Genomic markers for malignant progression in pulmonary adenocarcinoma with bronchioloalveolar features

Sarit Aviel-Ronen*, Bradley P. Coe^{†‡}, Suzanne K. Lau^{*§}, Gilda da Cunha Santos^{*¶}, Chang-Qi Zhu*, Dan Strumpf*, Igor Jurisica^{*§||}, Wan L. Lam^{†‡}, and Ming-Sound Tsao^{*§¶**}

*University Health Network, Ontario Cancer Institute and Princess Margaret Hospital Site, Toronto, ON, Canada M5G 2M9; [†]Department of Cancer Genetics and Developmental Biology, British Columbia Cancer Research Centre, Vancouver, BC, Canada V5Z 1L3; Departments of [§]Medical Biophysics, [¶]Laboratory Medicine and Pathobiology, and ^{||}Computer Science, University of Toronto, Toronto, ON, Canada M5G 2C1; and [‡]University of British Columbia, Vancouver, BC, Canada V6T 289

Edited by John D. Minna, University of Texas Southwestern Medical Center, Dallas, TX, and accepted by the Editorial Board April 30, 2008 (received for review October 9, 2007)

Bronchioloalveolar carcinoma (BAC), a subtype of lung adenocarcinoma (ADC) without stromal, vascular, or pleural invasion, is considered an *in situ* tumor with a 100% survival rate. However, the histological criteria for invasion remain controversial. BAC-like areas may accompany otherwise invasive adenocarcinoma, referred to as mixed type adenocarcinoma with BAC features (AWBF). AWBF are considered to evolve from BAC, representing a paradigm for malignant progression in ADC. However, the supporting molecular evidence remains forthcoming. Here, we have studied the genomic changes of BAC and AWBF by array comparative genomic hybridization (CGH). We used submegabase-resolution tiling set array CGH to compare the genomic profiles of 14 BAC or BAC with focal area suspicious for invasion with those of 15 AWBF. Threshold-filtering and frequency-scoring analysis found that genomic profiles of noninvasive and focally invasive BAC are indistinguishable and show fewer aberrations than tumor cells in BAC-like areas of AWBF. These aberrations occurred mainly at the subtelomeric chromosomal regions. Increased genomic alterations were noted between BAC-like and invasive areas of AWBF. We identified 113 genes that best differentiated BAC from AWBF and were considered candidate marker genes for tumor invasion and progression. Correlative gene expression analyses demonstrated a high percentage of them to be poor prognosis markers in early stage ADC. Quantitative PCR also validated the amplification and overexpression of *PDCD6* and *TERT* on chromosome 5p and the prognostic significance of *PDCD6* in early stage ADC patients. We identified candidate genes that may be responsible for and are potential markers for malignant progression in AWBF.

array comparative genomic hybridization | bronchioloalveolar carcinoma | microarray | non-small-cell lung carcinoma | prognostic markers

Lung adenocarcinoma (ADC) accounts for $\approx 35\%$ of all lung cancers and has an overall 5-year survival of 17% (1). The recent World Health Organization (WHO) classification recognized a particular subtype, bronchioloalveolar carcinoma (BAC), for its noninvasive features and excellent prognosis (2). BAC has a distinct histological pattern of tumor cells growing along preexisting alveolar framework without evidence of stromal, pleural, or vascular invasion. Yet, some invasive ADC, classified as mixed type, may have components or large areas of BAC-like pattern. Multistage development of adenocarcinoma putatively involves progression from atypical adenomatous hyperplasia (AAH) through BAC to invasive mixed type ADC with BAC features (AWBF) (3–5). Mice that express oncogenic *KRAS* develop histological changes that range from mild hyperplasia/dysplasia analogous to atypical adenomatous hyperplasia to alveolar adenomas and ultimately display overt ADC (6, 7). BAC-associated tumors have gained significant attention for their potentially greater sensitivity to treatment by

epidermal growth factor receptor (EGFR) inhibitors (8). The initial studies that recognized BAC as a distinct entity reported 5-year survival rates of 100% (9–11), but more recent studies have reported lower 5-year survival rates of 83–86% for resected stage I BAC patients (12–14). These rates possibly reflect difficulties in applying the histological criteria of invasion in BAC or AWBF. Some studies have also reported that BAC with focal areas of microinvasion may also have excellent prognosis similar to noninvasive BAC (11, 15). The identification of genes/proteins that may distinguish BAC from AWBF and are predictors of ADC with poor prognosis could be useful for the establishment of molecular pathological classification of lung ADC. In this study, we have used array comparative genomic hybridization (CGH) to test our hypothesis that BAC is molecularly distinguishable from AWBF by their differential genomic profiles and that marker genes for invasion and/or poor prognosis may be identified.

Results

Most chromosomal changes in both BAC and AWBF were subtle, indicating low levels of genomic alteration as well as partial attenuation by contaminating nonneoplastic host cells. The profiles of BAC and BAC with focal areas suspicious for invasion were indistinguishable and showed low copy gains at 1p, 2q, 5p, 7p, 11p, 11q, 12q, 16p, 16q, 17q, 20q, and 21q (Fig. 1B). Copy gains typically occurred at the subtelomeric regions. AWBF had similar chromosomal changes but with greater variability and frequency and longer segmental alterations. Deletions were also more common in AWBF and were observed mainly on 3p and 5q and to a lesser extent on 4q and 6q. In two patients with synchronous BAC and invasive AWBF, the BAC-like area of the latter showed greater aberrations than the BAC (Fig. 2A). In two other AWBF, greater alterations were also noted in BAC-like areas compared with invasive areas (Fig. 2B). Normal lung samples showed no alteration of these regions.

Author contributions: W.L.L. and M.-S.T. designed research; S.A.-R., B.P.C., S.K.L., and G.d.C.S. performed research; S.A.-R., B.P.C., S.K.L., and I.J. contributed new reagents/analytic tools; S.A.-R., B.P.C., S.K.L., G.d.C.S., C.-Q.Z., and D.S. analyzed data; and S.A.-R., B.P.C., S.K.L., C.-Q.Z., D.S., and M.-S.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.D.M. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE11945).

**To whom correspondence should be addressed at: Ontario Cancer Institute, Toronto, ON, Canada M5G 2M9. E-mail: ming.tsao@uhn.on.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0709618105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

There have been many attempts to classify lung carcinomas at the molecular level through various techniques, including metaphase CGH studies (16), gene expression (17–19), and array CGH analyses (20). Our whole-genome submegabase array study shows that BAC is characterized by alterations at subtelomeric chromosomal regions. Such aberrations could be explained mechanistically by the breakage/fusion/bridge (BFB) cycle (21). Normally telomeres protect chromosome ends and prevent their fusion; however, telomere loss may lead to chromosome instability and cyclic fusion, the formation of a chromosomal bridge and its breakage in proximity to the site of initial fusion, of sister chromatids during replication. This self-perpetuating process resolves through net gain of telomere by translocation of the ends of another chromosome, by small subtelomeric duplications of the end of the same chromosome, or by direct telomere addition. Consequently, gene amplification and

Fig. 4. *TERT* validation by QPCR and FISH. (A) *TERT* content measured by array CGH and QPCR on genomic DNA (relative to normal control) and FISH (mean gene copy number per nucleus) show high correlation between the different methods. The black lines connecting copy number of *TERT* (filled circles) and 5q (open squares) are drawn to highlight the difference in copy number between the two probes. The blue boxes mark the AWBF sampled in invasive areas. Samples T41A, T44A, and T46A represent the BAC-like area, and T41B, T44B, and T46B represent the invasive area of AWBF. (B) FISH performed on AWBF in BAC area (sample T195) using the dual-color FISH probe mix that contains the *hTERT* locus (5p15, green signal) and the control D5589 probe (5q31, red signal). Gain of *TERT* is reflected by the increased number of green compared with red foci.

deletions occur preferentially at the subtelomeric chromosomal regions. BFB also predicts progressive accumulation of genetic alterations, as observed in both BAC and AWBF. It has been shown that, as BFB progresses and more chromosomal abnormalities accumulate, the breakpoints are more interstitial (21).

The differential genomic changes noted between BAC and invasive AWBF provide important evidence for a better understanding of the pathogenesis of ADC. We have used two independent algorithms to enhance the certainty of the profile that distinguishes BAC from invasive AWBF. Our inability to clearly differentiate BAC from BAC with focal area of invasion at the genomic level suggests that both may have a similar behavior with low metastatic potential and that early invasion is likely determined at gene expression levels by epigenetic mechanisms. The finding also suggests that BAC or BAC with focal invasion, which are negative for the overexpression of identified marker genes, could potentially be grouped into a single diagnostic entity with excellent prognosis (11, 15).

The 113 candidate marker genes that we identified may represent part of the “signature of chromosomal instability” for invasion and malignant progression in AWBF (22). The correlative gene expression validation rate ($\approx 35\%$) in the Harvard and Michigan datasets was limited by the low number of probe sets in the microarray

platform that matched the genomic gene list (less than half). Nevertheless, it confirms the importance of some of our candidate markers in lung carcinoma (Table S1) and the overexpression of others such as *SAR1A* (23), *SYCP1* (24), and *MCM7* (22), which have been linked to other malignancies as well as lung cancer. We have previously reported the poor prognostic significance of *TERT* gene amplification in NSCLC (25). Our present findings extend the importance of *TERT* amplification to AWBF and increased *TERT* gene copy because of chromosome 5 polysomy.

PDCD6, programmed cell death 6, or apoptosis-linked gene 2 (*ALG-2*) is located on chromosome 5pter–5p15.2 and is in close proximity to *TERT*. It encodes a 191-aa protein that was originally considered pro-apoptotic (26). *PDCD6* belongs to the penta-EF hand Ca^{2+} -binding protein family (27) and is ubiquitously expressed in the body. *PDCD6* is required for T cell receptor-, glucocorticoid- (26), and FAS- (28) induced cell death. It interacts with the SH3-binding domain containing pro-apoptotic protein AIP1 (ALG-2-interacting protein-1) (29), peflin (30), and annexin XI (31) in a Ca^{2+} -dependent way as well as with DAPK1 (death-associated protein kinase 1) (32). During FAS-induced apoptosis, *PDCD6*, which is a 22-kDa protein, is cleaved in its N-terminus to yield a 19-kDa protein and translocates from the cytoplasmic membrane to the cytosol (28). More recent work questioned the need of *PDCD6* for apoptosis, as it may be compensated by other functionally redundant proteins (33). Immunohistochemical staining has revealed high expression of *PDCD6* in primary tumors compared with normal tissues of the breast, liver, and lung (34, 35). Both nuclear and cytoplasmic overexpression have been reported for lung cancer, especially metastatic ADC, indicating that it plays a role in survival pathways (35). We confirm that *PDCD6* is significantly overexpressed in lung ADC (35). Moreover, we have also demonstrated that *PDCD6* is a poor prognostic factor in both early stage NSCLC and ADC and thus may serve as one of the markers to differentiate more indolent from aggressive AWBF.

Potti *et al.* (19) reported a genomic strategy to refine prognosis for early stage NSCLC and identify patients at high risk of relapse after initial surgery. They constructed a lung metagene model based on gene expression data and showed that its prognostic accuracy surpasses that of a model based on traditional clinical data. Their model was applied to all histologic types of early stage disease but did not consider BAC as a special entity. Although none of the 122 genes in the published metagenes matched our 113 genes, analysis of our genes in their dataset showed that the overexpression of 27 genes (24%) was associated with poor prognosis in early stage ADC patients. Significantly higher gene copy number in AWBF compared with BAC was confirmed by QPCR on genomic DNA in 74% of these genes (20 of 27 genes, Table S2).

Table 1. Correlative gene expression validation results for the candidate amplified genes

	Toronto	Harvard (17)	Michigan (18)
No. of ADC samples	39	127	86
No. of normal lung samples	10	17	10
Array type	U133A	U95A	HuGeneFL
No. of genes/probe sets	13,840/22,215	9,513/12,625	5,945/7,129
No. of genes/probe sets up-regulated	4,885/6,251	3,175/3,635	1,692/1,909
No. of genes/probe sets down-regulated	3,076/3,998	1,887/2,119	929/946
% of genes up-regulated	35.29	33.37	28.46
% FDR used	4.06	3.98	4.98
No. of genes from 113 gene list that are present on array	87	59	42
No. of up-regulated genes that match 113 gene list	38	22	15
No. of genes that match by mistake based on FDR	1.54	0.88	0.75
Expected no. of up-regulated genes based on observed % of up-regulated genes	30	19	12
Validation rate, %	41.90	35.80	33.93

FDR, false discovery rate.

Validation by RT-QPCR. Gene copy numbers were evaluated for the DNA used in the array CGH studies by RT-QPCR, using primer sets for target and housekeeping genes. The evaluation of 33 genes, including *TERT* and *PDCD6*, was performed on all of the array CGH samples except two BACs (Tables S2 and S6). The mRNA expression study was carried out on two groups of samples: 10 pairs of matched ADC and their adjacent normal lung tissue and 85 NSCLC samples. Primer sets design, reaction conditions and analysis description, and patients' demographic information are included in *SI Text* and Tables S7 and S8.

Validation by FISH. The 21 cases studied by FISH included 7 BAC with or without suspicion for invasion and 14 AWWF; 3 of the latter were scored in both their BAC and invasive areas. An additional case of AWWF was scored only in the invasive area. Among these cases was one with synchronous BAC and invasive AWWF sampled from the BAC area. FISH failed in 6 samples. The FISH protocol is detailed in the *SI Text*.

The Toronto DNA Microarray Dataset. RNA was extracted by the phenol/chloroform method from 39 adenocarcinomas (Table S9) and 10 normal lung tissue samples. RNA quality was assessed by gel electrophoresis and Bioanalyzer (Agilent). cRNA synthesis, hybridization, and scanning were performed according to the manufacturer's protocol. The adenocarcinoma RNA was profiled on the U133A chip (Affymetrix) and the normal lung RNA on the U133A2 chip (Affymetrix). To ensure the compatibility of these two platforms, four of the 39 adenocarcinomas were reprofiled on the U133A2 chip.

Correlative Gene Expression Study. We validated the 113 amplified genes and the 149 deleted genes from array CGH analysis on the Toronto microarray dataset and on two publicly available lung cancer microarray expression datasets (17, 18),

referred to as "Harvard" and "Michigan," respectively. For a detailed description of the analytic process and a summary of the validation see the *SI Text* and Tables S1 and S4.

In addition, univariate analysis was performed on microarray expression data of stage I ADC patient samples from a third dataset referred to as "Duke" (19) to identify prognostic markers and compare them with the 113 candidate markers, as detailed in the *SI Text* and Table S7.

Statistical Analysis. The Mann-Whitney test was used to compare the genomic copy number of 33 genes including *TERT* and *PDCD6* (Table S2). Pearson correlation coefficients assessed the correlation between array CGH, QPCR, and FISH results. The Wilcoxon signed rank test was used to compare *PDCD6* expression in the paired ADC-normal samples. Survival analysis of *PDCD6* mRNA of 85 NSCLC patients and 34 stage I ADC patients from the Duke dataset is described in the *SI Text*.

ACKNOWLEDGMENTS. We thank Ni Liu, Olga Ludkovski, Paul Boutros, and Dr. Jeremy Squire for assistance and advice. This work was supported by the Canadian Cancer Society and National Cancer Institute of Canada Grant 015184, Genome Canada/British Columbia, Ontario Institute of Cancer Research, and the Canadian Institutes for Health Research. S.A.-R. is a fellow of the Canadian Institutes for Health Research Training Program for Clinician Scientists in Molecular Oncologic Pathology (STP-53912) and a recipient of the Ontario Cancer Institute Knudson Research Fellowship and the National Cancer Institute of Canada Terry Fox Foundation Clinical Research Fellowship. B.P.C. is supported by a scholarship from the Michael Smith Foundation for Health Research. M.-S.T. is the M. Qasim Choksi Chair in Lung Cancer Translational Research. I.J. is the recipient of the Canada Research Chair in Integrative Computational Biology.

- Travis WD, Travis LB, Devesa SS (1995) Lung Cancer. *Cancer* 75:191–202.
- Travis WD, Brambilla E, Müller-Hermelink HK, Harris CC, eds (2004) *WHO Classification of Tumors: Pathology and Genetics of Tumors of the Lung, Pleura, Thymus and Heart*. (IARC Press, Lyon), pp 35–44.
- Kitamura H, Kameda Y, Ito T, Hayashi H (1999) Atypical adenomatous hyperplasia of the lung. Implications for the pathogenesis of peripheral lung adenocarcinoma. *Am J Clin Pathol* 111:610–622.
- Kim CF, et al. (2005) Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* 121:823–835.
- Jackson EL, et al. (2001) Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev* 15:3243–3248.
- Johnson L, et al. (2001) Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* 410:1111–1116.
- Guerra C, et al. (2003) Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. *Cancer Cell* 4:111–120.
- Miller VA, et al. (2004) Bronchioalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. *J Clin Oncol* 22:1103–1109.
- Yokose T, et al. (2000) Favorable and unfavorable morphological prognostic factors in peripheral adenocarcinoma of the lung 3 cm or less in diameter. *Lung Cancer* 29:179–188.
- Suzuki K, et al. (2000) Prognostic significance of the size of central fibrosis in peripheral adenocarcinoma of the lung. *Ann Thorac Surg* 69:893–897.
- Noguchi M, et al. (1995) Small adenocarcinoma of the lung. Histologic characteristics and prognosis. *Cancer* 75:2844–2852.
- Rena O, et al. (2003) Stage I pure bronchioalveolar carcinoma: Recurrences, survival, and comparison with adenocarcinoma of the lung. *Eur J Cardiothorac Surg* 23:409–414.
- Ebriht MI, et al. Clinical pattern and pathologic stage but not histologic features predict outcome for bronchioalveolar carcinoma. *Ann Thorac Surg* 74:1640–1647.
- Breathnach OS, et al. (2001) Bronchioalveolar carcinoma of the lung: Recurrences and survival in patients with stage I disease. *J Thorac Cardiovasc Surg* 121:42–47.
- Sakurai H, et al. (2004) Grade of stromal invasion in small adenocarcinoma of the lung: Histopathological minimal invasion and prognosis. *Am J Surg Pathol* 28:198–206.
- Petersen I, Petersen S (2001) Towards a genetic-based classification of human lung cancer. *Anal Cell Pathol* 22:111–121.
- Beer DG, et al. (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8:816–824.
- Bhattacharjee A, et al. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98:13790–13795.
- Potti A, et al. (2006) A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355:570–580.
- Shibata T, et al. (2005) Genetic classification of lung adenocarcinoma based on array-based comparative genomic hybridization analysis: Its association with clinicopathologic features. *Clin Cancer Res* 11:6177–6185.
- Murnane JP, Sabatier L (2004) Chromosome rearrangements resulting from telomere dysfunction and their role in cancer. *BioEssays* 26:1164–1174.
- Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z (2006) A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38:1043–1048.
- Difilippantonio S, et al. (2003) Gene expression profiles in human non-small and small-cell lung cancers. *Eur J Cancer* 39:1936–1947.
- Tureci O, et al. (1998) Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens. *Proc Natl Acad Sci USA* 95:5211–5216.
- Zhu CQ, et al. (2006) Amplification of telomerase (hTERT) gene is a poor prognostic marker in non-small-cell lung cancer. *Br J Cancer* 94:1452–1459.
- Vito P, Lacana E, D'Adamo L (1996) Interfering with apoptosis: Ca²⁺-Binding protein ALG-2 and Alzheimer's disease gene ALG-3. *Science* 271:521–525.
- Maki M, Narayana SV, Hitomi K (1997) A growing family of the Ca²⁺-binding proteins with five EF-hand motifs. *Biochem J* 328:718–720.
- Jung YS, et al. (2001) Apoptosis-linked gene 2 binds to the death domain of Fas and dissociates from Fas during Fas-mediated apoptosis in Jurkat cells. *Biochem Biophys Res Commun* 288:420–426.
- Vito P, Pellegri L, Guet C, D'Adamo L (1999) Cloning of AIP1, a novel protein that associates with the apoptosis-linked gene ALG-2 in a Ca²⁺-dependent reaction. *J Biol Chem* 274:1533–1540.
- Kitaura Y, Matsumoto S, Satoh H, Hitomi K, Maki M (2001) Peflin and ALG-2, members of the penta-EF-hand protein family, form a heterodimer that dissociates in a Ca²⁺-dependent manner. *J Biol Chem* 276:14053–14058.
- Satoh H, Shibata H, Nakano Y, Kitaura Y, Maki M (2002) ALG-2 interacts with the amino-terminal domain of annexin XI in a Ca²⁺-dependent manner. *Biochem Biophys Res Commun* 291:1166–1172.
- Lee JH, Rho SB, Chun T (2005) Programmed cell death 6 (PDCD6) protein interacts with death-associated protein kinase 1 (DAPK1): Additive effect on apoptosis via caspase-3 dependent pathway. *Biotechnol Lett* 27:1011–1015.
- Jang IK, Hu R, Lacana E, D'Adamo L, Gu H (2002) Apoptosis-linked gene 2-deficient mice exhibit normal T-cell development and function. *Mol Cell Biol* 22:4094–4100.
- Krebs J, Saremaslani P, Caduff R (2002) ALG-2: A Ca²⁺-binding modulator protein involved in cell proliferation and in cell death. *Biochim Biophys Acta* 1600:68–73.
- la Cour JM, et al. (2003) Up-regulation of ALG-2 in hepatomas and lung cancer tissue. *Am J Pathol* 163:81–89.
- Czekay RP, Loskutoff DJ (2004) Unexpected role of plasminogen activator inhibitor 1 in cell adhesion and detachment. *Exp Biol Med (Maywood)* 229:1090–1096.
- Andreasen PA, Kjoller L, Christensen L, Duffy MJ (1997) The urokinase-type plasminogen activator system in cancer metastasis: A review. *Int J Cancer* 72:1–22.
- Dorsam RT, Gutkind JS (2007) G-protein-coupled receptors and cancer. *Nat Rev Cancer* 7:79–94.
- Nollen EA, et al. (2001) Modulation of in vivo HSP70 chaperone activity by Hip and Bag-1. *J Biol Chem* 276:4677–4682.
- Ravagnan L, et al. (2001) Heat-shock protein 70 antagonizes apoptosis-inducing factor. *Nat Cell Biol* 3:839–843.
- Barsyte-Lovejoy D, et al. (2006) The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res* 66:5330–5337.
- Coe BP, et al. (2006) Gain of a region on 7p22.3, containing MAD1L1, is the most frequent event in small-cell lung cancer cell lines. *Genes Chromosomes Cancer* 45:11–19.
- Chi B, DeLeeuw RJ, Coe BP, MacAulay C, Lam WL (2004) SeeGH-A software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics* 5:13.
- Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: Cluster analysis of microarray data. *Bioinformatics* 18:207–208.
- Jong K, Marchiori E, Meijer G, Vaart AV, Ylstra B (2004) Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics* 20:3636–3637.