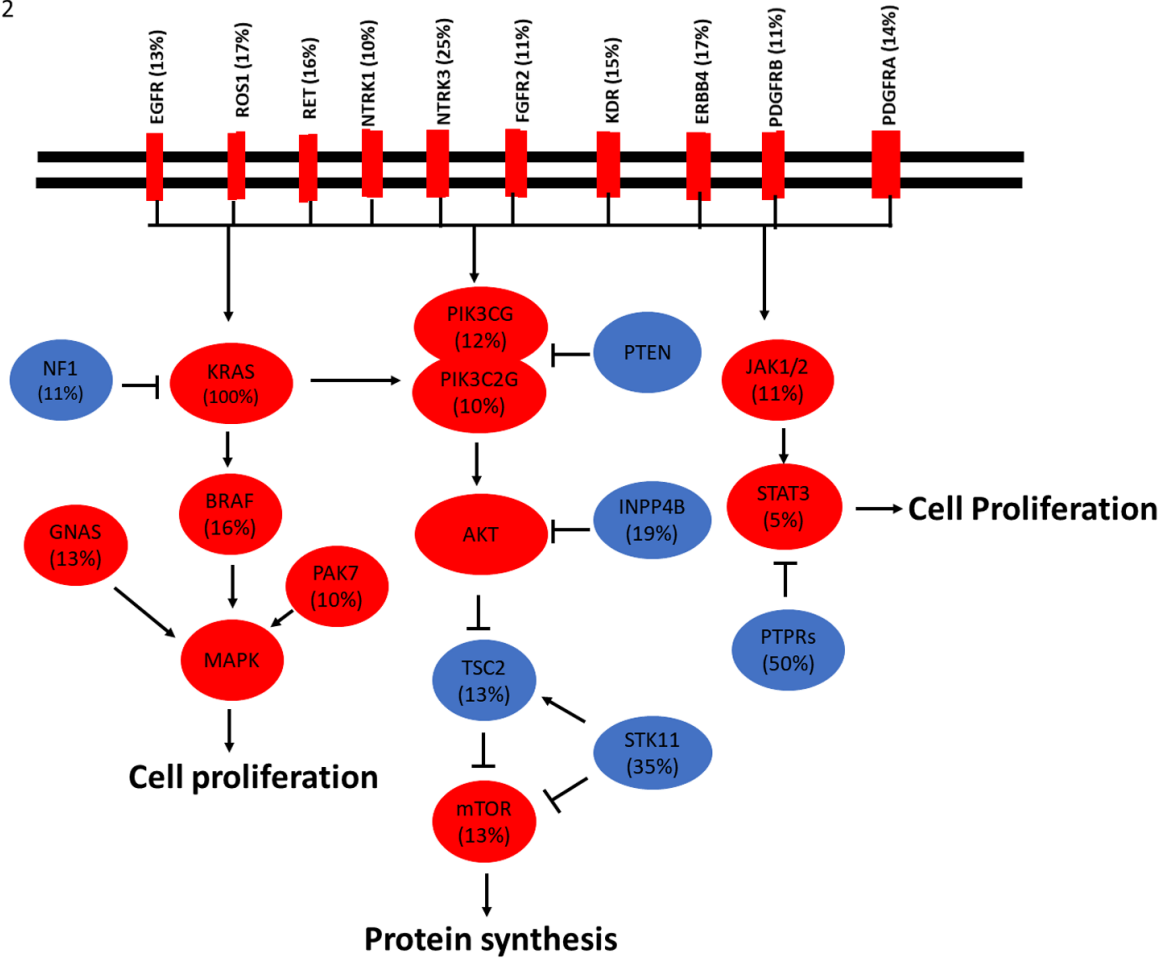
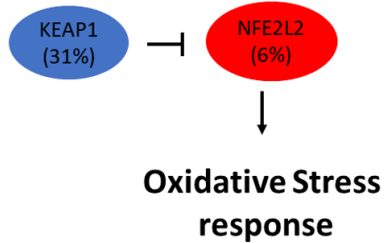


Figure 2

A.



B.



C.

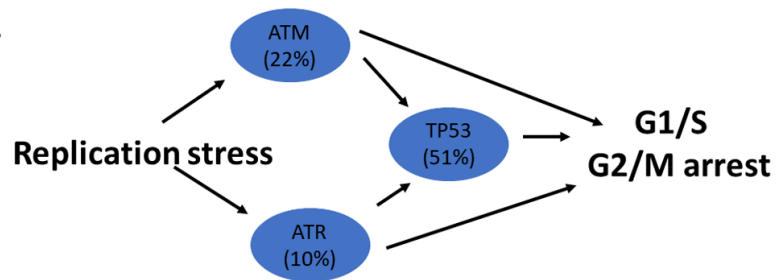
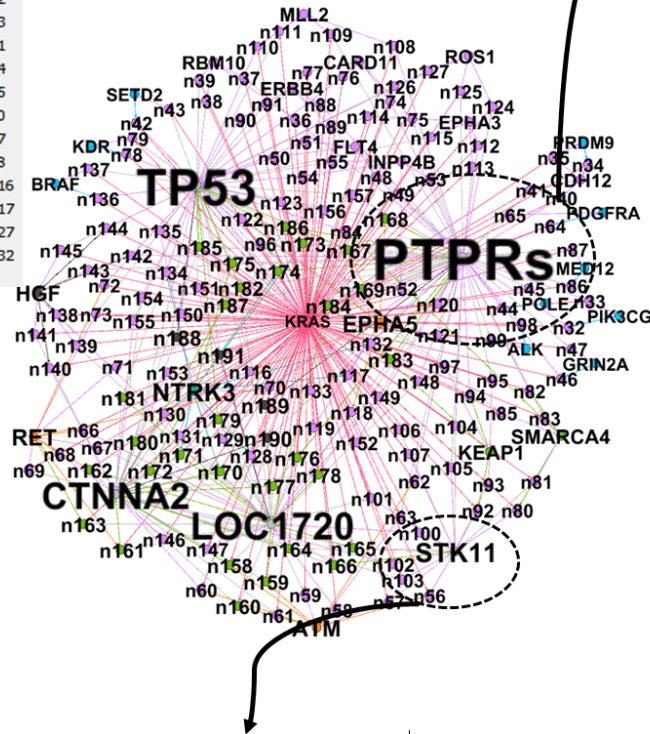


Figure 3

A.

In-degree



B.

Gene set	% of KRAS-mutant samples
{KRAS, PTPRS}	50.0%
{KRAS, PIK3CG, PTPRS}	10.1%
{KRAS, RBM10, PTPRS}	10.7%
{CDH12, KRAS, PTPRS}	10.1%
{KRAS, POLE, PTPRS}	10.1%
{GRIN2A, KRAS, PTPRS}	10.1%
{FLT4, KRAS, PTPRS}	12.5%
{INPP4B, KRAS, PTPRS}	15.5%
{ATM, KRAS, PTPRS}	13.1%
{KRAS, PDGFRA, PTPRS}	10.1%
{KRAS, RET, PTPRS}	11.9%
{CARD11, KRAS, PTPRS}	13.1%
{KRAS, SMARCA4, PTPRS}	11.9%
{KRAS, MED12, PTPRS}	11.3%
{ERBB4, KRAS, PTPRS}	12.5%
{KEAP1, KRAS, PTPRS}	13.7%
{ALK, KRAS, PTPRS}	10.1%
{KRAS, STK11, PTPRS}	15.5%
{KRAS, MLL2, PTPRS}	10.7%
{EPHA3, KRAS, PTPRS}	16.1%
{EPHA5, KRAS, PTPRS}	15.5%
{KRAS, ROS1, PTPRS}	11.3%
{KRAS, NTRK3, PTPRS}	16.7%
{HGF, KRAS, PTPRS}	11.3%
{KRAS, LOC1720, PTPRS}	22.0%
{CTNNA2, KRAS, PTPRS}	21.4%
{KRAS, TP53, PTPRS}	29.8%
{EPHA5, KRAS, TP53, PTPRS}	11.3%
{KRAS, NTRK3, TP53, PTPRS}	11.3%
{CTNNA2, KRAS, LOC1720, PTPRS}	20.8%
{KRAS, LOC1720, TP53, PTPRS}	16.1%
{CTNNA2, KRAS, TP53, PTPRS}	15.5%
{CTNNA2, KRAS, LOC1720, TP53, PTPRS}	14.9%

C.

Gene set	% of KRAS mutant samples
{KRAS, STK11}	35.1%
{ATM, KRAS, STK11}	10.7%
{KRAS, SMARCA4, STK11}	10.1%
{KEAP1, KRAS, STK11}	22.6%
{KRAS, LOC1720, STK11}	12.5%
{CTNNA2, KRAS, STK11}	11.9%
{KRAS, PTPRS, STK11}	15.5%
{KRAS, TP53, STK11}	10.7%
{CTNNA2, KRAS, LOC1720, STK11}	11.9%

	rules	support	confidence	lift		rules	support	confidence	lift
1	{KRAS → ERBB4}	0.04	0.17	1.35	35	{KRAS → SETD2}	0.04	0.16	1.43
2	{KRAS → EGFR}	0.03	0.13	0.46	36	{KRAS → MRE11A}	0.02	0.11	1.95
3	{KRAS → ROS1}	0.04	0.17	1.00	37	{KRAS → FANCA}	0.02	0.10	1.32
4	{KRAS → ALK}	0.03	0.14	0.94	38	{KRAS → BAGE}	0.03	0.14	1.39
5	{KRAS → RET}	0.04	0.16	1.35	39	{KRAS → RBM10}	0.04	0.17	1.62
6	{KRAS → FGFR2}	0.02	0.11	1.95	40	{KRAS → MLL2}	0.03	0.14	1.12
7	{KRAS → FLT1}	0.04	0.16	1.33	41	{KRAS → MLL3}	0.03	0.14	1.05
8	{KRAS → FLT3}	0.02	0.10	1.57	42	{KRAS → TBX3}	0.03	0.11	1.50
9	{KRAS → FLT4}	0.04	0.19	1.77	43	{KRAS → ATRX}	0.03	0.14	1.09
10	{KRAS → KDR}	0.03	0.15	1.45	44	{KRAS → SMARCA4}	0.04	0.19	1.32
11	{KRAS → EPHA3}	0.05	0.22	1.57	45	{KRAS → CREBBP}	0.03	0.13	1.36
12	{KRAS → EPHA5}	0.05	0.22	1.57	46	{KRAS → ZFXH3}	0.03	0.13	1.66
13	{KRAS → EPHA7}	0.03	0.12	1.52	47	{KRAS → TERT}	0.03	0.13	1.17
14	{KRAS → EPHB1}	0.02	0.11	1.06	48	{KRAS → ARID1A}	0.03	0.13	1.14
15	{KRAS → NTRK1}	0.02	0.10	1.15	49	{KRAS → PBRM1}	0.02	0.10	1.29
16	{KRAS → NTRK3}	0.06	0.25	1.84	50	{KRAS → MED12}	0.04	0.16	1.25
17	{KRAS → PDGFRA}	0.03	0.14	1.41	51	{KRAS → PRDM9}	0.02	0.10	1.37
18	{KRAS → PDGFRB}	0.03	0.11	1.29	52	{KRAS → CDH12}	0.03	0.13	1.45
19	{KRAS → HGF}	0.05	0.20	1.27	53	{KRAS → FAT1}	0.03	0.15	1.09
20	{KRAS → BRAF}	0.04	0.16	0.95	54	{KRAS → CTNNA2}	0.08	0.34	1.91
21	{KRAS → PAK7}	0.02	0.10	1.34	55	{KRAS → TRAF7}	0.02	0.10	2.11
22	{KRAS → NF1}	0.03	0.11	0.92	56	{KRAS → RFWD2}	0.03	0.11	1.83
23	{KRAS → GNAS}	0.03	0.13	1.57	57	{KRAS → PARK2}	0.03	0.12	1.77
24	{KRAS → PIK3C2G}	0.02	0.10	0.95	58	{KRAS → NOTCH2}	0.02	0.10	1.25
25	{KRAS → PIK3CG}	0.03	0.12	1.38	59	{KRAS → NOTCH3}	0.03	0.11	1.37
26	{KRAS → TSC2}	0.03	0.13	1.30	60	{KRAS → KEAP1}	0.07	0.31	1.73
27	{KRAS → MTOR}	0.03	0.13	1.24	61	{KRAS → LOC1720}	0.08	0.34	1.94
28	{KRAS → STK11}	0.08	0.35	2.02	62	{KRAS → RNU6-19P}	0.02	0.10	0.93
29	{KRAS → INPP4B}	0.04	0.19	1.73	63	{KRAS → CARD11}	0.04	0.17	1.56
30	{KRAS → JAK2}	0.02	0.11	1.63	64	{KRAS → BTK}	0.02	0.11	1.63
31	{KRAS → PTPRT}	0.12	0.50	1.49	65	{KRAS → STAG2}	0.02	0.10	1.34
32	{KRAS → TP53}	0.12	0.51	0.82	66	{KRAS → POLE}	0.03	0.14	1.63
33	{KRAS → ATM}	0.05	0.22	1.69	67	{KRAS → GRIN2A}	0.04	0.18	1.69
34	{KRAS → ATR}	0.02	0.10	0.79					

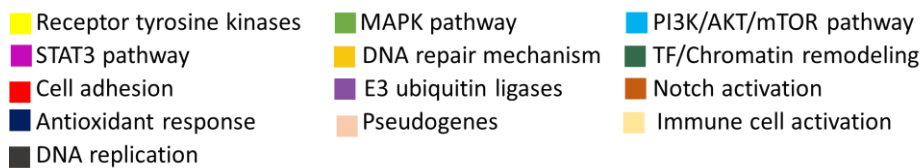


Table 1

References

- [1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman and F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012, *International journal of cancer* **136** (2015), E359-86.
- [2] R.L. Siegel, K.D. Miller and A. Jemal, Cancer statistics, 2015, *CA: a cancer journal for clinicians* **65** (2015), 5-29.
- [3] M.B. Carper and P.P. Claudio, Clinical potential of gene mutations in lung cancer, *Clinical and translational medicine* **4** (2015), 33-015-0074-1. Epub 2015 Nov 24.
- [4] A. Sandler, R. Gray, M.C. Perry, J. Brahmer, J.H. Schiller, A. Dowlati, R. Lilienbaum and D.H. Johnson, Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer, *The New England journal of medicine* **355** (2006), 2542-2550.
- [5] T.S. Mok, Y.L. Wu, S. Thongprasert, C.H. Yang, D.T. Chu, N. Saijo, P. Sunpaweravong, B. Han, B. Margono, Y. Ichinose, Y. Nishiwaki, Y. Ohe, J.J. Yang, B. Chewaskulyong, H. Jiang, E.L. Duffield, C.L. Watkins, A.A. Armour and M. Fukuoka, Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma, *The New England journal of medicine* **361** (2009), 947-957.
- [6] L.V. Sequist, J.C. Yang, N. Yamamoto, K. O'Byrne, V. Hirsh, T. Mok, S.L. Geater, S. Orlov, C.M. Tsai, M. Boyer, W.C. Su, J. Bannouna, T. Kato, V. Gorbunova, K.H. Lee, R. Shah, D. Massey, V. Zazulina, M. Shahidi and M. Schuler, Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations, *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31** (2013), 3327-3334.
- [7] R. Katayama, C.M. Lovly and A.T. Shaw, Therapeutic targeting of anaplastic lymphoma kinase in lung cancer: a paradigm for precision cancer medicine, *Clinical cancer research : an official journal of the American Association for Cancer Research* **21** (2015), 2227-2235.
- [8] T. Regad, Targeting RTK Signaling Pathways in Cancer, *Cancers* **7** (2015), 1758-1784.
- [9] Z. Lohinai, T. Klikovits, J. Moldvay, G. Ostoros, E. Raso, J. Timar, K. Fabian, I. Kovalszky, I. Kenessey, C. Aigner, F. Renyi-Vamos, W. Klepetko, B. Dome and B. Hegedus, KRAS-mutation incidence and prognostic value are metastatic site-specific in lung adenocarcinoma: poor prognosis in patients with KRAS mutation and bone metastasis, *Scientific reports* **7** (2017), 39721.
- [10] A.E. Karnoub and R.A. Weinberg, Ras oncogenes: split personalities, *Nature reviews.Molecular cell biology* **9** (2008), 517-531.

- [11] D.M. Feldser and S.E. Kern, Oncogenic levels of mitogen-activated protein kinase (MAPK) signaling of the dinucleotide KRAS2 mutations G12F and GG12-13VC, *Human mutation* **18** (2001), 357.
- [12] P.J. Roberts and C.J. Der, Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer, *Oncogene* **26** (2007), 3291-3310.
- [13] N. Mitin, K.L. Rossman and C.J. Der, Signaling interplay in Ras superfamily function, *Current biology : CB* **15** (2005), R563-74.
- [14] A. Matikas, D. Mistriotis, V. Georgoulas and A. Kotsakis, Targeting KRAS mutated non-small cell lung cancer: A history of failures and a future of hope for a diverse entity, *Critical reviews in oncology/hematology* **110** (2017), 1-12.
- [15] R. Agrawal, T. Imieliński and A. Swami, Mining association rules between sets of items in large databases, *SIGMOD Rec.* **22** (1993), 207-216.
- [16] R. Agrawal and R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, (1994), 487-499.
- [17] A. Zehir, R. Benayed, R.H. Shah, A. Syed, S. Middha, H.R. Kim, P. Srinivasan, J. Gao, D. Chakravarty, S.M. Devlin, M.D. Hellmann, D.A. Barron, A.M. Schram, M. Hameed, S. Dogan, D.S. Ross, J.F. Hechtman, D.F. DeLair, J. Yao, D.L. Mandelker, D.T. Cheng, R. Chandramohan, A.S. Mohanty, R.N. Ptashkin, G. Jayakumar, M. Prasad, M.H. Syed, A.B. Rema, Z.Y. Liu, K. Nafa, L. Borsu, J. Sadowska, J. Casanova, R. Bacares, I.J. Kiecka, A. Razumova, J.B. Son, L. Stewart, T. Baldi, K.A. Mullaney, H. Al-Ahmadie, E. Vakiani, A.A. Abeshouse, A.V. Penson, P. Jonsson, N. Camacho, M.T. Chang, H.H. Won, B.E. Gross, R. Kundra, Z.J. Heins, H.W. Chen, S. Phillips, H. Zhang, J. Wang, A. Ochoa, J. Wills, M. Eubank, S.B. Thomas, S.M. Gardos, D.N. Reales, J. Galle, R. Durany, R. Cambria, W. Abida, A. Cercek, D.R. Feldman, M.M. Gounder, A.A. Hakimi, J.J. Harding, G. Iyer, Y.Y. Janjigian, E.J. Jordan, C.M. Kelly, M.A. Lowery, L.G.T. Morris, A.M. Omuro, N. Raj, P. Razavi, A.N. Shoushtari, N. Shukla, T.E. Soumerai, A.M. Varghese, R. Yaeger, J. Coleman, B. Bochner, G.J. Riely, L.B. Saltz, H.I. Scher, P.J. Sabbatini, M.E. Robson, D.S. Klimstra, B.S. Taylor, J. Baselga, N. Schultz, D.M. Hyman, M.E. Arcila, D.B. Solit, M. Ladanyi and M.F. Berger, Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients, *Nature medicine* **23** (2017), 703-713.
- [18] C. Scoccianti, A. Vesin, G. Martel, M. Olivier, E. Brambilla, J.F. Timsit, L. Tavecchio, C. Brambilla, J.K. Field, P. Hainaut and European Early Lung Cancer Consortium, Prognostic value of TP53, KRAS and EGFR mutations in nonsmall cell lung cancer: the EUELC cohort, *The European respiratory journal* **40** (2012), 177-184.
- [19] L. Ding, G. Getz, D.A. Wheeler, E.R. Mardis, M.D. McLellan, K. Cibulskis, C. Sougnez, H. Greulich, D.M. Muzny, M.B. Morgan, L. Fulton, R.S. Fulton, Q. Zhang, M.C. Wendl, M.S. Lawrence, D.E. Larson, K. Chen, D.J. Dooling, A. Sabo, A.C. Hawes, H. Shen, S.N. Jhangiani, L.R.

Lewis, O. Hall, Y. Zhu, T. Mathew, Y. Ren, J. Yao, S.E. Scherer, K. Clerc, G.A. Metcalf, B. Ng, A. Milosavljevic, M.L. Gonzalez-Garay, J.R. Osborne, R. Meyer, X. Shi, Y. Tang, D.C. Koboldt, L. Lin, R. Abbott, T.L. Miner, C. Pohl, G. Fewell, C. Haipek, H. Schmidt, B.H. Dunford-Shore, A. Kraja, S.D. Crosby, C.S. Sawyer, T. Vickery, S. Sander, J. Robinson, W. Winckler, J. Baldwin, L.R. Chirieac, A. Dutt, T. Fennell, M. Hanna, B.E. Johnson, R.C. Onofrio, R.K. Thomas, G. Tonon, B.A. Weir, X. Zhao, L. Ziaugra, M.C. Zody, T. Giordano, M.B. Orringer, J.A. Roth, M.R. Spitz, I.I. Wistuba, B. Ozenberger, P.J. Good, A.C. Chang, D.G. Beer, M.A. Watson, M. Ladanyi, S. Broderick, A. Yoshizawa, W.D. Travis, W. Pao, M.A. Province, G.M. Weinstock, H.E. Varmus, S.B. Gabriel, E.S. Lander, R.A. Gibbs, M. Meyerson and R.K. Wilson, Somatic mutations affect key pathways in lung adenocarcinoma, *Nature* **455** (2008), 1069-1075.

[20] H.O. Kilgoz, G. Bender, J.M. Scandura, A. Viale and B. Taneri, KRAS and the Reality of Personalized Medicine in Non-small Cell Lung Cancer, *Molecular medicine (Cambridge, Mass.)* **22** (2016), 10.2119/molmed.2016.00151.

[21] M. Paolo, S. Assunta, R. Antonio, S.P. Claudia, B.M. Anna, S. Clorinda, C. Francesca, C. Fortunato and G. Cesare, Selumetinib in advanced non small cell lung cancer (NSCLC) harbouring KRAS mutation: endless clinical challenge to KRAS-mutant NSCLC, *Reviews on recent clinical trials* **8** (2013), 93-100.

[22] H. Davies, G.R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, J. Teague, H. Woffendin, M.J. Garnett, W. Bottomley, N. Davis, E. Dicks, R. Ewing, Y. Floyd, K. Gray, S. Hall, R. Hawes, J. Hughes, V. Kosmidou, A. Menzies, C. Mould, A. Parker, C. Stevens, S. Watt, S. Hooper, R. Wilson, H. Jayatilake, B.A. Gusterson, C. Cooper, J. Shipley, D. Hargrave, K. Pritchard-Jones, N. Maitland, G. Chenevix-Trench, G.J. Riggins, D.D. Bigner, G. Palmieri, A. Cossu, A. Flanagan, A. Nicholson, J.W. Ho, S.Y. Leung, S.T. Yuen, B.L. Weber, H.F. Seigler, T.L. Darrow, H. Paterson, R. Marais, C.J. Marshall, R. Wooster, M.R. Stratton and P.A. Futreal, Mutations of the BRAF gene in human cancer, *Nature* **417** (2002), 949-954.

[23] R. Seth, S. Crook, S. Ibrahim, W. Fadhil, D. Jackson and M. Ilyas, Concomitant mutations and splice variants in KRAS and BRAF demonstrate complex perturbation of the Ras/Raf signalling pathway in advanced colorectal cancer, *Gut* **58** (2009), 1234-1241.

[24] I.H. Sahin, S.M. Kazmi, J.T. Yorio, N.A. Bhadkamkar, B.K. Kee and C.R. Garrett, Rare Though Not Mutually Exclusive: A Report of Three Cases of Concomitant KRAS and BRAF Mutation and a Review of the Literature, *Journal of Cancer* **4** (2013), 320-322.

[25] C.H. Wilson, R.E. McIntyre, M.J. Arends and D.J. Adams, The activating mutation R201C in GNAS promotes intestinal tumourigenesis in Apc(Min/+) mice through activation of Wnt and ERK1/2 MAPK pathways, *Oncogene* **29** (2010), 4567-4575.

[26] S. Donovan, K.M. Shannon and G. Bollag, GTPase activating proteins: critical regulators of intracellular signaling, *Biochimica et biophysica acta* **1602** (2002), 23-45.

- [27] B.A. Cutts, A.K. Sjogren, K.M. Andersson, A.M. Wahlstrom, C. Karlsson, B. Swolin and M.O. Bergo, Nf1 deficiency cooperates with oncogenic K-RAS to induce acute myeloid leukemia in mice, *Blood* **114** (2009), 3629-3632.
- [28] H. Cheng, M. Shcherba, G. Pendurti, Y. Liang, B. Piperdi and R. Perez-Soler, Targeting the PI3K/AKT/mTOR pathway: potential for lung cancer treatment, *Lung cancer management* **3** (2014), 67-75.
- [29] S. Umemura, S. Mimaki, H. Makinoshima, S. Tada, G. Ishii, H. Ohmatsu, S. Niho, K. Yoh, S. Matsumoto, A. Takahashi, M. Morise, Y. Nakamura, A. Ochiai, K. Nagai, R. Iwakawa, T. Kohno, J. Yokota, Y. Ohe, H. Esumi, K. Tsuchihara and K. Goto, Therapeutic priority of the PI3K/AKT/mTOR pathway in small cell lung cancers as revealed by a comprehensive genomic analysis, *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* **9** (2014), 1324-1331.
- [30] C. Li Chew, A. Lunardi, F. Gulluni, D.T. Ruan, M. Chen, L. Salmena, M. Nishino, A. Papa, C. Ng, J. Fung, J.G. Clohessy, J. Sasaki, T. Sasaki, R.T. Bronson, E. Hirsch and P.P. Pandolfi, In Vivo Role of INPP4B in Tumor and Metastasis Suppression through Regulation of PI3K-AKT Signaling at Endosomes, *Cancer discovery* **5** (2015), 740-751.
- [31] J.A. Gasser, H. Inuzuka, A.W. Lau, W. Wei, R. Beroukhim and A. Toker, SGK3 mediates INPP4B-dependent PI3K signaling in breast cancer, *Molecular cell* **56** (2014), 595-607.
- [32] P. Karuman, O. Gozani, R.D. Odze, X.C. Zhou, H. Zhu, R. Shaw, T.P. Brien, C.D. Bozzuto, D. Ooi, L.C. Cantley and J. Yuan, The Peutz-Jegher gene product LKB1 is a mediator of p53-dependent cell death, *Molecular cell* **7** (2001), 1307-1319.
- [33] S.C. Stein, A. Woods, N.A. Jones, M.D. Davison and D. Carling, The regulation of AMP-activated protein kinase by phosphorylation, *The Biochemical journal* **345 Pt 3** (2000), 437-443.
- [34] A. Woods, S.R. Johnstone, K. Dickerson, F.C. Leiper, L.G. Fryer, D. Neumann, U. Schlattner, T. Wallimann, M. Carlson and D. Carling, LKB1 is the upstream kinase in the AMP-activated protein kinase cascade, *Current biology : CB* **13** (2003), 2004-2008.
- [35] N. Pecuchet, P. Laurent-Puig, A. Mansuet-Lupo, A. Legras, M. Alifano, K. Pallier, A. Didelot, L. Gibault, C. Danel, P.A. Just, M. Riquet, F. Le Pimpec-Barthes, D. Damotte, E. Fabre and H. Blons, Different prognostic impact of STK11 mutations in non-squamous non-small-cell lung cancer, *Oncotarget* **8** (2017), 23831-23840.
- [36] M.B. Schabath, E.A. Welsh, W.J. Fulp, L. Chen, J.K. Teer, Z.J. Thompson, B.E. Engel, M. Xie, A.E. Berglund, B.C. Creelan, S.J. Antonia, J.E. Gray, S.A. Eschrich, D.T. Chen, W.D. Cress, E.B. Haura and A.A. Beg, Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma, *Oncogene* **35** (2016), 3209-3216.

- [37] F. Skoulidis, M.D. Hellmann, M.M. Awad, H. Rizvi, B.W. Carter, W. Denning, Y. Elamin, J. Zhang, G.C. Leonardi, D. Halpenny, A. Plodkowski, N. Long, J.J. Erasmus, V. Papadimitrakopoulou, K. Wong, I.I. Wistuba, P.A. Janne, C.M. Rudin and J. Heymach, STK11/LKB1 co-mutations to predict for de novo resistance to PD-1/PD-L1 axis blockade in KRAS-mutant lung adenocarcinoma. *Journal of Clinical Oncology* **35** (2017), 9016-9016.
- [38] S.G. Julien, N. Dube, S. Hardy and M.L. Tremblay, Inside the human cancer tyrosine phosphatome, *Nature reviews.Cancer* **11** (2011), 35-49.
- [39] S. Zhao, D. Sedwick and Z. Wang, Genetic alterations of protein tyrosine phosphatases in human cancers, *Oncogene* **34** (2015), 3885-3894.
- [40] X. Zhang, A. Guo, J. Yu, A. Possemato, Y. Chen, W. Zheng, R.D. Polakiewicz, K.W. Kinzler, B. Vogelstein, V.E. Velculescu and Z.J. Wang, Identification of STAT3 as a substrate of receptor protein tyrosine phosphatase T, *Proceedings of the National Academy of Sciences of the United States of America* **104** (2007), 4060-4064.
- [41] T.A. Chan and A. Heguy, The protein tyrosine phosphatase receptor D, a broadly inactivated tumor suppressor regulating STAT function, *Cell cycle (Georgetown, Tex.)* **8** (2009), 3063-3064.
- [42] N.D. Peyser, Y. Du, H. Li, V. Lui, X. Xiao, T.A. Chan and J.R. Grandis, Loss-of-Function PTPRD Mutations Lead to Increased STAT3 Activation and Sensitivity to STAT3 Inhibition in Head and Neck Cancer, *PloS one* **10** (2015), e0135750.
- [43] B. Ortiz, A.W. Fabius, W.H. Wu, A. Pedraza, C.W. Brennan, N. Schultz, K.L. Pitter, J.F. Bromberg, J.T. Huse, E.C. Holland and T.A. Chan, Loss of the tyrosine phosphatase PTPRD leads to aberrant STAT3 activation and promotes gliomagenesis, *Proceedings of the National Academy of Sciences of the United States of America* **111** (2014), 8149-8154.
- [44] V.W. Lui, N.D. Peyser, P.K. Ng, J. Hritz, Y. Zeng, Y. Lu, H. Li, L. Wang, B.R. Gilbert, I.J. General, I. Bahar, Z. Ju, Z. Wang, K.P. Pendleton, X. Xiao, Y. Du, J.K. Vries, P.S. Hammerman, L.A. Garraway, G.B. Mills, D.E. Johnson and J.R. Grandis, Frequent mutation of receptor protein tyrosine phosphatases provides a mechanism for STAT3 hyperactivation in head and neck cancer, *Proceedings of the National Academy of Sciences of the United States of America* **111** (2014), 1114-1119.
- [45] D. Harada, N. Takigawa and K. Kiura, The Role of STAT3 in Non-Small Cell Lung Cancer, *Cancers* **6** (2014), 708-722.
- [46] B.D. Looyenga, D. Hutchings, I. Cherni, C. Kingsley, G.J. Weiss and J.P. Mackeigan, STAT3 is activated by JAK2 independent of key oncogenic driver mutations in non-small cell lung carcinoma, *PloS one* **7** (2012), e30820.

- [47] P. Dutta, N. Sabri, J. Li and W.X. Li, Role of STAT3 in lung cancer, *Jak-Stat* **3** (2015), e999503.
- [48] K.C. Arbour, E.J. Jordan, H.R. Kim, J. Dienstag, H. Yu, F. Sanchez-Vega, P. Lito, M.F. Berger, D.B. Solit, M.D. Hellmann, M.G. Kris, C.M. Rudin, A. Ni, M.E. Arcila, M. Ladanyi and G.J. Riely, Effects of co-occurring genomic alterations on outcomes in patients with KRAS-mutant non-small cell lung cancer, *Clinical cancer research : an official journal of the American Association for Cancer Research* (2017).
- [49] E.B. Krall, B. Wang, D.M. Munoz, N. Ilic, S. Raghavan, M.J. Niederst, K. Yu, D.A. Ruddy, A.J. Aguirre, J.W. Kim, A.J. Redig, J.F. Gainor, J.A. Williams, J.M. Asara, J.G. Doench, P.A. Janne, A.T. Shaw, R.E. McDonald Iii, J.A. Engelman, F. Stegmeier, M.R. Schlabach and W.C. Hahn, KEAP1 loss modulates sensitivity to kinase targeted therapy in lung cancer, *eLife* **6** (2017), 10.7554/eLife.18970.
- [50] R. Romero, V.I. Sayin, S.M. Davidson, M.R. Bauer, S.X. Singh, S.E. LeBoeuf, T.R. Karakousi, D.C. Ellis, A. Bhutkar, F.J. Sanchez-Rivera, L. Subbaraj, B. Martinez, R.T. Bronson, J.R. Prigge, E.E. Schmidt, C.J. Thomas, C. Goparaju, A. Davies, I. Dolgalev, A. Heguy, V. Allaj, J.T. Poirier, A.L. Moreira, C.M. Rudin, H.I. Pass, M.G. Vander Heiden, T. Jacks and T. Papagiannakopoulos, Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis, *Nature medicine* **23** (2017), 1362-1368.
- [51] J. Zhang, Q. Dai, D. Park and X. Deng, Targeting DNA Replication Stress for Cancer Therapy, *Genes* **7** (2016), 10.3390/genes7080051.
- [52] P.M. Reaper, M.R. Griffiths, J.M. Long, J.D. Charrier, S. McCormick, P.A. Charlton, J.M. Golec and J.R. Pollard, Selective killing of ATM- or p53-deficient cancer cells through inhibition of ATR, *Nature chemical biology* **7** (2011), 428-430.
- [53] H.C. Reinhardt, A.S. Aslanian, J.A. Lees and M.B. Yaffe, p53-deficient cells rely on ATM- and ATR-mediated checkpoint signaling through the p38MAPK/MK2 pathway for survival after DNA damage, *Cancer cell* **11** (2007), 175-189.
- [54] A. Al-Hendy, A. Laknaur, M.P. Diamond, N. Ismail, T.G. Boyer and S.K. Halder, Silencing Med12 Gene Reduces Proliferation of Human Leiomyoma Cells Mediated via Wnt/beta-Catenin Signaling Pathway, *Endocrinology* **158** (2017), 592-603.
- [55] C. Lu and C.D. Allis, SWI/SNF complex in cancer, *Nature genetics* **49** (2017), 178-179.
- [56] L.N. Micel, J.J. Tentler, P.G. Smith and G.S. Eckhardt, Role of ubiquitin ligases and the proteasome in oncogenesis: novel targets for anticancer therapies, *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31** (2013), 1231-1238.
- [57] S.B. Lee, J.J. Kim, H.J. Nam, B. Gao, P. Yin, B. Qin, S.Y. Yi, H. Ham, D. Evans, S.H. Kim, J. Zhang, M. Deng, T. Liu, H. Zhang, D.D. Billadeau, L. Wang, E. Giaime, J. Shen, Y.P. Pang, J. Jen,

J.M. van Deursen and Z. Lou, Parkin Regulates Mitosis and Genomic Stability through Cdc20/Cdh1, *Molecular cell* **60** (2015), 21-34.

[58] S. Lee, J. She, B. Deng, J. Kim, M. de Andrade, J. Na, Z. Sun, J.A. Wampfler, J.M. Cunningham, Y. Wu, A.H. Limper, M.C. Aubry, C. Wendt, P. Biterman, P. Yang and Z. Lou, Multiple-level validation identifies PARK2 in the development of lung cancer and chronic obstructive pulmonary disease, *Oncotarget* **7** (2016), 44211-44223.

[59] D. Dornan, I. Wertz, H. Shimizu, D. Arnott, G.D. Frantz, P. Dowd, K. O'Rourke, H. Koeppen and V.M. Dixit, The ubiquitin ligase COP1 is a critical negative regulator of p53, *Nature* **429** (2004), 86-92.

[60] K.S. Nair, R. Naidoo and R. Chetty, Expression of cell adhesion molecules in oesophageal carcinoma and its prognostic value, *Journal of clinical pathology* **58** (2005), 343-351.

[61] S. Kase, K. Sugio, K. Yamazaki, T. Okamoto, T. Yano and K. Sugimachi, Expression of E-cadherin and beta-catenin in human non-small cell lung cancer and the clinical significance, *Clinical cancer research : an official journal of the American Association for Cancer Research* **6** (2000), 4789-4796.

[62] M. Fanjul-Fernandez, V. Quesada, R. Cabanillas, J. Cadinanos, T. Fontanil, A. Obaya, A.J. Ramsay, J.L. Llorente, A. Astudillo, S. Cal and C. Lopez-Otin, Cell-cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas, *Nature communications* **4** (2013), 2531.

[63] J.P. Koivunen, J. Kim, J. Lee, A.M. Rogers, J.O. Park, X. Zhao, K. Naoki, I. Okamoto, K. Nakagawa, B.Y. Yeap, M. Meyerson, K.K. Wong, W.G. Richards, D.J. Sugarbaker, B.E. Johnson and P.A. Janne, Mutations in the LKB1 tumour suppressor are frequently detected in tumours from Caucasian but not Asian lung cancer patients, *British journal of cancer* **99** (2008), 245-252.

[64] H. Ji, M.R. Ramsey, D.N. Hayes, C. Fan, K. McNamara, P. Kozlowski, C. Torrice, M.C. Wu, T. Shimamura, S.A. Perera, M.C. Liang, D. Cai, G.N. Naumov, L. Bao, C.M. Contreras, D. Li, L. Chen, J. Krishnamurthy, J. Koivunen, L.R. Chirieac, R.F. Padera, R.T. Bronson, N.I. Lindeman, D.C. Christiani, X. Lin, G.I. Shapiro, P.A. Janne, B.E. Johnson, M. Meyerson, D.J. Kwiatkowski, D.H. Castrillon, N. Bardeesy, N.E. Sharpless and K.K. Wong, LKB1 modulates lung cancer differentiation and metastasis, *Nature* **448** (2007), 807-810.