

# Discordant Protein and mRNA Expression in Lung Adenocarcinomas\*

Guoan Chen‡, Tarek G. Gharib‡, Chiang-Ching Huang§, Jeremy M. G. Taylor§, David E. Misek¶, Sharon L. R. Kardia||, Thomas J. Giordano\*\*, Mark D. Iannettoni‡, Mark B. Orringer‡, Samir M. Hanash¶, and David G. Beer‡ ††

The relationship between gene expression measured at the mRNA level and the corresponding protein level is not well characterized in human cancer. In this study, we compared mRNA and protein expression for a cohort of genes in the same lung adenocarcinomas. The abundance of 165 protein spots representing 98 individual genes was analyzed in 76 lung adenocarcinomas and nine non-neoplastic lung tissues using two-dimensional polyacrylamide gel electrophoresis. Specific polypeptides were identified using matrix-assisted laser desorption/ionization mass spectrometry. For the same 85 samples, mRNA levels were determined using oligonucleotide microarrays, allowing a comparative analysis of mRNA and protein expression among the 165 protein spots. Twenty-eight of the 165 protein spots (17%) or 21 of 98 genes (21.4%) had a statistically significant correlation between protein and mRNA expression ( $r > 0.2445$ ;  $p < 0.05$ ); however, among all 165 proteins the correlation coefficient values ( $r$ ) ranged from  $-0.467$  to  $0.442$ . Correlation coefficient values were not related to protein abundance. Further, no significant correlation between mRNA and protein expression was found ( $r = -0.025$ ) if the average levels of mRNA or protein among all samples were applied across the 165 protein spots (98 genes). The mRNA/protein correlation coefficient also varied among proteins with multiple isoforms, indicating potentially separate isoform-specific mechanisms for the regulation of protein abundance. Among the 21 genes with a significant correlation between mRNA and protein, five genes differed significantly between stage I and stage III lung adenocarcinomas. Using a quantitative analysis of mRNA and protein expression within the same lung adenocarcinomas, we showed that only a subset of the proteins exhibited a significant correlation with mRNA abundance. *Molecular & Cellular Proteomics* 1:304–313, 2002.

Lung cancer is the leading cause of cancer death for both men and women in the United States. Adenocarcinomas of the lung comprise ~40% of all new cases of non-small cell

lung cancer and are now the most common histologic type. Functional genomics, broadly defined as the comprehensive analysis of genes and their products, have become a recent focus of the life sciences (1). Application of these approaches to lung adenocarcinomas has the potential to aid in the identification of high risk patients with resectable early stage lung cancer that may benefit from adjuvant therapy, as well as to identify new therapeutic targets. In human lung cancer, however, little is currently understood regarding the relationship between gene expression as determined by measuring mRNA levels and the corresponding abundance of the protein products.

A number of powerful techniques for analysis of gene expression have been used including differential display (2), serial analysis of gene expression (3), DNA microarrays (4), and proteomics via two-dimensional polyacrylamide gel electrophoresis and mass spectrometry (5). Bioinformatics tools have also been developed to help determine quantitative mRNA/protein expression profiles of all types of cells and tissues (6) and now can be applied to benign and malignant tumors. DNA microarrays (cDNA and oligonucleotide) permit the parallel assessment of thousands of genes and have been utilized in gene expression monitoring (7), polymorphism analysis (8), and DNA sequencing (9). Recent studies have focused on classification or identification of subgroups of lung tumors using DNA microarrays (10, 11). The use of mRNA expression patterns by themselves, however, is insufficient for understanding the expression of protein products, as additional post-transcriptional mechanisms, including protein translation, post-translational modification, and degradation, may influence the level of a protein present in a given cell or tissue. Proteomic analyses, a complementary technology to DNA microarrays for monitoring gene expression, involves protein separation and quantitative assessment of protein spots using 2D<sup>1</sup>-PAGE and protein identification using mass spectrometry. By combining proteomic and transcriptional analyses of the same samples, however, it may be possible to understand the complex mechanisms influencing protein expression in human cancer.

In this study, we determined mRNA and protein levels for 165 proteins (98 genes) in 76 lung adenocarcinomas and nine

From the Departments of ‡Surgery, §Biostatistics, ||Epidemiology, \*\*Pathology, and ¶Pediatrics, University of Michigan, Ann Arbor, Michigan 48109

Received, January 21, 2002, and in revised form, March 4, 2002

Published, MCP Papers in Press, March 12, 2001, DOI 10.1074/mcp.M200008-MCP200

<sup>1</sup> The abbreviations used are: 2D, two-dimensional; MALDI-MS, matrix-assisted laser desorption/ionization mass spectrometry.

## Protein and mRNA Correlation in Lung Adenocarcinomas

TABLE I

*Correlation coefficients of protein and mRNA where only one spot was present on 2D gels*

*r\**, correlation coefficient value > 0.2445; *p* < 0.05. Values in boldface are significant at *p* < 0.05.

Spot	Unigene	Gene name	<i>r*</i>	Protein name
1104	Hs.184510	SFN	<b>0.4337</b>	14-3-3 $\sigma$
0994	Hs.77840	ANXA4	<b>0.4219</b>	Annexin IV
1314	Hs.10958	DJ-1	<b>0.3982</b>	DJ-1 protein/MER5
1454	Hs.75428	SOD1	<b>0.3863</b>	Superoxide dismutase (Cu-Zn)
1638	Hs.227751	LGALS1	<b>0.3318</b>	Galectin 1
0264	Hs.129548	HNRPK	<b>0.3034</b>	Transformation up-regulated nuclear protein
1405	Hs.111334	FTL	<b>0.2849</b>	Ferritin light chain
0963	Hs.300711	ANXA5	<b>0.2468</b>	Annexin V
1252	Hs.4745	PSMC	<b>0.2445</b>	26 S proteasome p28
0906	Hs.234489	LDHB	0.4420	L-lactate dehydrogenase H chain (LDH-B)
1171	Hs.241515	COX11	0.2310	COX 11
1160	Hs.181013	PGAM1	0.2023	Phosphoglycerate mutase
0759	Hs.74635	DLD	0.1965	Dihydroliipoamide dehydrogenase precursor
1193	Hs.83383	AOE372	0.1932	Antioxidant enzyme AOE372
0172	Hs.3069	HSPA9B	0.1872	GRP75
0777	Hs.979	PDHB	0.1855	Pyruvate dehydrogenase E1- $\beta$ subunit precursor
1249	Hs.226795	GSTP1	0.1773	Glutathione S-transferase pi (GST-pi)
1685	Hs.76136	TXN	0.1732	Thioredoxin
1205	Hs.82314	HPRT1	0.1588	HG phosphoribosyltransferase
1230	Hs.279860	TPT1	0.1466	Translationally controlled tumor protein (TCTP)
0603	Hs.181357	LAMR1	0.1463	LAMR
1358	Hs.28914	APRT	0.1399	Adenine phosphoribosyl transferase
1410	Hs.82113	DUT	0.1213	dUTP pyrophosphatase (dUTPase)
1825	Hs.112378	LIMS1	0.1213	Pinch-2 protein
0871	Hs.250502	CA8	0.1122	Carbonic anhydrase-related protein; Syntaxin
0289	Hs.82916	CCT6A	0.1106	Chaperonin-like protein
1143	Hs.11465	GSTTLp28	0.0997	Glutathione S-transferase homolog (GST homolog)
1456	Hs.118638	NME1	0.0932	Nm23 (NDPKA)
1598	Hs.278503	RIG	0.0905	RILG (U32331)
1354	Hs.89761	ATP5D	0.0904	FIFO-type ATP synthase subunit d
1445	Hs.155485	HIP2	0.0843	Huntingtin interacting protein 2 (HIP2)
1479	Hs.177486	APP	0.0746	Amyloid B4A
0608	Hs.182265	KRT19	0.0439	Cytokeratin 19
1071	Hs.10842	RAN	0.0277	GTP-binding nuclear protein RAN(TC4)
0991	Hs.297939	CTSB	0.0254	Cathepsin B
0842	Hs.77274	PLAU	0.0248	Urokinase plasminogen activator
0823	Hs.198248	B4GALT1	0.0183	$\beta$ 1,4-galactosyl transferase
0613	Hs.1247	APOA4	0.0176	Apolipoprotein A4 (ApoA4)
1338	Hs.104143	CLTA	0.0123	Clathrin light chain A
0902	Hs.5123	SID6-306	0.0117	Cytosolic inorganic pyrophosphatase
1688	Hs.1473	GRP	-0.0040	Preprogastrin-releasing peptide
0265	Hs.274402	HSPA1B	-0.0071	Heat shock-induced protein
1414	Hs.77541	ARF5	-0.0096	ADP-ribosylation factor 1
0710	Hs.97206	HIP1	-0.0114	Huntingtin interacting protein 1 (HIP1)
0532	Hs.170328	MSN	-0.0132	Moesin/E
0525	Hs.284255	ALPP	-0.0148	Alkaline phosphate, placental
0513	Hs.76901	PDIR	-0.0289	Protein disulfide isomerase-related protein 5
1659	Hs.256697	HINT	-0.0312	Protein kinase C inhibitor
1262	Hs.7016	RAB7	-0.0362	Rab 7 protein
0190	Hs.184411	ALB	-0.0470	Albumin
0948	Hs.2795	LDHA	-0.0549	Lactate dehydrogenase-A (LDHA)
0502	Hs.180532	GPI	-0.0575	Hsp89
0152	Hs.75410	HSPA5	-0.0640	GRP78
1054	Hs.74276	CLIC1	-0.0686	Nuclear chloride channel (RNCC protein)
0709	Hs.253495	SFTPD	-0.0936	Pulmonary surfactant protein D
0867	Hs.78996	PCNA	-0.0982	PCNA
0165	Hs.180414	HSPA8	-0.1014	Heat shock cognate protein, 71 kDa
1109	Hs.75103	YWHAZ	-0.1018	14-3-3 $\zeta/\Delta$
0137	Hs.554	SSA2	-0.1032	Ro/ss-A antigen

TABLE I—continued

Spot	Unigene	Gene name	r*	Protein name
0278	Hs.4112	TCP1	-0.1237	T-complex protein I, $\alpha$ subunit
1769	Hs.9614	NPM1	-0.1738	B23/numatrin
0089	Hs.74335	HSPCB	-0.2049	Hsp90
2511	Hs.153179	FABP5	-0.2109	E-FABP/FABP5
1739	Hs.16488	CALR	-0.2344	Calreticulin 32
1138	Hs.301961	GSTM4	-0.2438	Glutathione S-transferase M4 (GST m4)
2533	Hs.77060	PSMB6	-0.2512	Macropain subunit $\Delta$

non-neoplastic lung tissues. Protein levels were determined using quantitative 2D-PAGE analysis, and the separated protein polypeptides were identified using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). The corresponding mRNA levels for the identified proteins within the same samples were determined using oligonucleotide microarrays. Correlation analyses showed that protein abundance is likely a reflection of the transcription for a subset of proteins, but translation and post-translational modifications also appear to influence the expression levels of many individual proteins in lung adenocarcinomas.

EXPERIMENTAL PROCEDURES

**Tissues**—Fifty-seven stage I and 19 stage III lung adenocarcinomas, as well as nine non-neoplastic lung tissue samples, were used for protein and mRNA analyses. Patient consent was obtained, and the project was approved by the Institutional Review Board. All tissues were obtained after resection at the University of Michigan Health System between May 1991 and July 1998. Tissues were all snap-frozen in liquid nitrogen and then stored at  $-80^{\circ}\text{C}$ . The patients included 46 females and 30 males ranging in age from 40.9 to 84.6 (average 63.8) years. Most patients (66/76) demonstrated a positive smoking history. Sixty-one tumor samples were classified as bronchial-derived, 14 were classified as bronchoalveolar, and one had both features. Eighteen tumor samples were classified as well differentiated, 38 were classified as moderate, and 19 were classified as poorly differentiated adenocarcinomas. Hematoxylin-stained cryostat sections ( $5\ \mu\text{m}$ ), prepared from the same tumor pieces to be utilized for protein and mRNA isolation, were evaluated by a pathologist and compared with hematoxylin- and eosin-stained sections made from paraffin blocks of the same tumors. Specimens were excluded from analysis if they showed unclear or mixed histology (e.g. adenocarcinomas), tumor cellularity less than 70%, potential metastatic origin as indicated by previous tumor history, extensive lymphocytic infiltration, or fibrosis or if the patient had received prior chemotherapy or radiotherapy.

**Oligonucleotide Array Hybridization**—The HuGeneFL oligonucleotide arrays (Affymetrix, Santa Clara, CA) containing 6800 genes were used in this study. Total RNA was isolated from all samples using Trizol reagent (Invitrogen). The resulting RNA was then subjected to further purification using RNeasy spin columns (Qiagen). Preparation of cRNA, hybridization, and scanning of the HuGeneFL arrays were performed according to the manufacturer’s protocol (Affymetrix, Santa Clara, CA). Data analysis was performed using GeneChip 4.0 software. The gene expression profile of each tumor was normalized to the median gene expression profile for the entire sample. Details of data trimming and normalization are described elsewhere (11).

**2D-PAGE and Quantitative Protein Analysis**—Tissue for both protein and mRNA isolation came from contiguous areas of each sample. Protein separation using 2D-PAGE, silver staining, and digitization

were performed as described previously (12, 13). Our 2D-PAGE system allows us to run 20 gels at one time (one batch). Spot detection and quantification were accomplished utilizing Bio Image Visage System software (Bioimage Corp., Ann Arbor, MI). The integrated intensity of each spot was calculated as the measured optical density units  $\times\ \text{mm}^2$ . Of the total possible 2000 spots detectable on each gel, 820 spots on the gel of each sample were matched using a Gel-ed match program with the same spots on a chosen “master” gel. In each sample, 250 ubiquitously expressed reference spots were used to adjust for variations between gels, such as that created by subtle differences in protein loading or gel staining. Slight differences because of batch were corrected after spot-size quantification.

**Mass Spectrometry and 2D Western Blotting**—Preparative 2D gels were run using extracts from A549 lung adenocarcinoma cells (obtained from ATCC) and using the identical experimental conditions as the analytical 2D gels, except 30% more protein was loaded. The resolved protein gels were silver-stained using successive incubations in 0.02% sodium thiosulfate for 2 min, 0.1% silver nitrate for 40 min, and 0.014% formaldehyde plus 2% sodium carbonate for 10 min. For protein identification, protein polypeptides underwent trypsin digestion followed by MALDI-MS using a MALDI-TOF Voyager-DE mass spectrometer (Perseptive Biosystems, Framingham, MA). The masses were compared with known trypsin digest databases using the MS-FIT database (University of California, San Francisco; prospector.ucsf.edu/ucsfhtml3.2/msfit.htm). Some of the polypeptides included in the analysis had been identified prior to this study on the basis of sequencing (14). The identified protein spots used in this paper are shown in Fig. 1A. The method for 2D-PAGE Western blot verification was as described previously (15). The 2D Western blot of GRP58 and Op18 are shown in Fig. 1, C and E; the others, such as GRP78, GRP75, HSP70, HSC70, KRT8, KRT18, KRT19, Vimentin, ApoJ, 14-3-3, Annexin I, Annexin II, PGP9.5, DJ-1, GST-pi, and PGAM, are described elsewhere.<sup>2</sup>

**Statistical Analysis**—Missing values were replaced with the mean value of the protein spot. The transform  $x \rightarrow \log(1 + x)$  was applied to normalize all protein expression values. The relationship between protein and mRNA expression levels within the same samples was examined using the Spearman correlation coefficient analysis (16). To identify potentially significant correlations between gene and protein expression, we used an analytical strategy similar to SAM (significance analysis of microarrays) (17), which uses a permutation technique to determine the significance of changes in gene expression between different biological states. To obtain permuted correlation coefficients between gene and protein expression, genes were exchanged first in such a way that permuted correlation coefficient were calculated based on pseudo pairs of genes and proteins. The distribution of permuted correlation coefficients became stable after 60 permutations. This procedure was then repeated 60 times to obtain 60 sets of permuted correlation coefficients. For each of the 60 permutations, the correlations of genes and proteins were ranked

<sup>2</sup> Chen *et al.*, submitted for publication.

## Protein and mRNA Correlation in Lung Adenocarcinomas

TABLE II

*Correlation coefficients of protein and mRNA where multiple isoforms were present on 2D gels*

*r\**, correlation coefficient value > 0.2445; *p* < 0.05. Values in boldface are significant at *p* < 0.05.

Spot	Unigene	Gene name	<i>r*</i>	Protein name
1494	Hs.81915	LAP18	<b>0.4003</b>	OP18 (Stathmin)
0957	Hs.77899	TPM1	<b>0.3930</b>	Tropomyosins 1–5
0353	Hs.289101	GRP58	<b>0.3802</b>	Protease disulfide isomerase (GRP58)
0855	Hs.169476	GAPD	<b>0.3693</b>	Glyceraldehyde-3-phosphate dehydrogenase
1198	Hs.41707	HSPB3	<b>0.3668</b>	Hsp27
1203	Hs.83848	TPI1	<b>0.3395</b>	Triose phosphate isomerase (TPI)
0523	Hs.65114	KRT18	<b>0.3335</b>	Cytokeratin 18
1492	Hs.81915	LAP18	<b>0.3234</b>	OP18 (Stathmin)
1493	Hs.81915	LAP18	<b>0.3154</b>	OP18 (Stathmin)
1181	Hs.78225	ANXA1	<b>0.3102</b>	Annexin variant I
0439	Hs.242463	KRT8	<b>0.3049</b>	Cytokeratin 8
0505	Hs.297753	VIM	<b>0.2939</b>	Vimentin
0593	Hs.297753	VIM	<b>0.2809</b>	Vimentin
1874	Hs.75313	AKR1B1	<b>0.2790</b>	Aldose reductase
0935	Hs.75544	YWHAH	<b>0.2775</b>	14-3-3 $\eta$
2524	Hs.78225	ANXA1	<b>0.2612</b>	Annexin I
2324	Hs.65114	KRT18	<b>0.2601</b>	Cytokeratin 18
1192	Hs.41707	HSPB3	<b>0.2558</b>	Hsp27
0350	Hs.289101	GRP58	<b>0.2516</b>	Phospholipase C (GRP58)
0992	Hs.75313	AKR1B1	–0.2460	Aldose reductase
0861	Hs.75313	AKR1B1	0.0761	Aldose reductase
0853	Hs.75313	AKR1B1	–0.0675	Aldose reductase
2503	Hs.76392	ALDH1	–0.0565	Aldehyde dehydrogenase
0381	Hs.76392	ALDH1	–0.0371	Aldehyde dehydrogenase
0371	Hs.76392	ALDH1	–0.0680	Aldehyde dehydrogenase
1179	Hs.78225	ANXA1	0.2052	Annexin variant I
0762	Hs.78225	ANXA1	–0.0739	Annexin I
0760	Hs.78225	ANXA1	–0.0228	Annexin I
2506	Hs.217493	ANXA2	0.2223	Lipocotin (annexin II)
0772	Hs.217493	ANXA2	0.2080	Lipocotin (annexin II)
0723	Hs.217493	ANXA2	0.0701	Lipocotin
1239	Hs.93194	APOA1	0.1133	Apolipoprotein A1 (ApoA1)
1237	Hs.93194	APOA1	–0.0373	Apolipoprotein A1 (ApoA1)
1234	Hs.93194	APOA1	–0.0894	Apolipoprotein A1 (ApoA1)
0428	Hs.25	ATP5B	0.0080	ATP synthase $\beta$ subunit precursor
0427	Hs.25	ATP5B	0.0122	ATP synthase $\beta$ subunit precursor
0424	Hs.25	ATP5B	–0.0992	ATP synthase $\beta$ subunit precursor
0863	Hs.75106	CLU	–0.0483	Apolipoprotein J (ApoJ)
0780	Hs.75106	CLU	–0.0443	Apolipoprotein J (ApoJ)
1527	Hs.119140	EIF5A	–0.0726	eIF-5A
1484	Hs.119140	EIF5A	–0.0376	eIF-5A
1728	Hs.5241	FABP1	–0.1916	L-FABP
1712	Hs.5241	FABP1	–0.0473	L-FABP
0947	Hs.169476	GAPD	0.1745	Glyceraldehyde-3-phosphate dehydrogenase
1232	Hs.75207	GLO1	0.2249	Glyoxalase-I
1229	Hs.75207	GLO1	0.0450	Glyoxalase-1
1595	Hs.158300	HAP1	–0.0137	Huntingtin-associated protein 1 (neuroan 1)
1810	Hs.75990	HP	–0.4672	$\alpha$ -Haptoglobin
1459	Hs.75990	HP	0.0802	$\alpha$ -Haptoglobin
1458	Hs.75990	HP	–0.0305	$\alpha$ -Haptoglobin
0619	Hs.75990	HP	0.0461	B-haptoglobin
0615	Hs.75990	HP	–0.0034	B-haptoglobin
1250	Hs.41707	HSPB3	–0.1024	Hsp27
0549	Hs.79037	HSPD1	0.1074	Hsp60
0338	Hs.79037	HSPD1	0.2265	Hsp60
0333	Hs.79037	HSPD1	0.1383	Hsp60
0331	Hs.79037	HSPD1	0.1603	Hsp60
2381	Hs.65114	KRT18	0.2016	Cytokeratin 18
0535	Hs.65114	KRT18	0.1106	Cytokeratin 18



## Protein and mRNA Correlation in Lung Adenocarcinomas

TABLE II—continued

Correlation coefficients of protein and mRNA where multiple isoforms were present on 2D gels

$r^*$ , correlation coefficient value  $> 0.2445$ ;  $p < 0.05$ . Values in boldface are significant at  $p < 0.05$ .

Spot	Unigene	Gene name	$r^*$	Protein name
0529	Hs.65114	KRT18	0.1279	Cytokeratin 18
0528	Hs.65114	KRT18	0.0414	Cytokeratin 18
0527	Hs.65114	KRT18	0.0436	Cytokeratin 18
0514	Hs.65114	KRT18	0.0733	Cytokeratin 18
0451	Hs.242463	KRT8	-0.0111	Cytokeratin 8
0446	Hs.242463	KRT8	0.0347	Cytokeratin 8
0444	Hs.242463	KRT8	-0.1311	Cytokeratin 8
0443	Hs.242463	KRT8	0.0942	Cytokeratin 8
1488	Hs.81915	LAP18	0.0495	OP18 (Stathmin)
0321	Hs.75655	P4HB	-0.0546	PDI (proly-4-OH-B)
0320	Hs.75655	P4HB	-0.0041	PDI (proly-4-OH-B)
1063	Hs.75323	PHB	0.0441	Prohibitin
0837	Hs.75323	PHB	0.1402	Prohibitin
0326	Hs.297681	SERPINA1	-0.0227	$\alpha$ -1-Antitripsin
0322	Hs.297681	SERPINA1	-0.0277	$\alpha$ -1-Antitripsin
0241	Hs.297681	SERPINA1	-0.0148	$\alpha$ -1-Antitripsin
1280	Hs.301254	SFTPA1	-0.1488	Pulmonary surfactant-associated protein
1278	Hs.301254	SFTPA1	-0.2040	Pulmonary surfactant-associated protein
0866	Hs.73980	TNNT1	0.1162	Troponin T
0778	Hs.73980	TNNT1	0.0740	Troponin T
1213	Hs.83848	TPI1	0.0024	Triose phosphate isomerase (TPI)
1210	Hs.83848	TPI1	0.0490	Triose phosphate isomerase (TPI)
1207	Hs.83848	TPI1	-0.1615	Triose phosphate isomerase (TPI)
1204	Hs.83848	TPI1	0.0209	Triose phosphate isomerase (TPI)
1202	Hs.83848	TPI1	0.0721	Triose phosphate isomerase (TPI)
1161	Hs.83848	TPI1	0.2265	Triose phosphate isomerase (TPI)
1052	Hs.77899	TPM1	-0.1040	Tropomyosin clean-product
1039	Hs.77899	TPM1	-0.2999	Cytoskeletal tropomyosin
1035	Hs.77899	TPM1	-0.3821	Tropomyosin
0783	Hs.77899	TPM1	0.0757	Tropomyosins 1–5
1574	Hs.194366	TTR	-0.0065	Transthyretin
0809	Hs.194366	TTR	0.0399	Transthyretin multimer
2202	Hs.76118	UCHL1	-0.0220	Ubiquitin carboxyl-terminal hydrolase isozyme L1
1246	Hs.76118	UCHL1	-0.1261	Ubiquitin carboxyl-terminal hydrolase isozyme L1
1242	Hs.76118	UCHL1	0.1473	Ubiquitin carboxyl-terminal hydrolase isozyme L1
0606	Hs.297753	VIM	0.0951	Vimentin
0594	Hs.297753	VIM	-0.2664	Vimentin-derived protein (vid4)
0508	Hs.297753	VIM	0.1008	Vimentin-derived protein (vid2)
0419	Hs.297753	VIM	0.0032	Vimentin-derived protein (vid1)
1279	Hs.75544	YWHAH	0.0059	14-3-3 $\eta$

such that  $\rho_p(i)$  denotes the  $i$ th largest correlation coefficient for  $p$ th permutation. Hence, the expected correlation coefficient,  $\rho_E(i)$ , was the average over the 60 permutations,  $\rho_E(i) = \sum_{p=1}^{60} \rho_p(i)/60$ . A scatter plot of observed correlations ( $\rho(i)$ ) versus the expected correlations is shown in Fig. 2D. For this study, we chose threshold  $\Delta = 0.115$  so that correlation would be considered significant if absolute value of difference between  $\rho(i)$  and  $\rho_E(i)$  was greater than the threshold. Twenty-nine (including one with observed correlation coefficient  $-0.4672$ ) of 165 pairs of gene and protein expression were called significant in such criteria, and the permuted data generated an average of 5.1 falsely significant pairs of gene and protein expression. This provided an estimated false discovery rate (the percentage of pairs of gene and protein expression identified by chance) for our data set.

### RESULTS

**Correlation of Individual Proteins and mRNA Expression within Each Tumor**—We have examined quantitatively 165

protein spots on 2D gels representing 98 genes and compared protein levels with mRNA levels for a cohort of 85 lung adenocarcinomas and normal lung samples. Of the 165 protein spots, 69 proteins were represented by only one known spot on 2D gels for an individual gene, whereas 96 protein spots showed multiple protein products from 29 different genes. 2D Western blotting verified the proteins identified by mass spectrometry when specific antibodies were available. Spearman correlation coefficients of the proteins and their associated mRNA for each protein spot were generated using all 76 lung adenocarcinomas and nine non-neoplastic lung tissues (see Tables I and II, and see Figs. 1 and 2). The correlation coefficients ( $r$ ) ranged from  $-0.467$  to  $0.442$  (Fig. 2D). A total of 28 protein spots (21 genes) were found to have a statistically significant correlation between expression of

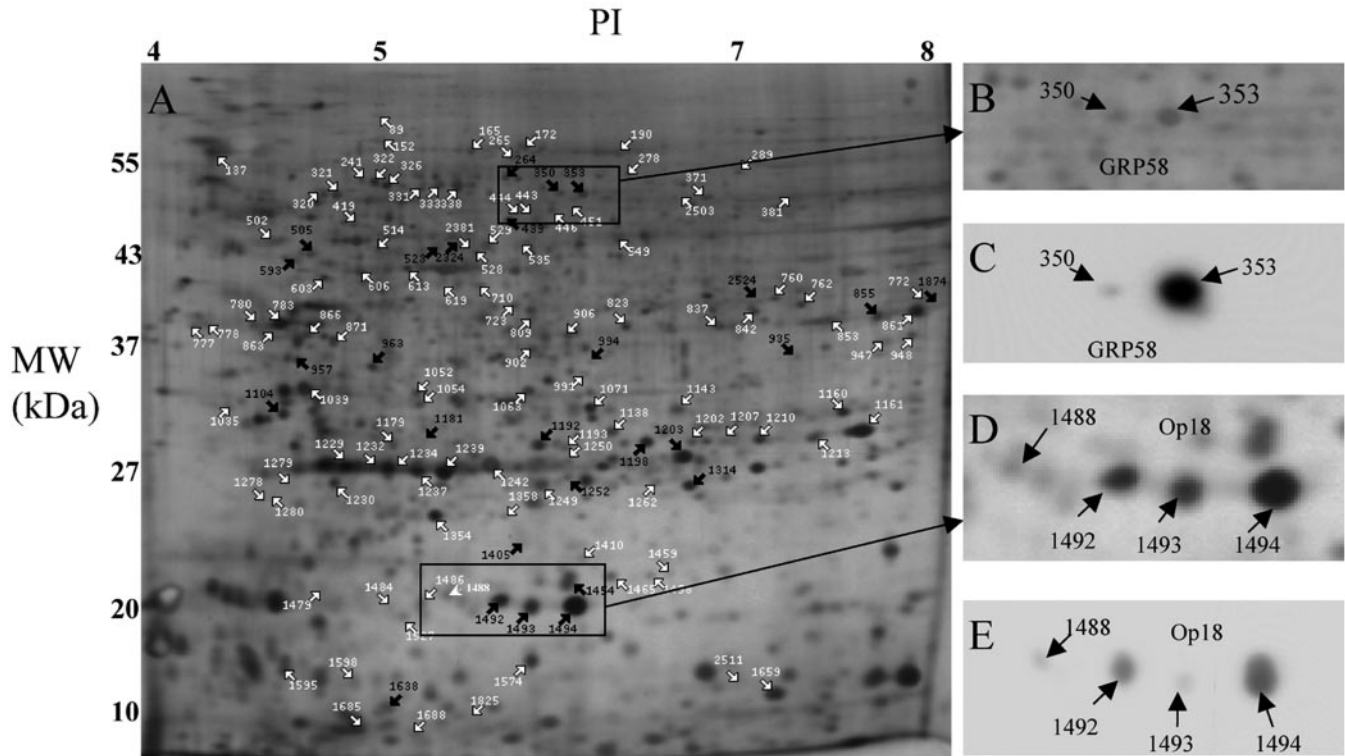


FIG. 1. A, digital image of a silver-stained 2D-PAGE separation of a stage I lung adenocarcinoma showing protein spots separated by molecular mass (*MW*) and isoelectric point (*PI*). Twenty-eight protein spots whose expression levels are correlated with mRNA abundance are indicated by the *black arrows*. B, the outlined areas of A showing protein GRP58. C, 2D Western blot of GRP58 from the A549 lung adenocarcinoma cell line. D, the outlined areas of A showing the protein isoforms of Op18. E, 2D Western blot of Op18 from A549 cells.

their protein and mRNA ( $r > 0.2445$ ;  $p < 0.05$ ). This accounts for 17% (28/165) of the 165 protein spots. Among the 69 genes for which only a single protein spot was known (Table I), nine genes (9/69, 13%) were observed to show a statistically significant relationship between protein and mRNA abundance ( $r > 0.2445$ ;  $p < 0.05$ ). The proteins whose expression levels were correlated with their mRNA abundance included those involved in signal transduction, carbohydrate metabolism, apoptosis, protein post-translational modification, structural proteins, and heat shock proteins (Table III).

**Individual Isoforms of the Same Protein Have Different Protein/mRNA Correlation Coefficients**—Of the 165 protein spots, 96 represent protein products of 29 genes with at least two isoforms. Among these 96 protein spots, 19 (19/96 protein spots, 20%) showed a statistically significant correlation between their protein and mRNA expression ( $r > 0.2445$ ;  $p < 0.05$ ) (Table II) and represented 12 genes (12/29, 41%). Individual isoforms of the same protein demonstrated different protein/mRNA correlation coefficients. For example, 2D-PAGE/Western analysis revealed four isoforms of OP18 differing in regards to isoelectric point but similar in molecular weight. Three of the four isoforms (spots 1492, 1493, and 1494) showed a statistically significant correlation between their protein and mRNA abundance ( $r = 0.3234$ ,  $0.3154$ , and  $0.4003$ , respectively). The fourth isoform (spot 1488) showed no correlation be-

tween protein and mRNA expression ( $r = 0.0495$ ). Similarly, just one of five quantified isoforms of cytokeratin 8 (spot 439) demonstrated a statistically significant correlation between protein and mRNA abundance ( $r = 0.3049$ ;  $p < 0.05$ ) (Table II).

In addition to differences in the relationship between mRNA levels and protein expression among separate isoforms, some genes with very comparable mRNA levels showed a 24-fold difference in their protein expression. Genes with comparable protein expression levels also showed up to a 28-fold variance in their mRNA levels.

**Lack of Correlation for mRNA and Protein Expression when Using Average Tumor Values across All 165 Protein Spots (98 Genes)**—The relationship between mRNA and protein expression was also examined by using the average expression values for all samples. To analyze this relationship using this approach, the average value for each protein or mRNA was generated using all 85 lung tissue samples. The range of normalized average protein values ranged from  $-0.0646$  to  $0.0979$  (raw value  $0.0036$  to  $4.1947$ ), and the range for mRNA was from  $0$  to  $15260.5$  for all 165 individual protein spots. The Spearman correlation coefficient for the whole data set (165 protein spots/98 genes) was  $-0.025$  (Fig. 3A). Even for the 28 protein spots (Fig. 2D) that were found to have a statistically significant correlation between their mRNA and protein, use of

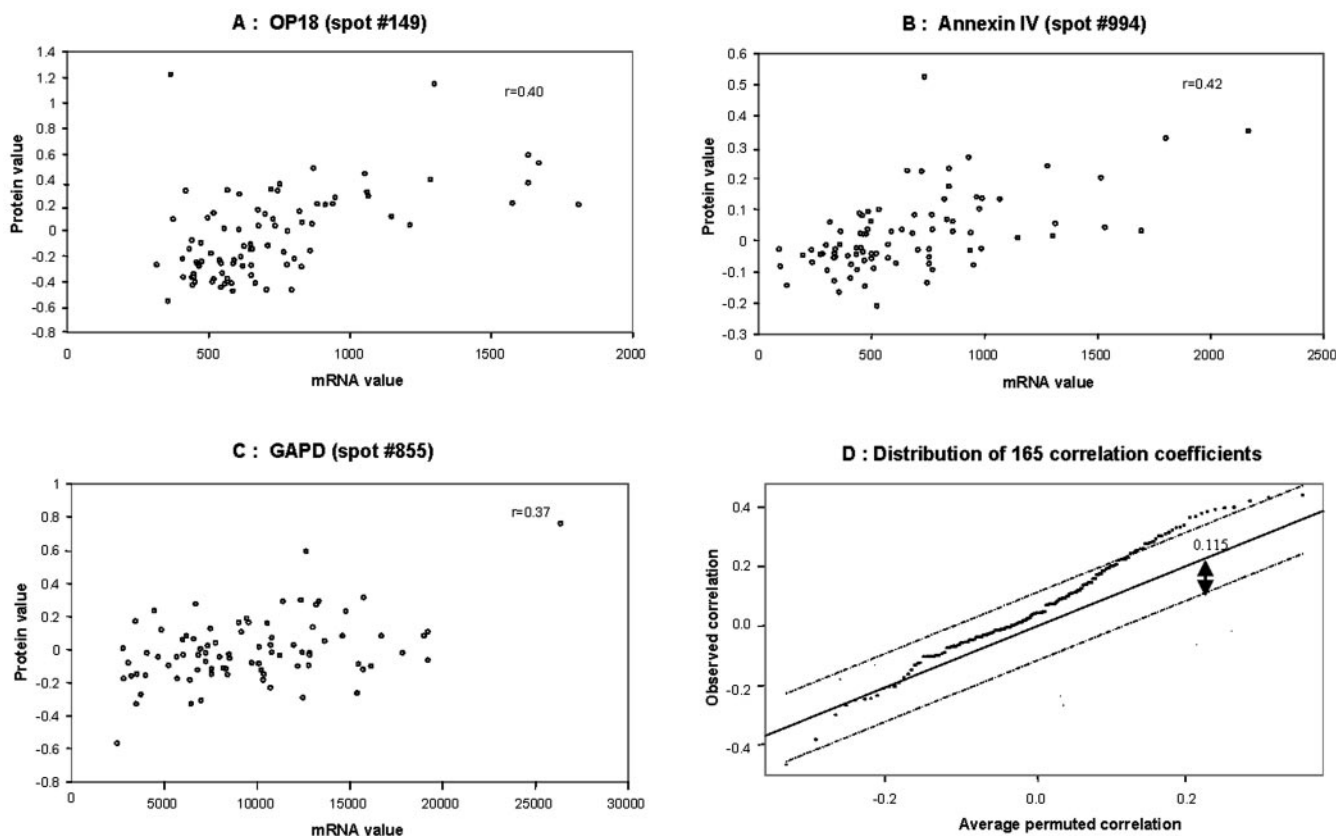


FIG. 2. A–C, plots showing the correlation between mRNA and protein for the three selected genes Op18, Annexin IV, and GAPD for all 76 lung adenocarcinomas and nine non-neoplastic lung samples ( $p < 0.05$ ). D, distribution of all 165 Spearman correlation coefficients ( $r$ ) and verification analysis using SAM. A more detailed description of the method is provided under “Experimental Procedures.” Approximately 17% of the 165 proteins demonstrate a significant correlation between mRNA and protein levels as demonstrated by the values shown beyond the outer range of threshold  $\Delta = 0.115$ . Normalized protein values were used, thus negative values for some proteins are observed.

the average value resulted in a correlation coefficient value of  $-0.035$ , which was not significant (Fig. 3B).

**Lack of a Relationship between Protein/mRNA Correlation Coefficients and Average Protein Abundance**—To determine whether an absolute protein level might influence the correlation with mRNA, the mean value of each protein (relative abundance) and the Spearman protein/mRNA correlation coefficients among all 85 samples were examined. No relationship between the protein abundance and the correlation coefficients was observed ( $r = 0.039$ ;  $p > 0.05$ ). A detailed analysis of separate subsets of proteins with differing levels of abundance (less than  $-0.0014$ , larger than  $-0.0014$ , or larger than  $0.0077$ ) also showed a lack of correlation between mRNA and protein expression among the 83 (50%), 82 (50%), and 41 (25%) of 165 total protein spots, respectively ( $r = 0.016, 0.08$ , and  $0.172$ , respectively).

**Stage-related Changes in the Protein/mRNA Correlation Coefficients**—To determine whether the 21 genes (28 protein spots) showing a significant correlation between the protein and mRNA expression among all samples demonstrate changes in this relationship during tumor progression, the correlations were examined separately for stage I ( $n = 57$ ) and

stage III ( $n = 19$ ) lung adenocarcinomas (Table III). The number of non-neoplastic lung samples ( $n = 9$ ) was insufficient for a separate correlation analysis of this group. Many of the protein spots represent one of several known protein isoforms for a given gene. The majority of genes (16/21) did not differ in the protein/mRNA correlation between stage I and stage III tumors indicating a similar regulatory relationship between the mRNA and protein spot. GRP-58, PSMC, SOD1, TPI1, and VIM, however, were found to demonstrate significant differences in the correlation coefficients between stage I and stage III lung adenocarcinomas. For GRP-58, PSMC, and VIM the change in the correlation coefficient was because of a relative increase in protein expression in stage III tumors. For SOD and TPI the change resulted from a relative decrease in expression of this specific protein in stage III tumors.

#### DISCUSSION

Relatively little is known about the regulatory mechanisms controlling the complex patterns of protein abundance and post-translational modification in tumors. Most reports concerning the regulation of protein translation have focused on

TABLE III  
Stage-dependent analysis of protein-mRNA correlation coefficients

r, correlation coefficient. Values in boldface indicate a significant difference between stage I and stage III.

Spot	Gene name	r (Stage I)	r (Stage III)	Function
1874	AKR1B1	0.269	0.106	Carbohydrate metabolism; electron transporter
2524	ANXA1	0.184	0.572	Phospholipase inhibitor; signal transduction
0994	ANXA4	0.660	0.362	Phospholipase inhibitor
0963	ANXA5	0.241	0.390	Phospholipase inhibitor; calcium binding; phospholipid binding
1314	DJ-1	0.363	0.354	Signal transduction
1405	FTL	0.126	0.358	Iron storage protein
0855	GAPD	0.243	0.581	Carbohydrate metabolism (glycolysis regulation)
0350	GRP58	0.327	<b>-0.087</b>	Signal transduction; protein disulfide isomerase
0264	HNRPK	0.360	0.243	RNA-binding protein (RNA processing/modification)
1192	HSPB3	0.457	0.633	Heat shock protein
0523	KRT18	0.115	0.371	Structural protein
0439	KRT8	0.323	0.436	Structural protein
1492	LAP18	0.483	0.663	Signal transduction; cell growth and maintenance
1638	LGALS1	0.200	0.528	Apoptosis; cell adhesion; cell size control
1252	PSMC	0.253	<b>0.060</b>	Protein degradation
1104	SFN	0.465	0.475	Signal transduction (protein kinase C inhibitor)
1454	SOD1	0.352	<b>0.079</b>	Oxidoreductase
1203	TPI1	0.378	<b>0.009</b>	Carbohydrate metabolism
0957	TPM1	0.475	0.225	Structural protein (muscle); control of heart
0593	VIM	<b>-0.054</b>	0.556	Structural protein
0935	YWHAH	0.283	0.210	Signal transduction

one or several protein products (18). Celis *et al.* (19) found a good correlation between transcript and protein levels among 40 well resolved, abundant proteins using a proteomic and microarray study of bladder cancer. By comparing the mRNA and protein expression levels within the same tumor samples, we found that 17% (28/165) of the protein spots (21/98 genes) show a statistically significant correlation between mRNA and protein. These proteins appear to represent a diverse group of gene products and include those involved in signal transduction, carbohydrate metabolism, protein modification, cell structure, heat shock, and apoptosis. These results suggest that expression of this subset of 165 proteins is likely to be regulated at the transcriptional level in these tissues. The majority of the protein isoforms, however, did not correlate with mRNA levels, and thus their expression is regulated by other mechanisms. We also observed a subset of proteins that demonstrated a negative correlation with the mRNA expression values; for example  $\alpha$ -haptoglobin demonstrated a strong negative correlation with its mRNA expression values. This may reflect negative feedback on the mRNA or the protein or the presence of other regulatory influences that are not understood currently.

Post-translational modification or processing will result in individual protein products of the same gene migrating to different locations on 2D-PAGE gels (20). Because the identity of all possible isoforms for each protein examined has not been characterized completely, this may influence the correlation analyses performed in this study. This is partly because of limitations of the 2D-PAGE and mass spectrometry technologies (21, 22). Potential inconsistencies between mRNA and protein correlations that have been reported may also be because of differences, even in the same gene, in the mech-

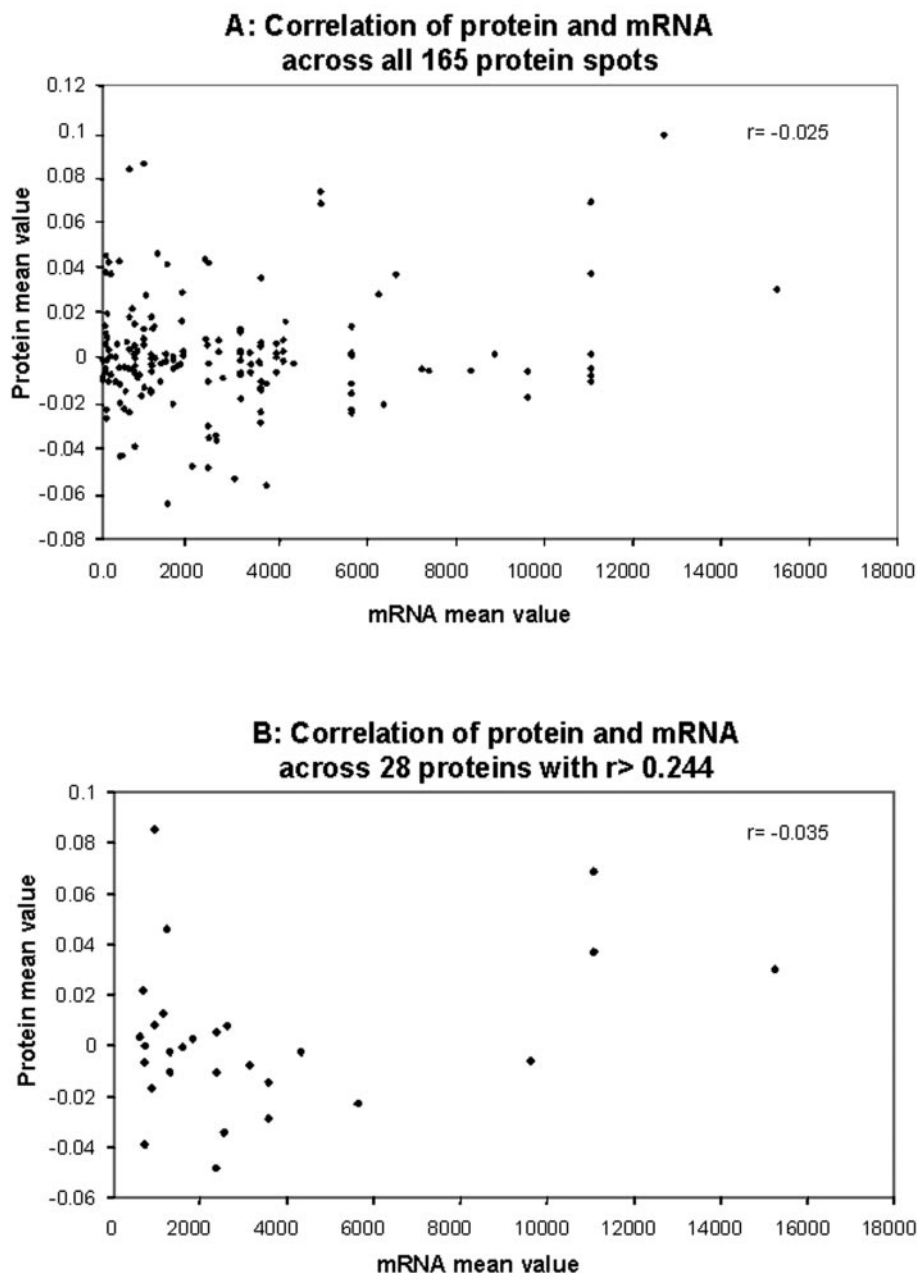
anisms of protein translation among different cells or as measured in different laboratories (23).

In this study, we examined 165 protein spots identified in lung adenocarcinomas. Ninety-six protein spots, representing the products of 29 genes, contained at least two protein isoforms. Nineteen of 96 protein spots, representing 12 genes, were shown to have a statistically significant correlation between their protein and mRNA expression, suggesting that the levels of these proteins reflects the transcription of the corresponding genes. Differences in protein/mRNA correlations were found among the individual isoforms of a given protein. For example, of the four OP18 isoforms, three showed a statistically significant correlation between the protein and mRNA expression levels. The lack of relationship for the one isoform, however, indicates that individual protein isoforms of the same gene product can be regulated differentially. This is not unexpected and likely reflects other post-translational mechanisms that can influence isoform abundance in tissues and cancer.

In addition to the analyses of the correlation of mRNA/protein within the same tumor samples, we also tested the global relationship between mRNA and the corresponding protein abundance across all 165 protein spots in the lung samples. A protein and mRNA average value for each gene was generated using all 85 lung tissues samples. We observed a very wide range of normalized average protein and mRNA values. The correlation coefficient generated using this average value data set was  $-0.025$ , and even for the 28 protein spots that showed a statistically significant correlation between individual mRNA and proteins, the correlation value was only  $-0.035$ . This suggests that it is not possible to predict overall protein expression levels based on average



FIG. 3. The overall correlation of mRNA and protein levels across all 165 protein spots (A) and across 28 protein spots that contained individual  $r$  values larger than 0.244 (B) are shown. Each protein or mRNA mean value was calculated based on all 76 lung adenocarcinomas and nine non-neoplastic lung samples using quantitative 2D-PAGE and Affymetrix oligonucleotide microarrays. The Spearman correlation coefficients for the two data sets (A and B) were  $-0.025$  and  $-0.035$ , respectively, indicating a lack of correlation if mean values for mRNA and protein for all samples is used.



mRNA abundance in lung cancer samples. This conclusion is also supported by previous results from Anderson and Seilhamer (24), who examined 19 genes in human liver cells, and by Gygi *et al.* (25), who examined 106 genes in yeast. Both studies found a lack of correlation between mRNA and protein expression when average or overall levels were used.

A good correlation was reported when the 11 most abundant proteins were examined in yeast (25), suggesting that the level of protein abundance may be a factor that may influence the correlation between mRNA and protein. In the present study, a fairly wide range of mean protein values among 165 protein spots in lung adenocarcinomas was observed, and the correlation coefficients also varied from  $-0.467$  to  $0.442$ .

A comparison between the mean value of each protein and the correlation coefficient generated using all 85 tissue samples did not reveal a strong relationship between the overall protein abundance and the correlation coefficients ( $r = 0.039$ ;  $p > 0.05$ ). Detailed analysis of different subsets of protein abundance also failed to show a correlation between mRNA and protein expression. Thus in contrast to yeast, a relationship between mRNA/protein correlation coefficient and protein abundance in human lung adenocarcinomas was not observed.

The results of this study indicate that the level of protein abundance in lung adenocarcinomas is associated with the corresponding levels of mRNA in 17% (28 proteins) of the total 165 protein spots examined. This was substantially

higher than the amount predicted to result by chance alone (which was 5.1) and suggests that a transcriptional mechanism likely underlies the abundance of these proteins in lung adenocarcinomas. We also demonstrate that the expression of individual isoforms of the same protein may or may not correlate with the mRNA, indicating that separate and likely post-translational mechanisms account for the regulation of isoform abundance. These mechanisms may also account for the differences in the correlation coefficients observed between stage I and stage III tumors, indicating that specific protein isoforms show regulatory changes during tumor progression. Further studies in lung adenocarcinomas will examine the relationship between the expression of individual protein isoforms and specific clinical-pathological features of these tumors, such as the presence of angiolymphatic invasion, and nodal or pleural surface involvement. The potential to identify specific protein isoforms associated with biological behavior in lung adenocarcinomas would be of considerable interest and will add to our understanding of the regulation of gene products by transcriptional, translational, and post-translational mechanisms.

**Acknowledgments**—We thank Kerby A. Shedden, Rork D. Kuick, Eric Puravs, Robert Hinderer, Melissa C. Krause, and Christopher Wood for assistance in this study.

\* This work was supported by NCI, National Institutes of Health Grant U19 CA-85953. 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.

‡‡ To whom correspondence should be addressed: General Thoracic Surgery, SRB II, B560, Box 0686, University of Michigan Medical School, Ann Arbor, MI 48109-086; E-mail: dgbeer@umich.edu.

## REFERENCES

- Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934
- Liang, P., Pardee, A. B. (1998) Differential display. A general protocol. *Mol. Biotechnol.* **10**, 261–267
- Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaelin, C. M., Marks, J. R., Riggins, G., Polyak, K. (2001) A sage (serial analysis of gene expression) view of breast tumor progression. *Cancer Res.* **61**, 5697–5702
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V. (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406**, 536–540
- Fung, E. T., Wright, G. L., Jr., Dalmasso, E. A. (2000) Proteomic strategies for biomarker identification: progress and challenges. *Curr. Opin. Mol. Ther.* **2**, 643–650
- Davidson, D., Baldock, R. (2001) Bioinformatics beyond sequence: mapping gene function in the embryo. *Nat. Rev. Genet.* **2**, 409–417
- Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., Fodor, S. P. (1996) Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614
- Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lander, E. S. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082
- Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P., Fodor, S. P. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5022–5026
- 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., Meyerson, M. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13790–13795
- Giordano, T. J., Shedden, K. A., Schwartz, D. R., Kuick, R., Taylor, J. M. G., Lee, N., Misek, D. E., Greenson, J. K., Kardia, S. L. R., Beer, D. G., Rennert, G., Cho, K. R., Gruber, S. B., Fearon, E. R., Hanash, S. (2001) Organ-specific molecular classification of lung, colon and ovarian adenocarcinomas using gene expression profiles. *Am. J. Pathol.* **159**, 1231–1238
- Strahler, J. R., Kuick, R., Hanash, S. M. (1989) in *Protein Structure: A Practical Approach* (Creighton, T., ed) pp. 65–92, IRL Press, Oxford
- Merril, C. R., Dunau, M. L., Goldman, D. (1981) A rapid sensitive silver stain for polypeptides in polyacrylamide gels. *Anal. Biochem.* **101**, 201–207
- Hanash, S. M., Strahler, J. R., Chan, Y., Kuick, R., Teichroew, D., Neel, J. V., Hailat, N., Keim, D. R., Gratiot-Deans, J., Ungar, D., Richardson, B. C. (1993) Data base analysis of protein expression patterns during T-cell ontogeny and activation. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3314–3318
- Brichory, F. M., Misek, D. E., Yim, A. M., Krause, M. C., Giordano, T. J., Beer, D. G., Hanash, S. M. (2001) An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9824–9829
- Lavens-Phillips, S. E., MacGlashan, D. W., Jr. (2000) The tyrosine kinases p53/56lyn and p72syk are differentially expressed at the protein level but not at the messenger RNA level in nonreleasing human basophils. *Am. J. Respir. Cell Mol. Biol.* **23**, 566–571
- Tusher, V. G., Tibshirani, R., Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5116–5121
- Tew, K. D., Monks, A., Barone, L., Rosser, D., Akerman, G., Montali, J. A., Wheatley, J. B., Schmidt, D. E., Jr. (1996) Glutathione-associated enzymes in the human cell lines of the National Cancer Institute Drug Screening Program. *Mol. Pharmacol.* **50**, 149–159
- Celis, J. E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., Thykjaer, T., Gromov, P., Yu, J., Palsdottir, H., Magnusson, N., Orntoft, T. F. (2000) Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **480**, 2–16
- Anderson, N. L., Anderson, N. G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**, 1853–1861
- Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9390–9395
- Fey, S. J., Larsen, P. M. (2001) 2D or not 2D. Two-dimensional gel electrophoresis. *Curr. Opin. Chem. Biol.* **5**, 26–33
- McBride, S., Walsh, D., Meleady, P., Daly, N., Clynes, M. (1999) Bromodeoxyuridine induces keratin protein synthesis at a posttranscriptional level in human lung tumor cell lines. *Differentiation* **64**, 185–193
- Anderson, L., Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537
- Gygi, S. P., Rochon, Y., Franz, B. R., Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730