[CANCER RESEARCH 62, 3244–3250, June 1, 2002]

Laser Capture Microdissection and Microarray Expression Analysis of Lung Adenocarcinoma Reveals Tobacco Smoking- and Prognosis-related Molecular Profiles[1]

Koh Miura, Elise D. Bowman, Richard Simon, Amy C. Peng, Ana I. Robles, Raymond T. Jones, Toyomasa Katagiri, Ping He, Hiroki Mizukami, Lu Charboneau, Takefumi Kikuchi, Lance A. Liotta, Yusuke Nakamura, and Curtis C. Harris [2]

Laboratory of Human Carcinogenesis [K. M., E. D. B., A. I. R., P. H., C. C. H.], Biometric Research Branch, Division of Cancer Treatment and Diagnosis [R. S.] and Laboratory of Pathology [L. C., L. A. L.], National Cancer Institute, and Genetics of Development and Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases [H. M.], NIH, Bethesda, Maryland 20892; EMMES Corporation, Rockville, Maryland 20850 [A. C. P.]; Department of Pathology and Program in Oncology, University of Maryland, Baltimore, Maryland 21201 [R. T. J.]; and Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan [To. K., Ta. K., Y. N.]

ABSTRACT

Recent expression profile analyses revealed that lung adenocarcinomas can be divided into several subgroups with diverse pathological features. Because cellular heterogeneity of tumors can confound these analyses, we used laser capture microdissection and microarray expression analysis to characterize the molecular profiles of lung adenocarcinomas. We found 45 genes delineating smokers and nonsmokers that were located at chromosomal loci frequently altered in non-small cell lung cancers, and 27 genes, which were differentially expressed between survivors and nonsurvivors 5 years after surgery. These results are consistent with the hypothesis that the abnormal expression of genes involved in maintaining the mitotic spindle checkpoint and genomic stability, *e.g.*, *hBUB3*, *hZW10*, and *APC2*, contribute to the molecular pathogenesis and tumor progression of tobacco smoke-induced adenocarcinoma of the lung.

INTRODUCTION

Lung cancer is the leading cause of male and female cancer deaths in the United States. Lung carcinoma is classified usually as small cell lung carcinoma or NSCLC[3] (adenocarcinoma, squamous cell carcinoma, and large cell carcinoma). In particular, adenocarcinoma is the most common type of lung cancer in women and nonsmokers, and it is increasingly associated with tobacco smoking as well.[4] Since the 1950s, steady rises in the incidence of adenocarcinoma of the lung have been observed in many developed countries with ethnically diverse populations (1). Recent reports (2, 3) used the gene expression profiling to divide lung adenocarcinomas into several subgroups and to discriminate primary cancers from metastases of extrapulmonary origin. Lung adenocarcinomas, when compared with squamous cell lung carcinomas or small cell lung carcinomas, show striking differences in the expression patterns (2, 3). In the analysis of clinical tumor specimen in these and other studies, a significant confounder is the cellular heterogeneity of normal and diseased tissue. To overcome this problem, LCM was developed to analyze clinical samples (4–7). We combined the LCM with microarray gene expression analysis to identify the genes differentially expressed in lung adenocarcinoma associated with the following clinical phenotypic subgroups: prognosis, smoking, and gender, and to generate the hypotheses concerning the molecular pathogenesis and tumor progression of lung carcinoma.

MATERIALS AND METHODS

Patients and Surgically Resected Frozen Tissues. The study was approved by the Institutional Review Boards of the University of Maryland, Baltimore, and the National Cancer Institute. Three pathologists diagnosed the tumor tissues with adenocarcinomas (8). We chose lung adenocarcinoma cases with tumor stages I or II by Tumor-Node-Metastasis classification to minimize any secondary or tertiary effects related to tumor stages III or IV. In Table 1, we summarized the characteristics of patients. Among the 19 patients with adenocarcinoma, 6 survived 5 years or more after surgery, 12 were nonsurvivors that died of a recurrence of lung cancer, and we could not find the exact cause of death for 1 patient. After the classical criteria, patients were defined as nonsmokers if they smoked <100 cigarettes in their life time. Five were nonsmokers, and 14 were smokers, and the pack-years among 14 smokers were 20–125 (mean $\pm$ SD; 48 $\pm$ 27). Ten were males and 9 were females.

RNA Extraction, LCM, and T7 Amplification. From each flash-frozen tumor tissue, we prepared 8-μm thick frozen sections. Total RNA was extracted from one section, and the quality was evaluated (Fig. 1*A*). Serial frozen sections ($\sim$18/case) were used for the subsequent analysis. We used the PixCell II LCM System from Acturus Engineering (Mountain View, CA) for laser capture and followed the manufacturer's protocol with several modifications. The sections were immersed in the relevant fixatives or staining solutions (70%, 95%, or 100% ethanol, hematoxylin, eosin, or RNase-free water) for $\sim$10 s each, followed by dehydration with xylene for 1 min. We recorded the pathological images for each case (Fig. 1, *C–M*). We repeated reverse transcription and T7 amplification twice to get two-round aRNA as probes for microarray analysis as reported previously (5).

RNA Reference Pool. As an experimental control, we used the BEAS-2B cell line that was isolated from normal human bronchial epithelium obtained from an autopsy of a noncancerous individual, which was immortalized with an adenovirus 12-SV40 virus hybrid (9).

Microarray and Acquisition of Data. We used cDNA microarray slides generated by Amersham Pharmacia Biotech (Piscataway, NJ) and fabricated in Laboratory of Molecular Medicine, University of Tokyo. Four sets of array slides contained 18,432 cDNAs selected from the UniGene database with 15,737 unique genes. The cDNAs were amplified by reverse transcription-PCR without any repetitive or poly(A) sequences. The PCR products were spotted in duplicate using an Array Spotter Generation III (Amersham). Each slide commonly contained 48 housekeeping genes. We labeled 2.5 μg of two-round aRNA from LCM-captured tissues or BEAS-2B by reverse transcription with Cy3-dCTP or Cy5-dCTP, respectively. Probes were hybridized to the microarrays in the Automated Slide Processor (Amersham), and signal intensities were quantified using ArrayVision software (Imaging Research, Inc., St. Catharines, Ontario, Canada). Four sets of signal intensities (Cy3 or Cy5, left panel or right panel) from duplicate spots were transformed to log2 scale, and the duplicate log-ratios were averaged. We excluded clones for which the Cy3:Cy5 ratios for duplicate spots differed by more than 2-fold. The log-ratios of nonexcluded clones were normalized by a median centering the 48 housekeeping genes on each array. We selected 6,216 clones with

Received 2/4/02; accepted 4/4/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported by Japan Society for the Promotion of Science Research Fellowship for Japanese Biomedical and Behavioral Researchers at NIH, 69811 (to K. M.).

[2] To whom requests for reprints should be addressed, at Laboratory of Human Carcinogenesis, National Cancer Institute, NIH, 37 Convent Drive, Building 37, Bethesda, MD 20892. Phone: (301) 496-2048; Fax: (301) 496-0497; E-mail: Curtis_Harris@nih.gov.

[3] The abbreviations used are: NSCLC, non-small cell lung carcinoma; hBUB3, human BUB3; LCM, laser capture microdissection; T7 amplification, T7-based RNA amplification; aRNA, amplified antisense RNA; LOH, loss of heterozygosity; TK2, thymidine kinase-2; hZW10, human ZW10; APC/C, anaphase-promoting complex or the cyclosome.

[4] Internet address: http://www.cdc.gov/tobacco/sgr_forwomen.htm.

Table 1 *Characteristics of patients*

ID	Age	Gender	Prognosis[a]	Smoking	Pack-years
1	75	M	D[b]	N	0
2	54	F	A	Y	40
3	48	M	A	Y	50
4	66	F	D	Y	47
5	74	M	D	Y	60
6	59	M	D	Y	125
7	72	F	D	Y	60
8	53	F	D	Y	20
9	55	M	A	Y	80
10	66	F	D	Y	40
11	49	M	D	Y	25
12	73	M	D	N	0
13	78	F	D	Y	60
14	61	M	D	Y	46
15	70	M	Unknown	Y	60
16	57	F	A	N	0
17	38	F	A	N	0
18	63	F	A	N	0
19	63	M	D	Y	30

[a] The status 5 years after surgery.
[b] D, dead; A, alive; N, no; Y, yes.

consistent duplicate measurements on at least 15 of the 19 arrays for additional analysis. The analyses were performed using BRB-ArrayTools.[5]

Hierarchical Clustering. We selected 2270 clones with consistent log-ratios present in at least 15 of the 19 samples, and which showed more variant expression patterns among the 19 samples. Then, we performed average linkage hierarchical clustering for 19 samples and 2270 clones with centered correlation as the distance metric (10).

Genes That Distinguish Phenotype Classes. Among 19 cancer cases, to identify genes that discriminate between two phenotype subclasses (*e.g.*, smoker *versus* nonsmoker), the F test with 0.0025 as a nominal significance level (P) was computed on log2-expression ratios of each clone. Next, phenotype labels were randomly permuted among the 19 samples, the F statistic for each gene in the permuted data set was recomputed, and the number of clones significant at the nominal 0.0025 level was counted. This process was repeated 2000 times, and the proportion of the 2000 replications with at least as many genes significant at the 0.0025 level, provided a global test of the null hypothesis that the expression profile is unrelated to phenotype class.

RESULTS

Following the strategy described in Fig. 1*A*, we recorded the tumor pathology (Fig. 1*C*), and the number of sections and spots used for LCM, and calculated the yield of aRNA after two rounds of amplification (Fig. 1*B*). Within the context of hierarchical clustering among 19 samples and 2270 clones (Fig. 2), 4 of the adenocarcinomas (cases 3, 8, 9, and 15) that clustered separately were diffusely invasive types and the population of cancer cells in LCM-captured tissues was relatively low.

When we performed the F test ($P = 0.0025$) for smokers and nonsmokers, 45 candidate genes were identified. The probability of obtaining 45 or more such genes in random partitions of the 19 expression profiles into two groups of 14 and 5 samples was only 0.045 (Table 2; Fig. 3*A*). Among the significant genes, 30 were known genes and 15 were ESTs or genes of unknown function. To compare the expression levels of genes on the chromosomal band 3p21.3 region between 5 nonsmokers and 14 smokers, we selected 21 known genes that were reported to be in this region and passed the same filtering as our other analyses (AbsLogRatioDifference <1, missing value <4). For those genes, we calculated the fold-difference in geometric means of 14 smokers divided by 5 nonsmokers (Table 3).

Because the exact cause of death for case 15 was unknown, we eliminated this case from the molecular profiling of genes related to prognosis. Using the 18 tumor cases, we performed the univariate F test between survivors and nonsurvivors 5 years after surgery. We found 27 genes differentially expressed between survivors and nonsurvivors with $P < 0.0025$ (Table 4; Fig. 3*B*). Under the null hypothesis of no relationship between the expression profile and survival, the probability of obtaining at least 27 genes significant at the 0.0025 level individually is 0.07. Among the significant genes, 18 were known genes and 9 were unknown genes.

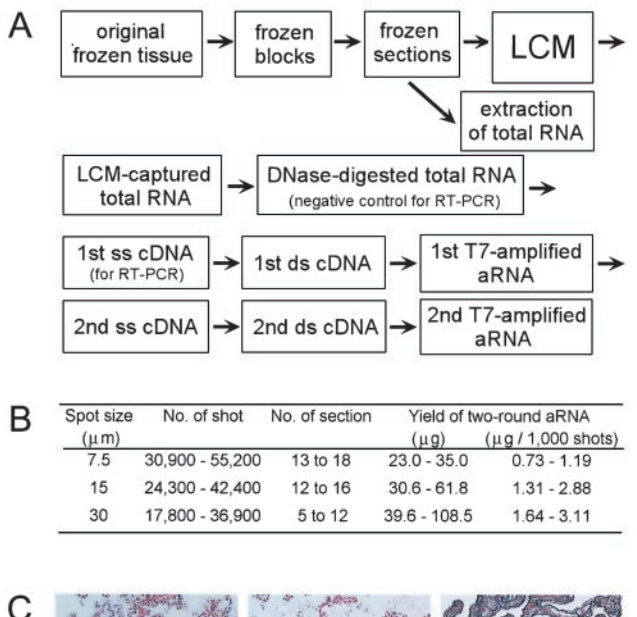
Additionally, to show the reliability of the F test and to estimate the number of false positives with our experimental set, we divided our 19 cases into 10 pairs of random subgroups and repeated the F test. With these random sets, we did not get any significant results, and the number of false positives ranged from 6 to 10 ($P = 0.0025$). We also used the F test to identify genes that classify gender or specific morphological features, but we have not been able to find a definitive result with our current experimental set.

DISCUSSION

Analysis of clinical samples is difficult because of the heterogeneity of cellular components in the tissues and their diverse pathological features. In analyzing genetic alterations or expression profiles in cancer cells, the contamination of normal epithelia or stromal cells may confound the analysis, especially when using tissues such as lung, brain, kidney, ovary, mammary gland, or prostate (5, 7, 11, 12), which are unlike the more homogenous fluorescence-activated cell sorted blood cells (13). The use of LCM should improve the sample preparation for microarray expression analysis (4). One disadvantage of microdissection is that RNA amplification is essential for the preparation of an adequate amount of RNA for probe generation. The reproducibility and usefulness of T7 amplification in gene expression profiling has been reported previously (14, 15). In our hands, the gene expression profile from two independent two-rounds of amplification of BEAS-2B RNA was highly correlated ($r^2 = 0.91$; data not shown).

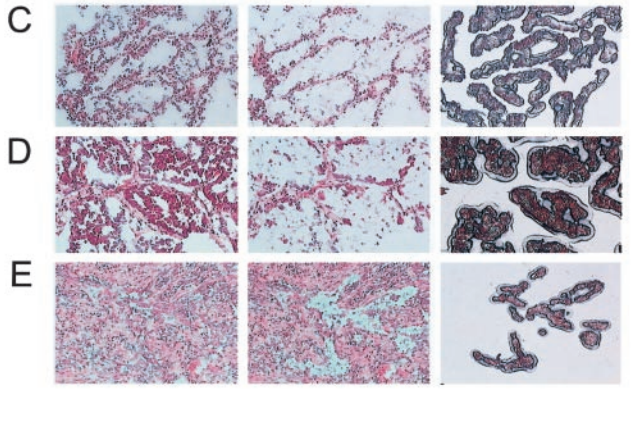
Several of the genes with lower expression among smokers are located in the chromosomal regions where genomic imbalance in NSCLC has been observed previously with high frequency. In Table 2, we summarized the reported data of genomic imbalance including homozygous deletion, which was analyzed with comparative genomic hybridization, LOH analysis, and other methods. For example, chromosome 3p21.3 is a well-known region for a frequent homozygous deletion in lung cancers. Lerman and Minna (16) analyzed 25 genes localized in the 630-kb homozygous deletion region on 3p21.3. Using the F test for smoking, we found two genes, *101F6* and *CACT*, located on 3p21 region. In addition to 3p21.3, 11q23–24, where *NCAM1* is also located, is known for a high frequency of LOH, and putative tumor suppressor genes may be found in this region (17). In addition, a high frequency of LOH in lung cancers has been reported in chromosomes 19p12–13, 19q13.3, and 4q, where *C19ORF3*, *EDG4*, *SIGLEC5*, and *RRH* are located (18, 19). Mori *et al.* (20) showed recently that microsatellite instability was observed frequently in the coding region of *PA2G4* on chromosome 12q13. The inactivation of these genes may be related to tobacco carcinogenesis. The high expression of *RAB4*, *DJ1*, *MCT*, and ribosomal protein L22 (RPL22) also may be related to the tobacco smoke-induced adenocarcinoma. *RAB4* gene shares biochemical properties with the *Ras* gene superfamily and encodes GTP-binding proteins. DJ1 cooperates with H-RAS in transforming NIH3T3 cells (21). The MCT protein is a novel candidate oncogene sharing a homology with cyclin H, which increases G_1 cyclin/cyclin-dependent kinase activity (22). DNA amplification of chromosome 3q26.1-q26.3 is detected frequently in lung squamous cell carcinomas, where *RPL22* is also located (23).

[5] Simon, R., and Peng, A. BRB-ArrayTools Users Guide. National Cancer Institute. Internet address: http://linus.nci.nih.gov/BRB-ArrayTools.html.



B

Spot size (μm)	No. of shot	No. of section	Yield of two-round aRNA (μg)	(μg / 1,000 shots)
7.5	30,900 - 55,200	13 to 18	23.0 - 35.0	0.73 - 1.19
15	24,300 - 42,400	12 to 16	30.6 - 61.8	1.31 - 2.88
30	17,800 - 36,900	5 to 12	39.6 - 108.5	1.64 - 3.11



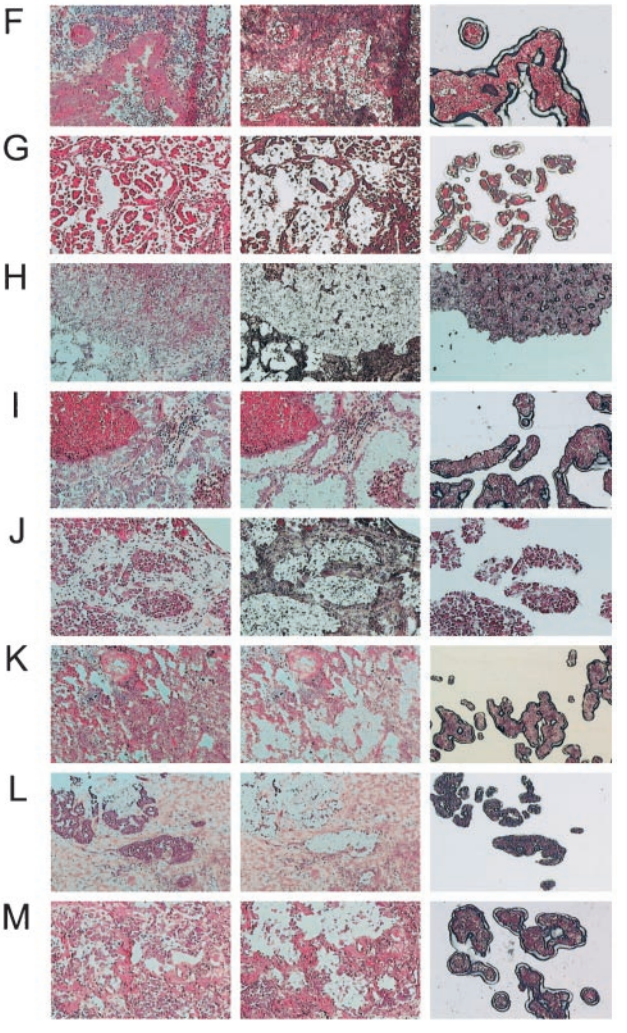


Fig. 1. Experimental course of LCM and T7 amplification. *A*, flow chart of tissue preparation, LCM, and two-round T7 amplification. ss cDNA, single-strand cDNA; ds cDNA, double-strand cDNA. *B*, the summary of LCM and two-round aRNA yields. *C–M*, pathological images of LCM. *Left panels*, before LCM; *middle panels*, after LCM; and *right panels*, the tissues captured on LCM caps. *C*, case 1; *D*, case 2; *E*, case 3; *F*, case 5; *G*, case 7; *H*, case 8; *I*, case 11; *J*, case12; *K*, case 15; *L*, case 18; and *M*, case 19.

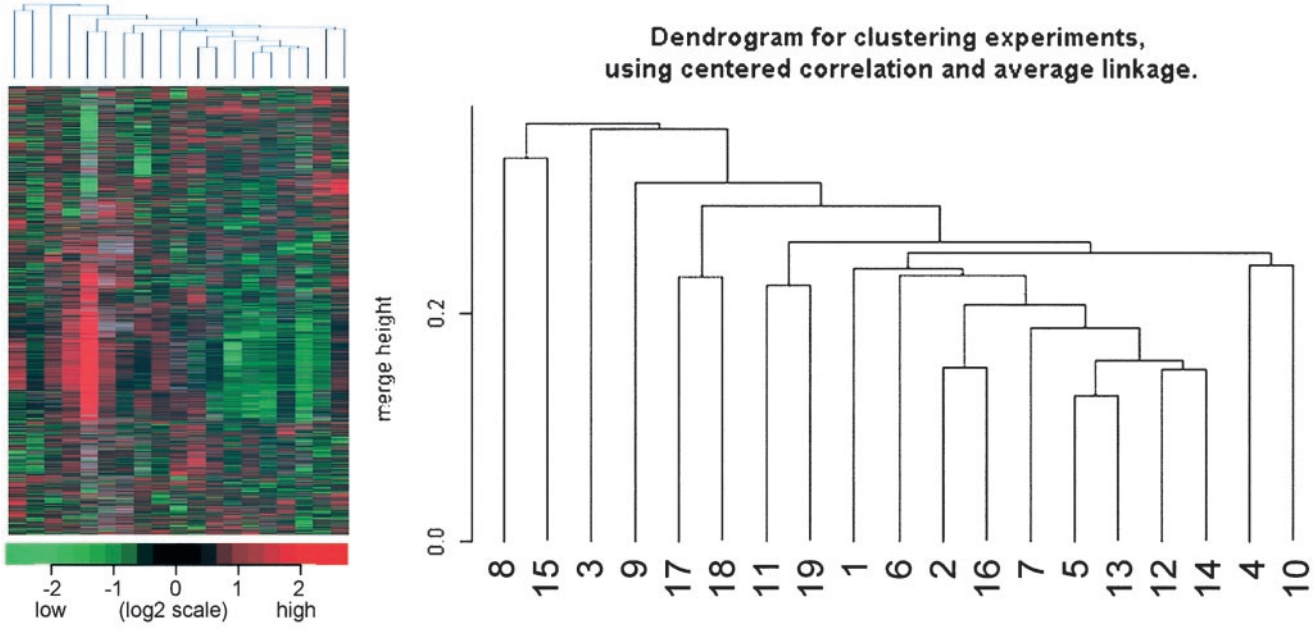


Fig. 2. A hierarchical clustering and gene expression cluster view of 19 samples and 2270 cDNA clones. *Green*, transcript levels below the median; *black*, equal to the median; *red*, greater than median; and *gray*, missing data for filtering.

Table 2 *Forty-five genes that classify smokers and nonsmokers and the information of genomic imbalance*

ID	Description	*P*	GenBank accession no.	Cytoband	Genomic imbalance at cytoband in NSCLC[a]	References
High in nonsmokers and low in smokers						
1	thymidine kinase 2, mitochondrial	0.00004	AA429631	16q22	gain(16q22, 45%)	19
2	putative tumor suppressor, 101F6	0.00010	AC002481	3p21.3	homozygous deletion(3p21.3), loss(3cen-p26, 95%), loss (3cen-p26, >98%)	16, 37, 38
3	UDP-Gal:βGlcNAc β 1,3-galactosyltransferase, polypeptide 4	0.00034	AA987754	6p21.3	loss(6p21.3, 42–69%), gain(6p11-p22, 27%), loss(6cen-p25, 100%)	18, 19, 37
4	integrin, α 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41B	0.00081	M34480	17q21.32	gain(17q11-q24, 82%), gain(17q21, 37%), loss(17q21, 53%), loss(42%)	19, 39–41
5	ESTs	0.00113	H24302			
6	neural cell adhesion molecule 1	0.00122	S71824	11q23.1	loss(11q23-q24, 71%), loss(11q23.2, 54%), loss(11q23)	42–44
7	retinal pigment epithelium-derived rhodopsin homolog, RRH	0.00122	AF012270	4cen-4q12	loss(4q33-q35, 47%), loss(43%), loss(4cen-q35, 100%)	18, 38, 45
8	SELENOPHOSPHATE SYNTHETASE; Human selenium donor protein	0.00129	U34044	Xq13	loss(Xp-q21, 67%)	45
9	sialic acid binding Ig-like lectin 5, SIGLEC5	0.00191	U71383	19q13.3	loss(19q13.3, 39%), loss(19cen-q13.4, 73%), gain(19cen-q13.4, 39%)	18, 19, 46
10	EST	0.00200	AA663323			
11	cytochrome P450, subfamily IID, polypeptide 7a	0.00224	X16866	22q13.2-q13.31	loss(22q13-qter, 57%), loss(22q11.2-q13.1, 75%)	18, 47
12	proliferation-associated 2G4 38kD, PA2G4	0.00227	U59435	12q13	loss(12q24.1, 63%)	47
13	carnitine/acylcarnitine translocase, CACT	0.00228	AA521247	3p21.31	homozygous deletion(3p21.3), loss(3cen-p26, 95%), loss(3cen-p26, >98%)	16, 37, 38
14	hypothetical protein FLJ23443	0.00230	W94201	3p22-p21.33	homozygous deletion(3p21.3), loss(3cen-p26, 95%), loss(3cen-p26, >98%)	16, 37, 38
15	ESTs	0.00234	AA532514			
16	chromosome 19 open reading frame 3, C19ORF3	0.00242	AF028824	19p13.1	loss(19p13.2, 75–83%), gain(82%), loss(19p12-p13.2, 77%), loss(19p13.3, 58%)	18, 19, 45, 48
17	EDG4	0.00243	AF011466	19p12	loss(19p13.2, 75–83%), gain(82%), loss(19p12-p13.2, 77%), loss(19p13.3, 58%)	18, 19, 45, 48
High in smokers and low in nonsmokers						
18	maternal G10 transcript	0.00025	S77329	7q22.1		
19	MCT protein	0.00033	AI336165	Xq22-q24	gain(Xq22-q25, 45%), homozygous deletion(Xq22, 67%)	19, 45
20	3-hydroxyisobutyrate dehydrogenase	0.00041	W95267	7p11.2-p21		
21	RAB4, member RAS oncogene family	0.00048	F05353	1q42-q43	gain(1q21.1-q44, 100%), gain(43%)	38, 46
22	ribosomal protein L22, RPL22	0.00055	AA962580	3q26	gain(3q24-qter, 91%), gain(3cen-q29, 35%), gain(3cen-q29, 35%)	19, 37, 46
23	destrin (actin depolymerizing factor)	0.00059	S65738	20p11.23		
24	ESTs	0.00069	W23535			
25	TATA-binding protein-binding protein	0.00072	N64689	6p21.33	loss(6p21.3, 42–69%), gain(6p11-p22, 27%), loss(6cen-p25, 100%)	18, 19, 37
26	TERF1 (TRF1)-interacting nuclear factor 2	0.00078	AA041364	14q12-q21.3	loss(14q12, 31%), loss(14q21, 18%), loss(14q11-q12, 38%)	18, 19, 40
27	ESTs	0.00096	AA478650			
28	tubulin, α, ubiquitous	0.00106	AA492218	12q12-q14.3	loss(12q24.1, 63%)	47
29	ESTs	0.00116	H95968			
30	Homo sapiens cDNA: FLJ22712 fis, clone HSI13435	0.00128	AI192127			
31	Homo sapiens cDNA: FLJ21880 fis, clone HEP02743	0.00149	T71056	9q34.13	loss(90%), loss(9q34, 50%), loss(40%)	37, 38, 45
32	NADH dehydrogenase (ubiquinone) Fe-S protein 4 (18kD)	0.00178	AI131435	5q11.1	loss(5q21, 35%), loss(5cen-q35, 100%)	38, 45
33	ESTs	0.00180	AI206756			
34	ESTs	0.00180	N54851			
35	glioma tumor suppressor candidate region gene 2, GLTSCR2	0.00182	AA459728	19q13.3	loss(19cen-q13.4, 73%), gain(19cen-q13.4, 39%), loss(19q13.3, 39%)	18, 19, 46
36	hypothetical protein FLJ20585	0.00189	AI336134	2p14-q23.3		
37	general transcription factor IIA, 2 (12kD subunit)	0.00192	U21242	15q11.25-q22.33		
38	RNA-binding protein regulatory subunit, DJ1	0.00196	AA576779	1p36.33-p36.12	loss(1p36, 40%)	38
39	ASB protein	0.00196	H40863	2q37	loss(2q33, 31%), loss(2q36-q37, 27%), loss(2q32, 63%)	18, 19, 47
40	vesicle-associated soluble NSF attachment protein receptor	0.00211	AA777761	14q23.3		
41	DKFZP547E1010 protein	0.00212	AI039693	1p36.1	loss(10q23-q25, 43%), loss(1p36-pter, 64%)	18, 45
42	hypothetical protein FLJ20027	0.00228	AA679314	10pter-q21.2	loss(10p11-q11, 36%), loss(10cen-q25, 100%)	38, 45
43	signal peptidase complex (18kD)	0.00228	AA742432	15q11.2	loss(15q11, 50%), gain(15cen-q26, 52%)	45, 46
44	ESTs	0.00232	AI276082			
45	CGI0 protein	0.00246	AA889621	14q22.1-q22.3		

[a] If the minimal region and the percentage of genomic imbalance involvement were known, they were written in the parentheses. The data of references 37 and 38 was analyzed with adenocarcinomas only. The data of references 18, 40–43, 45, 47, and 48 was analyzed with LOH analysis with polymorphic markers. The data of references 19, 37, 38, and 46 was analyzed with comparative genomic hybridization.

Decreased expression of glioma tumor suppressor candidate region gene 2 (GLTSCR2) and high expression of candidate oncogene *EDG4* among nonsmokers may contribute to the carcinogenesis of nonsmokers. The mitochondrial enzyme TK2 is also higher expressed in adenocarcinomas in nonsmokers than in smokers. TK2 is active throughout the cell cycle and is most likely correlated with mitochondrial content in the tumors. TK2 contributes to the metabolic activation of 2′,2′-difluorodeoxycytidine (gemcitabine), which has significant activity against non-small cell cancers (24). These data suggest the hypothesis that gemcitabine may be more active in the TK2 higher-expressing adenocarcinoma in nonsmokers when compared with smokers.

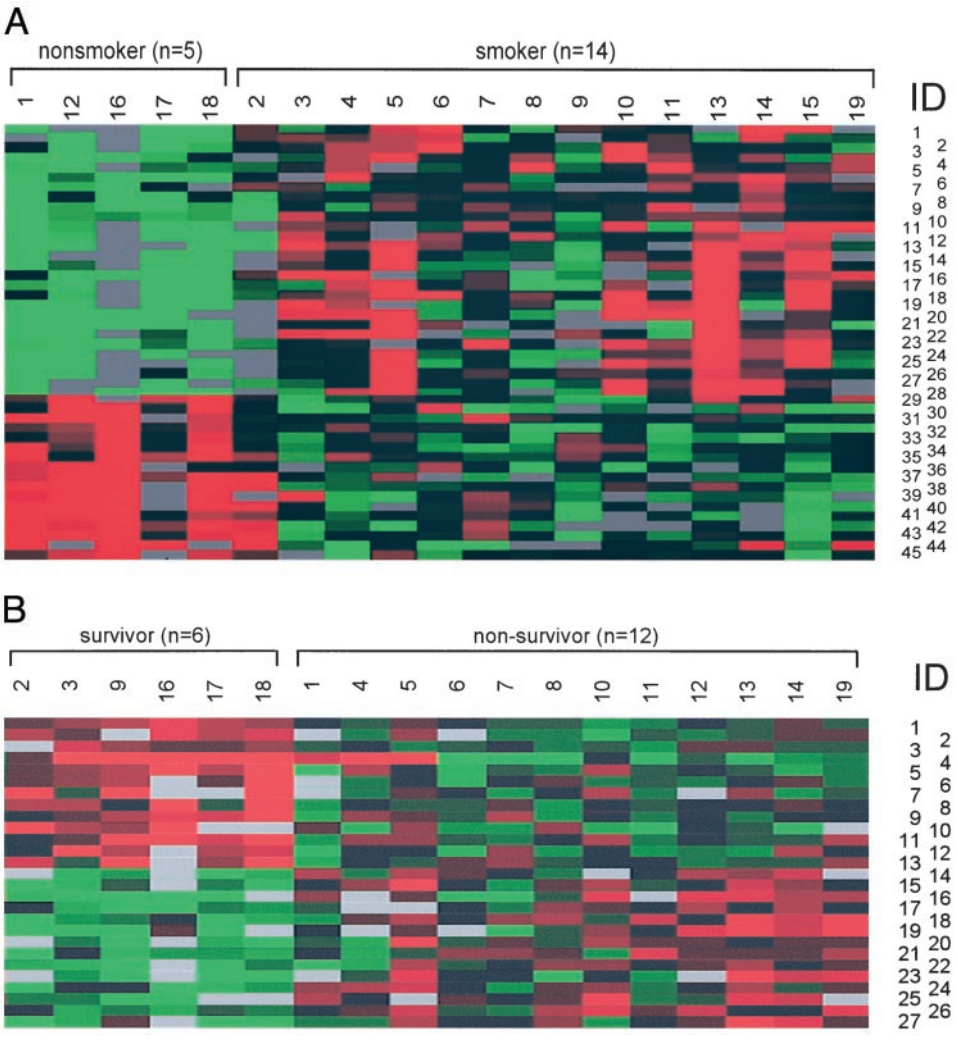


Fig. 3. *A*, an image plot view of 45 genes listed in Table 2. ID numbers of genes are identical to Table 2. *Green*, transcript levels below the median; *black*, equal to the median; *red*, greater than median; and *gray*, missing data for filtering. *B*, an image plot view of 27 genes listed in Table 4. ID numbers of genes are identical to Table 4. *Green*, transcript levels below the median; *black*, equal to the median; *red*, greater than median; and *gray*, missing data for filtering.

Among the 21 genes on chromosome 3p21.3 listed in Table 3, 16 genes including *PL6*, *101F6*, *HYAL2*, *FUS2*, and *RBM6*, which are located in 630 kb with a high frequency of homozygous deletions (16), showed lower expression in smokers. This result supported our hypothesis that genes with lower expression among smokers are located in the chromosomal regions with a high frequency of LOH, including 3p21.3.

Among the 27 genes that were differentially expressed between

Table 3 *Genes on chromosomal band 3p21.3 and the fold-difference of their expression levels*

Fold[a]	Description of genes on 3p21.3	GenBank accession no.	Symbol
High in nonsmokers and low in smokers			
0.32	protein tyrosine phosphatase, non-receptor type 23	C75369	PTPN23
0.46	carnitine/acylcarnitine translocase	AA521247	CACT
0.48	g20 protein	AI300189	LOC51161
0.48	hypothetical protein FLJ23443	W94201	FLJ23443
0.51	putative tumor suppressor	AC002481	101F6
0.56	ubiquitin specific protease 4 (proto-oncogene)	U20657	USP4
0.60	endonuclease G-like 1	AA743829	ENDOGL1
0.68	mitogen-activated protein kinase-activated protein kinase 3	U09578	MAPKAPK3
0.68	PL6 protein	U09584	PL6
0.69	hyaluronoglucosaminidase 2	U09577	HYAL2
0.76	glutathione peroxidase 1	M21304	GPX1
0.77	putative tumor suppressor	AF040704	101F6
0.82	putative tumor suppressor	AF040705	FUS2
0.86	ubiquitin specific protease 4 (proto-oncogene)	AA255621	USP4
0.87	ras homolog gene family, member A	L25080	ARHA
0.88	interferon-related developmental regulator 2	U09585	IFRD2
0.91	RNA binding motif protein 6	AF069517	RBM6
High in smokers and low in nonsmokers			
1.16	HYA22 protein	D88153	HYA22
1.31	SH3 protein	AA702730	AF3P21
1.56	laminin receptor 1 (67kD, ribosomal protein SA)	J03799	LAMR1
2.04	villin-like	D88154	VILL

[a] Fold-difference in geometric means of 14 smokers divided by 5 nonsmokers.

Table 4 *Twenty-seven genes that classify survivors and nonsurvivors, and the information of genomic imbalance*

ID	Description	*P*	GenBank accession no.	Cytoband	Genomic imbalance at cytoband in NSCLC[a]	References
High in survivors and low in nonsurvivors						
1	BUB3 (budding uninhibited by benzimidazoles 3, yeast) homolog	0.00011	AF047472	10q26	loss(10q26.13-qter, 36%)	45
2	ZW10 (Drosophila) homolog, centromere/kinetochore protein	0.00054	U54996	11q23.2	loss(11q23-q24, 71%), loss(11q23.2, 54%), loss(11q23)	42–44
3	dCMP deaminase	0.00109	L39874	4q35.1	loss(4q31-qter, 64%), loss(4q34-q35, 50%), loss(4q32-qter, 100%)	19, 38, 45
4	keratin 9 (epidermolytic palmoplantar keratoderma)	0.00113	Z29074	17q21.1-q21.2	gain(17q11-q24, 82%), gain(17q21, 37%), loss(17q21, 53%), loss(42%)	19, 39–41
5	KIAA1008 protein	0.00145	AA156488	13q21.23-q22.2	loss(13q11-q14, 62%), loss(13cen-qter, 90%), loss(13cen-qter, 21%)	38, 45, 46
6	proteasome 26S subunit, ATPase, 5	0.00155	D44467	17q23-q25	gain(17q11-q24, 82%), gain(17q21, 37%), loss(17q21, 53%), loss(42%)	19, 39–41
7	ESTs	0.00166	AA394152			
8	eyes absent (Drosophila) homolog 2	0.00184	AF055015	20q13.1		
9	nuclear RNA export factor 1	0.00191	U80073	11q12-q13	gain(11q13, 27%), gain(11q13, 100%)	19, 38
10	hypothetical protein KIAA1165	0.00219	AA037467	13q22.2	loss(13cen-qter, 90%), loss(13cen-qter, 21%)	38, 46
11	ancient ubiquitous protein 1	0.00226	AF100746	2p13	gain(2p11-p15, 45%)	19
12	NADH dehydrogenase (ubiquinone) Fe-S protein 3 (30kD)	0.00228	AF067139	11p11.11		
13	proteasome 26S subunit, ATPase, 1	0.00232	L02426	19p13.3	loss(19p13.3, 77%), loss(19p13.3, 58%)	45, 48
High in nonsurvivors and low in survivors						
14	anaphase-promoting complex 2	0.00021	AB037827			
15	acetylserotonin O-methyltransferase-like	0.00041	Y15521	Xp22.3, Yp11.3	loss(Yp, 44%), loss(Xp-q21, 67%)	19, 45
16	ribosomal protein L24	0.00051	M94314	3q12	gain(50%), gain(35%), gain(3cen-qter, 35%)	37, 41, 46
17	ESTs	0.00067	BC007849			
18	ESTs, Weakly similar to KIAA0748 protein [H. sapiens]	0.00091	BAA34468			
19	α-B glycoprotein	0.00094	W25099	19cen-q13.2		
20	hypothetical protein FLJ20548	0.00096	AK000555	9q34.13	loss(90%), loss(9q34, 50%), loss(40%)	37, 38, 45
21	seven transmembrane domain orphan receptor	0.00117	AB037108	3q21.3	gain(50%), gain(35%), gain(3cen-qter, 35%)	37, 39, 46
22	KIAA0747 protein	0.00125	BC004998	12q13.13	loss(12q24.1, 63%)	47
23	nucleolar protein 1 (120kD)	0.00145	M32110	12p13	gain(12p12, 27%)	19
24	ribosomal protein L15	0.00161	M77875	3p24.1	loss(3cen-p26, 95%), loss(3cen-p26, >98%)	37, 38
25	SH3 protein	0.00192	AF178432	3p21.31	homozygous deletion(3p21.3), loss(3cen-p26, 95%), loss(3cen-p26, >98%)	16, 37, 38
26	ESTs	0.00209	W02794			
27	hypothetical protein AL 133206	0.00247	AL133206	1p35.3-p34.1	gain(1p33-p35, 45%)	19

[a] If the minimal region and the percentage of genomic imbalance involvement were known, they were written in the parentheses. The data of references 36 and 45 was analyzed with adenocarcinomas only. The data of references 38–41, 43, 46, and 47 was analyzed with LOH analysis with polymorphic markers. The data of references 19, 36, 38, and 45 was analyzed with comparative genomic hybridization.

5-year survivors or more and nonsurvivors (Table 4), several of the genes are involved in the metaphase-anaphase transition and the mitotic spindle checkpoint. Whereas *APC2* was expressed at a higher level in cases with a poor prognosis, *hBUB3*, *hZW10*, and proteasome 26S subunit, *ATPase* 1 and 5 (*PSMC1 and PSMC5*) were expressed higher in cases with a better prognosis. The APC/C is activated by *Cdc20* and *Cdh1*, and promotes metaphase-anaphase transition by ubiquitinating cyclins and the anaphase inhibitor, which are subsequently degraded by the 26S proteasome, in which *PSMC1* and *PSMC5* gene products are subunits (25). *APC2* is an evolutionarily conserved component of the APC/C (26). In *Saccharomyces cerevisiae*, temperature-sensitive *APC2* mutants arrest in metaphase (27). *hBUB3* is a component of the mitotic checkpoint complex that inhibits APC/C (28), a multisubunit E3 ubiquitin ligase that targets proteins of which the proteolytic cleavage of cyclins and anaphase inhibitors is necessary for sister chromatid separation and the exit from mitosis (29, 30). *BUB3* is evolutionarily conserved in eukaryotes (31–33). Mutation of *BUB3* inactivates the mitotic spindle cell checkpoint in both the budding yeast *S. cerevisiae* (34) and the mouse (35). Murine *BUB3* gene knockout embryos accumulate mitotic errors including lagging chromosomes and micronuclei by 4.5 days and do not survive past 7.5 days (35). *ZW10* recruits dynactin and dynein to the centromeric kinetochore (36), a multiprotein-DNA complex responsible for attachment to spindle microtubules, prometaphase chromosome congression, anaphase initiation, and poleward movement. Mutation in the *Drosophila ZW10* gene results in both lagging chromosomes and aneuploidy (36). The positive correlation between poor prognosis and the reduced expression of *hBUB3* and *hZW10* is consistent with the hypothesis that a defective mitotic spindle checkpoint, and the resultant abnormal segregation and increasing aneuploidy would enhance tumor progression. Bhattacharjee *et al.* (3) also have found that increased expression of *hBUB3* is associated with a better prognosis in lung adenocarcinoma cases.

Our study and others (2, 3) generate hypotheses for both clinical and laboratory investigations. The combined use of LCM with microarray expression analysis should decrease the confounder of tissue heterogeneity. Laboratory studies of differentially expressed genes associated with prognosis may identify genes and their function involved in metastasis.

ACKNOWLEDGMENTS

We thank Drs. Hiroshi Murakami, Ryuichi Wada, Atul Kumar, Ivan Rosas, Masayuki Shiseki, Remy Pedeux, Shu Okamura, Stefan Ambs, Hasan Seker, Xin Wang, and Makoto Nagashima for technical advice and suggestions, Dr. John Gillespie for helping with the pathological diagnosis, Dr. Joanna Shih for helpful discussion on the statistical analysis, the microarray group of Dr. Nakamura's laboratory for the fabrication of microarrays, Mary G. McMenamin for cell culture, Audrey Salabes for collecting the clinical information, and Dorothea Dudek for editorial assistance.

REFERENCES

1. Franceschi, S., and Bidoli, E. The epidemiology of lung cancer. Ann. Oncol. *10* (Suppl. 5)*:* S3–S6, 1999.
2. Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., Van De, R. M., Rosen, G. D., Perou, C. M., Whyte, R. I., Altman, R. B., Brown, P. O., Botstein, D., and Petersen, I. Diversity of gene expression in adenocarcinoma of the lung. Proc. Natl. Acad. Sci. USA, *98:* 13784–13789, 2001.
3. Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., and Meyerson, M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA, *98:* 13790–13795, 2001.
4. Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A., and Liotta, L. A. Laser capture microdissection. Science (Wash. DC), *274:* 998–1001, 1996.
5. Luo, L., Salunga, R. C., Guo, H., Bittner, A., Joy, K. C., Galindo, J. E., Xiao, H., Rogers, K. E., Wan, J. S., Jackson, M. R., and Erlander, M. G. Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat. Med., *5:* 117–122, 1999.
6. Sgroi, D. C., Teng, S., Robinson, G., LeVangie. R., Hudson, J. R., Jr., and Elkahloun, A. G. *In vivo* gene expression profile analysis of human breast cancer progression. Cancer Res., *59:* 5656–5661, 1999.
7. Ono, K., Tanaka, T., Tsunoda, T., Kitahara, O., Kihara, C., Okamoto, A., Ochiai, K., Takagi, T., and Nakamura, Y. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res., *60:* 5007–5011, 2000.
8. WHO. International Histological Classification of Tumors, 2nd edition. Geneva: WHO, 1982.
9. Reddel, R. R., Ke, Y., Gerwin, B. I., McMenamin, M. G., Lechner, J. F., Su, R. T., Brash, D. E., Park, J. B., Rhim, J. S., and Harris, C. C. Transformation of human bronchial epithelial cells by infection with SV40 or adenovirus-12 SV40 hybrid virus, or transfection via strontium phosphate coprecipitation with a plasmid containing SV40 early region genes. Cancer Res., *48:* 1904–1909, 1988.
10. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA, *95:* 14863–14868, 1998.
11. Kohda, Y., Murakami, H., Moe, O. W., and Star, R. A. Analysis of segmental renal gene expression by laser capture microdissection. Kidney Int., *57:* 321–331, 2000.
12. Ornstein, D. K., Gillespie, J. W., Paweletz, C. P., Duray, P. H., Herring, J., Vocke, C. D., Topalian, S. L., Bostwick, D. G., Linehan, W. M., Petricoin, E. F., III, and Emmert-Buck. M. R. Proteomic analysis of laser capture microdissected human prostate cancer and *in vitro* prostate cell lines. Electrophoresis, *21:* 2235–2242, 2000.
13. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.
14. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T., and Marincola, F. M. High-fidelity mRNA amplification for gene profiling. Nat. Biotechnol., *18:* 457–459, 2000.
15. Poirier, G. M., Pyati, J., Wan, J. S., and Erlander, M. G. Screening differentially expressed cDNA clones obtained by differential display using amplified RNA. Nucleic Acids Res., *25:* 913–914, 1997.
16. Lerman, M. I., and Minna, J. D. The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumor suppressor genes. The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium. Cancer Res., *60:* 6116–6133, 2000.
17. Kuramochi, M., Fukuhara, H., Nobukuni, T., Kanbe, T., Maruyama, T., Ghosh, H. P., Pletcher, M., Isomura, M., Onizuka, M., Kitamura, T., Sekiya, T., Reeves, R. H., and Murakami, Y. TSLC1 is a tumor-suppressor gene in human non-small-cell lung cancer. Nat. Genet., *27:* 427–430, 2001.
18. Virmani, A. K., Fong, K. M., Kodagoda, D., McIntire, D., Hung, J., Tonk, V., Minna, J. D., and Gazdar, A. F. Allelotyping demonstrates common and distinct patterns of chromosomal loss in human lung cancer types. Genes Chromosomes Cancer, *21:* 308–319, 1998.
19. Michelland, S., Gazzeri, S., Brambilla, E., and Robert-Nicoud, M. Comparison of chromosomal imbalances in neuroendocrine and non-small-cell lung carcinomas. Cancer Genet. Cytogenet., *114:* 22–30, 1999.
20. Mori, Y., Yin, J., Rashid, A., Leggett, B. A., Young, J., Simms, L., Kuehl, P. M., Langenberg, P., Meltzer, S. J., and Stine, O. C. Instabilotyping: comprehensive identification of frameshift mutations caused by coding region microsatellite instability. Cancer Res., *61:* 6046–6049, 2001.
21. Nagakubo, D., Taira, T., Kitaura, H., Ikeda, M., Tamai, K., Iguchi-Ariga, S. M., and Ariga, H. DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation with ras. Biochem. Biophys. Res. Commun., *231:* 509–513, 1997.
22. Prosniak, M., Dierov, J., Okami, K., Tilton, B., Jameson, B., Sawaya, B. E., and Gartenhaus, R. B. A novel candidate oncogene, MCT-1, is involved in cell cycle progression. Cancer Res., *58:* 4233–4237, 1998.
23. Brass, N., Racz, A., Heckel, D., Remberger, K., Sybrecht, G. W., and Meese, E. U. Amplification of the genes BCHE and SLC2A2 in 40% of squamous cell carcinoma of the lung. Cancer Res., *57:* 2290–2294, 1997.
24. Abratt, R. P., Bezwoda, W. R., Falkson, G., Goedhals, L., Hacking, D., and Rugg, T. A. Efficacy and safety profile of gemcitabine in non-small-cell lung cancer: a phase II study. J. Clin. Oncol., *12:* 1535–1540, 1994.
25. Pintard, L., and Peter, M. Mitotic exit: closing the gap. Mol. Cell, *8:* 1155–1156, 2001.
26. Yu, H., Peters, J. M., King, R. W., Page, A. M., Hieter, P., and Kirschner, M. W. Identification of a cullin homology region in a subunit of the anaphase-promoting complex. Science (Wash. DC), *279:* 1219–1222, 1998.
27. Kramer, K. M., Fesquet, D., Johnson, A. L., and Johnston, L. H. Budding yeast RSI1/APC2, a novel gene necessary for initiation of anaphase, encodes an APC subunit. EMBO J., *17:* 498–506, 1998.
28. Sudakin, V., Chan, G. K., and Yen, T. J. Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1. BUB3, CDC20, and MAD2. J. Cell Biol., *154:* 925–936, 2001.
29. King, R. W., Deshaies, R. J., Peters, J. M., and Kirschner, M. W. How proteolysis drives the cell cycle. Science (Wash. DC), *274:* 1652–1659, 1996.
30. Hershko, A., and Ciechanover, A. The ubiquitin system. Annu. Rev. Biochem., *67:* 425–479, 1998.
31. Efimov, V. P., and Morris, N. R. A screen for dynein synthetic lethals in Aspergillus nidulans identifies spindle assembly checkpoint genes and other genes involved in mitosis. Genetics, *149:* 101–116, 1998.
32. Taylor, S. S., Ha, E., and McKeon, F. The human homologue of Bub3 is required for kinetochore localization of Bub1 and a Mad3/Bub1-related protein kinase. J. Cell Biol., *142:* 1–11, 1998.
33. Martinez-Exposito, M. J., Kaplan, K. B., Copeland, J., and Sorger, P. K. Retention of the BUB3 checkpoint protein on lagging chromosomes. Proc. Natl. Acad. Sci. USA, *96:* 8493–8498, 1999.
34. Li, R., and Murray, A. W. Feedback control of mitosis in budding yeast. Cell, *66:* 519–531, 1991.
35. Kalitsis, P., Earle, E., Fowler, K. J., and Choo, K. H. Bub3 gene disruption in mice reveals essential mitotic spindle checkpoint function during early embryogenesis. Genes Dev., *14:* 2277–2282, 2000.
36. Williams, B. C., Karr, T. L., Montgomery, J. M., and Goldberg, M. L. The Drosophila l(1)zw10 gene product, required for accurate mitotic chromosome segregation, is redistributed at anaphase onset. J. Cell Biol., *118:* 759–773, 1992.
37. Pei, J., Balsara, B. R., Li, W., Litwin, S., Gabrielson, E., Feder, M., Jen, J., and Testa, J. R. Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. Genes Chromosomes Cancer, *31:* 282–287, 2001.
38. Petersen, I., Bujard, M., Petersen, S., Wolf, G., Goeze, A., Schwendel, A., Langreck, H., Gellert, K., Reichel, M., Just, K., du, M. S., Cremer, T., Dietel, M., and Ried, T. Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. Cancer Res., *57:* 2331–2335, 1997.
39. Kristiansen, G., Yu, Y., Petersen, S., Kaufmann, O., Schluns, K., Dietel, M., and Petersen, I. Overexpression of c-erbB2 protein correlates with disease-stage and chromosomal gain at the c-erbB2 locus in non-small cell lung cancer. Eur. J. Cancer, *37:* 1089–1095, 2001.
40. Abujiang, P., Mori, T. J., Takahashi, T., Tanaka, F., Kasyu, I., Hitomi, S., and Hiai, H. Loss of heterozygosity (LOH) at 17q and 14q in human lung cancers. Oncogene, *17:* 3029–3033, 1998.
41. Fong, K. M., Kida, Y., Zimmerman, P. V., Ikenaga, M., and Smith, P. J. Loss of heterozygosity frequently affects chromosome 17q in non-small cell lung cancer. Cancer Res., *55:* 4268–4272, 1995.
42. Wang, S. S., Virmani, A., Gazdar, A. F., Minna, J. D., and Evans, G. A. Refined mapping of two regions of loss of heterozygosity on chromosome band 11q23 in lung cancer. Genes Chromosomes Cancer, *25:* 154–159, 1999.
43. Iizuka, M., Sugiyama, Y., Shiraishi, M., Jones, C., and Sekiya, T. Allelic losses in human chromosome 11 in lung cancers. Genes Chromosomes Cancer, *13:* 40–46, 1995.
44. Monaco, C., Negrini, M., Sozzi, G., Veronese, M. L., Vorechovsky, I., Godwin, A. K., and Croce, C. M. Molecular cloning and characterization of LOH11CR2A, a new gene within a refined minimal region of LOH at 11q23. Genomics, *46:* 217–222, 1997.
45. Girard, L., Zochbauer-Muller, S., Virmani, A. K., Gazdar, A. F., and Minna, J. D. Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. Cancer Res., *60:* 4894–4906, 2000.
46. Luk, C., Tsao, M. S., Bayani, J., Shepherd, F., and Squire, J. A. Molecular cytogenetic analysis of non-small cell lung carcinoma by spectral karyotyping and comparative genomic hybridization. Cancer Genet. Cytogenet., *125:* 87–99, 2001.
47. Shiseki, M., Kohno, T., Nishikawa, R., Sameshima, Y., Mizoguchi, H., and Yokota, J. Frequent allelic losses on chromosomes 2q, 18q, and 22q in advanced non-small cell lung carcinoma. Cancer Res., *54:* 5643–5648, 1994.
48. Sobottka, S. B., Haase, M., Fitze, G., Hahn, M., Schackert, H. K., and Schackert, G. Frequent loss of heterozygosity at the 19p13.3 locus without LKB1/STK11 mutations in human carcinoma metastases to the brain. J. Neuro-Oncol., *49:* 187–195, 2000.