

ORIGINAL ARTICLE

# Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach

R Li<sup>1</sup>, H Wang<sup>1</sup>, BN Bekele<sup>2</sup>, Z Yin<sup>3</sup>, NP Caraway<sup>1</sup>, RL Katz<sup>1</sup>, SA Stass<sup>4</sup> and F Jiang<sup>4</sup>

<sup>1</sup>Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>2</sup>Department of Biostatistics & Applied Mathematics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>3</sup>Department of Internal Medicine, The University of Texas Houston Medical School, Houston, TX, USA and <sup>4</sup>Department of Pathology, The University of Maryland School of Medicine, Baltimore, MD, USA

**Amplification and overexpression of putative oncogenes confer growth advantages for tumor development. We used a functional genomic approach that integrated simultaneous genomic and transcript microarray, proteomics, and tissue microarray analyses to directly identify putative oncogenes in lung adenocarcinoma. We first identified 183 genes with increases in both genomic copy number and transcript in six lung adenocarcinoma cell lines. Next, we used two-dimensional polyacrylamide gel electrophoresis and mass spectrometry to identify 42 proteins that were overexpressed in the cancer cells relative to normal cells. Comparing the 183 genes with the 42 proteins, we identified four genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – in which elevated protein expression correlated with both increased DNA copy number and increased transcript levels (all  $r > 0.84$ , two-sided  $P < 0.05$ ). These findings were validated by Southern, Northern, and Western blotting. Specific inhibition of *EEF1A2* and *KCIP-1* expression with siRNA in the four cell lines tested suppressed proliferation and induced apoptosis. Parallel fluorescence *in situ* hybridization and immunohistochemical analyses of *EEF1A2* and *KCIP-1* in tissue microarrays from patients with lung adenocarcinoma showed that gene amplification was associated with high protein expression for both genes and that protein overexpression was related to tumor grade, disease stage, Ki-67 expression, and a shorter survival of patients. The amplification of *EEF1A2* and *KCIP-1* and the presence of overexpressed protein in tumor samples strongly suggest that these genes could be oncogenes and hence potential targets for diagnosis and therapy in lung adenocarcinoma.** *Oncogene* advance online publication, 12 December 2005; doi:10.1038/sj.onc.1209289

**Keywords:** lung cancer; microarrays; proteomics; tissue microarray

## Introduction

In lung adenocarcinoma, as in other types of cancer, gene amplification and the consequent overexpression of the amplified oncogene play an important role in the development of tumors, because their overexpression confers a growth advantage. The ability to identify putative oncogenes that are activated during tumorigenesis could facilitate the choice of molecular genetic targets for diagnosis and therapy of the disease. This concept has been exemplified by *HER-2*, which was first found to be amplified in neuroblastomas and subsequently shown to be associated with poor prognosis in breast cancer (Ross and Fletcher, 1999). Now, *HER-2* aberrations are used as a predictor of response to therapy, and treatment of *HER-2*-positive breast cancer with the monoclonal anti-*HER-2* antibody trastuzumab has been shown to improve prognosis (Ross and Fletcher, 1999). Emerging evidence of common amplicons in lung adenocarcinomas (Luk *et al.*, 2001; Jiang *et al.*, 2004; Tonon *et al.*, 2005) suggests that additional oncogenes remain to be identified; however, conventional techniques are ineffective in pinpointing such oncogenes. Parallel measurement of DNA copy number and mRNA levels in cDNA microarrays permits changes in copy number to be compared with transcription levels on a gene-by-gene basis to generate lists of candidate genes within the defining amplicons (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, use of transcript patterns does not allow assessment of the expression of protein products or identification of proto-oncogenes. Another approach, identifying differentially expressed proteins by proteomic analysis and then comparing the proteins present with mRNA expression in cDNA microarrays from the same specimens, can clarify the extent to which changes in transcript patterns reflect changes in their cognate proteins and post-transcriptional mechanisms (Chen *et al.*, 2002), but this approach cannot be used to identify oncogenes driven by extensive increases of their gene copy number. Moreover, using individual microarrays or proteomic approaches alone cannot distinguish the cancer-driving oncogenes that directly propel tumor progression from the larger number of passenger genes that may be concurrently over-represented but are not biologically relevant in tumor development.

Correspondence: Assistant Professor F Jiang, Department of Pathology, The University of Maryland School of Medicine, 10 South Pine Street, MSTF 7th floor, Baltimore, MD 21201-1192, USA.  
E-mail: fjiang@som.umaryland.edu  
Received 27 July 2005; revised 19 October 2005; accepted 27 October 2005

In this study, we used a comprehensive approach that integrated simultaneous comparative genomic hybridization (CGH) and transcript microarray with proteomic analyses of six lung adenocarcinoma cell lines. We directly and specifically identified four putative oncogenes that could have been activated through amplification and consequent elevation of transcript expression. We used small interfering RNA (siRNA) to inhibit the expression of two of these four genes in the lung cancer cell lines, which further implicated them in oncogenesis. We then explored the clinical significance of these findings by assessing the expression of these two genes in tissue microarrays of human lung cancer specimens. Our findings underscore the power of integrated functional genomic analyses for identifying putative oncogenes in tumorigenesis; such activated genes could be useful as targets for diagnosis or therapy in lung cancer.

## Results

### *Simultaneous global genomic and transcript analyses identify 183 genes with increases in genomic copy numbers and transcript expression levels*

To identify genes in which increased DNA copy number might contribute to increased transcript in lung adenocarcinomas, first we used CGH with microarrays of six lung adenocarcinoma cell lines. We identified 587 genes showing increases in DNA copy number across all six cell lines (Supplementary Table 1S), which were distributed as 90 amplicons on all chromosomes except for chromosomes 13 and Y (Supplementary Table 2S). A subsequent transcript test with the identical arrays of the same cell lines revealed 275 genes that showed increased mRNA levels (Supplementary Table 3S). Using random permutation tests across all cancer cell lines, we identified 183 genes (31%) that showed elevated transcript levels from the 587 genes that were over-represented in the genome (Table 1), suggesting that elevated transcript levels of the 183 genes may reflect their genomic over-representation in the cancer cells. These findings are consistent with previous reports linking genomic changes with altered transcript patterns in breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, our finding that only 31% of the genes showing increased DNA copy numbers had cognate increases in transcript expression in lung adenocarcinomas is different from the overall rates of 40–60% reported for breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). This discordance may reflect methodologic differences between studies or biological differences between breast cancer and lung adenocarcinoma.

### *Proteomic analyses identify four genes for which protein abundance was associated with increases in the cognate gene and transcript levels*

Analysis of transcript patterns is insufficient for understanding the expression of protein products and the effect of genomic over-representation on the expression

of their cognate proteins. To extend these findings beyond genomic over-representation to expression of the protein products of those genes, we next assessed protein expression in the same cell lines by two-dimensional polyacrylamide gel electrophoresis (PAGE) and found that 42 different proteins, representing 42 individual genes, were significantly increased in the cancer cell lines (Table 2; Supplementary Figures 1S and 2S). Some of these proteins were identified as having multiple isoforms, and all individual isoforms exhibited increases in expression ranging from 4.6 to 12.8 times their expression in normal lung tissue cells. In comparing protein level of the 42 genes with changes in their cognate genomic and mRNA expression from the global microarray analyses, we found that four (9.5%) of those 42 genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – showed statistically significant correlations between elevated protein expression and increases in both copy number and mRNA expression (all  $r > 0.84$ ;  $P < 0.05$ ) (Table 2) in the cancer cell lines. These findings imply that the abundance of these four proteins is attributable to the amplification and consequent elevated transcription of their cognate genes.

### *Validation of copy number, transcript, and protein expression of PRDX1, EEF1A2, CALR, and KCIP-1 in lung cancer cell lines*

To confirm our findings from the high-throughput analyses, we next used Southern, Northern, and Western blotting to assess DNA, RNA, and protein levels for the four genes identified in the six cell lines. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increase in protein level did not correlate with genetic changes. Overall, we found excellent concordance between the CGH microarray and Southern blotting analyses, transcript array and Northern blotting analyses, and proteomic and Western blotting analyses for all five genes (Figure 1). For example, *KCIP-1* showed fivefold amplification in five of the six cancer cell lines, whereas *NFKB1* showed no such increase in any of the cell lines. As for transcript expression, Northern blotting of *EEF1A2* showed high expression in five of the six cancer cell lines; again, levels of *NFKB1* transcript were not increased in any cancer cell line as compared with normal bronchial epithelial cells. The results of Western blotting were also consistent with the results of the proteomic experiments; for example, five of the cancer cell lines exhibited strong protein bands for *PRDX1* as compared with normal cells. These findings provide strong support for the validity of the results derived from the high-throughput techniques in this study.

These parallel analyses also revealed close correlations in the extent of changes in gene copies, transcript, and protein of each of the four genes in the cancer cell lines. For example, in the five cancer cell lines that showed at least fourfold increases in *EEF1A2* copy number, expression of transcript and protein was also increased by at least a factor of four as well (relative to their expression in normal cells) (Supplementary Figure 3S). The protein abundance of the four genes showing

**Table 1** List of 183 genes with statistically significant correlation (0.05) between genomic copy number and transcript level

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
ENO1	1	8.5	0.0085
DDOST	1	20.1	0.0111
SFN	1	26.4	0.0113
MLP	1	32.2	0.0114
AKR1A1	1	45.4	0.0128
PRDX1	1	45.4	0.0122
UQCRH	1	46.2	0.0125
RPL7	1	96.4	0.0127
COL11A1	1	102.6	0.0129
MCL1	1	147.3	0.0222
PSMB4	1	148.1	0.0131
JTB	1	150.7	0.0134
RPS27	1	150.7	0.0135
HAX1	1	151	0.0266
MUC1	1	151.9	0.0143
CCT3	1	153.1	0.0167
CRABP2	1	153.4	0.0148
TKT	1	159.3	0.0152
ATP1B1	1	165.8	0.0234
CHIT1	1	199.7	0.0154
SNRPE	1	200.2	0.0165
YWHAQ	2	9.6	0.0159
ODC1	2	10.60	0.0119
RPL31	2	101.20	0.0161
BENE	2	110.40	0.0169
STAT1	2	191.80	0.0175
HSPD1	2	198.30	0.0277
HSPE1	2	198.30	0.0185
RPL37A	2	217.30	0.0388
IGFBP2	2	217.50	0.0189
RPS7	2	3.30	0.0193
RAB1A	2	65.30	0.0204
IGKC	2	89.00	0.0285
LTF	3	46.3	0.0455
PFN2	3	151	0.0207
KPNA4	3	161.5	0.0211
S100P	4	6.7	0.1122
UGDH	4	39.3	0.0215
UCHL1	4	41.1	0.0222
SPP1	4	89.3	0.0227
TRIM2	4	154.7	0.0231
FGB	4	156	0.0235
FGG	4	156	0.0441
SDHA	5	0.251	0.0243
PDCD6	5	0.305	0.0245
CCT5	5	10.3	0.0446
PTPRF	5	14.2	0.0248
RPL37	5	40.8	0.0251
ENC1	5	74	0.0336
QP-C	5	132.2	0.0466
SPINK1	5	147.2	0.0256
CANX	5	179.2	0.0263
SOX4	6	21.7	0.0321
HDGF	6	22.6	0.0362
RPS10	6	34.6	0.0177
RPL10A	6	35.4	0.0369
VEGF	6	43.7	0.0372
OSF-2	6	45.4	0.0173
FSCN1	7	5.3	0.0378
CYCS	7	24.9	0.0381
CBX3	7	25.9	0.0289
IGFBP3	7	45.7	0.0389
CLDN4	7	72.7	0.0403
HSPB1	7	75.5	0.0433
CALR	7	92.7	0.0425
COL1A2	7	93.6	0.0457
ATP5J2	7	98.7	0.0475
AKR1B10	7	133.6	0.0481

**Table 1** (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
RPS20	8	56.7	0.0482
TCEB1	8	74.6	0.0486
LAPTM4B	8	98.5	0.0497
RPL30	8	98.7	0.0054
KCIP-1	8	101.6	0.0093
PABPC1	8	101.78	0.0119
EEF1D	8	144.4	0.0121
TSTA3	8	144.5	0.0122
RPL8	8	145.6	0.0128
TRA1	9	117.1	0.0136
RPL35	9	121.1	0.0133
HSPA5	9	121.5	0.0135
LCN2	9	124.4	0.0137
DPP7	9	133.4	0.0139
PFKP	10	3.2	0.0223
AKR1C1	10	5.1	0.0146
PLAU	10	75.6	0.0356
DSP	10	76.7	0.0289
TALDO1	11	0.434	0.0143
SLC22A1L	11	2.9	0.0151
TSSC3	11	2.9	0.0611
RPL27A	11	8.7	0.0156
ST5	11	8.8	0.0162
LDHA	11	18.5	0.0168
MDK	11	46.4	0.0162
DOC-1R	11	67.5	0.0167
MMP12	11	102.8	0.0177
HYOU1	11	118.9	0.0183
SCNN1A	12	6.3	0.0185
LDHB	12	21.7	0.0193
KRT7	12	52.3	0.0196
KRT5	12	52.6	0.0197
KRT6E	12	52.6	0.0201
ERBB3	12	56.2	0.0212
NACA	12	56.8	0.0218
TM4SF3	12	71.2	0.0401
NTS	12	86.2	0.0215
ASCL1	12	103.3	0.0219
TXNRD1	12	104.6	0.0223
CKAP4	12	106.6	0.0124
COX6A1	12	120.7	0.0435
BGN	12	122.5	0.0235
RAN	12	129.88	0.0238
RPL36A	14	48.1	0.0243
PGD	14	50.7	0.0248
THBS2	15	37.5	0.0251
TRAF4	15	38.3	0.0253
SPINT1	15	38.7	0.0254
RPL17	15	45.26	0.0411
PKM2	15	70.1	0.0258
IDH2	15	88.2	0.0211
RPL23A	16	0.377	0.0264
MSLN	16	0.753	0.0366
UBE2I	16	1.3	0.0271
RPS2	16	1.95	0.0281
CLDN9	16	3.1	0.0329
ARL6IP	16	18.7	0.0412
EIF3S8	16	28.3	0.0336
TUFM	16	28.9	0.0377
ALDOA	16	30.1	0.038
NME4	16	53.6	0.0381
GPR56	16	57.4	0.0386
CDH1	16	68.5	0.0289
NQO1	16	69.5	0.0396
SLC7A5	16	87.6	0.0397
APRT	16	88.6	0.0411
GALNS	16	88.6	0.0255
RPL13	16	89.3	0.0431
MCP	17	32.4	0.0465

Table 1 (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
ERBB2	17	35.11	0.0483
JUP	17	39.8	0.0495
CRF	17	40.39	0.0505
RPL27	17	41.1	0.0046
NME1	17	46.59	0.0082
COL1A1	17	48.6	0.0108
ABCC3	17	49.1	0.0326
NME2	17	49.6	0.0111
RPL38	17	72.7	0.0117
SMT3H2	17	73.6	0.0119
SYNGR2	17	76.6	0.0122
LGALS3BP	17	77.4	0.0127
P4HB	17	80.3	0.0126
PPAP2C	19	0.221	0.0228
GPI	19	39.55	0.0145
HPN	19	40.2	0.0129
ZNF146	19	41.4	0.0131
SPINT2	19	43.4	0.0238
PSMD8	19	43.5	0.0132
YIF1P	19	43.5	0.0135
RPS16	19	44.6	0.0144
CEACAM5	19	46.9	0.0145
CEACAM6	19	46.9	0.0143
GIPR	19	50.8	0.0259
SNRPD2	19	50.9	0.0413
KDELR1	19	53.6	0.0152
RPL28	19	60.6	0.0156
RPS5	19	63.6	0.0267
TRIM28	19	63.7	0.0158
DAP	20	35.6	0.0166
TOP1	20	40.3	0.0172
UBE2C	20	45.1	0.0174
RPS21	20	61.6	0.0268
EEF1A2	20	62.8	0.0185
TFF3	21	42.6	0.0186
TFF1	21	42.7	0.0192
CSTB	21	44.1	0.0201
MIF	22	22.6	0.0202
XBP1	22	27.5	0.0204
PRDX4	X	22.9	0.0198
SYN1	X	46.3	0.0204
TIMP1	X	46.3	0.0209
PLP2	X	47.8	0.0212
MAGED1	X	50.3	0.0331
RPS4X	X	71	0.0124
SSR4	X	152.6	0.0232

corresponding increases in both DNA copy number and mRNA provides further evidence that these could be oncogenes, the activation of which is reflected by genomic amplification and consequent increases in transcript level in lung adenocarcinoma cell lines.

*Specific inhibition of EEF1A2 and KCIP-1 expression by siRNAs led to decreased cell proliferation and induction of apoptosis*

To further prove the oncogenic function of the identified genes in lung tumorigenesis, we used siRNAs to inhibit the endogenous expression of EEF1A2 and KCIP-1 protein in four lung cancer cell lines (H1563, H229, H522, and SK-LU). Transfection of the cancer cells with specific siRNAs reduced the level of EEF1A2 and KCIP-1 protein by 70–90% 48 h after transfection

(Supplementary Figure 4S). In contrast, EEF1A2 and KCIP-1 protein levels remained unchanged in mock-treated control cells and in cells transfected with a scrambled siRNA sequence. At 48 h after siRNA transfection, the percentage of proliferation of the transfected cancer cells was reduced to 15–30% as compared with 91–100% of cell proliferation of the same cell lines treated with PBS or scrambled siRNA (Supplementary Figure 5S). Apoptosis of siRNA-transfected cells was 27–34%, whereas only 4% of the same cell lines treated with PBS or scrambled siRNA showed apoptosis. These results strongly support an oncogenic role for the identified genes in lung cancer and confirm their potential usefulness as therapeutic targets for the disease.

*Amplification and protein expression of KCIP-1 and EEF1A2 in lung tissue*

To further validate these findings and to assess the possible clinical significance of the four potential putative oncogenes identified from the cell lines, we first applied fluorescence *in situ* hybridization and immunohistochemical analysis, in parallel, to commercially available human lung tissue microarrays (Ambion, Austin, TX, USA) to evaluate the status of two of these four genes in lung cancer tissue specimens. (Commercially available antibodies to PRDX1 or CALR were not suitable for use in immunohistochemical analysis when this report was written.) Overexpression of KCIP-1 and EEF1A2 protein in the tumors was concordant with amplification of the corresponding genes ( $P=0.0003$  for KCIP-1 and  $P=0.0011$  for EEF1A2). For example, 16 (35%) of the 46 lung adenocarcinomas in the microarray showed amplification of *KCIP-1*, and strong cytoplasmic staining for KCIP-1 protein was seen in 18 tumors (39%) (Figure 2). We next examined whether overexpression of these genes was associated with increased cell proliferation by analysing Ki-67 expression in contiguous sections of the tissue microarrays. Positive Ki-67 expression was found to correlate with positive expression of both KCIP-1 ( $P=0.02$ ) and EEF1A2 ( $P=0.01$ ). To extend these findings, we then studied 11 tissue microarray blocks comprising normal and tumor tissue specimens from 113 patients with pathologic stage I non-small-cell lung cancer who had undergone curative surgery (Wang *et al.*, 2005). Immunohistochemical analysis showed that EEF1A2 was expressed in 32 cases (28%) and KCIP-1 in 29 cases (26%). Univariate and multivariate Cox proportional hazards models were used to detect possible associations between EEF1A2 and KCIP-1 expression and clinicopathologic variables. Expression of EEF1A2 or KCIP-1 was associated with short overall survival time ( $P=0.0012$  for EEF1A2 and  $P=0.0026$  for KCIP-1) (Supplementary Figure 6S). Age at diagnosis, histologic type of cancer, degree of tumor differentiation, and smoking history were not associated with survival time.

Although only two genes were validated in the lung tissue microarrays (because available antibodies to the other two genes were not suitable for use in

**Table 2** Proteins showing significant overexpression in cancer cell lines relative to those in normal bronchial epithelial cell lines and their correlation coefficients with increased DNA copy number or mRNA values<sup>a</sup>

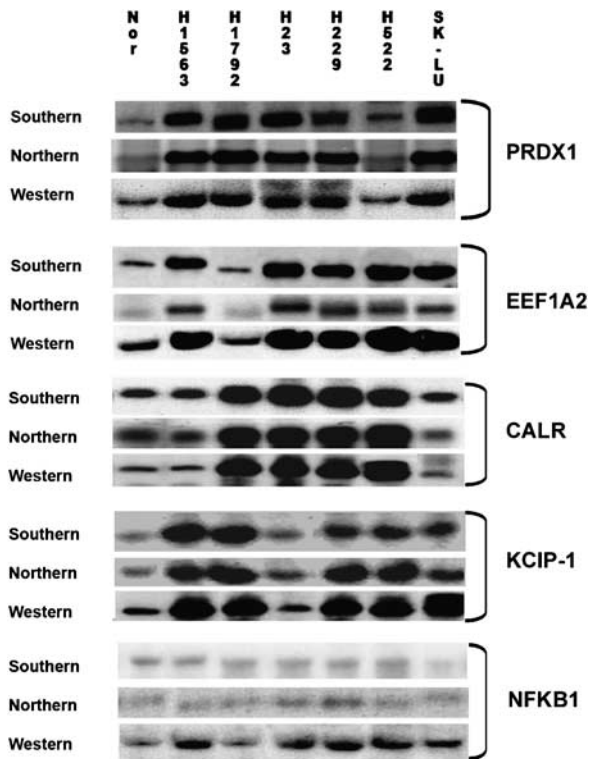
Acc. no.	Gene ID	Gene	Mw/pI	Description	r with genomic copy changes <sup>b</sup>	r with mRNA changes <sup>b</sup>
Q06830	5052	PRDX1	48.4/5.4	Peroxiredoxin 1	<b>0.92364</b>	<b>0.91892</b>
Q05639	1917	EEF1A2	50.5/5.7	Eukaryotic translation elongation factor 1 alpha 2	<b>0.90218</b>	<b>0.89456</b>
P27797	811	CALR	61/5.5	Calreticulin	<b>0.84128</b>	<b>0.86434</b>
P63104	7534	KCIP-1	27/6.5	Tyrosine 3-monooxygenase activation protein, zeta	<b>0.84467</b>	<b>0.85499</b>
P07237	5034	P4HB	54/6.2	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	<b>0.91884</b>	0.76786
Q04695	3872	KRT17	48.0/4.9	Keratin 17	0.00236	<b>0.86892</b>
P09211	2950	GSTP1	23.2/4.7	Glutathione S-transferase pi	<b>0.84218</b>	0.69456
P17936	3486	IGFBP-3	31.6/5.8	Insulin-like growth-factor binding protein 3	0.06412	0.16434
P26641	1937	EEF1G	50/6.4	Eukaryotic translation elongation factor 1 gamma	0.00446	<b>0.85549</b>
P08727	3880	KRT19	44.1/5.2	Keratin 19	-0.04884	<b>0.86786</b>
P04792	3315	HSPB1	22/6.5	Heat shock 27 kDa protein 1	0.00364	0.31892
P00558	5230	PGK1	44.5/4.2	Phosphoglycerate kinase 1	0.50402	0.79456
Q01995	5876	TAGLN	22.5/4.3	Transgelin	-0.34128	-0.26434
P08631	3055	JTK9	59.5/6.8	Hemopoietic cell kinase	-0.01446	0.02549
P09382	3956	LGALS1	16/5.5	Galectin-1, galactoside-binding, soluble, 1	0.026623	0.01123
Q92784	8110	DPF3	25.8/4.8	D4, zinc and double PHD fingers, family 3	0.094884	-0.03214
P54257	9001	HAP1	75.5/6.5	Huntington-associated protein 1	0.12364	-0.08108
P05783	3875	KRT18	48/5.3	Keratin 18	0.010218	0.60544
P05787	3856	KRT8	9.2/4.4	Keratin 8	0.041280	<b>0.84566</b>
P00738	3240	HP	55.2/6.2	Haptoglobin	0.044679	-0.14501
P09769	2268	FGR	59.5/5.2	Gardner-Rasheed feline sarcoma viral oncogene homolog	0.031264	-0.13789
P19838	4790	NFKB1	50.4/6.3	Nuclear factor of kappa light gene enhancer in B-cells 1	0.04467	-0.14501
P29034	6273	S100A2	10.9/4.6	S100 calcium-binding protein A2	<b>0.87964</b>	0.243214
Q13105	7709	ZBTB17	87.9/5.3	Zinc-finger and BTB domain containing 17	-0.17636	0.048108
Q00987	4193	MDM2	75.2/4.8	Transformed 3T3 cell double minute 2	-0.19782	-0.50544
P27816	4134	MAP4	111/5.4	Microtubule-associated protein 4	0.25872	-0.05356
P52732	3832	KIF11	119.2/6.2	Kinesin family member 11	-0.25778	-0.53444
P25205	4172	MCM3	90.9/5.5	Minichromosome maintenance deficient 3	0.25644	0.053666
P08631	3055	HCK	59.5/5.7	Hemopoietic cell kinase	0.65533	0.054501
P09237	4316	MMP7	22.6/5.8	Matrix metalloproteinase 7	0.234987	<b>0.876820</b>
P30305	994	CDC25B	64.9/4.5	Cell division cycle 25B	0.045116	0.283214
P50290	998	CDC42	21.3/6.1	Cell division cycle 42 (GTP-binding protein, 25 kDa)	-0.47636	0.088108
P61586	387	RHOA	19.8/6.9	Ras homolog gene family, member A	-0.49782	-0.00544
P63000	5879	RAC1	21.5/6.8	Ras-related C3 botulinum toxin substrate 1	-0.05583	-0.03566
P07437	203068	TUBB	49.6/6.5	Tubulin, beta polypeptide	0.255533	0.145010
P24864	898	CCNE1	47.1/4.3	Cyclin E1	-0.65116	0.232149
P04141	1437	CSF2	16.9/6.3	Colony stimulating factor 2 (granulocyte-macrophage)	-0.64636	-1.28108
P28072	5694	PSMB6	25.3/5.2	Proteasome (prosome, macropain) subunit, beta type, 6	-0.69782	-1.30544
P00352	216	ALDH1A1	54.7/4.3	Aldehyde dehydrogenase 1 family, member A1	-0.75872	0.03356
Q03013	2948	GTM4	25.3/5.0	Glutathione S-transferase M4	-0.78533	0.134501
P63241	1984	EIF5A	10/4.4	Eukaryotic translation initiation factor 5A	-0.97893	-1.44321
Q01469	2171	EFABP	18.0/4.2	Fatty acid-binding protein 5	0.25684	-0.36432

<sup>a</sup>Only the gene showing statistically significant increased protein expression with increases in both genomic copy number and transcript simultaneously will be considered as potential putative oncogene in lung adenocarcinoma cells. <sup>b</sup>r, Spearman correlation coefficients between proteins and genomic or mRNA values are based on all six cancer cell lines; bold indicates  $P < 0.05$ , if  $r > 0.84000$ . Mw, molecular weight; pI, isoelectric point.

immunohistochemical analysis), these findings are consistent with those from our cell lines, demonstrating again that genomic amplification and consequent increases in amounts of transcript may be, at least in part, driving the abundance of proteins in these lung tumors. The association between expression of these genes and that of Ki-67, a known indicator of poor prognosis in lung cancer (Martin *et al.*, 2004), suggests that activation of these genes may be an indicator of tumor aggressiveness. These results also suggest that expression of EEF1A2 and KCIP-1 proteins in stage I non-small-cell lung cancer may be useful as a marker for distinguishing patients with relatively poor prognosis from those who might benefit from adjuvant treatment.

## Discussion

Our current study illustrates the power of integrated functional genomic analyses for identifying putative oncogenes and for evaluating their potential clinical significance. Among the four identified oncogenes, three genes (*PRDX1*, *CALR*, and *KCIP-1*) have been implicated in lung tumorigenesis. *PRDX1* is an antioxidant protein involved in regulating cell proliferation, differentiation, and apoptosis. Kim *et al.* (2003) found *PRDX1* expression to be elevated in both lung cancer and adjacent normal lung tissue, suggesting that activation of *PRDX1* may enhance proliferation in lung cancer. *CALR* has a major role in  $Ca^{2+}$  binding and the



**Figure 1** Confirmation by Southern, Northern, and Western blot analyses of increased DNA copies, transcript levels, and protein levels in the four genes identified in high-throughput analyses. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increased protein level did not correlate with genetic changes. The blotting results are consistent with the results from the CGH array, transcript array, and proteomic analyses. Nor, indicates normal bronchial epithelial cell line. All the experiments were repeated at least three times with each cell line. Means of normalized to  $\beta$ -actin signal intensities on Southern, Northern, and Western blots, along with 95% confidence intervals, were calculated ( $\beta$ -actin signals are not shown in the figure; two different normal bronchial epithelial cell lines were used in the confirmation and only one normal cell line is shown in the figure).

transcriptional regulation of other genes and was recently found to be overexpressed in 73% of 40 lung adenocarcinomas (Oates and Edwards, 2000). *KCIP-1* belongs to the 14-3-3 family, which participates via the MAPK and Wnt signaling pathways in the regulation of many cellular processes including cell proliferation and differentiation as well as tumorigenesis (Thomas *et al.*, 2005). *KCIP-1* was recently found to be expressed in all 12 lung tumors tested in a single-institution study (Qi *et al.*, 2005). Interestingly, *EEF1A2* was originally considered a putative oncogene in ovarian cancer on the basis of its being amplified in 25% and overexpressed in 30% of the same set of ovarian tumors (Anand *et al.*, 2002); functional analyses have established its oncogenic role in cellular transformation (Lee, 2003). Our discovery that *EEF1A2* may be a putative oncogene in lung adenocarcinoma demonstrates the power of our functional genomic strategy for rapidly identifying potential oncogenes.

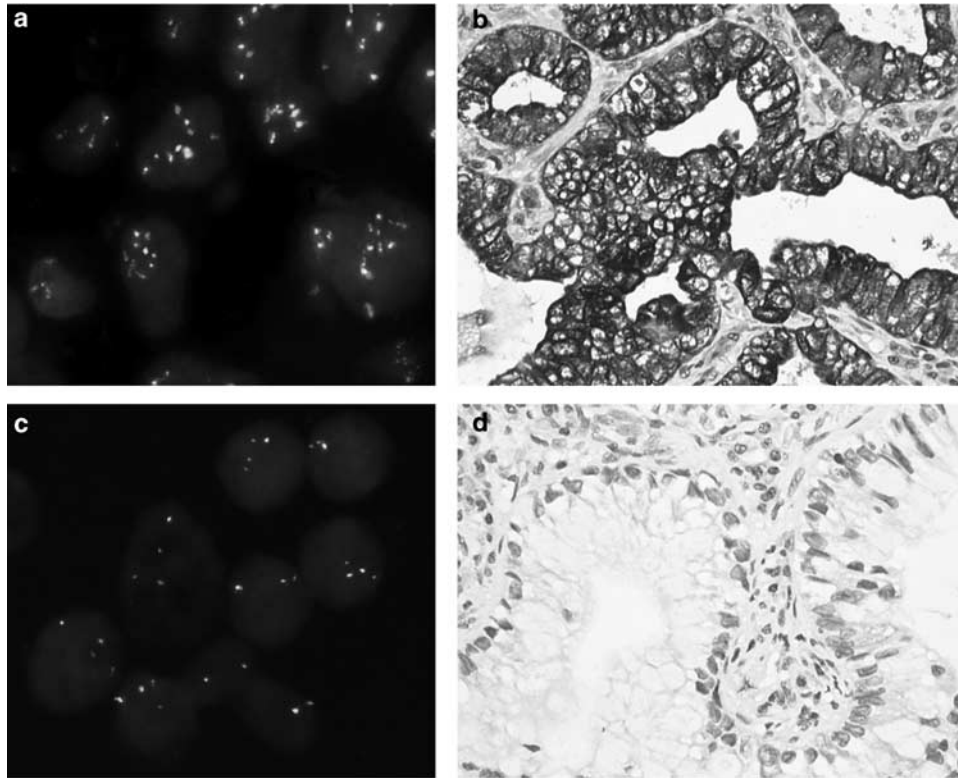
Although the main focus of this study was to specifically identify putative oncogenes, it should be

noted that 90.7% of the genes showing high protein expression did not show corresponding increases in both DNA copy number and transcript, a finding consistent with that of others that transcriptional, translational, and post-translational regulatory mechanisms can greatly influence the abundance of protein in lung tumorigenesis (Chen *et al.*, 2002). For example, *NFKB1* is a critical arbiter of immune responses, cell survival, and transformation and is often activated in several types of tumors (Chen *et al.*, 2002). Deregulation of *NFKB1* is thought to be modulated through phosphorylation of Ser337 by protein kinase A (Chen *et al.*, 2002). In our study, 68.8% of the genes showing over-representation in the genome did not show elevated transcript levels, implying that at least some of these genes are ‘passenger’ genes that are concurrently amplified because of their location with respect to amplicons but lack biological relevance in terms of the development of lung adenocarcinoma.

Although the potential oncogenes we identified here are likely to be important, certainly other oncogenes could be involved in the development of lung adenocarcinoma. The oligo microarray we used consists of 22 000 probes, which represent only about 60% of the human genome. Moreover, each probe was designed for the 3' region of expressed sequence tags of the selected genes. Also, our results were initially derived from cancer cell lines, although the findings were later confirmed in human tissue samples. Our ongoing study using microarrays with information on more genes and the development of high-resolution proteomic analyses for use with larger numbers of specimens will allow more comprehensive analyses of the molecular consequences of gene amplifications. Such expanded analyses will very likely lead to the identification of additional oncogenes.

Some of the results of our current study were comparable to those of other studies of lung cancer. For example, genomic copy number and protein levels of *KCIP-1* were previously found to be amplified and overexpressed in primary lung cancers by cDNA clone-based CGH array analysis (Jiang *et al.*, 2004) and proteomic analysis (Chen *et al.*, 2002), respectively. Our functional genomic approach, which integrates simultaneous CGH, transcript microarrays, proteomic analyses, and siRNA, allows us not only to quickly identify potential oncogenes but also to explore their significance as diagnostic and therapeutic targets in tumor progression – more than could be achieved by any technique alone.

Genes identified in this way may serve as promising targets for diagnosis and therapy in lung adenocarcinoma. Further research on the clinical implications of such genes is needed; experiments now underway in our laboratory include overexpression of the genes in normal cells, disruption of the function of these genes in cancer cells, and investigation of how interactions among these genes (or interactions with other known oncogenes) may mediate the expression of the transformed phenotype.



**Figure 2** *EEF1A2* amplification is associated with high *EEF1A2* protein expression in lung adenocarcinomas. (a) Cells from a lung adenocarcinoma sample in which *EEF1A2* is amplified show more green signals (*EEF1A2*) than red signals (chromosome 20 centromeric probe) (original magnification,  $\times 400$ ). (b) Immunohistochemical staining of cells from the same tissue sample as in panel a shows strong *EEF1A2* staining in the cytoplasm. (c) A lung adenocarcinoma sample with two copies of *EEF1A2* and chromosome 20 centromeric probe, indicating no *EEF1A2* amplification (original magnification,  $\times 400$ ). (d) Immunohistochemical staining of cells from the same tissue sample as in panel c shows negative staining for *EEF1A2*.

## Materials and methods

### Cell lines

Six human lung adenocarcinoma cell lines (H23, H229, H1792, SK-LU-1, H522, and H1563) were obtained from the American Type Culture Collection (Manassas, VA, USA). Two normal bronchial epithelial cell lines were obtained from Clontech (Palo Alto, CA, USA). Genomic DNA, mRNA, and protein were derived from a single harvest of these cells.

### DNA and RNA profiles by microarray analysis

Genomic DNA labeling and hybridization were performed as described previously (Barrett *et al.*, 2004) with Agilent's Human 1A Oligo Microarray (V2) (Agilent Technologies, Palo Alto, CA, USA), which contains 22 000 unique 60-mer oligos. Details of the protocol for analysing transcripts are available at <http://www.chem.agilent.com>. Map positions for arrayed genes were assigned by identifying the DNA sequence represented in the UniGene cluster and matching it with the Golden Path genome assembly (<http://genome.ucsc.edu/>; Mat 7, 2004 Freeze). Microarray images of DNA copy number and expression were analysed by using AgilentCGH Analytics and Feature Extraction software. DNA copy number profiles that deviated significantly from background signal ratios (measured from normal control cell hybridization, as described elsewhere; Barrett *et al.*, 2004) were interpreted as evidence of true differences in DNA copy number. The criteria for defining genomic over-representation and amplicons are described elsewhere (Hyman *et al.*, 2002); details are given in the

Supplementary Information. An increase in mRNA level was defined as a twofold increase in signal ratio relative to that of the control ( $\log_2 > 1$ ).

### Quantitative two-dimensional PAGE and mass spectrometry

Analysis of proteins by two-dimensional PAGE and their identification by mass spectrometry were performed as previously described (Shen *et al.*, 2004). Briefly, protein pellets were solubilized in rehydration buffer, after which the first-dimension isoelectric focusing was carried out with a Protean IEF Cell (Bio-Rad Laboratories) and the second-dimension separation was carried out with Bio-Rad's Ready Gel Precast Gels and the Bio-Rad Criterion Cell apparatus. Protein spots were visualized by silver-based staining, and all gels were assessed with Bio-Rad's PDQuest 2D gel image analysis software. Selected spots were subjected to in-gel tryptic digestion and analysed on a Voyager-DE PRO matrix-assisted laser desorption ionization/time-of-flight mass spectrometer (Applied Biosystems, Foster City, CA, USA). The mass list of the 20 most intense monoisotopic peaks for each sample was entered in the MS-Fit search program (v3.2.1) (<http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm>) and searched in the National Center for Biotechnology Information protein database.

### Southern, Northern, and Western blot analyses

Southern, Northern, and Western blot hybridizations were performed according to standard protocols. cDNA clones for the tested genes were purchased from Invitrogen (Carlsbad,

CA, USA) and prepared as probes for the blot hybridizations. Antibodies used were obtained as follows: PRDX1, CALR, NFKB1, KCIP-1, and  $\beta$ -actin from Santa Cruz Biotechnology (Santa Cruz, CA, USA); and EEF1A2 from Upstate Biotechnology (Waltham, MA, USA).

#### *Fluorescence in situ hybridization and immunohistochemical analyses of lung tissue microarrays*

Fluorescence *in situ* hybridizations and immunohistochemical analyses of KCIP-1 and EEF1A2 were carried out as described elsewhere (Jiang *et al.*, 2002; Wang *et al.*, 2005) with Lung Tissue Microarrays (Ambion, Austin, TX, USA) and 11 homemade microarray blocks containing tissue samples from 113 patients with pathologic stage I non-small-cell lung cancer (Wang *et al.*, 2005). DNA probes specific for *KCIP-1* and *EEF1A2* were obtained by screening a Human BAC Clone library (Invitrogen) by polymerase chain reaction as described previously (Jiang *et al.*, 2002). The antibodies used for the immunohistochemical analyses were the same as those used for the Western blotting. Cell proliferation of the lung tissues was assessed with a Ki-67 monoclonal antibody from Santa Cruz Biotechnology. Definitions of the cutoff value for a positive result of each antibody are shown in Supplementary Information.

#### *siRNA transfection, cellular proliferation assay, and apoptosis analysis*

Transfections were carried out by using siPORT Lipid Transfection Agent (Ambion) with siRNAs targeting KCIP-1 or EEF1A2 or with a scrambled siRNA duplex (siControl) (Dharmacon Inc., Lafayette, CO, USA), with PBS used as a negative control (Jiang *et al.*, 2002). Cells were fixed 24, 48, or 96 h later and subjected to further tests. All siRNAs were prepared by using a transcription-based method with Silencer siRNA according to the manufacturer's instructions (Ambion). Sequences of the individual siRNAs are listed in Supplementary Table 4S. Inhibition of cell growth by the

siRNAs was determined by MTT staining, and cell growth rate was plotted against the percentage of viable cells in the saline-treated controls (a value arbitrarily set at 100%) (Jiang *et al.*, 2002). Apoptosis was analysed by fluorescence cell cycle analysis of terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling with FITC-labeled dUTP (Boehringer Mannheim Biochemicals, Mannheim, Germany) (Jiang *et al.*, 2005).

#### *Statistical analyses*

Relationships between gene copy number and mRNA level were examined as described elsewhere (Hyman *et al.*, 2002, Supplementary Information). Correlations between protein abundance and DNA copy number and mRNA expression of the corresponding genes were evaluated with the Spearman correlation coefficient. Fisher's exact test and  $\chi^2$ -tests were used to analyse associations between amplification and expression of the candidate genes with various histopathologic variables of the samples in the tissue microarrays. Univariate and multivariate analyses were carried out with Cox's proportional hazards model to determine which independent factors might have a joint significant influence on survival. A *P*-value  $\leq 0.05$  was considered statistically significant; all statistical tests were based on a two-sided significance level.

#### **Acknowledgements**

This work was supported by National Institutes of Health Grant CA113707-01, an institutional research grant from The University of Texas MD Anderson Cancer Center, a Developmental Project/Career Development Award from The University of Texas Specialized Programs of Research Excellence in Lung Cancer P50 CA70907, and an M Keck Center for Cancer Gene Therapy Award (all to FJ). We thank Christine F Wogan of the Department of Scientific Publications for editorial review of this manuscript.

#### **References**

- Anand N, Murthy S, Amann G, Wernick M, Porter LA, Cukier IH *et al.* (2002). *Nat Genet* **31**: 301–305.
- Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R *et al.* (2004). *Proc Natl Acad Sci USA* **101**: 17765–17770.
- Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardina SL *et al.* (2002). *Mol Cell Proteomics* **1**: 304–313.
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E *et al.* (2002). *Cancer Res* **62**: 6240–6245.
- Jiang F, Caraway NP, Li RY, Katz RL. (2005). *Oncogene* **24**: 3409–3418.
- Jiang F, Lin F, Price R, Gu J, Medeiros LJ, Zhang HZ *et al.* (2002). *J Mol Diagn* **4**: 144–149.
- Jiang F, Yin Z, Caraway NP, Li R, Katz RL. (2004). *Neoplasia (New York)* **6**: 623–635.
- Kim HJ, Chae HZ, Kim YJ, Kim YH, Hwang TS, Park EM *et al.* (2003). *Cell Biol Toxicol* **19**: 285–298.
- Lee JM. (2003). *Reprod Biol Endocrinol* **1**: 69–73.
- Luk C, Tsao MS, Bayani J, Shepherd F, Squire JA. (2001). *Cancer Genet Cytogenet* **125**: 87–99.
- Martin B, Paesmans M, Mascaux C, Berghmans T, Lothaire P, Meert AP *et al.* (2004). *Br J Cancer* **91**: 2018–2025.
- Oates J, Edwards C. (2000). *Histopathology* **36**: 341–347.
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE *et al.* (2002). *Proc Natl Acad Sci USA* **99**: 12963–12968.
- Qi W, Liu X, Qiao D, Martinez JD. (2005). *Int J Cancer* **113**: 359–363.
- Ross JS, Fletcher JA. (1999). *Semin Cancer Biol* **9**: 125–138.
- Shen J, Person MD, Zhu J, Abbruzzese JL, Li D. (2004). *Cancer Res* **64**: 9018–9026.
- Thomas D, Guthridge M, Woodcock J, Lopez A. (2005). *Curr Top Dev Biol* **67**: 285–303.
- Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y *et al.* (2005). *Proc Natl Acad Sci USA* **102**: 9625–9630.
- Wang H, Zhang Z, Li R, Ang KK, Zhang H, Caraway NP *et al.* (2005). *Int J Cancer* **116**: 285–290.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>).