There are two regulatory single nucleotide polymorphisms (rSNPs) at the beginning of the second intron of the mouse K-ras gene that are strongly associated with lung cancer susceptibility. We performed functional analysis of three SNPs (rs12228277: T>A, rs12226937: G>A, and rs61761074: T>G) located in the same region of human KRAS. We found that rs12228277 and rs61761074 result in differential binding patterns of lung nuclear proteins to oligonucleotide probes corresponding to alternative alleles; in both cases, the transcription factor NF-Y is involved. G>A substitution (rs12226937) had no effect on the binding of lung nuclear proteins. However, all the nucleotide substitutions under study showed functional effects in a luciferase reporter assay. Among them, rs61761074 demonstrated a significant correlation with allele frequency in non-small-cell lung cancer (NSCLC). Taken together, the results of our study suggest that a T>G substitution at nucleotide position 615 in the second intron of the KRAS gene (rs61761074) may represent a promising genetic marker of NSCLC.

1. Introduction

Identification of regulatory single nucleotide polymorphisms (rSNPs) in a genome is one of the most important topics in modern molecular genetics. rSNPs are usually located in non-coding gene segments such as 5′ and 3′ flanks, introns, and distant regulatory regions. They influence the level and/or pattern of gene expression, and as a consequence may be associated with various diseases (Epstein 2009). According to recent genome-wide association study (GWAS) data, the vast majority of SNPs associated with disease traits (~90%) are situated within non-coding regions of the genome (Manolio et al. 2009; Maurano et al. 2012). However, SNPs located outside the coding regions are the least studied. Currently, the main content of databases on pathologically significant human gene mutations is represented by coding SNPs (cSNPs) that change the structure of the gene protein product. In particular, cSNPs account for 86% of the total

Keywords. K-RAS; lung cancer; NF-Y binding; regulatory SNPs (rSNP)
number of mutations compiled in HGMD, the central disease-associated human gene mutation database (Stenson et al. 2009; Cooper et al. 2010). The group of rSNPs, including mutations able to influence transcription initiation, elongation, translational characteristics and stability of mRNA, constitute only 3% of the HGMD dataset (Cooper et al. 2010). Therefore, identification of new rSNPs associated with disease is of great importance, since it can bring new perspectives to our understanding of the molecular basis of different pathologies.

Of special interest among rSNPs are polymorphisms localized to the binding sites of various transcription factors (Ponomarenko et al. 2002a; Bryzgalov et al. 2013). These rSNPs may eliminate an existing binding site, create a binding site for another transcription factor, or increase/decrease the affinity for a given factor, with a dramatic effect on the gene expression pattern. There are numerous examples of such rSNPs associated with various diseases. In particular, the reported GWAS SNP rs6801957: G>A in the SCN10A enhancer disrupts TBX3/TBX5 binding and reduces the tissue-specific activity of the enhancer in the heart (van den Boogaard et al. 2012). In contrast, A>C substitution (rs12553612) in the IFNα8 promoter creates a binding site for OCT1 and enhances the promoter activity in a luciferase reporter assay, and is associated with better overall survival of glioma patients (Kohanbash et al. 2012). Each of the two substitutions 663G>A and 666G>T in intron 2 of the human TDO2 gene, which are associated with a number of psychiatric disorders (Comings et al. 1996), not only leads to destruction of the YY-1 binding site, but simultaneously creates binding sites for unknown transcription factors (Vasiliev et al. 1999). The substitution of −30T>A in the TATA box of the human β-globin gene (HBB) promoter leads to a fourfold decrease in the TBP/TATA affinity (Drachkova et al. 2014), and a decrease in the β-globin mRNA content to 8–13% of the normal amount in β-thalassemia patients (Fei et al. 1988). Conversely, the AFP gene promoter in the case of hereditary persistence of α-fetoprotein carries two substitutions (−119G>A and −55C>A) in its HNF1 binding sites, which increase both the affinity to HNF1 and the level of reporter gene transcription (Alj et al. 2004).

We have previously demonstrated that two base variations at the beginning of the second intron of the mouse K-ras gene (G>C and C>A at nucleotide positions 288 and 296, respectively) disrupt the composite regulatory element consisting of two NF-Y binding sites and a single binding site for GATA-6 (Ponomarenko et al. 2002b; Gorshkova et al. 2006). These variations are strongly associated with susceptibility to lung cancer in mice. The haplotype GC was found exclusively in mice susceptible to spontaneous and chemically-induced lung carcinogenesis (more than 20 mouse strains), whereas the CA haplotype was observed in all four susceptible strains studied (A/J, A/WySnJ, GR and ICR) (Ryan et al. 1987; Chen et al. 1994; Timofeeva et al. 1999; Timofeeva et al. 2002).

KRAS is the main oncogene involved in the pathogenesis of human pancreatic, colorectal and pulmonary adenocarcinomas (Schubbert et al. 2007). The frequency of KRAS mutations located in codon 12 (or less frequently in codons 13 and 61) is approximately 15–40% in lung cancer (Rodenhuis and Slebos 1992; Huncharek et al. 1999). However, little is known to date about regulatory polymorphisms of KRAS associated with this malignancy. The most studied is U>G substitution (rs61764370) in the let-7 complementary site in the KRAS 3′-untranslated region, resulting in KRAS over-expression in vitro. This substitution is significantly associated with an increased risk of non-small cell lung cancer (NSCLC) (Chin et al. 2008; Kundu et al. 2012). Two other examples are A>C substitution (rs11836509) in intron 3, and T insertion (rs34176876) in the 3′-untranslated region of this gene. Both mutations show significantly increased odds ratios (ORs) for lung adenocarcinoma, accompanied by multiple atypical adenomatous hyperplasia (Kohno et al. 2008). Our finding of rSNPs at the beginning of the second intron of the mouse K-ras gene (Ponomarenko et al. 2002b; Gorshkova et al. 2006) encouraged us to investigate common SNPs located in the same region of human KRAS. Therefore, we studied the binding of lung nuclear proteins to oligonucleotides corresponding to different alleles, determined their influence on the luciferase reporter, and evaluated the frequency of the allelic variants under study in lung cancer patients and controls from the Western Siberian region.

2. Materials and methods

2.1 Patients and healthy controls

Patients (n=40) with non-small lung cancer (NSCLC) under the care of the Tomsk Cancer Research Institute were included in this study, among them, 25 patients were diagnosed with squamous cell lung cancer, 14 patients were diagnosed with adenocarcinoma and 1 patient had large cell carcinoma of lung. The control group consisted of 44 healthy volunteers. Ethical Committee of the Tomsk Cancer Research Institute approved the current research.

2.2 Experimental animals and preparation of nuclear extracts

Experiments were carried out on A/He male mice from the colony kept at the Institute of Cytology and Genetics, Novosibirsk, Russia. Nuclear extracts from the lungs were prepared according to the method of (Gorski et al. 1986) with the modifications described by (Shapiro et al. 1988). Protein concentrations were evaluated by the (Lowry et al. 1951).
2.3 Electrophoretic mobility shift assays

Table 1 lists the sequences of oligonucleotides used for EMSA tests. For each allelic version double stranded oligonucleotides V1 and V2 were synthesized with 5'-CAGT tetranucleotide overhangs and SNP position in the center. Introduction of labels (α-32P dATP) in the DNA probe was conducted using fill in of shortened 3' ends of Klenov fragments of DNA polymerase I. The protein nuclear extract was incubated with sheared salmon sperm DNA (1 μg of DNA per 7 μg of total protein) for 10 min at 0°C. After that, 4 μg of extract was added to the probes containing 50 pM of radioactive labeled oligonucleotide and incubated at room temperature for 15 min. 2 ng of GATA-6 or NF-Y antibodies (Santa Cruz Biotechnology) were added into the reaction mixture either during pre-incubation with salmon sperm DNA or immediately after addition of labeled oligonucleotide. Then the mixture was subjected to electrophoresis in 4.5% PAAG in 0.5 × TBE (89 mM Tris-borate, 89 mM H2BO3, 2 mM EDTA at 4°C). The gel was exposed with X-ray film. Two to three independent replicates were performed for each EMSA experiment.

Table 1. List of oligonucleotide probes tested in EMSA

<table>
<thead>
<tr>
<th>SNP identifier in dbSNP NCBI</th>
<th>SNP location in the gene from the beginning of intron 2</th>
<th>Sequences of oligonucleotides 5’→ 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12228277:T&gt;A</td>
<td>190 bp</td>
<td>V1: cagtACTTTCTCTGGGAGTGTCACGGGTCCATGA</td>
</tr>
<tr>
<td>rs12226937:G&gt;A</td>
<td>506 bp</td>
<td>V2: cagtACTTTCTCTGGGAGTGTCACGGGTCCATGA-3’</td>
</tr>
<tr>
<td>rs61761074:T&gt;G</td>
<td>615 bp</td>
<td>V1: cagtTGAGTAAATTTAATTGFAGTTCATGTTTTG</td>
</tr>
<tr>
<td>V2: cagtTGAGTAAATTTAATTGFAGTTCATGTTTTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Two allelic versions of oligonucleotides correspond to each SNP. V1 – major allele. V2 – minor allele. SNP position marked in bold style. Tetranucleotide overhangs are indicated in lowercase letters.

2.4 Transient transfection

A549 (Human lung adenocarcinoma epithelial cell line) cells were cultivated on the basis of the Center of Cellular Technologies, Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences in plastic cultural flasks ‘Greiner’ in a humidified atmosphere 5% CO2 and 95% air at 37°C in medium DMEM/F12 ‘GIBCO’ which contains 10% of fetal bovine serum ‘Thermo Scientific HYCLONE’, penicillin (100 U/mL) and streptomycin (100 mg/mL) produced by ‘Sigma’. Viability of cells was assessed both after defrosting and in the process of cultivation. The media was changed 2 times a week.

Oligonucleotides corresponding to major and minor alleles (table 1) were cloned into the pGL 4.23 vector (Promega, USA), and co-transfected with pRL-TK using Screen FectA (Incella GmbH, Germany). 48 h after transfection, luciferase activity was measured by the dual-luciferase reporter assay kit (Promega, USA). Firefly luciferase activity was normalized to Renilla luciferase activity.

2.5 Western blotting

A549 cells (10⁷) were washed with ice-cold PBS and then pelleted by centrifugation at 300g for 2 min. The cell pellet was resuspended in 1 mL of Buffer 1 [10 mM HEPES, pH 7.9, 10 mM KCl, 1 mM dithiothreitol, 0.5 mM spermidin, 0.15 mM spermin, 0.1 mM EDTA, 0.1 mM EGTA, 0.5 mM PMSF, 16 Halt protease inhibitor cocktail (Thermo Scientific)] and placed on ice to swell for 15 min. After addition of 62 ml of 10% (w/v) Nonidet P-40, the sample was gently vortexed for 10 s and then centrifuged at 400g for 5 min at 4°C. The nuclear pellet was resuspended in 100 mL of Buffer 2, 20 mM HEPES, pH 7.9, 25% (w/v) glycerol, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 1 mM dithiothreitol, 16Halt protease inhibitor cocktail (Thermo Scientific) and incubated for 20 min on ice followed by a centrifugation at 10,000g for 10 min at 4°C. The supernatant containing the nuclear proteins was stored at −70°C. 10 μg of protein was subjected to SDS PAGE and transferred to Immobilon – P membranes (MILLIPORE) in transfer buffer (25 mM Tris, 192 mM glycine, 0.1% and 20% ethanol). Membranes were blocked with 5% nonfat dried milk, 0.1% Tween 20 in PBS (PBS-T). The blotted membranes were incubated with monoclonal antibodies to NF-YA at 2:5000 dilution. The bands were treated with antibodies Goat Anti-Rabbit (Abcam) at 1:10000 dilution and detected with an ECL kit (Amersham) according to the instruction.

2.6 DNA extraction and genotyping

Equal volume of Buffer 3 (0.32 M Sucrose, 10 mM Tris-HCl (pH 7.5), 5 mM MgCl2, 1% v/v Triton X-100) was added to EDTA-anticoagulated peripheral blood samples (4ml) and incubated on ice for 3 to 5 min. Samples were...
then centrifuged at 250 g, 4°C for 10 min and the supernatant was decanted. Half volume of the Buffer 3 was then added to the pelleted white blood cells and mixed by gentle vortexing for 10 s. The tubes were again centrifuged at 250 g, 4°C for 10 min and 0.8 volume of Buffer 4 (25mM EDTA pH 8.0, 75mM NaCl, 1% SDS, 50 μg/mL proteinase K) was added to the pelleted leukocytes. The cells were lysed by gentle vortexing at 37°C until homogeneity was reached. The genomic DNA was purified by standard sequential phenol/chloroform/isomylalcohol extraction and salt/ethanol precipitation. Then DNA was re-suspended in nuclease free water. The concentration of genomic DNA was standardized to 30 ng/mL for genotyping. A fragment, containing rs12228277 (T>A), rs12226937 (G>A), and rs61761074 (T>G) was amplified in GeneAmp PCR System 9700 (Applied Biosystems) in a total reaction volume of 20 μL containing 1 μL of genomic DNA, 2 μL of 10x AS-buffer, 2 μL of each primers (C=3.5pmol/μL), 2 mM of dNTPs, 1 μL(C=1U/μL) of Taq-polymerase (SibEnzyme, Russia) using the following conditions: 95°C for 1 min, 40 cycles of 94°C for 15 s, 62°C for 30 s, 72°C for 1 min, than 72°C for 7 min. The primers used were: the forward 5′-GCCTT-TAGCC-GCCGC-AGAAC-3′ and the reverse 5′-TGCTT-GGGAT-GGAAG-TTCTA-CTCCA-3′. 1195 bp amplicon was purified by electrophoresis in a 0.8% GTG agarose (Ambion) gel. Sanger’s reaction was conducted in the amplificator GeneAmp PCR System 9700 (Applied Biosystems) in 10 μL of reaction mixture containing 1 μg of DNA matrix, 2 pM primer, 2 μL BigDye Terminator Cycle Sequencing Kit v.3.1 (Applied Biosystems), and 2 μL of BigDye buffer (Applied Biosystems) according to manufacturer instructions. Products were purified by Sephadex G100 gel-filtration. Purified samples were dried on the concentrator Eppendorf 5301 and analyzed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) in the Interinstitute DNA Sequencing Center.

2.7 Statistical analysis

Statistical analysis was performed using R-project version 3.1.2 (http://www.r-project.org/). A Student t-test was used to compare relative luciferase activity of reporter constructs containing minor and major alleles. Fisher F-test was used to compare allele frequency for sample of NSCLC patients and sample of controls from the Western Siberian region. The level of significance was set at p<0.05

3. Results

3.1 SNPs in intron 2 of the human KRAS gene and their influence on binding of lung nuclear proteins

According to the dbSNP NCBI database [http://www.ncbi.nlm.nih.gov/snp/ (Human Genome Build 37)], at the beginning of intron 2 of the human KRAS gene

![Figure 1](http://example.com/image1.png)

**Figure 1.** rs 1222877:T>A, and rs61761074:T>G alter binding patterns of lung nuclear proteins. Labelled probes match rs 1222877 (A), rs12226937 (B) and rs61761074 (C) major and minor alleles. The probes were incubated with nuclear extract in the absence or presence of antibodies specific to NF-Y and GATA-6. (A) Antibodies against NF-Y were added after mixing the nuclear extract with a DNA probe; (C) antibodies against NF-Y were added during pre-incubation of the nuclear extract with sheared salmon sperm DNA. Antibodies against GATA-6 were added after mixing the nuclear extract with a DNA probe in both cases. Changes in the binding of allelic variants with the nuclear proteins are indicated by arrows; S – supershift. The sequence of NF-Y binding site (5′ cagt AGACCCTACGGATGTGTTAATCCTT 3′) corresponds to NF-Y gel shift oligonucleotide (Santa Cruz Biotechnology, INC). Added tetranucleotide overhangs are indicated in lowercase letters.

*J. Biosci.* 40(5), December 2015
there are three SNPs with a minor allele frequency (MAF) above 3%: rs12228277 (T>A), rs12226937 (G>A), and rs61761074 (T>G), located at a distance of 190, 506, and 615 bp relative to the translation start, respectively. For each of these SNPs, double-stranded oligonucleotides corresponding to different alleles, V1 and V2, were synthesized (table 1). These were used as DNA probes in electrophoretic mobility shift assay (EMSA) experiments with lung nuclear extract proteins.

According to the data shown in figure 1, the T>A substitution at position 190 weakened the least mobile band (figure 1A, lanes 2 and 5), whereas the T>G substitution at position 615 led to the appearance of a new band in the mobility shift assay (figure 1C, lanes 2 and 5). The G>A substitution at position 506 did not cause any changes in the binding pattern of nuclear proteins (figure 1B).

Our previous research has shown that the lung cancer-related SNPs in intron 2 of the K-ras gene of mice affect the transcription factor binding sites for NF-Y and GATA-6. Both factors are promising candidates, as NF-Y is known to be involved in cell-cycle control (Ly et al. 2013) and GATA-6 to be involved in lung cell differentiation (Warburton et al. 1998). Therefore, we determined whether nucleotide changes in intron 2 of the human KRAS gene affected the binding of those transcription factors. The use of specific antibodies showed that both of the nucleotide substitutions (rs12228277 and rs61761074) which change the binding pattern of nuclear proteins affected the binding site of NF-Y. However, the results differed between rs12228277 and rs61761074; this situation was probably caused by differences in experimental protocols. In the first case (figure 1A, lanes 1 and 4) antibodies to NF-Y were added after mixing the nuclear extract with a DNA probe; in the second case (figure 1C, lanes 1 and 4) antibodies to NF-Y were added during pre-incubation of the nuclear extract with sheared salmon sperm DNA, and only after that was the DNA probe added. As a result, in figure 1A a supershift can be seen, whereas in figure 1C the disappearance of a mobility shift band can be seen. GATA-6 does not participate in the binding of the genomic region containing rs12228277, but GATA-6 is a part of the protein complex in the vicinity of rs61761074. However, a T>G substitution in this region did not affect the binding of GATA-6.

Overall, the above data suggest that the nucleotide changes of T>A at position 190 bp (rs12228277) and T>G at position 615 (rs61761074) are located in the region of the second intron of the human KRAS gene to which protein complexes bind, and these substitutions changed the binding pattern of nuclear proteins in this region. It should be noted that, as reported in mice, the above substitutions affected NF-Y binding sites.

3.2 rs12228277, rs12226937 and rs61761074 affect transcriptional activity of the luciferase reporter

To evaluate the effects of rs12228277, rs12226937 and rs61761074 on transcriptional activity of the plasmid
pGL4.23, we prepared six reporter plasmids, containing the same oligonucleotides (table 1) that were used for the EMSA assays. For transfections, we used the human lung adenocarcinoma A549 cells, after ensuring that NF-Y is expressed (figure 2).

According to the data in figure 3, insertion (before the promoter) of the oligonucleotides corresponding to the major variants of rs12226937 and rs61761074 exhibited the highest activity, similar to the activity of the construct pGL4.23. With respect to minor alleles, both significantly reduced the level of expression of the reporter gene. Notably, in the case of rs61761074, the effect was more pronounced. Insertion of the oligonucleotides corresponding to both the major and minor variants of rs1222877 reduced the level of expression of the reporter gene, and the effect of the minor allele was twofold stronger than the effect of the major allele ($p<0.05$). Therefore, in the case of T>A (rs12228277) and T>G (rs61761074), nucleotide substitutions led to alterations in the nuclear protein binding pattern and the luciferase activity. In contrast, G>A substitution (rs12226937) only showed a reduction of the reporter expression level. This may be because transfection experiments are performed with living cells, while EMSA is an in vitro technique. The nuclear extract preparation for EMSA is time-consuming, and some nuclear proteins may become inactive by the end of this procedure. These results demonstrate that the reporter assays were more informative.

In general, our experiments indicate that the genomic locus containing rs12228277, rs12226937 and rs61761074 may act as a regulatory region, thus supporting the functional significance of allelic variations in intron 2 of the human KRAS gene.

3.3 Analysis of ChIP-seq datasets revealed NF-Y binding and the enhancer marks H3K4me1 and H3K4me2 in intron 2 of the human KRAS gene

Our results indicate that there are NF-Y binding sites in the beginning of intron 2 of the K-ras gene, and that the region is potentially regulatory. However, both EMSA and luciferase reporter assays may not reflect the behavior of this segment of K-ras in the native chromatin context. Nowadays, ChIP-seq is the gold standard for studying transcription factor–chromatin interactions in vivo. We made use of publicly available genome-wide ChIP-seq datasets from the ENCODE project (Raney et al. 2011) to investigate NF-Y binding throughout the region of interest (chr12:25.394.714-

![Figure 3](image_url). Analysis of activity of reporter constructs. A549 cells were co-transfected with 100 ng of pRL-TK vector encoding Renilla luciferase as internal control and 1000 ng of pGL 4.23 vector encoding Firefly luciferase downstream oligonucleotides corresponding to major and minor alleles (table 1). 48 h after the co-transfection, luciferase activity was measured by the dual-luciferase reporter assay kit (Promega, USA). Firefly luciferase activity was normalized to Renilla luciferase activity. The bars indicated the mean ± SD of the luciferase activity ($n = 7$). Significant difference *$p<0.05$ was assessed by Student’s $t$-test, compared to major allele.
25.394,714). We found that rs1222877 falls into an NF-Y binding locus in human GM12878 lymphoblastoid cells and HeLa S3 cells, but not in K562 myelogenous leukemia cells (Supplementary figure 1A). The other two SNPs (rs61761074 and rs12226937) are situated in an NF-Y binding locus in HeLa S3 only. The same experiments are yet to be performed on lung or lung cancer cells; however, since many NF-Y binding loci are common to different cell types (Fleming et al. 2013), it can be assumed that NF-Y binds to the corresponding region in these cells also.

Further evidence for a regulatory function of this region in lung cancer cells comes from the analysis of ENCODE datasets on histone modifications (Supplementary figure 1B). The beginning of intron 2 appeared to be enriched for the enhancer marks H3K4me1 and H3K4me2 in the human lung adenocarcinoma epithelial cell line A589 in comparison with K562 and HeLa S3 (Supplementary figure 1B). The region in lung cancer cells was also significantly enriched for chromatin marks associated with active transcription, confirming K-ras expression in these cells.

3.4 Minor G allele of rs61761074 is associated with an increased risk of developing NSCLC

The frequency of allelic variants A and T (rs12228277), G and A (rs12226937) and T and G (rs61761074) in lung cancer patients and controls from the Western Siberian region are shown in table 2. For comparison, the table also contains the data from dbSNP NCBI. One can see that the frequency of the minor allele G (rs61761074) is substantially increased (more than three-fold) in our sample of NSCLC patients compared to the data from NCBI. Moreover, in our sample of healthy volunteers, this variant was not found. Further confirmation of the association of the G allele with lung cancer was obtained from the analysis of G (rs61761074) allele frequency in different populations. It was found that these frequencies may differ significantly. In a European population, the G allele occurrence is 7%, while in a Sub-Saharan African population it does not occur at all (table 3). Furthermore, the incidence of lung cancer in a European population is 29 cases per 100,000 people, while in a Sub-Saharan African population the incidence is 3.6 cases per 100,000 people.

4. Discussion

The main findings of this study show that rs61761074 is a putative regulatory element located in intron 2 of the human KRAS gene, and its minor allele G may be of importance as a preclinical marker of non-small-cell lung cancer (NSCLC). NSCLC is the main type of lung cancer (80–85% of cases) (Larsen & Minna 2011) and the leading cause of cancer-related deaths (Siegel et al. 2011; Henley et al. 2014). The molecular basis of NSCLC is complex and heterogeneous. It is known that numerous cell signaling pathways may be disrupted to promote NSCLC, including mutations in critical growth regulatory proteins (KRAS, EGFR, B-RAF, MEK-1, HER2, MET, EML-4-ALK, KIF5B-RET, and NKX2.1) and inactivation of growth inhibitory pathways (TP53, PTEN, p16, and LKB-1) (Johnson et al. 2012).

One of the main factors leading to NSCLC development is activating mutations in the KRAS gene (15–40% of cases), which are mostly located in codon 12 (less frequently in codons 13 and 61) (Rodenhuis & Slebos 1992; Huncharek et al. 1999). Mutation in codon 12 results in the KRAS protein becoming constitutively active. In this activated, GTP-bound state, KRAS signals downstream to MEK/

<table>
<thead>
<tr>
<th>SNP identifier</th>
<th>dbSNP NCBI*</th>
<th>Sample of NSCLC patients from the Western Siberian region</th>
<th>Sample of controls from the Western Siberian region</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12228277</td>
<td>T – 88%; A – 12%</td>
<td>T – 84%; A – 16%</td>
<td>T – 85%; A – 15%</td>
</tr>
<tr>
<td>rs61761074</td>
<td>T – 98%; G – 2%</td>
<td>T – 94%; G – 6% (p&lt;0.05)†</td>
<td>T – 100%; G – 0%</td>
</tr>
<tr>
<td>rs12226937</td>
<td>G – 65%; T – 35%</td>
<td>G – 83%; T – 17%</td>
<td>G – 69%; T – 31%</td>
</tr>
</tbody>
</table>

The values are presented as frequency in percentage (%) The frequencies comparison between groups was analyzed by Fisher test. The level of significance was set at p<0.05.

†Fisher test, compared to sample of controls from the Western Siberian region.
NSCLC, non-small-cell lung cancer.
ERK, PI3K or the RalGDS pathways to regulate cell proliferation, differentiation and death (Malumbres & Barbacid 2003). The activating mutations in codon 12 of the K-ras gene also cause lung cancer in mice (Meuwissen et al. 2001; Ramakrishna et al. 2000). In addition, two base variations have been discovered in mice at nucleotide positions 288 and 296 in the second intron of the K-ras gene, associated with susceptibility to lung tumors. Studies show that susceptible strains of mice have nucleotides C and A at those positions, whereas resistant mice have G and C (Ryan et al. 1987; Chen et al. 1994; Timofeeva et al. 1999; Timofeeva et al. 2002). These SNPs are thought to be regulatory because they change the binding pattern of lung nuclear proteins (Chen et al. 1994; Gorshkova et al. 2006). Among the proteins whose binding affinity is affected by the above substitutions, we previously identified NF-Y and GATA-6 (Gorshkova et al. 2006). In the light of these data, it was of great interest to reveal if SNPs located in the same region of humanKras are also regulatory.

In this study, we uncovered three potentially regulatory polymorphisms in intron 2 of the human Kras gene, rs12228277: T>A, rs12226937: G>A and rs61761074: T>G, located at positions 190, 506 and 615 bp relative to the translation start. Our data show that two of these substitutions (rs12228277 and rs61761074) change the binding patterns of lung nuclear proteins in EMSA, and in both cases, there are changes in binding of the NF-Y transcription factor. The T>A substitution (rs12228277) increases the affinity of NF-Y binding, whereas the T>G substitution (rs61761074) creates a new NF-Y-binding site. Our EMSA results are in agreement with ENCODE ChIP-seq data demonstrating NF-Y binding in vivo to the region of the SNP locations in some cell lines. NF-Y is a ubiquitous transcription factor that acts synergistically with other transcription factors for activation or repression of a number of genes (Romier et al. 2003; Milton et al. 2013; Lin et al. 2014). The remarkable feature of NF-Y is its ability to distort DNA at a degree varying between 62° and 82°, and so this factor is proposed to play an important architectural function in chromatin (Ronchi et al. 1995; Wenhu et al. 2010).

The potential regulatory function of all the SNPs under study was further confirmed by luciferase reporter assays. We have shown that the T>A substitution at position 190 bp (rs12228277), G>A substitution at position 506 bp (rs12226937) and T>G substitution at position 615 bp (rs61761074) influence luciferase activity, indicating that these SNPs may be located in a regulatory region. The additional evidence of the regulatory role of this region came from ENCODE ChIP-seq data demonstrating its enrichment for the enhancer marks H3K4me1 and H3K4me2.

---

Table 3. Allele frequencies of rs1222877, rs12226937 and rs61761074 in European population and Sub-Saharan African population compared to lung cancer incidence

<table>
<thead>
<tr>
<th>Allele frequency</th>
<th>European population</th>
<th>Sub-Saharan African population</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12228277</td>
<td>rs12228277</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>36%</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12226937</td>
<td>rs12226937</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs61761074</td>
<td>rs61761074</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2%</td>
<td></td>
</tr>
</tbody>
</table>

The percentage of the allele frequency from dbSNP NCBI

Lung Cancer incidence from GLOBOCAN 2012 (http://globocan.iarc.fr/Default.aspx)

*Age-standardised rate per 100,000 persons (both sexes) per year. [GLOBOCAN 2012 (http://globocan.iarc.fr/Default.aspx)]

---

J. Biosci. 40(5), December 2015
Our genotyping results suggest that rs61761074 may be linked to predisposition to lung cancer. The frequency of its minor allele is elevated in our sample of patients compared both to the data from NCBI and the control sample of healthy volunteers from the Western Siberian region. It is noteworthy that this allele does not occur in a Sub-Saharan African population characterized by low lung cancer incidence [GLOBCAN 2012 (http://globoCan.iarc.fr/Default.aspx)].

Although it is not possible at this time to further define the mechanism of lung cancer predisposition for patients who harbor the rs61761074 G allele, our *in vitro* studies suggest that altered regulation of K-ras expression may be one plausible mechanism. An additional hypothesis could be that rs61761074 somehow influences the expression of another cancer-related gene (or genes). Nowadays, there are numerous examples of such distant effects. In particular, trans-effector were detected between mean platelet volume SNP rs12485738 and seven well-known blood coagulation genes (*F13A1, GP1BB, GP9, ITGA2B, MMRN1, THBS1* and *VWF*), in both the peripheral blood data and the monocyte data (Fehrmann *et al.* 2011). Similarly, intron 19 of *CLEC16A*, containing many autoimmune disease-associated SNPs, appears to behave as a regulatory sequence, affecting the expression of a neighboring gene, *DEXI*. Moreover, using chromosome conformation capture (3C) the *DEXI* promoter region and intron 19 of *CLEC16A* were shown to be in physical proximity, although separated by a loop of >150 kb (Davison *et al.* 2012).

With this in mind, we have used publicly available genome-wide ChIA-PET datasets for five different cell lines from the ENCODE project to analyze chromatin regulatory contacts between the beginning of intron 2 of human *KRAS* and other genome regions. It was found that this region is in direct contact with the promoter regions of the genes *CASC1* and *LYRM5*, situated at a distance of 50 kb relative to the *KRAS* promoter. Moreover, these regions may interact indirectly via a potential enhancer that is located at a distance of 120 kb (chr12:25346001-25541636) relative to the *KRAS* start site (Supplementary figure 2). Both *CASC1* and *LYRM5* have been linked with murine lung tumorigenesis (Manenti *et al.* 2004; Liu *et al.* 2006). Recently, a functional role for *CASC1* in human lung cancer has also been demonstrated, as many cancer cell lines, including NSCLC, require *CASC1* for promoting escape from mitotic-stress induced tumor cell death (Sinnott *et al.* 2014).

Taken together, our results progress the understanding of the molecular and pathological significance of regulatory SNPs in the *KRAS* gene, and indicate the potential of rs61761074 as an important genetic marker in the development of NSCLC.

Acknowledgements

The research has been carried out with support of the grant from Russian Academy of Science Program Fundamental Science for Medicine #23, and Federal Targeted Programmer for Research and Development in Priority Areas of Development of the Russian Scientific and Technological Complex for 2014-2020, Development of Molecular Signatures for Early Detection of Lung Cancer (No. 14.575.21.0064 from 05.08.2014) and State Budget Project N VI.58.1.2.

References


Bryzgalov LO, Antontseva EV, Matveeva MY, Shilov AG, Kashina EV, Mordvinov VA and Merkulova TI 2013 Detection of regulatory SNPs in human genome using ChIP-seq ENCODE data. PLoS One **8** e78833


Cooper DN, Chen JM, Ball EV, Howells K, Mort M, *et al.* 2010 Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum. Mutat.* **31** 631–655


*J. Biosci.* **40**(5), December 2015


Kundu ST, Nallur S, Paranjape T, Boeke M, Weidhaas JB and Slack FJ 2012 KRAS alleles: the LCS6 3′UTR variant and KRAS coding sequence mutations in the NCI-60 panel. *Cell Cycle* 11 361–366


Ponomareno JV, Orlova GV, Frolow AS, Gelfand MS and Ponomareno MP 2002a SELEX_DB: a database on in vitro selected oligomers adapted for recognizing natural sites and for analyzing both SNPs and site-directed mutagenesis data. *Nucleic Acids Res.* 30 195–199


Romier C, Cocchiarella F, Mantovani R and Moras D 2003 The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J. Biol. Chem.* 278 1336–1345


Timofeeva OA, Filipenko ML and Kaledin VI 1999 The study of correlation between the K-ras genotype and murine susceptibility to chemically induced lung neoplasms. Genetika 35 1309–1312
Vasiliev GV, Merkulov VM, Kobzev VF, Merkulova TI, Ponomarenko MP and Kolchanov NA 1999 Point mutations within 663-666 bp of intron 6 of the human TDO2 gene, associated with a number of psychiatric disorders, damage the YY-1 transcription factor binding site. FEBS Lett. 462 85–88

MS received 04 April 2015; accepted 07 August 2015

Corresponding editor: SORAB N DALAL