Identification of candidate genes for lung cancer somatic mutation test kits

Yong Chen#, Jian-Xin Shi#, Xu-Feng Pan, Jian Feng and Heng Zhao
Thoracic Department, Shanghai Chest Hospital, Shanghai, China.

Abstract

Over the past three decades, mortality from lung cancer has sharply and continuously increased in China, ascending to the first cause of death among all types of cancer. The ability to identify the actual sequence of gene mutations may help doctors determine which mutations lead to precancerous lesions and which produce invasive carcinomas, especially using next-generation sequencing (NGS) technology. In this study, we analyzed the latest lung cancer data in the COSMIC database, in order to find genomic “hotspots” that are frequently mutated in human lung cancer genomes. The results revealed that the most frequently mutated lung cancer genes are EGFR, KRAS and TP53. In recent years, EGFR and KRAS lung cancer test kits have been utilized for detecting lung cancer patients, but they presented many disadvantages, as they proved to be of low sensitivity, labor-intensive and time-consuming. In this study, we constructed a more complete catalogue of lung cancer mutation events including 145 mutated genes. With the genes of this list it may be feasible to develop a NGS kit for lung cancer mutation detection.

Keywords: Lung cancer, Next-generation sequencing, Somatic mutation kit, COSMIC.

Received: January 15, 2013; Accepted: June 4, 2013.

Introduction

Lung cancer is the most common cancer in terms of incidence and mortality throughout the world, accounting for 13% of all cases and for 18% of deaths in 2008 (Jemal et al., 2011). In China, lung cancer rates are increasing because smoking prevalence continues to either rise or show signs of stability (Youlden et al., 2008; Jemal et al., 2010). Lung cancer is most often diagnosed at late stages, when it has already presented local invasion and distal metastases (Perez-Morales et al., 2011). Therefore, the identification of early molecular events inherent to lung tumorigenesis is an urgent need, so as to provide a basis for intervention in carcinogenesis.

All cancers arise as a result of the acquisition of a series of fixed DNA sequence abnormalities, mutations, many of which ultimately confer a growth advantage to the cells in which they have occurred. Several mutated genes related to tumor growth, invasion or metastasis have been identified in lung cancer, and new agents that inhibit the activities of these genes have been developed, aiming to improve the outcome of lung cancer treatment (Dy and Adjei, 2002). Among these genes, EGFR (epidermal growth factor receptor) is frequently overexpressed in non-small-cell lung cancer (NSCLC) (Rosell et al., 2009). EGFR tyrosine kinase inhibitors (e.g. Gefitinib and Erlotinib) have been tested in trials for treating NSCLC (Fukuoka et al., 2003; Kris et al., 2003; Giacccone et al., 2004; Spigel et al., 2011; Liu et al., 2012). Furthermore, KRAS and TP53 gene mutations have been found in up to 30% of lung cancer cases and have been considered as predictive factors of poor prognosis (Huncharek et al., 1999; Pao et al., 2005; Mogi and Kuwano, 2011). These frequently mutated genes can be used to design kits for early detection of carcinogenesis. For example, a kit from Life Technologies Corporation (Ion AmpliSeqTM) was designed to detect 739 COSMIC mutations in 604 loci from 46 oncogenes and tumor suppressor genes, with emphasis on the deep coverage of genes KRAS, BRAF and EGFR for the detection of somatic mutations in archived cancer samples.

In this study, we analyzed the latest data on lung cancer, aiming to identify frequently mutating genomic “hotspot” regions in human lung cancer genes. The results are significant and promising, once the ability to identify the actual sequence of mutations may help determining which mutations lead to precancerous lesions and which produce invasive carcinomas. Thus, our study may contribute to improve lung cancer diagnosis and design better prognosis kits.

Materials and Methods

Database of somatic mutations in cancer

The COSMIC (Catalogue of Somatic Mutations in Cancer) database (Forbes et al., 2011) was designed to store and display somatic mutation information and related
details and contains information on human cancers. The current release (v64) describes over 913,166 coding mutations of 24,394 genes from almost 847,698 tumor samples. To construct a complete dataset of cancer mutation information, we had to start by finding a complete catalogue of gene mutations in lung cancer patients. Therefore, we downloaded somatic mutation data from the COSMIC database. All genes selected for the COSMIC database came from studies in the literature and are somatically mutated in human cancer (Bamford et al., 2004). Based on this authority resource, we constructed a complete dataset of cancer mutation information for the analysis described in the following.

**Lung cancer mutation extraction**

As our aim was to collect data on lung cancer, we searched for mutation information in the web-software BioMart Central Portal. BioMart offers a one-stop shop solution to access a wide array of biological databases, such as the major biomolecular sequence, pathway and annotation databases such as Ensembl, Uniprot, Reactome, HGNC, Wormbase and PRIDE (Haider et al., 2009). We used the Cancer BioMart web-interface, with the following criteria: 1. Primary site = “lung”; 2. Mutation ID is not empty. The first criterion ensures that the mutation occurs in lung tissues, and the second criterion helps excluding the samples without mutation in a specific gene. Thereby we obtained the list of mutations in lung cancer.

**Mutation frequency calculation**

In order to identify the most important mutated genes in lung cancer, we calculated the mutation frequency for each mutated gene. In this calculation, we considered the same sample used in different experiments as a different sample. For example, if a gene AKT1 mutation was found in two different experiments, gene AKT1 was assigned a mutation frequency of 2, even if both experiments were performed with samples from the same tissue of the same patient. Sometimes, frequencies are presented as percentages. In this study, however, we did not divide the frequency of 2 by the whole sample, because we focused only on how common the mutation is and how many of these mutations were identified. For example, if the mutation percentage was 100%, but the number of samples with the mutation was only 3, this gene was not accepted in our diagnostic kit.

**Protein-Protein Interaction (PPI) network**

The number of mutation events in the list of lung cancer mutations is very high, but some of these mutations are not found in lung cancer only. So, in order to find the key genes of this list, we analyzed the relationship between those genes. We started with the intent of using KEGG for digging into these relations. However, KEGG shows the very putative gene in a specific biological pathway, and there are many genes which cannot be located in the accurate site in some pathways. For the past few years, PPI databases have become a major tool for digging into biological relations. The great protein-protein interaction source offers a possible way of guessing their function through the interacted protein. If an interacted gene has a lung-regulated mechanism, the anchor gene will always show a similar function. Then, if all genes inputted to PPI have similar functions, there will be a regulation network among them.

As there are so many public PPI databases and each database has its own features, we combined the following databases, introduced by a former paper (Mathivanan et al., 2006): HPRD, IntAct, MIPS, BIND, DIP, MINT, PDZBase and Reactome. Genes of the mutation list were mapped to these PPI databases and a PPI network was constructed. Thereafter, we found that some genes were isolated from the main network and could exclude them from our list of candidate genes for lung cancer. With this combined database, we were able to narrow down our lung cancer candidate gene list as much as possible.

**Results**

**The most complete catalogue of lung cancer mutation data**

Using the methods described above, we obtained a complete list of lung cancer mutations (data not shown) comprising a total of 21,135 mutation events. To our best knowledge, this is the most complete and detailed catalogue of mutation events associated with lung cancer. Almost all the 21,135 listed events are somatic mutations, with only two exceptions: mutation c.1334_1335ins17 in gene FLCN is a confirmed germline mutation, and mutation

![Mutation type in Lung cancer genes](image)
c.1579_1580GG > CT in gene SF3B1 is a nonspecified type of mutation. To obtain a profile of the mutation type distribution in lung cancer, we calculated the statistical frequency of each mutation type, presented in Figure 1, showing that there are many mutation subtypes, such as missense, nonsense, deletions and insertions. Among them, the missense mutations accounted for the largest proportion (61%).

Calculation of mutation frequency in lung cancer

The gene mutation list contains 21,135 mutation events related to 20,906 unique samples. In order to screen the most important mutated genes, we calculated the mutation frequency of each gene in the list. Figure 2 illustrates the top 23 genes found in lung cancer, clearly showing that the most frequently mutated genes in lung cancer are EGFR, KRAS and TP53, with a mutation frequency of

![Figure 2](image)

**Figure 2** - Top 23 most frequently mutated genes in lung cancer. The X-axis represents the top 23 genes and labels are the gene symbols. The Y-axis reflects the total amount of mutation events occurred in these genes.

![Figure 3](image)

**Figure 3** - Types of mutations in the top 23 most frequently mutated genes in lung cancer. The X-axis represents the top 23 genes and labels are the gene symbols. The Y-axis reflects the number of mutation types occurred in the corresponding genes.
10957, 3106 and 2034, respectively. Next, the mutation events in each gene were sorted (Figure 3), this showing that the mutation type of each gene varies dramatically, even in the top 23 mutated genes. As shown in Figure 3, gene TP53 was the one with the largest number of mutation types, amounting to more than ten times the number of mutation types of KRAS, although the mutation frequency of KRAS was higher than that of TP53.

**Figure 4** - Protein-protein interaction (PPI) network for somatic mutations in lung cancer genes (genes which have no relationship with this main network are not shown).
Construction of the PPI network

By mapping the mutated genes into PPI databases, we constructed a PPI network, shown in Figure 4. For a deep data-mining of this network, we calculated the interaction weight (numbers of neighbors) of each core node and visualized the relationships of weight and mutation event for each gene (Figure 5). Analyzing Table 1 and Figure 5, it becomes evident that genes with high mutation frequencies also had higher interaction weights. For example, the top 3 mutated genes EGFR, KRAS and TP53 also had higher interaction weights: 32, 37 and 41, respectively. On the other hand, we noticed that some genes with relative lower mutation frequencies were the core nodes in the PPI network. For example, AKT1 has a high PPI weight (41) but a low mutation frequency (6).

Candidate genes for sequencing kits

After mining the COSMIC database and analyzing the lung cancer PPI network, we screened the most important mutated genes in lung cancer based on one of the following criteria: PPI weight > 7 and mutation frequency > 5. After selection, 145 genes meeting the cutoff criteria were screened out (Table 2). We consider that these mutated genes could be used to design sequencing kits for diagnostic purposes.

Discussion

Many researchers have attempted to find a complete mutation profile of each cancer. In this study, we obtained a list of lung cancer mutations totaling 21,135 mutation events. We believe that to this date this list is the most complete and detailed catalogue of lung cancer mutation events available. Mutations from Stage I to Stage II, from cell line to biopsy, from small cell carcinoma to NSCLC, were almost all included in this list.

As expected, by calculating the mutation frequency for each gene in this list, EGFR, KRAS and TP53 were found to be the top 3 most frequently mutated genes in lung cancer. In addition, these three genes were the hub nodes in the PPI network. EGFR and KRAS have been proved to be lung cancer oncogenes for years. An investigation done in 2004 on the gefitinib therapy effect found somatic mutations of EGFR in 15 of 58 unselected tumors from Japan and in one out of 61 from the United States (Paez et al., 2004). EGFR has since been accepted as a target for lung cancer therapy, and EGFR mutations may predict sensitiv-

---

**Figure 5** - PPI core genes showing number of neighbor genes vs. somatic mutation frequency. Each dot represents a lung cancer gene; genes with more than 10 neighbors or with more than 10 COSMIC somatic mutation events are shown.
ity to gefitinib. In recent years, developing EGFR mutations into a diagnostic target has been a research hotspot. In 2008, Maheswaran et al. (2008) used molecular characterization of circulating tumor cells as a strategy for noninvasive serial monitoring of tumor genotypes during treatment. It is known that most lung adenocarcinoma-associated EGFR mutations confer sensitivity to specific EGFR tyrosine kinase inhibitors. Politi and Lynch (2012) found that EGFR exon 19 insertion mutations are also sensitive to this class of drugs. All these findings suggest that lung cancer patients should be tested for EGFR mutations.

After EGFR, the second most important gene in the development of lung cancer is KRAS. As early as in 2001, Johnson et al. (2001) found that mice carrying KRAS mutations were highly predisposed to a range of tumor types, predominantly early-onset lung cancer. Furthermore, mutations of KRAS and EGFR can be combined to predict prognosis. For example, Massarelli et al. (2007) found that patients with both EGFR mutation and increased EGFR copy number had a > 99.7% chance of objective response to EGFR-TKI therapy, whereas patients with KRAS mutation with or without increased EGFR copy number had a > 96.5% chance of disease progression. They concluded that the KRAS mutation should be included as an indicator of resistance in the panel of markers used to predict response to EGFR-TKI lung cancer therapy.

### Table 1 - Protein-protein interaction (PPI) network core genes, listed by gene symbol, weight and mutation frequency.

| Gene    | Weight | Mutation frequency | Gene    | Weight | Mutation frequency |
|---------|--------|--------------------|---------|--------|--------------------|
| PIK3CA  | 48     | 93                 | TYK2    | 7      | 5                  |
| AKT1    | 41     | 6                  | EPHB1   | 6      | 10                 |
| TP53    | 41     | 2034               | ITPR1   | 6      | 7                  |
| HRAS    | 39     | 9                  | MLL     | 6      | 7                  |
| CTNNB1  | 34     | 3048              | APLNR   | 5      | 5                  |
| KRAS    | 37     | 3016              | FGFR4   | 5      | 9                  |
| PTEN    | 36     | 62                 | IRAK2   | 5      | 5                  |
| NRAS    | 34     | 33                 | KIAA1804| 5      | 5                  |
| EGFR    | 32     | 10957             | PIK3C3  | 5      | 8                  |
| PTPN11  | 32     | 6                  | PTCH2   | 5      | 6                  |
| MTO1    | 31     | 6                  | ROS1    | 5      | 8                  |
| CBL     | 30     | 10                 | TLR4    | 5      | 10                 |
| PIK3CG  | 29     | 16                 | ABL2    | 4      | 7                  |
| ABL1    | 27     | 5                  | AKAP9   | 4      | 5                  |
| CDKN2A  | 26     | 310                | CDH11   | 4      | 10                 |
| ERBB2   | 26     | 83                 | EPHA5   | 4      | 19                 |
| RB1     | 26     | 90                 | EPHA7   | 4      | 8                  |
| BRAF    | 25     | 130                | EPHB6   | 4      | 8                  |
| NOTCH1  | 24     | 17                 | FBXW7   | 4      | 14                 |
| IRS1    | 22     | 5                  | GRM1    | 4      | 8                  |
| NF1     | 21     | 28                 | ITPR3   | 4      | 5                  |
| SMAD4   | 21     | 11                 | NRG3    | 4      | 5                  |
| CREBBP  | 19     | 6                  | NTRK2   | 4      | 11                 |
| PDGFRα  | 19     | 16                 | PAK3    | 4      | 6                  |
| BRCA2   | 18     | 7                  | ADAMTS13| 3      | 5                  |
| JAK2    | 18     | 5                  | EPHB3   | 3      | 5                  |
| KIT     | 18     | 11                 | ITK     | 3      | 12                 |
| STK11   | 18     | 172                | KEAP1   | 3      | 12                 |
| LPHN3   | 17     | 15                 | NCOA2   | 3      | 6                  |
| FGFR2   | 15     | 11                 | NFE2L2  | 3      | 44                 |

| Gene    | Weight | Mutation frequency | Gene    | Weight | Mutation frequency |
|---------|--------|--------------------|---------|--------|--------------------|
| LRP1B   | 15     | 17                 | NLRP3   | 3      | 8                  |
| MAP2K4  | 15     | 7                  | NOTCH3  | 3      | 8                  |
| JAK3    | 14     | 5                  | P2RY8   | 3      | 5                  |
| KDR     | 14     | 16                 | RUNX1T1 | 3      | 9                  |
| MET     | 14     | 44                 | TNI1    | 3      | 7                  |
| RET     | 14     | 9                  | CDC42BPA| 2      | 6                  |
| BAI3    | 13     | 27                 | DNER    | 2      | 5                  |
| EPHA3   | 13     | 17                 | HECW1   | 2      | 6                  |
| FLT3    | 13     | 7                  | HERC1   | 2      | 5                  |
| NTRK1   | 13     | 9                  | INHBA   | 2      | 5                  |
| PCH1    | 13     | 6                  | INSRR   | 2      | 11                 |
| TERT    | 13     | 7                  | LTK     | 2      | 6                  |
| ATM     | 12     | 22                 | NTRK3   | 2      | 20                 |
| FLT1    | 12     | 10                 | PAK6    | 2      | 5                  |
| GRM8    | 12     | 15                 | PIK3C2G | 2      | 8                  |
| FLT4    | 11     | 7                  | RBL1    | 2      | 5                  |
| ALK     | 10     | 16                 | CCKBR   | 1      | 5                  |
| ERBB4   | 10     | 20                 | CSMD3   | 1      | 11                 |
| MEN1    | 10     | 9                  | HEPH    | 1      | 5                  |
| MSH6    | 10     | 7                  | KIAA1618| 1      | 2                  |
| APC     | 9      | 19                 | LAT51   | 1      | 5                  |
| EVI1    | 9      | 1                  | LMTK2   | 1      | 8                  |
| HYRC    | 9      | 1                  | NPY5R   | 1      | 6                  |
| SMARCA4 | 9      | 26                 | PAK7    | 1      | 11                 |
| NOTCH2  | 8      | 10                 | PKHD1   | 1      | 15                 |
| NOTCH4  | 8      | 8                  | PPP1R3A | 1      | 9                  |
| PTPRD   | 8      | 14                 | ROR2    | 1      | 8                  |
| EPHA4   | 7      | 5                  | TTX22   | 1      | 7                  |
| GLI3    | 7      | 11                 | TTN     | 1      | 15                 |

Chen et al.
lated to lung cancer, we believe that this PPI network contains the most important genes related to lung cancer.

Many companies detect lung cancer by only four somatic gene mutations (EGFR, KRAS, BRAF and PI3K). As expected, these genes are all included in our list (Table 2; mutation frequency of BRAF = 130, weight = 25; mutation frequency of PI3K3A = 93, weight = 48). BRAF encodes a RAS-regulated kinase that mediates cell growth and the activation of the malignant transformation kinase pathway (Sithanandam et al., 1990). Brose et al. (2002) found that BRAF mutations in human lung cancers may identify a subset of tumors sensitive to targeted therapy. Furthermore, an in vivo study with the inhibitor of the last of the four genes, PI3K, aimed at testing its activity in lung cancer treatment (Engelman et al., 2008), this leading to the conclusion that inhibitors of the PI3K-mTOR pathway may be activated in cancers with PI3KCA mutations and, when combined with MEK inhibitors, may effectively treat KRAS mutated lung cancers.

As EGFR and KRAS kits are widely used, we listed our EGFR and KRAS mutation events in Tables 3 and 4. In these tables, we sorted the mutations in EGFR and KRAS by frequency, with “Y” meaning the typical mutation used in the detection kits supplied by many companies; and “-” meaning that the mutation has a location in the genome similar to some of the other detected mutations. But, first of all, “-” is an alert to the fact that there are many different kinds of mutation in the same region. Traditional methods such as PCR are unable to detect such complicated mutations. This is the first advantage that a Next-Generation Se-

| Gene symbol | Gene symbol | Gene symbol | Gene symbol | Gene symbol |
|-------------|-------------|-------------|-------------|-------------|
| EVI1        | EPHA3       | CBL         | MAP2K4      | PIAIA       |
| HYRC        | UBR5        | RUNX1T1     | ITPR1       | PAK6        |
| EGFR        | PIK3CG      | RUNX1T1     | FLT4        | P2RY8       |
| KRAS        | PDGFR1      | RET         | FLT3        | NRG3        |
| TP53        | KDR         | PPP1R3A     | BRCA2       | MECOM       |
| CDKN2A      | ALK         | NTRK1       | ABL2        | LMTK3       |
| STK11       | TTN         | MEN1        | USP29       | LATS1       |
| BRAF        | PKHD1       | KSR2        | RNF213      | KIAA1804    |
| PIK3CA      | LPHN3       | HRAS        | PTPN11      | JAK3        |
| RB1         | GRM8        | FGFR4       | PTCH2       | JAK2        |
| ERBB2       | PTPRD       | ROS1        | PTCH1       | ITPR3       |
| PTEN        | FBXW7       | ROR2        | PAK3        | IRS1        |
| NFE2L2      | KEAP1       | PIK3C3      | NPY5R       | IRAK2       |
| MET         | ITK         | PIK3C2G     | NCOA2       | INHBA       |
| CTNNB1      | SMAD4       | NOTCH4      | MTOR        | HERC1       |
| CDKN2a(p14) | PK7         | NOTCH3      | MKRN3       | HEPH        |
| NRAS        | NTRK2       | NLRP3       | MERTK       | FBXO10      |
| MUC16       | KIT         | MYO3B       | LTK         | ERCC6       |
| NF1         | INSRR       | LMTK2       | HECW1       | EPHB3       |
| BAI3        | GLI3        | GRM1        | FLT4_ENST00000261937 | EPHA4 |
| SMARCA4     | FGFR2       | EPHB6       | DOCK3_ENST0000266037 | DNER |
| MUC16_ENST00000331986 | CSMD3 | EPHA7       | CREBBP      | DGKB        |
| ATM         | TLR4        | ENSG00000121031 | CDC42BPA | CCKBR       |
| NTRK3       | PRKDC       | TLN1        | AKT1        | BAI2        |
| ERBB4       | NOTCH2      | TERT        | ZMYM2_ENST00000456228 | APLNR |
| EPHA5       | FLT1        | TBX22       | VEGFC       | ANKK1       |
| APC         | FBXW7_NM_018315_2 | TAF1L | TYK2 | AKAP9_ENST00000356239 |
| NOTCH1      | EPHB1       | MSH6        | ROBO2       | AKAP9       |
| LRPIB       | CDH11       | MLL         | RBL1        | ADAMTSL3    |
sequencing (NGS) technology can offer. Secondly, it is obvious that the frequent mutations represent a high percentage in the three kits and, on the other hand, many mutations with frequencies below 30 are not listed in the three kits. The cost of detecting more than 100 mutations at the same time by PCR is very high, conferring NGS another advantage over PCR detection.

It is really urgent to develop a NGS kit for detecting lung cancer mutations. Our genes for the sequencing kit can be designed for somatic mutation detection. The 145 gene set comprises all of the somatic mutation detecting purpose genes - *EGFR, KRAS, BRAF* and *PIK3CA* (Saal et al., 2005) - and may provide a feasible choice for a NGS kit. With the progresses in sequencing technology, mutations in lung cancer patients can be detected in one day or even less time. This technology applied to cancer genome sequencing can speed up cancer research, and the kit for diagnostic or recurrence evaluation should be introduced in clinical care as soon as possible, in order to offer patients a better chance of less suffering and a higher survival perspective.

Table 3 - Mutation events for *EGFR* mutation kits.

| Mutation in *EGFR* | Mutation frequency |
|--------------------|--------------------|
| L858R              | 2688               |
| c.2573T > G:L858R  | 1378               |
| E746_A750del       | 955                |
| c.2235_2249del15:E746_A750del | 609         |
| T790M              | 346                |
| c.2236_2250del15:E746_A750del | 326     |
| c.2240_2257del18:L747_P753 > S | 113  |
| c.2369C > T:T790M  | 102                |
| L747_P753 > S     | 81                 |
| c.2239_2248TTAAGAGAAG > C:L747_A750 > P | 72  |
| L861Q             | 70                 |
| L747_T751del      | 58                 |
| c.2237_2255 > T:E746_S752 > V | 47  |
| L747_A750 > P     | 47                 |
| G719S             | 45                 |
| c.2240_2254del15:L747_T751del | 43  |
| c.2582T > A:L861Q | 39                 |
| G719A             | 36                 |
| L747_S752del      | 36                 |
| S768I             | 34                 |
| c.2239_2256del18:L747_S752del | 31  |
| c.2155G > A:G719S | 24                 |
| c.2156G > C:G719A | 24                 |
| L747_P753del      | 24                 |
| E746_S752 > V     | 23                 |
| c.2155G > T:G719C | 20                 |
| c.2239_2253del15:L747_T751del | 20  |
| E746_T751del      | 20                 |
| c.2239_2251 > C:L747_T751 > P | 19  |
| c.2303G > T:S768I | 19                 |
| c.2237_2251del15:E746_T751 > A | 18  |
| G719T             | 16                 |
| c.2239_2247del9:L747_E749del | 15  |

| Mutation in *EGFR* | Mutation frequency |
|--------------------|--------------------|
| c.2497T > G:L833V  | 15                 |
| G719C             | 14                 |
| L747_A750del      | 13                 |
| L747_T751 > P     | 13                 |
| c.2125G > A:E709K  | 12                 |
| D770_N771insSVD   | 12                 |
| L747_P753 > Q     | 12                 |
| E746_T751 > A     | 11                 |
| P772_H773insPR    | 11                 |
| c.2126A > C:E709A  | 10                 |
| H773R             | 7                  |
| R776H             | 7                  |
| c.2233_2247del15:K745_E749del | 6  |
| c.2238_2252del15:L747_T751del | 6  |
| c.2254_2277del24:S752_I759del | 6  |
| c.2582T > G:L861R  | 6                  |
| H773_V774insNPH   | 6                  |
| L833V             | 6                  |
| c.2126A > G:E709G  | 5                  |
| c.2237_2253 > TTGCT:E746_T751 > VA | 5  |
| c.2237_2256 > TT:E746_S752 > V | 5  |
| c.2253_2276del24:S752_I759del | 5  |
| c.2311_2312ins9:D770_N771insSVD | 5  |
| c.2319_2320ins9:H773_V774insNPH | 5  |
| c.2504A > T:H835L  | 5                  |
| c.2543C > T:P848L  | 5                  |
| c.89_889del801:V30_R297 > G | 5  |
| K745_A750del      | 5                  |
| L747_T751 > S     | 5                  |
| L747P             | 5                  |
| L861R             | 5                  |
| V843I             | 5                  |
**Table 4** - Mutation events for KRAS mutation kits.

| Mutation in KRAS | Mutation frequency |
|------------------|--------------------|
| c.34G > T:G12C  | 1273               |
| c.35G > T:G12V  | 625                |
| c.35G > A:G12D  | 524                |
| c.35G > C:G12A  | 202                |
| c.34G > A:G12S  | 139                |
| c.37G > T:G13C  | 99                 |
| c.38G > A:G13D  | 76                 |
| c.34G > C:G12R  | 69                 |
| c.34_35GG > TT:G12F | 22             |
| c.183A > T:Q61H | 13                 |
| c.182A > T:Q61L | 12                 |
| c.183A > C:Q61H | 10                 |
| Q61H             | 10                 |
| c.182A > G:Q61R | 9                  |
| G12D             | 9                  |
| c.181C > A:Q61K | 6                  |
| c.37G > A:G13S  | 6                  |
| c.181C > G:Q61E | 4                  |
| c.37G > C:G13R  | 3                  |
| G12F             | 3                  |
| c.182A > C:Q61P | 2                  |
| c.198_199ins15:A66_M67insEEYSA | 2        |
| c.34_35GG > CT:G12L | 2             |
| c.35_36GT > AA:G12E | 2             |
| c.38_39GC > TT:G13V | 2             |
| G12A             | 2                  |
| G12C             | 2                  |
| G12V             | 2                  |
| L19F             | 2                  |
| c.15AT:K5N      | 1                  |
| c.180_181TC > CA:Q61K | 1            |
| c.205delG:D69fs*4 | 1              |
| c.27T > C:V9V   | 1                  |
| c.31G > C:A11P  | 1                  |
| c.32C > T:A11V  | 1                  |
| c.34_36GGT > TGC:G12C | 1          |
| c.35_36GT > TC:G12V | 1              |
| c.35delG:G12fs*3 | 1                |
| c.36T > C:G12G  | 1                  |
| c.38G > C:G13A  | 1                  |
| c.38G > T:G13V  | 1                  |
| c.52G > A:A18T  | 1                  |
| c.53C > A:A18D  | 1                  |
| G12_G13insG     | 1                  |
| G12S             | 1                  |
| G13D             | 1                  |

**References**

Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A, Teague J, Futreal P and Stratton M (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 91:355-358.

Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J and Nicholson A (2002) BRAF and RAS mutations in human lung cancer and melanoma. Cancer Res 62:6997.

Dy GK and Adjei AA (2002) Novel targets for lung cancer therapy: Part II. J Clin Oncol 20:3016-3028.

Engelman JA, Chen L, Tan X, Crosby K, Guimaraes AR, Upadhyay R, Maira M, McNamara K, Perera SA and Song Y (2008) Effective use of PI3K and MEK inhibitors to treat mutant Kras G12D and PIK3CA H1047R murine lung cancers. Nat Med 14:1351-1356.

Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K and Menzies A (2011) COSMIC: Mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 39:D945-D950.

Fukuoka M, Yano S, Giaccone G, Tamura T, Nakagawa K, Douillard JY, Nishiwaki Y, Vansteenkiste J, Kudoh S, Rischin D, et al. (2003) Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. J Clin Oncol 21:2237-2246.

Giaccone G, Herbst RS, Manegold C, Scagliotti G, Rosell R, Miller V, Natale RB, Schiller JH, Von Pawel J, Pluzanska A, et al. (2004). Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: A phase III trial-INTACT 1. J Clin Oncol 22:777-784.

Haider S, Ballester B, Smedley D, Zhang J, Rice P and Kasprzyk A (2009) BioMart Central Portal - Unified access to biological data. Nucleic Acids Res 37:W23-W27.

Huncharek M, Muscat J and Geschwind JF (1999) K-ras oncogene mutation as a prognostic marker in non-small cell lung cancer: A combined analysis of 881 cases. Carcinogenesis 20:1507-1510.

Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D (2011) Global cancer statistics. CA Cancer J Clin 61:69-90.

Jemal A, Center MM, DeSantis C and Ward EM (2010) Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev 19:1893-1907.

Johnson L, Mercer K, Greenbaum D, Bronson RT, Crowley D, Tuveson DA and Jacks T (2001) Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature 410:1111-1116.

Kris MG, Natale RB, Herbst RS, Lynch Jr TJ, Prager D, Belani CP, Schiller JH, Kelly K, Spiridonidis H, Sandler A, et al. (2003) Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: A randomized trial. JAMA 290:2149-2158.

Liu G, Cheng D, Ding K, Le Maitre A, Liu N, Patel D, Chen Z, Seymour L, Shepherd FA and Tsao MS (2012) Pharmacogenetic analysis of BR.21, a placebo-controlled randomized phase III clinical trial of erlotinib in advanced non-small cell lung cancer. J Thorac Oncol 7:316-322.

Maheswaran S, Sequist LV, Nagrath S, Uilks L, Brannigan B, Collura CV, Inserra E, Diederichs S, Iafrate AJ and Bell DW
(2008) Detection of mutations in EGFR in circulating lung-cancer cells. N Engl J Med 359:366-377.

Massarelli E, Varella-Garcia M, Tang X, Xavier AC, Ozburn NC, Liu DD, Bekele BN, Herbst RS and Wistuba II (2007) 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 13:2890-2896.

Mathivanan S, Periaswamy B, Gandhi T, Kandasamy K, Suresh S, Mohmood R, Ramachandra Y and Pandey A (2006) An evaluation of human protein-protein interaction data in the public domain. BMC Bioinformatics 7:S19.

Mogi A and Kuwano H (2011) TP53 mutations in nonsmall cell lung cancer. J Biomed Biotechnol 2011:583929.

Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, et al. (2004) EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. Science 304:1497-1500.

Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, Zakowski MF, Heelan RT, Kris MG and Varmus HE (2005) KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. PLoS Medicine 2:e17.

Perez-Morales R, Mendez-Ramirez I, Castro-Hernandez C, Martinez-Ramirez OC, Gonsebatt ME and Rubio J (2011) Polymorphisms associated with the risk of lung cancer in a healthy Mexican Mestizo population: Application of the additive model for cancer. Genet Mol Biol 34:546-552.

Politi K and Lynch TJ (2012) Two sides of the same coin: EGFR exon 19 deletions and insertions in lung cancer. Clin Cancer Res 18:1490-1492.

Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, Majem M, Lopez-Vivanco G, Isla D, Provencio M, et al. (2009) Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 361:958-967.

Saal LH, Holm K, Maurer M, Memeo L, Sù T, Wang X, Yu JS, Malmström PO, Mansukhani M and Enoksson J (2005) PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. Cancer Res 65:2554.

Sithanandam G, Kolch W, Duh FM and Rapp UR (1990) Complete coding sequence of a human B-raf cDNA and detection of B-raf protein kinase with isozyme specific antibodies. Oncogene 5:1775-1780.

Spigel DR, Burris 3rd HA, Greco FA, Shipley DL, Friedman EK, Waterhouse DM, Whorf RC, Mitchell RB, Daniel DB, Zangmeister J, et al. (2011) Randomized, double-blind, placebo-controlled, phase II trial of sorafenib and erlotinib or erlotinib alone in previously treated advanced non-small-cell lung cancer. J Clin Oncol 29:2582-2589.

Youlden DR, Cramb SM and Baade PD (2008) The International Epidemiology of Lung Cancer: Geographical distribution and secular trends. J Thorac Oncol 3:819-831.

Internet Resources

COSMIC,http://www.sanger.ac.uk/genetics/CGP/cosmic (Dec. 12, 2012).

BioMart Central Portal, http://www.biomaart.org (Dec. 20, 2012).

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.