The SRY-HMG box gene, SOX4, is a target of gene amplification at chromosome 6p in lung cancer

Pedro P. Medina1,†, Sandra D. Castillo1, Sandra Blanco3, Marta Sanz-Garcia3, Cristina Largo2, Sara Alvarez2, Jun Yokota4, Ana Gonzalez-Neira2, Javier Benitez2, Hans C. Clevers5, Juan C. Cigudosa2, Pedro A. Lazo3 and Montse Sanchez-Cespedes1,*

1Lung Cancer Group, Molecular Pathology Programme and 2Human Cancer Genetics Program, Centro Nacional de Investigaciones Oncologicas (CNIO), E-28029 Madrid, Spain, 3Programa de Oncologia Translacional, Instituto de Biologia Molecular y Celular del Cancer (CIC)—Consejo Superior de Investigaciones Científicas (CSIC)—Universidad de Salamanca, Salamanca, Spain, 4Biology Division, National Cancer Center Research Institute, Tokyo, Japan and 5Hubrecht Laboratory, Center for Biomedical Genetics, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

Received October 8, 2008; Revised and Accepted January 15, 2009

The search for oncogenes is becoming increasingly important in cancer genetics because they are suitable targets for therapeutic intervention. To identify novel oncogenes, activated by gene amplification, we analyzed cDNA microarrays by high-resolution comparative genome hybridization and compared DNA copy number and mRNA expression levels in lung cancer cell lines. We identified several amplicons (5p13, 6p22-21, 11q13, 17q21 and 19q13) that had a concomitant increase in gene expression. These regions were also found to be amplified in lung primary tumours. We mapped the boundaries and measured expression levels of genes within the chromosome 6p amplicon. The Sry-HMG box gene SOX4 (sex-determining region Y box 4), which encodes a transcription factor involved in embryonic cell differentiation, was overexpressed by a factor of 10 in cells with amplification relative to normal cells. SOX4 expression was also stronger in a fraction of lung primary tumours and lung cancer cell lines and was associated with the presence of gene amplification. We also found variants of SOX4 in lung primary tumours and cancer cell lines, including a somatic mutation that introduced a premature stop codon (S395X) at the serine-rich C-terminal domain. Although none of the variants increased the transactivation ability of SOX4, overexpression of the wildtype and of the non-truncated variants in NIH3T3 cells significantly increased the transforming ability of the weakly oncogenic RHOA-Q63L. In conclusion, our results show that, in lung cancer, SOX4 is overexpressed due to gene amplification and provide evidence of oncogenic properties of SOX4.

INTRODUCTION

Solid tumours are characterized by a complex pattern of cytogenetic and genetic changes (1). The use of high-resolution and genome-wide scanning platforms provides detailed information about the recurrent patterns of chromosomal abnormalities in cancer. In particular, the search for regions with a notable increase in copy number provides an opportunity to identify novel oncogenes, which constitute suitable targets for therapeutic intervention. The positive association of oncogene activation (somatic activating mutations or gene amplification) and the clinical response to treatments with specific small inhibitors or monoclonal antibodies underline the usefulness of oncogenes as key target molecules for the design of novel and specific drugs in cancer therapy.
Genome-wide screening for increased DNA copy number and gene expression in cancer cells helps us in our search to identify regions with high levels of gene amplification and overexpression. In lung tumours, some of the classical oncogenes altered by amplification include the MYC gene family, MDM2, or ERBB2 (2–6). More recently, the identification of somatic EGFR gene mutations, and possibly amplification, coupled with the ability of the EGFR-mutant tumours to respond to treatment with tyrosine kinase inhibitors of EGFR underscores the usefulness of oncogenes as key target molecules for the design of novel and specific drugs in cancer therapy (7,8). New amplified oncogenes in lung cancer, such as NKX2-1 and PI3KCA, have been unveiled and open new windows for novel targeted therapies (9,10). Other regions carrying recurrent gene amplification have been described, which are likely to contain oncogenes (10).

Our present work reports our analysis of cDNA microarrays by high-resolution comparative genome hybridization (CGH), by which we directly compared DNA copy number and mRNA expression levels as a means of identifying novel oncogenes altered by gene amplification in lung cancer. Our results provide evidence that the Sry-HMG box gene, SOX4 (sex-determining region Y box 4), which encodes a transcription factor, is the oncogene targeted in the chromosome 6p amplicon. The SOX4 gene belongs to a large family of transcription factors related by homology in their DNA-binding domains to the HMG-box region of the testis-determining gene, SRY, and has been proposed as an oncogene in leukemias and in solid tumours (11–19).

RESULTS
Analysis of DNA copy number in lung cancer cell lines and primary tumours

We used cDNA-based CGH microarrays to profile DNA copy number and cDNA microarrays to assess global gene expression of the lung cancer cell lines A427, A549, Calu-3, NCI-H23, NCI-H441, NCI-H522, NCI-H1299 and NCI-H2126. Increased DNA copy number was observed at the 5p13, 6p21-22, 11q13, 17q21 and 19q13 chromosomal regions (Fig. 1A) concomitantly with the increased gene expression of transcripts within the ampilcns. A known 17q21 amplicon in the Calu3 cells, which probably targets the ERBB2 oncogene, was used as a positive control. Interphase and metaphase FISH analysis of all the ampilcns verified the amplification, as indicated by the presence of at least 10 copies in all the ampilcns compared with only between two and four in the corresponding centromeric reference. In metaphase preparations, all ampilcns showed a pattern suggesting a mechanism for gene amplification based on intrachromosomal homogenously staining regions (HSRs) (Fig. 1B). The approximate size of ampilcns and copy number detected by FISH for each cell line are presented in Table 1. The candidate oncogenes in each region are also listed. Using FISH analysis on tissue microarrays (TMAs) we confirmed that these alterations were also present in a fraction of lung primary tumours. The frequency of gene amplification, as defined in the Materials and Methods, was 8% (two out of 25 lung tumours tested) for the 5p13 and 6p21-22 ampilcns and 4% (one out of 25 lung tumours tested) for the 11q13, 17q21 and 19q13 ampilcns. Twenty-eight percent of the tumours showed gene amplification of at least one of the ampilcns. Examples of the FISH analysis in lung primary tumours are depicted in Figure 1C.

SOX4 is highly overexpressed in cancer cell lines with chromosome 6p amplification and in lung primary tumours

Amplification at 5p and 6p was more commonly detected in primary tumours, so we decided to examine these in greater detail. First, boundary delimitation of the ampilcns was determined using real-time quantitative PCR (QPCR). Regarding chromosome 5p13, QPCR of several genes within the ampilcon confirmed the presence of 10 x more copies compared with the reference gene, PARP8, located at 5q11.2 (Supplementary Material, Fig. S1A). Gene amplification occurred in a region of about 5 Mb containing at least 25 known genes (from SKP2 to FYB, telomere to centromere). Using cDNA microarrays and reverse transcription QPCR (RT–QPCR), we tested the levels of 14 of these transcripts and found only a slight increase in gene expression in some of them (between 1.2 and 1.8 relative gene expression) compared with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and other cell lines (Supplementary Material, Fig. S1B). Similarly, we used QPCR to determine the boundaries and levels of gene expression within the chromosome 6p ampilcon in the NCI-H522 cells (hereafter referred to as H522). FISH analysis clearly reproduced previous CGH observations (20) of the presence of two normal and two derivative chromosomes 6 (Fig. 1B). One of the derivative chromosomes 6 shows a region of gene amplification in an HSR manner. In addition, the control probe, on 6q12, also appears as part of the ampilcon, suggesting the existence of a complex rearrangement. Delimitation of the amplified region both by CGH microarrays and QPCR demonstrated the presence of co-amplification of two regions (6p21.2 and 6p22.3) that are about 10 Mb apart (Fig. 2A), which suggests that there exists a mechanism of gene amplification coupled with a complex rearrangement. QPCR for DNA confirmed a 7-fold increase in the 6p21.2 and 6p22.3 regions relative to the control gene, NUP43, at 6q25.1 (Fig. 2A). The H522 cells have three long arms on chromosome 6, so the absolute copy number could be as high as 21. That the copy number at chromosome 6p21-22 in the H522 cells is greater than 20 is evident in Figure 1B. According to cDNA microarray analysis, gene expression within the ampilcns was notably higher in 6p22.3 than in 6p21.2. Thus, we selected the 6p22.3 region for further analysis. CGH microarrays defined the boundaries for gene amplification within a region of about 10 Mb containing at least 25 known genes and open reading frames (from MYLIP to TTRAP, telomere to centromere). We combined the RT–QPCR and cDNA microarray analyses to measure expression levels of 10 genes (MYLIP, FAM8A1, NUP153, DEX, ID4, E2F, CDKAL1, SOX4, ALDH5A1, TTRAP) within the region (Fig. 2B and C). Compared with cells without gene amplification, ID4 and SOX4 were considerably more strongly expressed than GAPDH (data not shown). However, when levels of ID4 and SOX4 were measured relative to normal lung tissue (the median of three specimens), only SOX4 showed a high level of expression in the H522 cells.
In fact, ID4 was found to be downregulated in the lung primary tumours and cancer cell lines tested (Supplementary Material, Fig. S2), similarly to what has been previously reported in gastric cancer (21). Thus, SOX4 is the most likely driver gene of the amplification at 6p. The genomic DNA of SOX4 is not interrupted by introns and would be amplified by our primers for RT–QPCR. To rule out the possibility that the high levels of SOX4 expression arose from contaminating genomic DNA in our cDNA preparations, we tested for the presence of PCR products of the genomic DNA of MYLIP and E2F3, which are also present in the 6p amplicon, but found none (data not shown). Further evidence of the overexpression of SOX4 came from the observation of higher levels of SOX4 protein in the H522 cells than in other lung cancer cell lines (Fig. 2E). Interestingly, in Figure 2E, it can be observed that the H2126 cells, which do not have SOX4 amplification or overexpression, carry high levels of SOX4 protein. This is intriguing and may indicate the existence of mechanisms for protein stabilization.

To examine further its contribution to lung carcinogenesis we first used real-time QPCR to compare the levels of the SOX4 transcript in a panel of normal lung tissues (n = 16), lung cancer cell lines (n = 85) and lung primary tumours (n = 70) (Fig. 3A). All primary tumours were of the non-small cell lung cancer (NSCLC) type, while lung cancer cell lines included 19 small cell lung cancer (SCLCs), 61 NSCLC and five other histopathologies. Three (3.5%) cell lines and five (6%) primary tumours showed levels of SOX4 expression at least three times higher than the mean value for normal tissues (mean = 10.25; SD ± 5.1). It was also evident that in some tumours SOX4 expression was downregulated in cancer cells compared with those of normal lung. Next, we determined the presence of SOX4 amplification in lung cancer cell lines and lung primary tumours using real-time QPCR on genomic DNA or FISH analysis. Six percent of the lung cancer cell lines (five out of 80 analyzed by DNA-QPCR) and 9% of the lung primary tumours (four out of the 42 that it was possible to analyze by either FISH or DNA-QPCR) had gene amplification (see cut-off values for FISH analysis in Materials and Methods) (Fig. 3A). As expected, most primary tumours with gene amplification had substantially increased gene expression. However, some tumour cells, especially among those of the SCLC type, had high levels of SOX4 transcript in the absence of gene

Figure 1. Screening for amplicons in lung cancer cell lines and primary tumours. (A) Schematic representation of the CGH-cDNA microarray-based results. The stars indicate the regions with the greatest increases in gene copy number (gene amplification) in the indicated cell line and chromosome. (B) FISH analysis confirming the presence of 6p gene amplification in interphase (above) and metaphase (below) in the H522 cell line (BACs:RP11-52L15 and BACs:RP11-631J2 for SOX4; BACs:RP11-2277K21 as control in 6q12). The derivative chromosome 6 is indicated with arrows. The red arrow points to the derivative chromosome 6 carrying gene amplification. (C) Examples of gene amplification at 5p (BACs: CTD-3092E20; pericentromeric 5q: RP11-91H22) and 19q (BACs:RP11-123C1 and pericentromeric RP11-20606) by FISH analysis in two different lung primary tumours.
amplification, suggesting that there are other mechanisms of SOX4 overexpression, particularly in this histopathological type of lung cancer. In fact, SOX4 expression was significantly greater in the SCLC cell lines (mean (pared with the NSCLC cell lines (mean

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Chromosome</th>
<th>Interval (Mb)</th>
<th>Candidate oncogene</th>
<th>Copy number</th>
</tr>
</thead>
<tbody>
<tr>
<td>H23</td>
<td>5p13</td>
<td>34.0–39.1</td>
<td>SKP2, GDNF, OSMR, RICTOR</td>
<td>&gt;15</td>
</tr>
<tr>
<td>H522</td>
<td>6p22.3</td>
<td>16.2–26.1</td>
<td>DEK, ID4, E2F3, SOX4, RHO, RPS6KB2, AIP, DOK-1R</td>
<td>&gt;21</td>
</tr>
<tr>
<td>H522</td>
<td>6p21.2</td>
<td>35.5–39.8</td>
<td>PIM1</td>
<td>&gt;20</td>
</tr>
<tr>
<td>H23</td>
<td>11q13.2</td>
<td>66.5–67.5</td>
<td>RHOD, RPS6KB2, AIP, DOK-1R</td>
<td>&gt;10</td>
</tr>
<tr>
<td>H441</td>
<td>11q13.4</td>
<td>72.0–73.3</td>
<td>RAB6</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Calu3</td>
<td>17q21</td>
<td>32.0–36.9</td>
<td>ERBB2, TOPO2A</td>
<td>&gt;20</td>
</tr>
<tr>
<td>H2126</td>
<td>19q13</td>
<td>54.1–54.6</td>
<td>PPFIA3</td>
<td>&gt;10</td>
</tr>
</tbody>
</table>

The candidate oncogenes in each group are also indicated.

**SOX4 gene mutations in lung tumours**

Oncogenes can also be activated by point mutations, usually missense mutations or in-frame deletions/insertions. To test for the presence of activating mutations at SOX4 in lung tumours we sequenced its entire coding exon in a panel of 50 lung cancer cell lines and 46 primary tumours with various histopathologies. In the lung primary tumours, there were three germline variants, which correspond to polymorphisms because they were found in a similar frequency in genomic DNA from control individuals (Table 2). Two novel variants, A316V and A204del, were identified only in the controls. In addition to these, there was a somatic mutation at the 395 residue that introduced a premature stop codon at the C-terminal domain, predicting a shorter SOX4 protein devoid of the region immediately after the serine-rich domain (Fig. 3C and D). This alteration was absent in the controls. In the cell lines, we also identified variants that cause changes in the protein sequence. All of these were heterozygous and clustered in the serine-rich domain. However, the lack of normal matched tissues in these cases prevented us from determining whether some of these were of somatic or germline origin. None of these variants has been detected in the genomic DNA from 70 control individuals tested (Table 2). We also found three neutral variants, S386S, A177A and L405L, in a primary tumour, in the NCI-H23 cell line and in a normal control, respectively.

**Transactivation and transforming ability of overexpressed wildtype and mutant forms of SOX4**

We cloned ΔS387, A350P and S395X mutations in the pCEFL expression vector (Fig. 4A). As depicted in Figure 4A, the ectopic expression of the S395X mutant leads to a substantial increase in the levels of the SOX4 shorter protein as compared with the wildtype or with the other variants. This effect, which was reproducible in several independent experiments and cell lines, may indicate that the lack of the C-terminal domain increases the stability of the SOX4 protein. To evaluate their possible effect on the transactivation ability of SOX4, the different constructs were co-transfected with a reporter-Luciferase construct [SOX optimal (SOP)] in (i) the mouse fibroblasts cells, NIH3T3 and (ii) a panel of cancer cells (NCI-H1299, HCT116, A549, MCF7, HeLa). Our results revealed that the ability of SOX4 to transactivate the reporter gene was not changed in A350P or ΔS387 variants but was completely abrogated in the S395X mutant form (Fig. 4B and Supplementary Material, Fig. S3). The lack of transactivation of the SOP reporter construct in the S395X was not due to a dominant negative effect since co-transfection of wildtype SOX4 together with increasing amounts of S395X did not inhibit the transactivation induced by the wildtype form (Supplementary Material, Fig. S4). The truncated S395X protein is devoid of the C-terminal domain, and a loss of transactivation ability after truncation of the SOX4 protein at amino acid position 219 has been reported (22). In addition, we evaluated the ability of wildtype and mutant SOX4 to induce oncogenic transformation of the NIH3T3 mouse fibroblast cells, as measured by the ability of the cells to form foci. To this end we transfected the plasmids driving the expression of the wildtype, and the A350P, ΔS387 and S395X variants into the NIH3T3 cells. Neither the wildtype nor the mutant forms of SOX4 induced focus formation (data not shown). Moreover, we co-transfected the different SOX4 constructs with limiting amounts of plasmids expressing the activated HRAS (HRAS-G12V) or RHOA (RHOA-Q63L) oncoproteins. Increasing amounts of wildtype or ΔS386, A350P variants significantly increased the oncogenic ability of RHOA-Q63L (Fig. 5A), implying that SOX4 overexpression has oncogenic potential. However, neither the wildtype SOX4 nor any of the mutants were able to increase the oncogenic ability of HRAS-G12V (Fig. 5B), maybe because HRAS-G12V oncogenic ability was always greater than that of RHOA-Q63L. Intriguingly, the S395X variant behaved in an opposite manner to the wildtype and the A350P and ΔS386 mutants. S395X did not affect the transforming ability of RHOA-Q63L, but massively reduced the number of foci formed by HRAS-G12V (Fig. 5), suggesting that SOX4 transcriptional activity supports HRAS transforming ability. This particular behaviour of the S395X mutant was not due to an aberrant subcellular localization, because it retained the ability to localize in the nucleus as determined using immunofluorescence and confocal microscopy analysis (Supplementary Material, Fig. S5).

**DISCUSSION**

Our cDNA-based CGH microarray analysis reveals the presence of several amplicons in lung cancer cells, some of them coincident or located close to others that have been...
The presence of these amplicons was consistently confirmed by FISH analysis, which underlines the robustness of cDNA microarray-based platforms for determining DNA copy number. Among the regions that showed consistent amplification was chromosome 6p22 in the H522 cells. This chromosome is amplified or rearranged in many types of tumours (20–23) and was also reported to be a common breakpoint of reciprocal translocations (24), which strongly implies the presence of an important oncogene in this region. Here we show that chromosome 6p22 amplification in H522 cells leads to substantial overexpression of $SOX4$ and $ID4$ genes. Our current observations are consistent with those previously noted, that $SOX4$ is amplified and highly expressed in H522 cells (5). It has been suggested that $E2F$ and $ID4$ could be the oncogene target of chromosome 6p22 amplification (25,26). However, $E2F$ was not strongly overexpressed in H522 cells and, although $ID4$ did exhibit strong overexpression, it was weaker than observed in normal lung tissues. This suggests that $ID4$ is, in fact, repressed during tumourigenesis. $ID4$ is one of four members of the basic helix-loop-helix family of transcription factors, which acts as transcription inhibitory proteins. Previous reports have described the common downregulation of $ID4$ by promoter hypermethylation in gastric cancer, breast cancer and leukaemia, implying a tumour suppressor role for the encoded protein (27,28). In contrast, $SOX4$ overexpression appears to be tumour-specific. $SOX4$ was overexpressed in a subset of lung primary tumours and additional lung cancer cell lines relative to normal lung. In agreement with our findings, global expression analysis performed in lung cancer also has previously shown that $SOX4$ is overexpressed in a subset of lung tumours (29–31). Moreover, in NSCLC, which accounts for 80% of diagnosed lung cancers, $SOX4$ overexpression is associated with the occurrence of gene amplification. It has recently been reported that $SOX4$ expression can be downregulated through the expression of miR-335, leading to suppression of metastasis and migration, and that miR-335 expression is lost in primary breast tumours (32). Thus, the decreased expression of miR-335 could account for those cases of lung cancer,
especially in SCLC tumours, where SOX4 is overexpressed in the absence of gene amplification.

SOX4 is a strong candidate oncogene. Independent studies have demonstrated that SOX4 (also known as ecotropic viral integration site 16, Evi16) is among the most frequently targeted genes in retroviral insertional mutagenesis, leading to neoplastic transformation in murine hematopoietic cells (11–15). Moreover, transfection of normal murine bone marrow cells with an Sox4-expressing retrovirus induces myeloid leukaemias (14). SOX4 may also be important in

Table 2. SOX4 variants found in lung cancer

<table>
<thead>
<tr>
<th>Tumour Id.</th>
<th>Histology</th>
<th>Change</th>
<th>Effect</th>
<th>Germline</th>
<th>Het/Hom.</th>
<th>Other mut.</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary tumours (n = 46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>SCC</td>
<td>c.1207C&gt;A</td>
<td>p.S395X</td>
<td>NO</td>
<td>nd</td>
<td>TP53</td>
<td>0/70</td>
</tr>
<tr>
<td>348</td>
<td>AC</td>
<td>c.82C&gt;G</td>
<td>p.L28V</td>
<td>YES</td>
<td>nd</td>
<td>KRAS</td>
<td>4/70</td>
</tr>
<tr>
<td>362</td>
<td>SCC</td>
<td>c.519insCGG</td>
<td>p.G173insG</td>
<td>YES</td>
<td>nd</td>
<td></td>
<td>0/70</td>
</tr>
<tr>
<td>Lung cancer cell lines (n = 50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H661</td>
<td>LCC</td>
<td>c.1161delCTC</td>
<td>p.S387del</td>
<td>nd</td>
<td>Het</td>
<td>BRG1,P16, TP53</td>
<td>0/70</td>
</tr>
<tr>
<td>H1618</td>
<td>SCLC</td>
<td>c.1109C&gt;T</td>
<td>p.S368L</td>
<td>nd</td>
<td>Het</td>
<td>P16</td>
<td>0/70</td>
</tr>
<tr>
<td>Ms18</td>
<td>SCLC</td>
<td>c.1054G&gt;C</td>
<td>p.A350P</td>
<td>nd</td>
<td>Het</td>
<td></td>
<td>0/70</td>
</tr>
</tbody>
</table>

SCC, squamous cell carcinoma; AC, adenocarcinoma; LCC, large cell carcinoma; SCLC, small cell lung cancer. nd, not determined; Het, heterozygous; Hom, homozygous.

*Refers to the frequency of the variant in normal Spanish population, from the analysis of 70 control individuals. Two new variants, p.A316V and p.A204del, were identified in 7 and 1% of the controls, respectively.

Figure 3. Analysis of levels of SOX4 expression, amplification and screening for mutations in lung primary tumours and cancer cell lines. (A) Quantitative RT-PCR to determine the levels of gene expression of SOX4 relative to GAPDH in normal lung tissues, lung primary tumours and lung cancer cell lines. Black bars indicate tumours or cell lines with SOX4 amplification, as revealed by FISH (primary tumours) or QPCR of genomic DNA (cell lines). Grey and punctuated bars indicate no amplification and no data, respectively. The inset shows a primary tumour with gene amplification by FISH. (B) Expression levels of SOX4 obtained from real-time quantitative RT–PCR analysis relative to GAPDH in SCLC (mean = 12.2; SD ± 14.6) compared with NSCLC (mean = 4.25; SD ± 7.8) cell lines (*P < 0.005, t-test). (C) Point mutation leading to a premature stop codon in a lung primary tumour (104-T) compared with matched normal DNA (104-N) and ΔS387 in the indicated cell line. (D) Amino acid composition of the SOX4 protein. Germline variants are indicated in bold, whereas somatic or variants with unknown germline status are indicated in red. The HMG domain is highlighted in light grey, and the serine-rich region of the SOX4 protein is illustrated in dark grey.
the development of solid tumours. SOX4 protein levels are increased in bladder, prostate, urothelial and adenoid cystic carcinomas (16,17–19) and knocking down SOX4 by small-interfering RNA transfection induces apoptosis of prostate and adenoid cystic cancer cells (18,19).

In this paper, we also show that SOX4 overexpression in NIH3T3 cells increases the number of foci induced by the mutant RHOA-Q63L, which by itself is weakly oncogenic compared with other small GTPases, such as HRAS (33–35). This occurred in the wildtype and the variants that produce amino acid substitutions or in-frame deletions, indicating that SOX4 overexpression promotes loss of contact inhibition in these cells. The lack of functional differences in the behaviour of the SOX4 non-truncated variants suggests that these are merely rare polymorphisms. The effect of SOX4 overexpression by which the number of foci induced by the mutant RHOA-Q63L is increased, but not by the action of HRAS-G12V, is intriguing. It was also surprising to identify a somatic nonsense mutation in SOX4 in a primary tumour, especially because it encodes a protein lacking transcriptional activity and because co-expression of this mutant form with non-truncated variants suggests that this gene has oncogenic properties.

In conclusion, we report here that gene amplification is one of the mechanisms that underlie the SOX4 overexpression observed in lung tumours and that SOX4 overexpression, specifically the active forms, synergizes the promotion of cell growth induced by the RHOA-Q63L oncogene. This represents genetic evidence of the relevance of SOX4 in tumour development and suggests that this gene has oncogenic properties.

MATERIALS AND METHODS

Cell lines and primary tumours

The NCI-H23, NCI-A427, NCI-H441, NCI-H522, NCI-A549, NCI-H1299, NCI-H2126 and Calu-3 cell lines were obtained from the American Type Culture Collection (Rockville, MD, USA) and grown under recommended conditions. DNA and RNA from additional lung cancer cell lines used for real-time quantitative RT–PCR analysis and for sequencing of SOX4 were kindly provided by Luis M. Montuenga and Ruben Pio of the Division of Oncology, Centro Para La Investigación Médica Aplicada (CIMA), University of Navarra, Spain, and from Jun Yokota, National Cancer Center Research Institute, Tokyo, Japan. Fresh frozen and paraffin-embedded lung primary tumours were provided by the CNIO Tumour Bank Network, CNIO, Madrid, Spain, in collaboration with the Pathology Department at the Hospital Universitario 12 de Octubre. Lung tumours for QPCR analysis were selected on the basis of the availability of frozen tissue and the good quality of total RNA. Lung tumours for sequencing SOX4 were selected on the basis of the availability of matched normal DNA. Paraffin-embedded tissues were used for the construction of three TMAs, as previously described (37).

CGH and analysis of global gene expression

DNA and RNA from cell lines and primary tumours were extracted following standard protocols, as previously described (37,38). Representative sections from primary tumour tissue used for DNA extraction were stained with H&E. Freshly frozen tissue from tumours was meticulously dissected to ensure that specimens contained at least 75% tumour cells. The CGH array and global gene expression analyses were conducted on the CNIO OncoChipTM, a cDNA microarray platform containing 7657 genes (I.M.A.G.E cDNA clones, Research Genetics; Huntsville, AL, USA), some of which are duplicated to give a total of 11 718 spots. A complete list can be found at http://bioinfo.cnio.es/data/oncochip/.

For gene expression, 35 μg of total RNA was used to synthesize and fluorescently label cDNA, as previously described (38). For CGH analysis, hybridizations were performed using genomic DNA extracted from each cell line. As a reference, we used a pool of genomic DNA from the blood of six unrelated control donors. Genomic DNAs were Alul- and Rsal-digested, labelled with Cy5 (huT78) and Cy3 (control) using a BioPrime labelling kit (Life Technologies, Inc., Gaithersburg, MD, USA) and hybridized on microarrays at 50°C for 14–16 h, as previously described (37). Slides were scanned for Cy3 and Cy5
fluorescence using the Scanarray 5000 XL apparatus (GSI Lumonics Kanata, Ontario, Canada). Images were quantified and the data analyzed as previously described (38).

FISH analysis
To determine whether there had been any gene amplification, we performed FISH analysis of cell lines and TMAs. We used the following BAC clones (Supplementary Material, Table S1), labelled with Spectrum Orange (Vysis, Downer’s Grove, IL, USA), and a reference BAC clone (Supplementary Material, Table S1), labelled in Spectrum Green (Vysis). The BACs were obtained from the BACPAC Resource Center of the Children’s Hospital Oakland Research Institute (Oakland, CA, USA). Dual-colour FISH was performed in 4 μm sections of the TMAs. The slides were deparaffinized, boiled in a pressure cooker with 1 mmol/L EDTA (pH 8.0) for 6 min, incubated with pepsin at 37°C for 30 min and then dehydrated. The probes were denatured at 75°C for 1. 5 min and left overnight to hybridize at 37°C in a humid chamber. Slides were washed with 0.4× SSC and 0.3% NP-40 for 2 min and 2× SCC and 0.1% NP-40 for 5 min. Finally, they were covered with 4,6-diamidino-2-phenylindole (DAPI II, Vysis) for chromatin counterstaining before microscopy. Copy number was determined with no previous knowledge of the genetic nature of the samples. Fluorescence signals were scored in each sample by counting the number of single-copy genes and control probe signals in an average of 130 (60–210) well-defined nuclei. Gene amplification was defined as the presence of at least five times as many copies of gene signals as pericentromeric signals in the same chromosome. FISH analysis of lung primary tumours was performed in TMAs. In those cases that were positive for gene amplification, the observations were verified repeating the analysis in a complete section of the tumour.

Antibodies and western blots
For western blotting, cells were scraped from the dishes into the lysis buffer. A total of 100 μg of total protein was separated by SDS–PAGE and blotted with rabbit anti-SOX4 (A574) (1:4000) (CS-129-100, Diagenode, Liège, Belgium) or mouse anti-HA antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA). The secondary antibodies anti-mouse-IgG:HRP (Santa Cruz Biotechnology) and goat anti-Rabbit HRP (DAKO, Glostrup, Denmark) were added to give a final dilution of 1:3000.

SOX4 mutation screening
Tumour DNA from primary lung carcinomas and lung cancer cell lines of various histological types was screened for SOX4 gene mutations. The entire coding region of SOX4, which consists of a single exon, was PCR-amplified in a single product.
Expression plasmids and reporters

Full-length human SOX4 and the different mutants were cloned into the BamHI and Xhol restriction sites of the vector pCEFL (a gift of Silvio Gutkind, NIH, Bethesda, MD, USA) to generate the expression plasmids pCEFL-SOX4, pCEFL-SOX4-S395X, pCEFL-SOX4-A350P and pCEFL-SOX4-ΔS387. All inserts were verified by automatic sequencing. HA epitope tags were introduced at the N-terminus of the SOP and not optimal (NOP) reporters, both of which possess the Luc gene, which is under the control of a fragment containing the AACAAGAAG concatamer. Cells were transfected with 0.75 μg of the vector, 0.15 μg of β-galactosidase and the indicated amount of SOX4 constructs with 6 μl of JetPEI. Luc activity was measured 48 h post-transfection with the Promega Kit (Promega, Madison, WI, USA) and normalized in all cases by β-galactosidase activity (40–42).

Immunofluorescence identification of SOX4 protein

To determine the subcellular localization of the SOX4 protein, we performed an immunofluorescence assay. The cells were fixed with acetone/methanol (1:1) for 5 min and kept frozen at −20°C until processing. Cells were transfected with the different SOX4 constructs and stained with DAPI and anti-HA. Immunofluorescence was performed as previously described (40). Fluorescence was analyzed by confocal microscopy (Zeiss, LSM150), and the co-localization of both markers was evaluated electronically.

Real-time QPCR assay

The genomic DNA and mRNA levels of genes were measured by SYBR green real-time PCR. To determine the mRNA levels, DNase-treated RNA was reverse-transcribed and subsequently amplified using an ABI Prism 7900 Sequence detector (Applied Biosystems, CA, USA). Reactions were performed in triplicate. As internal controls in the quantitative RT–PCR for screening SOX4 overexpression in the large panel of lung primary tumours and cancer cell lines, we used the human GAPDH, B-ACTIN and TATA box binding protein (TBP), since it is well known that endogenous levels of the internal controls can vary among individuals and tumours. Only tumours that demonstrated high levels of SOX4 expression when using each of them as controls were considered to be overexpressed. Those with discrepant results were discarded. To analyze genomic DNA levels, the control gene is indicated in each experiment. To determine SOX4 copy number in lung cancer cell lines, comparison with levels of gene expression allowed a cut-off value to be established, which defined SOX4 gene amplification (fold change of SOX4 genomic DNA relative to the control gene, NUP43, on chromosome 6q). The use of a control gene in the same chromosome as SOX4 helps us to distinguish polyplody from true gene amplification. The sequences of the primers used in each case can be provided upon request.

Focus formation and growth inhibition assays

Focus assays with the indicated mammalian expression vectors were performed as described (43). Briefly, 150,000 NIH3T3 mouse fibroblast cells were seeded at a density of 150,000 cells/100 mm diameter plate in DMEM media containing 10% (v/v) fetal bovine serum. After 24 h, cells were transfected with calcium phosphate, adding 20 μg salmon sperm DNA. Each assay was performed in triplicate. The cultures were maintained at 37°C in a humidified atmosphere of 5% CO₂/95% O₂. After 11–13 days, the cells were stained with Giems (Sigma-Aldrich, St Louis, MO, USA).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank our colleagues in the Tumour Bank of the CNIO and in the Pathology Department of the Hospital 12 de Octubre (Madrid). We also acknowledge the technical help of the CNIO’s Genomics Unit. P.P.M is a recipient of a Hope Funds Fellowship for Cancer Research.

Conflict of Interest statement. None declared.

FUNDING

The work was supported by grants from the Spanish Ministerio de Educacion (SAF2005-00626) and Instituto de Salud Carlos III (RTICC), to M.S.-C., and Spanish Ministerio de Educacion (SAF2007-60242 and CSD2007-0017), Junta de Castilla y Leon (Consejeria de Educacion, CSI-14A08 and SAF2005-00626) and Instituto de Salud Carlos III (RTICC), to M.S.-C., and Spanish Ministerio de Educacion (SAF2007-60242 and CSD2007-0017), Junta de Castilla y Leon (Consejeria de Educacion, CSI-14A08 and GR-15) to P.A.L. Fellowships to M.S.-G. and S.D.C. were granted by the CSIC and Comunidad Autonoma de Madrid, respectively.

REFERENCES


