Featured Article

Lung Cancers Detected by Screening with Spiral Computed Tomography Have a Malignant Phenotype when Analyzed by cDNA Microarray

Fabrizio Bianchi,[1,2,3] Jiangting Hu,[1] Giuseppe Pelosi,[4] Rosalia Cirincione,[5] Mary Ferguson,[1] Cathy Ratcliffe,[6] Pier Paolo Di Fiore,[2,3] Kevin Gatter,[1] Francesco Pezzella,[1] and Ugo Pastorino[5]

[1]Cancer Research UK Tumor Pathology Group, Nuffield Department of Clinical Laboratory Sciences, John Radcliffe Hospital, Oxford, United Kingdom; [2]Istituto Federazione Italiana Ricerca Cancro di Oncologia Molecolare, Milan, Italy; [3]Department of Experimental Oncology, and [4]Department of Pathology, Istituto Europeo di Oncologia, Milan, Italy; [5]Department of Thoracic Surgery, Istituto Nazionale dei Tumori, Milan, Italy; and [6]National Translational Cancer Research Network Coordinating Centre, Radcliffe Infirmary. Oxford, United Kingdom

ABSTRACT

Purpose: **Spiral computed tomography (CT) can detect lung cancer at an early stage, but the malignant potential is unknown. The question is, as follows: do these small lesions have the same lethal potential as do symptomatic tumors?**

Experimental Design: **We used a cDNA microarray platform and compared the gene expression profile of spiral CT-detected lung carcinomas with a matched case-control population of patients presenting with symptomatic lung cancer.**

Results: **CT-detected and symptomatic tumors have shown a comparable gene expression profile. Correspondence analysis has demonstrated that nine genes were differentially expressed, although with a high variability across the samples that prevented distinguishing the two groups of tumors. Analysis of these nine genes has suggested that early-detected tumors have higher levels of retinoic acid production and higher expression levels of caveolin 2, matrix Gla, and cystatin A, which are already known to be lost during tumor progression.**

Conclusions: **All of the tumors observed are histologically malignant according to the WHO Classification. Early lung cancers that are detected by screening have a gene expression pattern similar to, but not identical to, that of symptomatic lung carcinomas.**

INTRODUCTION

Spiral computed tomography (CT) has been tested as a powerful and fast technique to detect non–small cell lung cancer (NSCLC) of less than one centimeter in diameter, opening new possibilities for its early detection (1, 2). However, the efficacy of spiral CT has been questioned on the ground that earlier trials, in which radiography and sputum cytology were used, failed to reduce lung cancer mortality, possibly because they detected a substantial number of noninvasive or otherwise indolent cancers.

Because even smaller lung cancers might be detected in the future, the question as to their lethal potential will need to be properly answered to estimate the real benefit of early-detection programs in high-risk individuals (1, 3–5), and prospective clinical trials will be necessary to provide evidence of the efficacy of CT scan screening.

Along with clinical trials, cDNA microarray technology has been used to establish whether early-detected lung carcinomas have a full malignant potential or not (1, 6).

In 2000, we launched a prospective early-detection trial with spiral CT, positron emission tomography and molecular markers in a cohort of 1,035 heavy smokers, the two-year results of which have been published previously (7). In the present study, we aimed to establish whether microarray technology could identify a different genetic expression profile of CT-detected lung cancers, compared with symptomatically clinically detected tumors of similar pathological stage, whose potential lethality is recognized (8).

MATERIALS AND METHODS

Patients and Tissue Samples. A total of 1,035 volunteers ages 50 or older with a minimum of 20 pack-year index, no prior history of malignant diseases and adequate performance status, underwent annual low-dose single-slice spiral CT without contrast material. High resolution CT and three-dimensional analysis was done within 1 month in noncalcified lesions larger than 5 mm, with assessment of contrast enhancement and 18F-fluorodeoxyglucose-positron emission tomography with calcu-

Received 3/31/04; revised 5/11/04; accepted 5/18/04.
Grant support: Supported in part by Cancer Research UK and, in part, by the Italian Association for Cancer Research, and the Italian Department of Health. The microarray consortium is funded by the Wellcome Trust, Cancer Research UK, and the Ludwig Institute of Cancer Research.
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.
Note: Supplementary data for this article can be found at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org).
Requests for reprints: Francesco Pezzella, Cancer Research UK Tumor Pathology Group, Nuffield Department of Clinical Laboratory Sciences, John Radcliffe Hospital, Oxford, OX3 9DU, U.K. Phone: 44-01865-220497; Fax: 44-01865-222916; E-mail: francesco.pezzella@cancer.org.uk.

lation of standardized uptake value (SUV) in lesions of 7 mm or larger.

The first 20 consecutive CT-detected and surgically resected lung cancers were enrolled in the cDNA microarray study. The 20 controls were a consecutive series of primary lung cancers resected from 2000 to 2001, from patients presenting with a variety of respiratory and/or systemic symptoms. All these patients were not previously treated with chemoradiotherapy. Controls were matched by age, sex, smoking, cell type and tumor size.

All of the samples were snap-frozen in liquid nitrogen immediately after surgical resection and were stored at −140°C.

Sample Preparation and Microarray Hybridization. Frozen tissues were crushed with a mortar and pestle in liquid nitrogen, then transferred into TRIzol Reagent (Invitrogen Ltd, Paisley, United Kingdom) and homogenized with an Ultra-Turrax T25 (IKA Labortechik, Staufeny, Germany). Total RNA was extracted according to the TRIzol manufacturer's protocol. Total RNA, 5 μg, were amplified with Amino Allyl MessageAmp aRNA kit (Ambion, Austin, TX) and were indirectly labeled with Cy3 monofunctional dye (Amersham Biosciences United Kingdom Ltd, Buckinghamshire, United Kingdom) for the sample RNA and Cy5 monofunctional dye for the reference RNA, and then were hybridized onto the same microarray. The reference RNA that we used was the Universal Human Reference total RNA (Stratagene, Amsterdam, the Netherlands). The microarrays used in this study were human cDNA microarrays, form the Sanger Centre (Hixton, Cambridge, United Kingdom) as part of the LICR/CRUK Microarray Consortium, manufactured by depositing PCR products of cDNA clones onto three-dimensional-link–activated glass slides. Each array contains 10,750 spots representing 9,932 sequence-validated cDNA elements for 6,000 known, named human genes and/or expressed sequence tags (ESTs).[7]

The microarrays were hybridized for 16 hours in a 47°C incubator. Microarrays were washed and scanned immediately with a GenePix 4000B microarray scanner (Axon Instruments Inc., Union City, CA).

Data Analysis and Statistics. Data analysis was done with Quantarray 3.0 (Perkin-Elmer Life Sciences Inc., Boston, MA), GeneSpring 5.1 (Silicon Genetics, Redwood City, CA), and J-Express Pro (Molmine AS, Bergen, Norway). For each microarray, the median intensity of all spots in each channel was normalized to the median intensity of the spots in the control channel. The spot intensities used to calculate the median value for normalization were always background subtracted. The intensity-dependent lowess normalization was then applied (20.0% of data were used for smoothing), followed by a per-chip normalization for which all of the measurements on each array were divided by the 50th percentile of all of the measurements taken from the same array. On the log-transformed values, we applied a parametric test without assuming that variances were equal (Welch *t* test/Welch ANOVA). We used the Benjamini and Hochberg procedure (Multiple Testing Correction) to control the false discovery rate defined as the proportion of genes expected to occur by chance (assuming genes are independent) relative to the proportion of identified genes (9, 10). An agglomerative hierarchical clustering was applied with the cosine correlation and average linkage. Correspondence analysis was done with J-Express Pro software.

Quantitative Real-time PCR. The quantitative real-time PCR reactions were done on an ABI Prism 5700 Sequence Detection system (PE Biosystem, Foster City, CA). The reactions were done in triplicate for each gene (*stat1*, *osf-2*, *cav2*) including a non-template control, on a randomly chosen subset of CT-detected tumors ($n = 14$), symptomatic tumors ($n = 13$), and all 4 non-neoplastic tissues. These subsets of samples were chosen to use a single 96-well optical plate for each assay. The template cDNA was synthesized from 2 μg of total RNA, with the RETROscript kit (Ambion, Austin, TX). The predeveloped TaqMan assays used are commercially available (Assays-on-Demand gene expression products, Applied Biosystems, Foster City, CA). We used β-Actin Control Reagent (Applied Biosystems) as reference gene. PCR cycles were 40 cycles of 15 seconds at 95°C and 1 minute at 60°C.

RESULTS

The Expression Profile of 3,231 Genes in CT-Detected and Symptomatic Tumors Is Highly Similar. Two of the CT-detected cancers and one of the controls were not suitable for mRNA amplification because of the poor quality of the sample, leaving a total of 37 tumors for the analysis, including 4 non-neoplastic lung samples. Clinical and pathologic features of the two groups are illustrated in Table 1.

Data were normalized and background was subtracted. Genes found to have an absolute signal intensity of between 200 and 600 (within the range of negative spots) were excluded. Using this restriction, we got a list of 3,231 genes not biased by slide batch variability and local background. This set of genes

Table 1 Clinical and pathologic features of the two groups

		CT-detected			Non–CT-detected		
		18 cases	SD	%	19 controls	SD	%
Age	Median	59			67		
Sex	Male	12		67	15		79
Smoking	Median*	55			60		
Type	Adeno	12		67	13		68
	Squamous	5		28	5		26
pStage	1a	11		61	11		58
Diameter	Average mm	15	±8		29	±15	
Ki67	Average %	32	±22		36	±21	
EGFR	Average %	64	±39		46	±42	
Fascin	Average %	54	±41		55	±41	
Neu	>10%	5		28	6		32
p53	>10%	5		28	4		21

Abbreviation: Adeno, adenocarcinoma.
* Pack-year index.

[7] Information regarding the clone set and array preparation can be obtained from *http://www.sanger.ac.uk/Projects/Microarrays*. The technology platform for printing and processing slides can be viewed at http://www.biorobotics.com.

Table 2 Nine genes differentially expressed between screening-detected tumors and symptomatic tumors found by correspondence analysis

Gene name and description	Accession no.	% change in CT screened	% change in symptomatic	Function
cav2, Caveolin 2	NM_001233	−15	−57	Tumor suppressor
MGP, matrix Gla protein	X53331	−17.5	−50	Calcium binding; expressed at high levels in lung; structural protein of bone; extracellular matrix structural protein
Stf1, cystatin A	X05978	−8	−37.5	Cysteine protease inhibitor; such peptidases are increased in cancer metastasis and inflammation
adh2, alcohol dehydrogenase	D00137	−6	−46	Zinc binding; electron transporter; expressed and active in normal lung; retinoic acid synthesis
aldh1, aldheyde dehydrogenase 1	K03000	61	−7	Electron transporter; retinoic acid synthesis
adamdec1, ADAM-like, decysin 1	Y13323	149	80	Disintegrin protease; negative regulation of cell adhesion
Vcam	M30257	138.5	81	Cell adhesion; expressed by cytokine-activated endothelium; mediates the adhesion of monocytes and lymphocytes
HIF2A, hypoxia-induced factor 2a	AF052094	−19	−52.5	Positive regulation of angiogenesis
sfrp2, secreted frizzled-related	AF311912	111	18	Transmembrane receptor for Wnt family members; protects from apoptosis

was analyzed with a parametric *t* test on log-transformed data (*P* value cutoff of 0.01 and multiple testing correction) to seek differences between the CT-screened and the symptomatic tumors. No differentially expressed genes were found between the two groups.

Correspondence Analysis Identifies Nine Genes Differentially Regulated between CT-Detected and Symptomatic Cases. Bearing in mind a possible statistical bias in the analysis above, we used correspondence analysis to reanalyze the spiral CT-detected tumors and symptomatic cases. This method creates a low-dimensional projection from an original high-dimensional data set, revealing interdependencies between two variables (the genes and the conditions of the experiment; refs. 11, 12). Initially we filtered the data with a SD cutoff of ≥ 1 and discarded all those genes that did not pass the filter in all of the three conditions (CT-detected, symptomatic, and normal). From this new list of 2,918 genes, we calculated the median value for every gene in each group (CT-detected, symptomatic, and normal). All of the genes with a SD of <0.5 across the groups were eliminated; this restriction identifies genes with a substantial change in at least two of the three conditions. The results provided a list of 97 genes: 88 genes were differentially expressed between tumors and non-neoplastic samples, with no essential distinction between the asymptomatic (CT-detected) and the symptomatic tumors. The remaining nine genes (listed in Table 2) were differentially regulated in their median values between the CT-detected and the symptomatic specimens, but were still unable to divide into separate clusters the asymptomatic and symptomatic lung cancer cases (data not shown).

Some of these nine genes showed a similar expression in normal tissues *versus* CT-detected tumors but were down-regulated in the symptomatic tumors (see Table 2).

Tumor *versus* Normal Tissue. When we introduced the group of 4 normal samples and did a parametric ANOVA, *P* value 0.01 and multiple testing corrections ($\sim$1.0% of the identified genes would be expected to pass this restriction by chance), we found 319 genes (see list in Supplementary data 1)[8] be differentially expressed between the tumors (CT-screened and symptomatic) and the normal samples. Using this list of genes, we did a hierarchical clustering that failed again to define clusters that could separate the CT-screened or symptomatic samples. On the contrary, the 4 non-neoplastic samples clustered tightly together, and they were separated from all of the other samples (see image in Supplementary data 2).[8,9]

Because some of the early-detected tumors (and, therefore, the correspondent matched controls) were larger than stage 1A, we restricted our analysis to 11 stage-1A asymptomatic cases and the matched 11 stage-1A symptomatic cases, and we did not find any genes differentially expressed. We then analyzed the 4 tumors we had that were smaller than 10 mm *versus* those larger-than-or-equal-to 10 mm and still no difference was found.

On the contrary, we defined a set of 113 genes specific for the histologic type of tumor (data not shown), among which the strongest differences in gene expression were related to adenocarcinomas with predominant bronchioloalveolar features. The subset of squamous tumors showed a higher clustering distance than did the other specimens. The non-neoplastic samples appeared as a well-defined subcluster, alongside the larger cluster of adenocarcinomas.

Among the 319 genes differentially expressed between tumors and normal tissues, several were related to metabolic pathways, indicating that cells were active and dividing (Table 3). In addition, we also identified three genes *Znf198*, *Stat1*, and *Osf-2*, that were highly up-regulated in the tumor samples. *Znf198* was fused with FGFR1 in the chromosome translocation

[8] Supplementary Data 1, List of the 319 genes differentially expressed between normal tissue and tumors, can be found at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org).

[9] Supplementary Data 2, Agglomerative unsupervised hierarchical clustering of the 319 genes differentially expressed between normal tissue and tumors, can be found at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org).

Table 3 Selection of genes differentially expressed between normal lung and NSCLC tumors

Gene name	Accession no.	*P*	% change in tumor	Molecular function
paics	X53793	8.46E-05	65	phosphoribosylaminoimidazole-succinocarboxamide synthase
gpat	U00238	4.91E-05	49	amidophosphoribosyltransferase
hprt	M31642	2.94E-05	76	hypoxanthine phosphoribosyltransferase
gfat	M90516	0.009445	52	glutamine amidotransferase; glutamine-fructose-6-phosphate transaminase
prsi	D00860	0.008419	47	ribose-phosphate pyrophosphokinase
ldh1	X02152	8.09E-06	103	L-lactate dehydrogenase
p450c11	X55764	0.001824	−26	steroid 11-β-monooxygenase
ptsg1	U63846	0.003188	31	prostaglandin-endoperoxide synthase
pik3ca	Z29090	0.002801	56	phosphoinositide-3-kinase, catalytic, α polypeptide
tyms	X02308	0.003953	43	deoxyribonucleoside monophosphate biosynthesis; nucleobase, nucleoside, nucleotide and nucleic acid metabolism
znf198	AF060181	2.84E-06	109	protein binding
stat1	NM_007315	4.69E-08	206	DNA binding; transcription factor; transcription activating factor; hematopoeitin/interferon-class (D200-domain) cytokine receptor signal transducer
osf-2	D13665	0.002515	299	cell adhesion; osteoblast specific factor 2 (fasciclin I-like)
ube2b	NM_003337	0.009466	−35	ubiquitin conjugating enzyme
ezf, gklf	AF105036	4.60E-06	−49	transcription factor; transcription activating factor

t(8;13)(p11;q11–12) specifically found in patients afflicted with the stem cell leukemia/lymphoma (SCLL) syndrome. The fusion gene resulting from this translocation has been shown to work as a constitutively active protein tyrosine kinase phosphorylating STAT1, -3, and -5, but not STAT4 or -6 (13, 14). *Stat 1* is located in the gene cluster with *Znf198* in the 319 gene list. Other genes, also found to be down-regulated in the tumors, included *ube2b* (homologue of yeast RAD6 involved in DNA repair) and *klf4* (Kruppel like factor 4). *klf4* has been shown to be an essential mediator of p53 in controlling the G_1-S progression phase of the cell cycle after DNA damage (15, 16).

Furthermore, we considered only the adenocarcinomas in the cohort of patients. We reduced the dataset to 13 unscreened and 10 CT-detected adenocarinomas (the two bronchioloalveolar carcinomas were also excluded). We did not find any genes differentially regulated (*P* value < 0.01) among the 3,231-gene list. Moreover, if we reduce the main gene list to 75 genes with an average-fold change cutoff of 1.5 (calculated between the two groups) and then apply the Welch *t* test analysis (about less than 1 gene would be expected to pass the restriction by chance.), we found only five unique genes to be regulated, (*CREG*, *ALDH1*, *CD59*, *WTAP*, and *HIF2A*). Two of these genes, *HIF2A* and *ALDH1*, were already reported to be regulated by the correspondence analysis. This confirms the high similarity of the two groups even if we consider only the adenocarcinomas, ruling out the possibility that the variability introduced by different histologic subtypes can mask a bigger gene expression signature able to clearly divide the two cohorts of patients. Consistently, with the previous analysis, we found 153 genes differentially expressed between normal and all of the adenocarcinomas. In addition, 143 of the 153 genes were within the 319-gene list of genes differentially expressed between all of the tumors and the normal lung.

Confirmation by Real-time PCR. The regulation identified by the microarray analysis was confirmed by real-time PCR for three genes (*cav2, stat1*, *osf-2*; Fig. 1*A*). Moreover, *cav2*, which was originally selected by correspondence analysis, was confirmed by real-time PCR to have a similar profile of expression in the CT-screened tumors as in the non-neoplastic tissues (Fig. 1*B*).

Immunohistochemical Studies. In keeping with the microarray profiles, immunohistochemical analysis showed no difference in the expression of Ki67, p53, EGFR, Neu, or Fascin between spiral CT-detected and unscreened tumors (Table 1).

DISCUSSION

Our results showed that, as far as their mRNA expression pattern was concerned, the CT-detected tumors were not distinguishable from symptomatic cases. This result holds also when we limited the analysis to tumors at stage pT_{1A} and to adenocarcinomas (data not shown).

The lack of a marked difference in gene expression profile between smaller lesions (<10 mm) and the larger tumors in these series suggests that the former have already switched at the molecular level to a malignant phenotype.

In contrast, neoplastic and non-neoplastic samples clustered separately with a marked difference in gene expression profiles (Fig. 1), as published previously (17, 18).

We compared the list of 675 genes published by Bhattacharjiee et al. (17) and were able to distinguish adenocarcinoma subtypes and NSCLC histologic types among the 133 genes found by us. We used SOURCE database to map the two lists to Unigene clusters[10] and then compared for overlapping. We found only seven overlapped genes (*JUN*, *STAT1*, *ADE2H1*, *SDR1*, *CX43*, *SSP1*, and *AF1Q*). This is mainly due to a different array platform used (Affymetrix single-channel *versus* double-channel cDNA array) and has been described previously (19).

Microarray profiles were consistent with immunohistochemical analyses, in which CT-detected tumors showed similar expressions of Ki67, p53, EGFR, Neu, and Fascin. These results

[10] Internet address: (http://genome-www5.stanford.edu/cgi-bin/source/sourceBatchSearch.

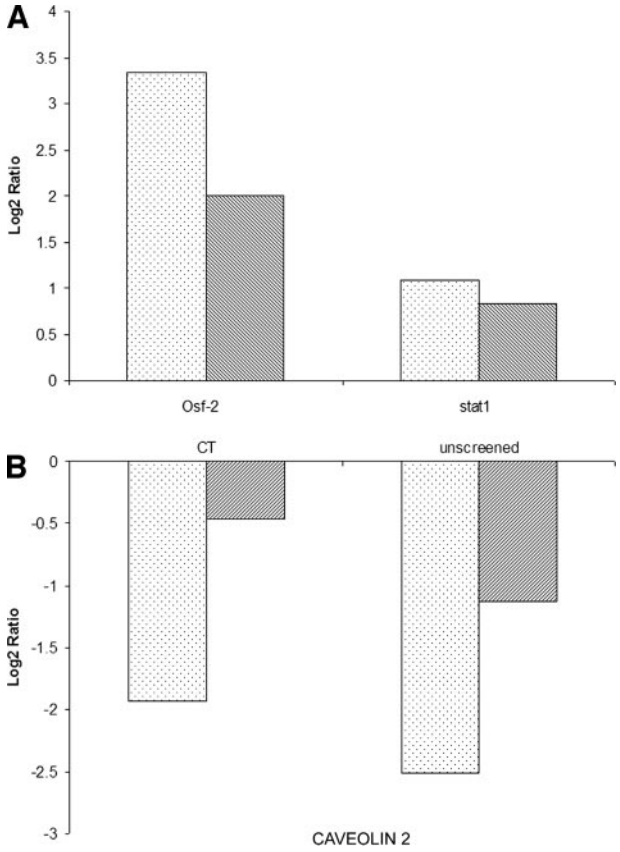

Fig. 1 Validation by real-time PCR. *A*, real-time PCR and cDNA microarray results comparison. The scale in the Y axes stands for the logarithm base 2 of the ratio between the mean of the gene expression values of the tumor samples ($n = 23$) and the mean of the gene expression values of all the non-neoplastic samples ($n = 4$). Two-fold of difference in the expression of a gene correspond to a unit in the Y axes. ▧, array; ▨, real time. *B*, real-time PCR and cDNA microarray results comparison. The scale in the Y axes, limited to the two bars labeled CT in the X axes, stands for the logarithm base 2 of the ratio between the mean of the gene expression values of the CT-detected tumors ($n = 14$) and the mean of the gene expression values of all the non-neoplastic samples ($n = 4$). Limited to the two bars labeled unscreened in the X axes, the scale in the Y axes stands for the logarithm base 2 of the ratio between the mean of the gene expression values of the symptomatic tumors ($n = 13$) and the mean of the gene expression values of all the non-neoplastic samples ($n = 4$). Two-fold of difference in the expression of a gene correspond to a unit in the Y axes. ▨, *array*; ▨, real time.

further support the hypothesis that CT-detected tumors have a malignant phenotype.

We also showed that, with a correspondence analysis, nine genes were differentially expressed across the samples between the CT-detected and symptomatic tumors (Table 1). Although it was not able to cluster the two types of patients into distinct groups because of the high variability, this analysis suggests some ongoing changes in the early phases of lung cancer. All of these nine genes were expressed in the symptomatic tumors at a lower level than in the CT scan-detected cases and loss of expression of five of these genes is known to be associated with tumor progression.

Caveolin 2 is involved, with caveolin 1, in the formation of caveolae which are pivotal in the storage and inactivation of signaling molecules (*e.g.*, EGFR; refs. 20–22), making caveolin 2 an important negative regulator of cell proliferation and tumor genesis. The down-regulation of caveolin 1 in NIH 3T3 cell lines was shown to be sufficient to hyperactivate the p42/44 mitogen-activated protein kinase cascade and to induce tumors in immunodeficient mice (23).

Matrix Gla (MGP) mRNA has been found to be down-regulated in colon adenocarcinoma compared with normal mucosa (24). In metastatic renal cell carcinoma and prostatic carcinoma, the loss of MGP expression compared with primary tumors has been associated with tumor progression and metastasis (25). Our data support these findings as we found lower levels in the tumors from symptomatic patients. However, this is probably not a general principle because levels of MGP have been found to be increased in relapsing gliomas (26).

We confirm that cystatin A may have a tumor-suppressor function (27) because it is already known to be expressed in a variety of normal tissues, but it is lost during the clinical course of a variety of malignant diseases, such as prostatic adenocarcinoma and squamous carcinomas of skin and esophagus (28–30).

The other two genes known to be lost during lung cancer progression were *ADH2* and *ALDH1*. They belong, respectively, to the alcohol dehydrogenase and aldehyde deydrogenase families, *ADH2* being expressed in normal lung (31). Genes belonging to these two families are usually regulated in parallel. *ADH2* catalyzes the transformation of all-*trans* retinol and 9-*cis* retinol to all-trans retinal and 9-*cis* retinal, and *ALDH1* transforms these latter to all-*trans* retinoic acid and 9-*cis* retinoic acid, the two retinoids involved in gene regulation (32). Our findings suggest that, during the early phases of lung cancer, there is a progressive loss in the production of retinoic acid within the neoplastic cells.

We also found higher levels of expression in CT scan-detected tumors of VCam and ADAMDEC1 (the first described member of a new subclass of mammalian disintegrin metalloproteinases expressed on dendritic cells and macrophages; 33). This suggests that, in the earlier stages of these lung tumors, there is more intense inflammation, possibly induced by an increased recruitment of macrophages. The higher levels of HIF2a observed support this hypothesis further, because this molecule is predominantly expressed by macrophages.

Beer *et al.* (18) have described a set of gene those expression levels are able to identify Stage 1 lung Adenocarcioma at high risk of relapse. Non of the nine genes differentially expressed between CT-detected and symptomatic tumors was present in this list. However two of our genes (aldeheyde dehydrogenase 1 and cystatin A) belonged to the same families of two genes present in the top 100 genes described by Beer *et al.* (aldeheyde dehydrogenase 8 and cystatin B), suggesting that these two families are relevant to lung cancer progression.

All of the tumors observed are histologically malignant according to the WHO classification. Early lung cancers detected by screening have a gene expression pattern similar but not identical to that of symptomatic lung carcinomas

The ongoing trials (2, 7) will tell us whether early lung cancer detection can prevent the occurrence of advanced and incurable lung cancer in these high-risk individuals, by stopping

the metastatic progression of very small tumors with already malignant genetic phenotype.

ACKNOWLEDGMENTS

We would like to thank the staff of the Sanger Institute Microarray Facility (http://www.sanger.ac.uk/Projects/Microarrays) for the supply of arrays, laboratory protocols, and technical advice (David Vetrie, Cordelia Langford, Adam Whittaker, Neil Sutton), Quantarray/GeneSpring datafiles and all data analysis and databases relating to elements on the arrays (Kate Rice, Rob Andrews, Adam Butler, Harish Chudasama). The human IMAGE cDNA collection was obtained from the MRC HGMP Resource Centre (Hinxton, United Kingdom). All cDNA clones resequencing was performed by Team 56 at the Sanger Institute. We also wish to thank Julian Downward, John Sgourous, and Simon Tomlinson (Cancer Research United Kingdom, Lincoln's Inn Field, London) for their continuous support.

REFERENCES

1. Mulshine JL. Screening for lung cancer: in pursuit of pre-metastatic disease. Nat Rev Cancer 2003;3:65–73.

2. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999;354:99–105.

3. Bach PB, Kelley MJ, Tate RC, McCrory DC. Screening for lung cancer: a review of the current literature. Chest 2003;123:72S–82S.

4. Mahadevia PJ, Fleisher LA, Frick KD, Eng J, Goodman SN, Powe NR. Lung cancer screening with helical computed tomography in older adult smokers: a decision and cost-effectiveness analysis. JAMA 2003; 289:313–22.

5. Miettinen OS, Henschke CI. CT screening for lung cancer: coping with nihilistic recommendations. Radiology 2001;221:592–6;discussion,7.

6. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (Wash DC) 1995;270:467–70.

7. Pastorino U, Bellomi M, Landoni C, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. Lancet 2003;362:593–7.

8. Motohiro A, Ueda H, Komatsu H, Yanai N, Mori T. Prognosis of non-surgically treated, clinical stage I lung cancer patients in Japan. Lung Cancer 2002;36:65–9.

9. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9:811–8.

10. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 2003;19:368–75.

11. Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M. Correspondence analysis applied to microarray data. Proc Natl Acad Sci USA 2001;98:10781–6.

12. Kishino H, Waddell PJ. Correspondence analysis of genes and tissue types and finding genetic links from microarray data. Genome Inform Ser Workshop Genome Inform 2000;11:83–95.

13. Baumann H, Kunapuli P, Tracy E, Cowell JK. The oncogenic fusion protein tyrosine kinase ZNF198/fibroblast growth factor receptor-1 has signaling function comparable to interleukin-6 cytokine receptors. J Biol Chem 2003;278:16198–208.

14. Xiao S, Nalabolu SR, Aster JC, et al. FGFR1 is fused with a novel zinc-finger gene, ZNF198, in the t(8;13) leukaemia/lymphoma syndrome. Nat Genet 1998;18:84–7.

15. Chen X, Whitney EM, Gao SY, Yang VW. Transcriptional profiling of Kruppel-like factor 4 reveals a function in cell cycle regulation and epithelial differentiation. J Mol Biol 2003;326:665–77.

16. Yoon HS, Chen X, Yang VW. Kruppel-like factor 4 mediates p53-dependent G1/S cell cycle arrest in response to DNA damage. J Biol Chem 2003;278:2101–5.

17. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA 2001;98:13790–5.

18. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002; 8:816–24.

19. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 2002;18:405–12.

20. Li S, Okamoto T, Chun M, et al. Evidence for a regulated interaction between heterotrimeric G proteins and caveolin. J Biol Chem 1995;270:15693–701.

21. Oka N, Yamamoto M, Schwencke C, et al. Caveolin interaction with protein kinase C. Isoenzyme-dependent regulation of kinase activity by the caveolin scaffolding domain peptide. J Biol Chem 1997;272: 33416–21.

22. Li S, Seitz R, Lisanti MP. Phosphorylation of caveolin by src tyrosine kinases. The alpha-isoform of caveolin is selectively phosphorylated by v-Src in vivo. J Biol Chem 1996;271:3863–8.

23. Galbiati F, Volonte D, Engelman JA, et al. Targeted downregulation of caveolin-1 is sufficient to drive cell transformation and hyperactivate the p42/44 MAP kinase cascade. EMBO J 1998;17:6633–48.

24. Fan C, Sheu D, Fan H, Hsu K, Allen Chang C, Chan E. Downregulation of matrix Gla protein messenger RNA in human colorectal adenocarcinomas. Cancer Lett 2001;165:63–9.

25. Levedakou EN, Strohmeyer TG, Effert PJ, Liu ET. Expression of the matrix Gla protein in urogenital malignancies. Int J Cancer 1992; 52:534–7.

26. van den Boom J, Wolter M, Kuick R, et al. Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. Am J Pathol 2003;163:1033–43.

27. Abrahamson M, Alvarez-Fernandez M, Nathanson CM. Cystatins. Biochem Soc Symp 2003;70:179–99.

28. Luo A, Kong J, Hu G, et al. Discovery of Ca(2+)-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. Oncogene 2004;23:1291–9.

29. Mirtti T, Alanen K, Kallajoki M, Rinne A, Soderstrom KO. Expression of cystatins, high molecular weight cytokeratin, and proliferation markers in prostatic adenocarcinoma and hyperplasia. Prostate 2003;54:290–8.

30. Palungwachira P, Kakuta M, Yamazaki M, et al. Immunohistochemical localization of cathepsin L and cystatin A in normal skin and skin tumors. J Dermatol 2002;29:573–9.

31. Smith M, Hopkinson DA, Harris H. Studies on the subunit structure and molecular size of the human alcohol dehydrogenase isozymes determined by the different loci, ADH1, ADH2, and ADH3. Ann Hum Genet 1973;36:401–14.

32. Duester G. Families of retinoid dehydrogenases regulating vitamin A function: production of visual pigment and retinoic acid. Eur J Biochem 2000;267:4315–24.

33. Mueller CG, Rissoan MC, Salinas B, et al. Polymerase chain reaction selects a novel disintegrin proteinase from CD40-activated germinal center dendritic cells. J Exp Med 1997;186:655–63.