Molecular classification of tumors with special reference to *EGFR* mutation in lung cancer

Yasushi Yatabe

Published online: 9 November 2006

© Springer-Verlag 2006

Abstract Unsupervised hierarchical clustering in expression profiling analyses allows molecular classification of tumors based on the similarity of genomewide expression patterns. This new molecular-based classification shares the current pathological classification in part, although it also provides additional clues to identify cancers by their biological groups. Herein, we introduce a novel means of molecular classification of lung cancer. When examining the gene expression profiling analyses of lung cancers published so far, the molecular classification differs from the current classification schema, dividing lung cancers into two distinct branches that do not segregate SCLC and NSCLC, as might be expected. One of the branches includes adenocarcinoma alone, which is associated with a normal expression profile, while the other branch includes all four histological subtypes. We further examined the adenocarcinoma subset of the first branch. This subset of adenocarcinomas is characterized by frequent development in females and non-smokers, expression of thyroid transcription factor-1 and surfactant proteins, and specific involvement of epidermal growth factor receptor (EGFR) gene mutation. Furthermore, this subset is highly distinctive not only among lung cancers but also

This work was presented at the 21st Bristol-Myers Squibb Nagoya International Cancer Treatment Symposium, "Lung Cancer: Novel Therapy against Malfunctioning Molecules", 24–25 February 2006, Nagoya, Japan.

Y. Yatabe (⊠) Department of Pathology and Molecular Diagnostics, Aichi Cancer Center, Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan e-mail: yyatabe@aichi-cc.jp among carcinoma arising in other tissue sites, in terms of *EGFR* gene mutation and expression profiles. Although further studies are needed to clarify this adenocarcinoma subset, the distinction should be taken into consideration in lung cancer research and clinical strategies for treatment.

Keywords $EGFR \cdot$ Molecular classification \cdot Expression profiling analysis \cdot Lung cancer \cdot Bronchioloalyeolar carcinoma

Introduction

The completion of the human genome project now makes it possible to perform genome-wide analyses. Expression profiling analysis is such an example, and by comparing genome-wide expression patterns we can classify tumors according to their biological characteristics (Fig. 1). Unsupervised hierarchical clustering is commonly used for this purpose, and therefore the clustering represents the molecular classification of tumors.

Using this clustering method, breast cancers have been analyzed [21]. Breast cancers have been molecularly classified into four distinct subtypes including normal breast-like, luminal cell-like, basal cell-like, and HER2 subtypes [21, 26–28]. This molecular classification schema has been reconstructed with various series, and specific involvement of *BRCA1* mutation in basal cell-like types supports the biological significance of the classification [6, 27]. Genome-wide status of loss of heterozygosity has also revealed that breast cancers can be divided into four subtypes that are identical to the molecular classification using expression profiling



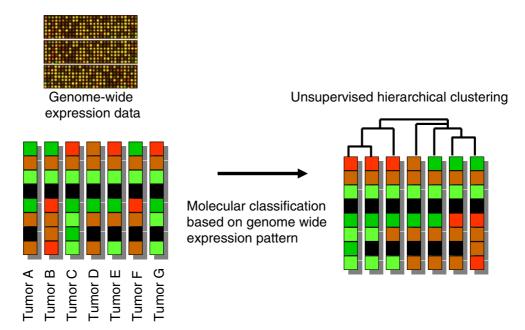


Fig. 1 Molecular classification based on genome-wide expression pattern. By comparing their genome-wide expression patterns tumors can be classified according to molecular-biological charac-

teristics. Unsupervised hierarchical clustering is commonly used for this purpose, and therefore the clustering represents the molecular classification of tumors

analyses [32]. Recently, it has been reported that the breast cancer subtypes defined molecularly are associated with the clinical response to neoadjuvant chemotherapy [23].

In expression profiling analysis of lung cancers, according to data published so far, four major histological subtypes—adenocarcinoma, large cell carcinoma, small cell carcinoma, and squamous cell carcinoma—are recognized as distinct clusters [2, 3, 8, 10, 31]. However, the molecular classification distinctly differs from that used in current therapeutic strategies, in that molecular classification highlights the high diversity of lung adenocarcinomas and places SCLC and NSCLC in close relation. In this article, we examine and attempt to interpret the biological significance of this diversity.

Diversity of lung adenocarcinomas

Based on the available data, the molecular classification schema for lung cancers is summarized in Fig. 2, although the proportion of each cluster varies among reports [2, 3, 8, 10, 31]. In general, lung cancers branch off into two groups then cluster into histological subtypes. The left branch of the schema includes adenocarcinoma alone, which is associated with the profile of normal lung tissue, while the group in the right branch is composed of all four histological subtypes. Namely, lung adenocarcinoma can be divided into two subsets

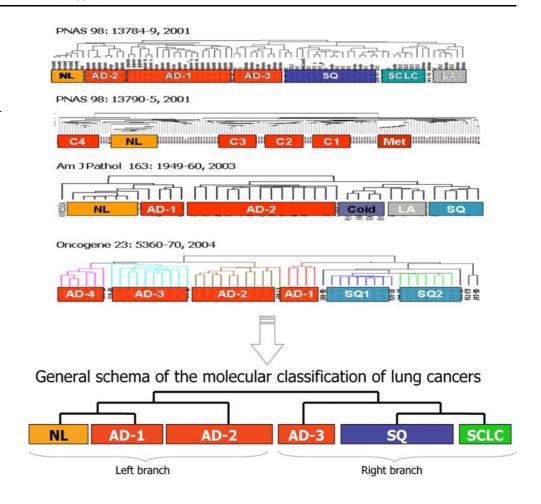
of right- and left-branched groups. The left branch corresponds to "adeno-group 1" and "adeno-group 2" in the study by Garber et al. [8], C3 and C4 according to Bhattacharjee et al. [2], group 2 in Borczuk's work [3], and AD3 and AD4 in our study [31]. All these studies describe this subset of adenocarcinoma as being characterized by molecules expressed in the peripheral lung parenchyma such as thyroid transcription factor (TTF)-1 and various isoforms of surfactant proteins that suggest a close association with peripheral lung parenchyma. They also support the idea that this subtype is always clustered together with normal lung tissues

In contrast, the adenocarcinoma group of the right branch is mixed with the clusters of squamous cell carcinoma and SCLC. Notably, Bhattacharjee et al. [2] include in their analysis the expression profiles of metastatic lung cancers from the colon. Surprisingly, this metastatic carcinoma is stationed within this right branch and adjacent to lung squamous cell carcinoma, but not as a single cluster isolated from the lung cancers, implying that genome-wide expression patterns of metastatic adenocarcinoma are rather similar to those of lung squamous cell carcinomas, and that the differences between metastatic carcinoma and lung squamous cell carcinoma are less pronounced than those between right and left groups of lung adenocarcinomas.

When adenocarcinomas arising from various organ sites were compared, lung adenocarcinoma was distinguished by expression of certain characteristic



Fig. 2 Molecular classification of lung cancers. Four representative molecular classifications, modified in order by the author, reveal almost identical features (summarized in lower panel). Lung cancers branch off into two groups then cluster into histological subtypes



molecules [9, 24]. These always included markers of peripheral lung parenchyma such as TTF-1 and surfactant proteins, which are also selected as markers of the left branch. These findings indicate that the adenocarcinoma subset of the left branch is highly distinctive not only among lung cancers but also among adenocarcinomas arising from various tissue sites. This relationship may be analogous to that of whales, dolphins, and fish living in the sea (Fig. 3). Although all these creatures live together, whales and dolphins are mammals, not fish. In comparison with the distinct subset of the left branch, the difference between lung squamous cell carcinoma and metastatic cancer may be very small.

Terminal respiratory unit and its adenocarcinomas

Anatomically, normal lung tissue is composed of an air-conducting system and peripheral lung parenchyma wherein gas is exchanged. Bronchial surface epithelium, bronchial glands, bronchial cartilage, and muscles make up the former system, while pneumocytes, unique cells responsible for gaseous exchange, and capillary vessels are the main components of the

peripheral parenchyma (Fig. 4). Two phases are observed in fetal/neonatal ontogeny of the lung [4, 16]. Branching morphogenesis from lung buds is initiated, and the conducting system is generated. Development and maturation of the parenchyma follow, and alveolar differentiation is completed in the first few weeks after birth. The adenocarcinoma subset seen in the left branch of the molecular classification appears to be derived from epithelium of the peripheral parenchyma.

Each adenocarcinoma subtype in the molecular classification is characterized by its expression pattern of certain molecules [2, 3, 8, 10, 31]. Interestingly, characteristic markers for the left branch subset are almost identical. Among them, we have focused on TTF-1. This molecule is a transcription factor that regulates functionally important molecules for respiration (e.g., surfactant proteins) [5, 19, 33]; TTF-1 knockout mice specifically lack peripheral lung parenchyma but not bronchial tree [11, 18]; and TTF-1 exhibits restricted expression in human pneumocytes and non-ciliated bronchial epithelium from early development to adult lung [20, 29, 36]. These findings suggest that TTF-1 functions as a master key regulator of the peripheral lung parenchyma. Indeed, when examined in a normal



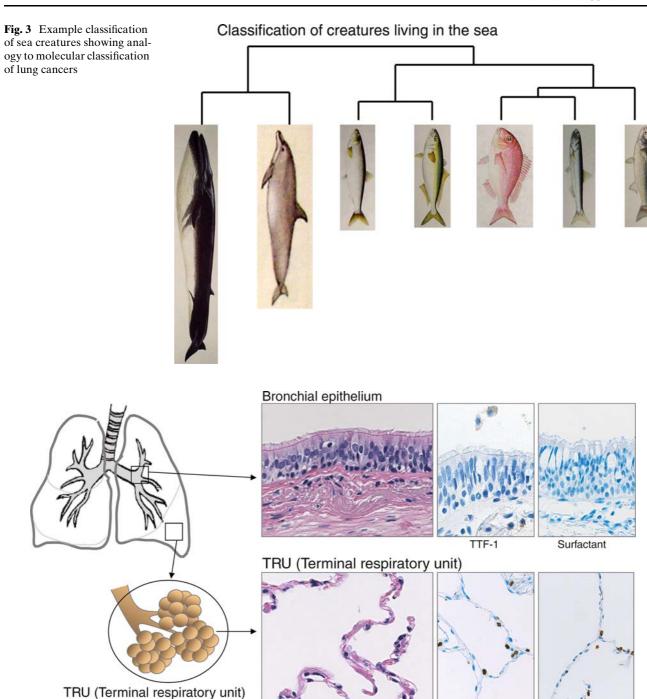


Fig. 4 The two major components of the lung: an air-conducting system and the terminal respiratory unit (TRU). TRU shares expression of the left branch of adenocarcinomas in the molecular classification

lung the expression is so uniform that it appears to represent a functional unit or cellular lineage. We therefore term this the "terminal respiratory unit (TRU)" [34–36] (Fig. 4).

We examined the clinicopathological characteristics of TRU-type adenocarcinoma using TTF-1 as a marker for the TRU feature. TRU-type adenocarcinoma is frequently observed in females and non-smokers [36].

Histologically, the cellular features of TTF-1-positive adenocarcinoma resemble TRU cells, whereas TTF-1-negative adenocarcinomas appear derived from bronchial glands and/or bronchial surface epithelium. Furthermore, cancer-associated genes including *p53*, *KRAS*, and *p27* show different expression patterns between TTF-1-positive and -negative adenocarcinomas, suggesting different molecular pathogenic mechanisms [34, 36].



In 2003, two phase II trials of gefitinib, Iressa Dose Evaluation in Advanced Lung Cancer (IDEAL) 1 and 2, demonstrated that certain subsets of patients (namely Japanese, females, and those with adenocarcinoma) appeared to have higher response rates [7, 13]. Subsequent analysis by Miller et al. [17] revealed that smoking history and bronchioloalveolar pathologic subtypes were predictive factors for positive response to gefitinib. These features overlapped with those of TRU-type adenocarcinoma, suggesting some association. Therefore we examined the correlation between epidermal growth factor receptor (EGFR) mutation and TRU-type adenocarcinoma [12]. In a total of 241 consecutive NSCLC samples, EGFR mutation was detected in 97 adenocarcinomas and one adenosquamous carcinoma. More than 90% of the EGFRmutated tumors expressed TRU markers including TTF-1 and surfactant proteins and showed morphological features of TRU cells. This specific involvement of EGFR mutation in TRU-type adenocarcinoma was supported by multivariate analysis. A logistic regression model revealed that only smoking status and cellular lineage (TRU vs. non-TRU type) were significant factors affecting EGFR mutation. Therefore it was concluded that EGFR mutation is specifically involved in TRU-type adenocarcinoma.

Interpretation and biological significance of the molecular classification

Gene mutation in the EGFR kinase domain is extremely rare in extra-pulmonary cancers; it was only detected in a case of colon cancer through the examination of >1,000 cancers arising from organs other than the lung (Table 1) [1, 14, 15, 25]. Although a large number of hematological cancers and sarcomas harbor characteristic gene alterations, such tumor-specific gene alterations are very exceptional in epitheliaderived cancers. However, this enigma may be explained by the characteristics of TRU cells and the specific involvement of EGFR mutation in TRU-type adenocarcinoma. As mentioned above, TRU cells are unique, being the only cells involved in gaseous exchange. Therefore only the lungs contain TRU-like cells. EGFR mutation is specifically involved in TRUtype adenocarcinoma, suggesting that TRU cells may have a specific susceptibility to acquisition of EGFR mutations. This explanation is consistent with the molecular classification findings.

Separation of these two distinct groups of lung cancers may also provide important insights for lung cancer research. The molecular difference between the two

Table 1 Gene mutation of EGFR kinase domain in extra-pulmonary cancers

Tumor(s) examined	References				Total
	[15]	[25]	[14]	[1]	
Breast	0/141	0/31	0/93	_	0/265
Head and neck	0/65	_	_	_	0/65
Pancreas	0/54	_	_	_	0/54
Liver	_	_	0/73	_	0/73
Stomach	_	0/54	0/185	_	0/239
Prostate	0/25	0/24	_	_	0/49
Colon	0/20	0/56	0/98	1/293	1/467
Kidney	0/15	_	_	_	0/15
Brain	0/4	_	_	0/58	0/62
Gall bladder	_	0/50	_	_	0/50
Bladder	_	0/28	_	_	0/28
Leukemia (adult)	_	_	0/88	_	0/88
Total	0/324	0/243	0/537	1/351	1/1455

groups is particularly distinct, while EGFR mutation is specific in the left TRU-type branch of the molecular classification. Therefore when the molecules differentially expressed in lung cancers with and without EGFR mutation are compared, it is likely that the result represents only the difference between the two distinct groups and not upregulated or downregulated molecules derived from EGFR gene mutation. Indeed, our recent article [30] reported that three of the top six genes differentially expressed in tumors with and without EGFR mutation are identical to those with TRU and non-TRU features. This indicates that selection of proper objective indicators is important to obtain biological significance in cancer research and clinical investigation. A recent article on breast cancers [22] exemplifies this significance: X-chromosome abnormalities were specifically found in basal-like breast cancers, and therefore it was suggested that this alteration plays an important role only in this subset. This abnormality might otherwise have been mistaken for a rare alteration if this subset was not considered.

Consideration of this remarkable distinction may help us develop better therapeutic strategies for lung cancers. It is quite reasonable that metastatic adenocarcinoma from the colon is treated differently from lung adenocarcinoma, because the tumors are of different organs. However, TRU and non-TRU-type adenocarcinomas are currently treated under the same therapeutic strategy, although differences between the expression profiles of these tumor types are much greater than those between metastatic cancer of the colon and squamous cell carcinoma of the lung. As in the case of breast cancers [23], it may be interesting to examine the therapeutic response by molecular subtypes using specimens under clinical trial conditions.



Summary

The aims of tumor classification include estimation of clinical outcomes and tumor characteristics, choice of treatment based on such estimation, understanding of cancer pathogenesis, and so on. On this basis, it is very important to discriminate SCLC from NSCLC in current therapeutic strategies, which are based on consensus resulting from many studies describing differences in clinicopathologic features, response to therapy, and treatment outcomes. On the other hand, comparison of genome-wide expression profiles of lung cancers suggests that there is a great distinction between adenocarcinomas, namely TRU- and non-TRU-type adenocarcinomas. Evidence for this distinction is supported by specific involvement of EGFR mutation and different impacts of smoking in these two subsets. Further studies are warranted to clarify the distinctions between lung adenocarcinomas, and consideration of these differences may provide us with greater understanding of lung cancer pathogenesis and allow us to develop better therapeutic strategies.

Acknowledgments I would like to acknowledge the invaluable support and encouragement of Dr. Tetsuya Mitsudomi, Dr. Takashi Takahashi, and all the collaborators in our research.

References

- Barber TD, Vogelstein B, Kinzler KW, Velculescu VE (2004) Somatic mutations of EGFR in colorectal cancers and glioblastomas. N Engl J Med 351:2883
- Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, Meyerson M (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA 98:13790–13795
- Borczuk AC, Gorenstein L, Walter KL, Assaad AA, Wang L, Powell CA (2003) Non-small-cell lung cancer molecular signatures recapitulate lung developmental pathways. Am J Pathol 163:1949–1960
- Cardoso WV (2000) Lung morphogenesis revisited: old facts, current ideas. Dev Dyn 219:121–130
- Cassel TN, Suske G, Nord M (2000) C/EBP alpha and TTF-1 synergistically transactivate the Clara cell secretory protein gene. Ann NY Acad Sci 923:300–302
- Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, Trudel M, Akslen LA (2003) Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 95:1482–1485
- Fukuoka M, Yano S, Giaccone G, Tamura T, Nakagawa K, Douillard JY, Nishiwaki Y, Vansteenkiste J, Kudoh S, Rischin D, Eek R, Horai T, Noda K, Takata I, Smit E, Averbuch S, Macleod A, Feyereislova A, Dong RP, Baselga J (2003) Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer. J Clin Oncol 21:2237–2246

- Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, Van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D, Petersen I (2001) Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci USA 98:13784– 13789
- Giordano TJ, Shedden KA, Schwartz DR, Kuick R, Taylor JM, Lee N, Misek DE, Greenson JK, Kardia SLR, Beer DG, Rennert G, Cho KR, Gruber SB, Fearon ER, Hanash S (2001) Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. Am J Pathol 159:1231–1238
- 10. Kikuchi T, Daigo Y, Katagiri T, Tsunoda T, Okada K, Kakiuchi S, Zembutsu H, Furukawa Y, Kawamura M, Kobayashi K, Imai K, Nakamura Y (2003) Expression profiles of nonsmall cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. Oncogene 22:2192–2205
- 11. Kimura S, Hara Y, Pineau T, Fernandez-Salguero P, Fox CH, Ward JM, Gonzalez FJ (1996) The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. Genes Dev 10:60–69
- Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T (2004) Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res 64:8919–8923
- 13. Kris MG, Natale RB, Herbst RS, Lynch TJ Jr, Prager D, Belani CP, Schiller JH, Kelly K, Spiridonidis H, Sandler A, Albain KS, Cella D, Wolf MK, Averbuch SD, Ochs JJ, Kay AC (2003) Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. JAMA 290:2149–2158
- 14. Lee JW, Soung YH, Kim SY, Park WS, Nam SW, Lee JY, Yoo NJ, Lee SH (2005) Absence of *EGFR* mutation in the kinase domain in common human cancers besides non-small cell lung cancer. Int J Cancer 113:510–511
- 15. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350:2129–2139
- Mendelson CR (2000) Role of transcription factors in fetal lung development and surfactant protein gene expression. Annu Rev Physiol 62:875–915
- 17. Miller VA, Kris MG, Shah N, Patel J, Azzoli C, Gomez J, Krug LM, Pao W, Rizvi N, Pizzo B, Tyson L, Venkatraman E, Ben-Porat L, Memoli N, Zakowski M, Rusch V, Heelan RT (2004) Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. J Clin Oncol 22:1103–1109
- Minoo P, Su G, Drum H, Bringas P, Kimura S (1999) Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos. Dev Biol 209:60-71
- Nord M, Cassel TN, Braun H, Suske G (2000) Regulation of the Clara cell secretory protein/uteroglobin promoter in lung. Ann NY Acad Sci 923:154–165
- Ordonez NG (2000) Thyroid transcription factor-1 is a marker of lung and thyroid carcinomas. Adv Anat Pathol 7:123–127
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Ress CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergammenschlkov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale A-L, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406:747–752



- Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, Ganesan S (2006) X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell 9:121–132
- Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P, Morandi P, Fan C, Rabiul I, Ross JS, Hortobagyi GN, Pusztai L (2005) Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 11:5678– 5685
- 24. Shedden KA, Taylor JM, Giordano TJ, Kuick R, Misek DE, Rennert G, Schwartz DR, Gruber SB, Logsdon C, Simeone D, Kardia SLR, Greenson JK, Cho KR, Beer DG, Fearon ER, Hanash S (2003) Accurate molecular classification of human cancers based on gene expression using a simple classifier with a pathological tree-based framework. Am J Pathol 163:1985–1995
- 25. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, Fong KM, Lee H, Toyooka S, Shimizu N, Fujisawa T, Feng Z, Roth JA, Herz J, Minna JD, Gazdar AF (2005) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 97:339–346
- 26. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale A (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 98:10869–10874
- 27. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale A-L, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100:8418–8423

- Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci USA 100:10393–10398
- Stahlman MT, Gray ME, Whitsett JA (1996) Expression of thyroid transcription factor-1 (TTF-1) in fetal and neonatal human lung. J Histochem Cytochem 44:673–678
- Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, Mitsudomi T, Takahashi T (2006) Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 24:1679–1688
- Tomida S, Koshikawa K, Yatabe Y, Harano T, Ogura N, Mitsudomi T, Some M, Yanagisawa K, Takahashi T, Osada H, Takahashi T (2004) Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. Oncogene 23:5360–5370
- 32. Wang ZC, Lin M, Wei LJ, Li C, Miron A, Lodeiro G, Harris L, Ramaswamy S, Tanenbaum DM, Meyerson M, Iglehart JD, Richardson A (2004) Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. Cancer Res 64:64–71
- Whitsett JA, Glasser SW (1998) Regulation of surfactant protein gene transcription. Biochim Biophys Acta 1408:303–311
- 34. Yatabe Y (2004) Role of expression of thyroid transcription factor-1 in pulmonary adenocarcinoma. In: Hayat MA (eds) Immunohistochemistry and in situ hybridization of human carcinomas. Elsevier Science/Academic, New York, pp 169–179
- Yatabe Y, Kosaka T, Takahashi T, Mitsudomi T (2005) EGFR mutation is specific for terminal respiratory unit type adenocarcinoma. Am J Surg Pathol 29:633–639
- Yatabe Y, Mitsudomi T, Takahashi T (2002) TTF-1 expression in pulmonary adenocarcinomas. Am J Surg Pathol 26:767–773

