

Lung cancer progression and metastasis from the prognostic point of view

Kentaro Inamura · Yuichi Ishikawa

Received: 7 March 2009 / Accepted: 16 February 2010 / Published online: 12 March 2010
© Springer Science+Business Media B.V. 2010

Abstract Lung cancer is the leading cause of cancer death in men and women worldwide. Since the occurrence of metastases in distant organs is the major reason for mortality of cancer patients, we need to elucidate the underlying mechanisms. Many studies featuring analysis of gene expression, comparative genomic hybridization and loss of heterozygosity analysis have been performed and generated support for the hypothesis that metastatic potential is acquired early in tumorigenesis. Furthermore, it is now clear that the majority of tumor cells have the potential to metastasize. Although many changes in gene expression profiles have been established retrospectively, translational research is now a high priority to enable clinical application and treatment based on laboratory findings.

Keywords Lung cancer · Metastasis · Prognosis · Gene expression · Comparative genomic hybridization · Loss of heterozygosity analysis

Introduction

Lung cancer, the leading cause of cancer death in men and women worldwide and continuing to rise in frequency, is generally classified as of either small-cell lung carcinoma (SCLC) or non-SCLC (NSCLC) types. Within these groups further distinctions are made, with NSCLCs sub-divided into adenocarcinomas, squamous cell carcinomas (SCCs), and

large cell carcinomas (LCCs). The occurrence of metastases in distant organs is the major cause of death for the vast majority of lung cancer patients. Clinical outcomes can be roughly predicted by pathological-Stage (p-Stage) and 5 year survival for p-Stage I cases, pathologically lacking metastases, is relatively good, ranging from 60 [1] to 90% [2]. Even when cancer lesions have been fully removed and no metastasis is found at surgery, however, some patients with p-Stage I lesions suffer recurrence and die of cancer relapse. Presumably, these already had micrometastases at the time of tumor removal. To avoid unnecessary lymph node dissection in low-risk cases but ensure that postoperative adjuvant therapy is performed for high-risk patients, we need a clinically useful approach to better stratify patients with respect to the risk of recurrence. Towards a rational treatment, we need to elucidate metastatic gene signatures and molecular mechanisms of lung cancer progression. The aim of the present review is to survey findings on lung cancer progression and metastasis from the prognostic point of view, especially emphasizing our study [3] paying attention to heterogeneity of lung cancers, an important characteristic.

The beginning of gene expression profiling in lung cancers

In November, 2001, pioneering studies of gene expression profiling in lung cancers were reported at the same time by a Stanford University group [4] and a Harvard University group [5]. Subdivision of the tumors based on gene expression faithfully recapitulated their histological classification and characteristic expression profiles for each histological type could be identified. The insulinoma-associated gene 1 (IA-1) and the human achaetes homolog 1 (hASH1) were found to be neuroendocrine SCLC markers, shared also

K. Inamura · Y. Ishikawa (✉)
Department of Pathology, The Cancer Institute, Japanese
Foundation for Cancer Research (JFCR), 3-10-6 Ariake,
135-8550 Koto-ku, Tokyo, Japan
e-mail: ishikawa@jfcrc.or.jp

by carcinoid tumors. Identified as SCC markers were Keratin 5 (KRT5), KRT17 and Tumor protein p63, which is associated with development of squamous epithelium. Supporting the traditional view that lung adenocarcinomas are a heterogeneous group, distinct subclasses were evident. One adenocarcinoma subgroup was comprised of tumors expressing neuroendocrine markers, such as hASH1 and IA-1, associated with a significant decrease in patient survival when compared to other adenocarcinomas [5]. Another subgroup appeared to express markers of alveolar type II pneumocytes and was characterized by high relative expression of TTF1 or surfactant protein genes.

Gene expression profiling predicting survival of patients with lung adenocarcinomas

As mentioned above, the Stanford University group [4] and the Harvard University group [5] first identified prognostically different subgroups of adenocarcinomas by gene expression profiling. One subgroup with a poor prognosis was revealed to have neuroendocrine features. Subsequently, many further studies using gene expression profiling have been reported. Beer et al. [6] described development of a risk index, compiling the relative expression of 50 genes, to identify high or low risk groups of Stage I adenocarcinomas that correlated with patient survival.

Ramaswamy et al. [7] compared gene expression profiles of adenocarcinoma metastases and unmatched primary adenocarcinomas and found patterns that allowed distinction between the two, but also reported that a subset of primary tumors had similar expression to metastases. This finding led them to challenge “the notion that metastases arise from rare cells within the primary tumor.” They suggested that the majority of tumor cells have the potential to metastasize, but this remains controversial and Liotta and Kohn have argued against their conclusions [8]. When lists of genes are examined, it is unclear whether the expression profile is a cause or a local consequence of the metastatic process. Ramaswamy et al. did not microdissect tumor cells for analysis of their tissue specimens and consequently the gene-expression pattern data reflect contributions from multiple cell populations. Thus, the expression pattern of the genes in the authors’ signature set may be at least partially due to activated host stromal elements. Indeed, two of the important upregulated genes in the list encode stromal collagen.

Although gene expression profiles that can classify cancer patients according to the risk of recurrence have been found, most studies have been retrospective. Very recently, Potti et al. [9] documented a “lung metagene model” that can identify individuals at increased risk for disease recurrence with stage IA NSCLC, which they now plan to use for a prospective randomized clinical trial. Translational research

is now an urgent priority to enable clinical application of basic research findings.

Gene expression profiling using hierarchical clustering and non-negative matrix factorization in squamous cell carcinomas

After the adenocarcinoma, the SCC is the most frequent lung cancer histology, accounting for approximately 30% of the total. Its development is the most strongly related to smoking. For adenocarcinomas, subclassification by differentiation grade [10] or histological pattern [2] is useful to predict prognosis. For SCCs, differentiation grade is used for pathological subclassification, but it correlates poorly with prognosis. Although SCCs demonstrate some histological variation, such as with the basaloid variant, this does not allow good prediction of prognosis. The present system used to subclassify SCC is thus insufficient and we have therefore attempted to make a clinically useful classification based on gene expression profiling [11]. By hierarchical clustering, we subclassified SCCs into two prognostically significant subclasses. Furthermore, consensus clustering with a non-negative matrix factorization (NMF) approach indicated the robustness of this classification (Fig. 1). NMF appears to be more accurate for choice of input genes than hierarchical clustering and can be combined with a quantitative evaluation of the robustness with numbers of clusters [12]. Both hierarchical clustering and NMF approaches (Fig. 1) indicated that SCCs can be divided into two groups, SCC-A and SCC-B, with prognostic variation (Fig. 2a). The cophenetic correlation coefficient, k , quantitatively indicated the two-centroid clustering to be the most robust with the highest value, as attested by clear block diagonal patterns (Fig. 2b). Up-regulation of cell-proliferation-related genes was evident in the subclass with poor survival. In the subclass with better survival, genes involved in differentiated intracellular functions, such as the MAPKKK cascade, ceramide metabolism, or regulation of transcription, were upregulated.

Histological typing and gene expression profiles in high-grade neuroendocrine tumors

The current WHO classification of high-grade neuroendocrine tumors (HGNTs) currently recognizes large-cell neuroendocrine carcinoma (LCNEC), a subclass of LCC, and SCLC as a distinct group [13]. Since LCNEC and SCLC share several histological features, a consensus differential diagnosis between LCNEC and SCLC is sometimes difficult, even among experienced lung pathologists. Hence, by the microarray technique, we analyzed gene

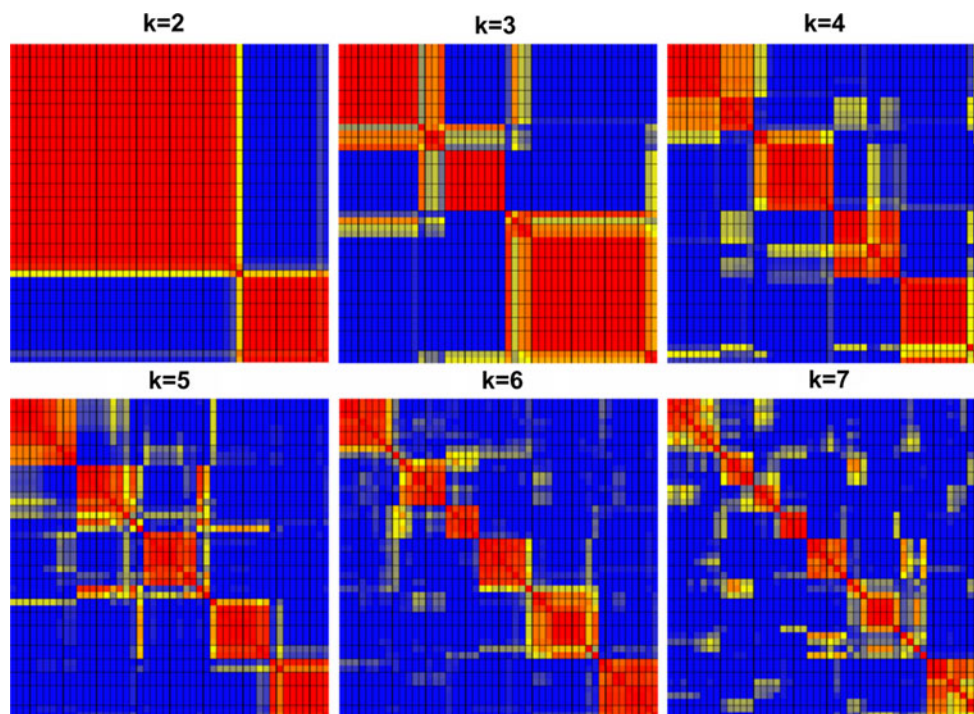
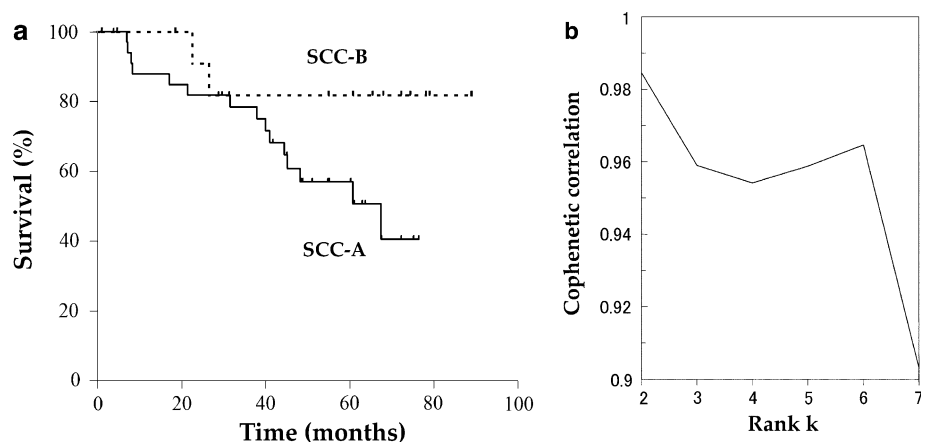


Fig. 1 Reordered consensus averaging 50 connectivity matrices computed at $k = 2-7$ for all SCC samples with 3,344 genes. Samples were hierarchically clustered, colored from 0 (samples never in the same cluster) to 1 (samples always in the same cluster)

Fig. 2 a Kaplan–Meier survival curves for the 48 SCC patients (SCC-A vs. SCC-B). **b** Cophenetic correlation coefficients for the hierarchically clustered matrices



expression profiles of HGNTs with other histological tumor groups and normal lung tissue [14]. By hierarchical clustering, we could readily identify distinct groups for carcinoids, LCC, adenocarcinoma, and normal lung (Fig. 3a). While we could not subclassify SCLC and LCNEC by gene expression profiling, two prognostically significant subtypes of HGNT were evident, independent of SCLC and LCNEC ($P = 0.0094$). Many genes distinguished the HGNT groups. There was no significant difference in survival between SCLC and LCNEC samples (Fig. 3b; $P = 0.37$).

Integrated classification of lung tumors and cell lines by expression profiling

The utility of cancer cell lines depends largely on their accurate classification, commonly based on histopathological diagnosis of the cancers from whom they were derived. However, because cancers are often heterogeneous, cell lines, which also have a propensity to alter in vitro, may not be truly representative. We therefore performed gene expression profiling, which can faithfully recapitulate histological classification of tumors, to examine different cell

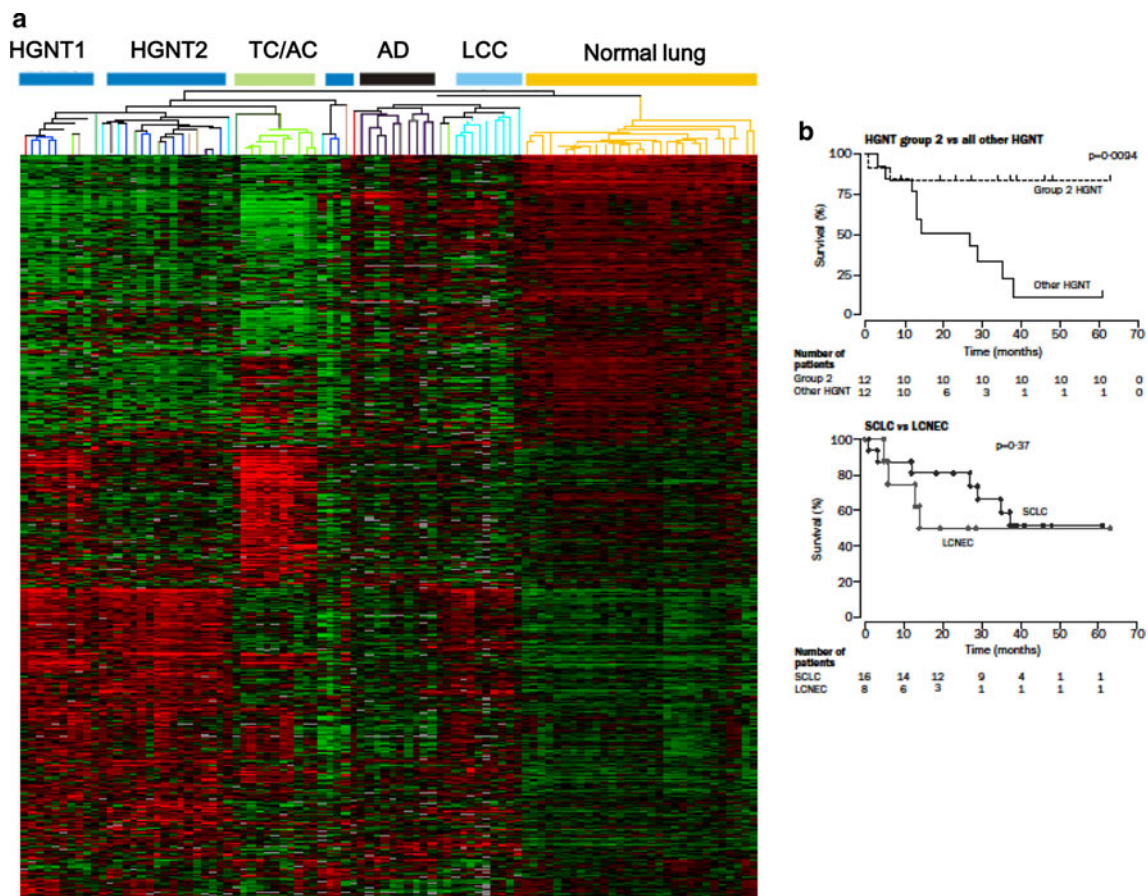


Fig. 3 a Unsupervised hierarchical clustering of 64 lung cancer and 30 normal lung samples against 2,803 genes with expression differentially regulated in neuroendocrine tumors. *HGNT* high grade neuroendocrine tumor, *TC/AC* typical carcinoids/atypical carcinoids, *AD* adenocarcinoma, *LCC* large cell carcinoma. **b** Kaplan-Meier

survival curves for patients with HGNTs in group 2 and all other HGNTs and for histopathologically diagnosed SCLC versus LCNEC. SCLC, small-cell lung carcinoma; LCNEC, large-cell neuroendocrine carcinoma

lines [15]. After excluding genes which show clear distinction between fresh and cell-line samples, hierarchical clustering resulted in a large degree of integration of cell lines into four main tumor branches, an SCLC branch, a SCC branch, a cell-line branch, and a branch containing normal tissue, adenocarcinoma and LCC (Fig. 4). As a result, most of SCC cell lines or SCLC cell lines grouped with fresh SCC tumors or fresh SCLC tumors, respectively. In contrast, although none of adenocarcinoma cell lines clustered with fresh adenocarcinoma tumors, some of them clustered with fresh SCC tumors or fresh SCLC tumors. Adenocarcinomas may ultimately progress toward one of two poorly differentiated phenotypes with expression profiles resembling SCC or SCLC. Our observations suggest that adenocarcinoma cell lines either dedifferentiate toward molecular pathologies resembling SCLC or SCC, or that clonal expansion of SCC or SCLC subcomponents occurs frequently. Analysis of larger numbers of adenocarcinoma samples taken at the time of surgery and autopsy will be

required to verify that adenocarcinomas develop similarly in situ.

Comparison of accumulated allele loss between primary lung tumors and lymph node metastases

Sasatomi et al. [16] have compared loss of heterozygosity (LOH) at microsatellites between primary NSCLCs and their lymph node metastases, calculating fractional allele loss (FAL), defined as the ratio of chromosomes affected by LOH in the informative chromosomes, for each sample. With Stage II NSCLCs, the FAL was found to be significantly less in the metastatic sites compared with the primary neoplasms. The authors advanced the theory that this phenomenon was the result of early metastatic spread of the carcinoma, with the primary neoplasm then acquiring additional genetic changes. This concept should be borne in mind when comparing molecular profiles of

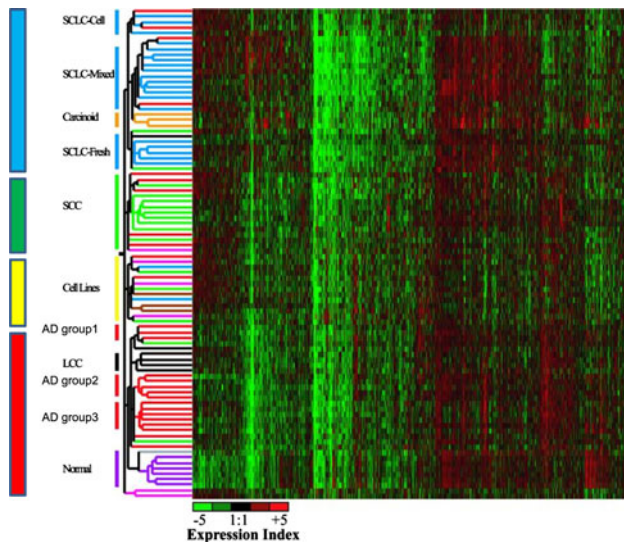


Fig. 4 Dendrogram of the reduced data set of 4,253 genes after filtering for commonly regulated genes in either fresh or cell-line samples. Groupings indicated on the *left* represent distinct clusters of particular carcinoma types. AD, adenocarcinoma; Normal, normal lung tissue

primary neoplasms with those of metastatic or recurrent sites.

Gene expression profile for tissue-specific metastasis

Kang et al. [17] have identified, in a human breast cancer cell line, a specific set of genes that mediates metastasis to bone. They suggested that primary tumors with metastatic capacity possess the poor-prognosis signature but, additional functions, provided by a set of bone metastasis genes, must be expressed in order to achieve an overt, tissue-specific metastasis phenotype. Organ-specific expression profiles for human small-cell lung cancer metastases in mice have also been reported by Kakiuchi et al. [18], but it remains unclear whether these might already be present in the parental cells.

Heterogeneity of primary tumors and metastatic potential

Introduction

Recent microarray experiments have suggested that the majority of cancer cells have the potential to metastasize, with obvious clinical and therapeutic implications, not only in breast cancers [19], but also in lung cancers [7]. Ramaswamy et al. [7] identified a molecular signature of metastatic potential within the bulk of each primary lung cancer, suggesting that metastatic potential is in fact

acquired early and is a feature of the majority of lung cancer cells. To test this hypothesis, we adopted the lung adenocarcinoma, which characteristically shows widespread intratumoral heterogeneity, as a model [3].

The mixed type adenocarcinoma in the lung, which shows a variety of histological subtypes, is the most frequent subtype in the WHO classification criteria [13], accounting for approximately 80% of resected lung adenocarcinomas [20]. An invasive component with high cellular and structural atypia is often included but peripherally well-differentiated components with low atypia may also be present (Fig. 5). Is the metastatic signature detected only in the aggressive component with high atypia? Or is it present in the entire tumor irrespective of morphological heterogeneity? If the latter is true, then it follows that the metastatic potential is acquired early in tumor progression and the entire tumor, including the morphologically less malignant component, may have metastatic potential.

Using lymph node-positive lung adenocarcinomas, we compared gene expression profiles among moderately-differentiated components with aggressive appearance, peripheral well-differentiated components with less malignant appearance, and patient-matched lymph node metastases. Node-negative lung adenocarcinomas, which are morphologically indistinguishable from node-positive tumors, were included for comparison and differential diagnosis.

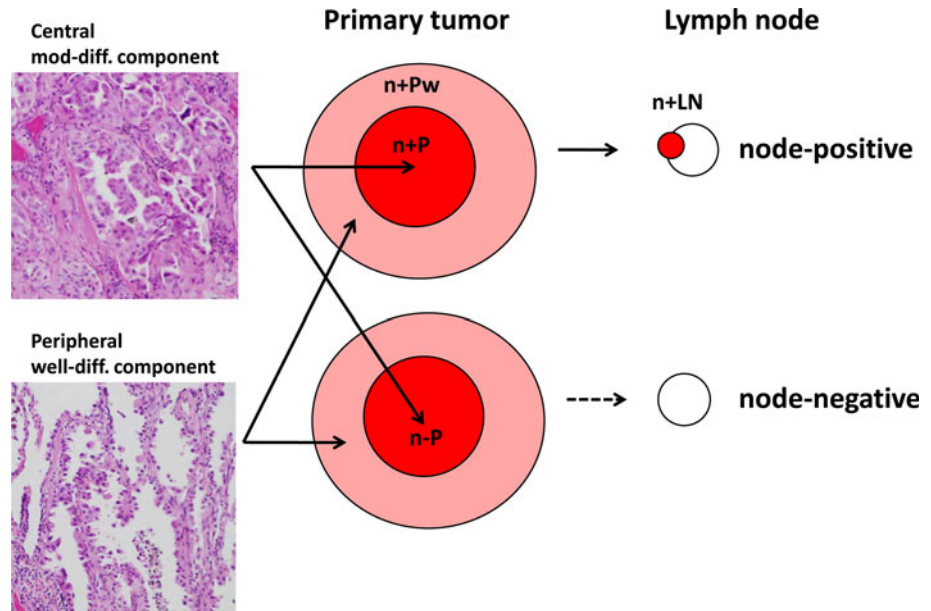
Schematic design

The schematic design for this study is shown in Fig. 5. We analyzed 10 pairs of primary lung adenocarcinomas and their synchronous lymph node metastases and 11 primary lung adenocarcinomas without lymph node or organ metastases. With the 21 primary lung adenocarcinomas, we isolated tumor cells from predominant moderately-differentiated components. In five of the 10 node-positive primary lung adenocarcinomas, we additionally isolated peripheral well-differentiated components. To focus selectively on cancer cells, we applied laser capture microdissection (LCM). For comparison, we included six samples of macrodissected normal lung tissue.

Experiments and results

Firstly, we wished to identify the overall gene expression signature in all 42 samples by unsupervised hierarchical clustering with a set of highly variable genes. Two-way hierarchical clustering was performed using a Pearson correlation (Fig. 6). Eight of the 10 pairs of primary and metastatic tumors clustered next to each other. In these cases, the metastatic tumors had a higher similarity to their matching primary tumors than to all other tumors. In only two of the node-positive cases (cases 5 and 9), the metastatic tumors did

Fig. 5 Schematic design of this study. n + P, node-positive primary tumor; n-P, node-negative primary tumor; n + LN, node-positive lymph node; n + Pw, well-differentiated component of the node-positive primary tumor. Representative microscopic images of primary lung adenocarcinomas used in this study are shown, comprising moderately-differentiated components dominating large portions of the lesions and well-differentiated components evident in peripheral portions



not cluster with their matching primary tumors, but a substantial similarity was still observed. In addition, all the peripheral well-differentiated components showed tight clustering with the predominant moderately-differentiated components from the same primary tumors. In the overall gene expression signatures, the central and the peripheral components from the same primary tumor were strikingly similar to each other.

To identify the gene expression signature of lymph node metastases, we performed a statistical comparison between the 10 pairs of primary and metastatic tumors using a paired *t*-statistic. Only 12 genes were yielded. For 11 of these genes, no significant differences were found between 10 lymph node metastases and 11 node-negative primary tumors, suggesting that they are not part of a metastatic expression signature. Only one gene, *Chromosome 4 open reading frame 7 (C4orf7)* was significantly higher in the lymph node metastases than both node-positive and -negative primary tumors. However, as *C4orf7* is expressed characteristically by follicular dendritic cells found in lymph nodes, the high expression presumably resulted from contamination in the LCM process. It follows that no significant metastatic changes were detected between primary tumors and their lymph node metastases. The marked similarity between primary tumors and their lymph node metastases drove us to consider that the gene expression signature of lymph node metastasis might be acquired by the majority of primary tumor cells.

Next, to identify the metastatic expression signature detected in the primary tumors, we performed a statistical comparison (Welch's *t*-test) between 10 node-positive primary tumors and 11 node-negative primary tumors. This

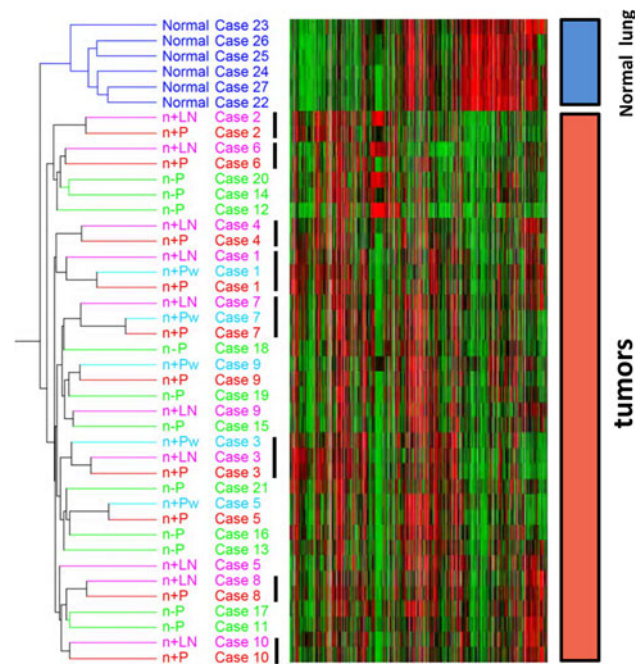


Fig. 6 Unsupervised hierarchical clustering for 2,451 genes against 42 samples comprising 10 pairs of node-positive primary tumors (n + P) and node-positive lymph nodes (n + LN), 11 node-negative primary tumors (n-P), 5 well-differentiated components of the node-positive primary tumors (n + Pw), and 6 normal tissues (Normal). Columns represent gene and rows represent samples. Note that eight of the 10 pairs of primary and metastatic tumors clustered next to each other (*bars*)

yielded 75 genes, comprising 37 with significantly higher expression in node-positive than node-negative primary tumors and 38 with significantly lower expression. The top discriminating gene was *homeobox B2 (HOXB2)* with

higher expression in node-positive primary tumors. Malignant potential associated with ectopic *HOXB2* expression has been reported recently [21]. Down-regulated genes in node-positive primary tumors include *VAMP-associated protein A* (VAPA), involved in vesicle trafficking, and *Zinc finger protein 36 homolog* (ZFP36), which is also known as *tristetraprolin* (TTP) and is involved in degradation of tumor necrosis factor α . The *IQ motif containing GTPase activating protein 1* (*IQGAP1*), one of the molecular markers for lymph node metastasis identified by a microarray study using LCM [22], was also found to be included in the down-regulated genes.

The next issue was whether this 75-gene signature might be maintained throughout the metastatic process, and also be present in peripheral well-differentiated components of primary tumors. We therefore performed supervised hierarchical clustering of all the 42 samples against these 75 genes using a Pearson correlation (Fig. 7). Node-positive cases formed a distinct independent group, except one case (case 9), separate from node-negative tumors and normal lung tissues. The latter two clustered together. The node-positive group included the metastatic tumors and the primary well-differentiated components. Also in this metastatic gene expression signature, as with the overall gene

expression signature, samples from the same case showed tight clustering. In leave-one-out cross-validation analysis, the 75 genes predicted their groups with 100% accuracy. Using real-time RT-PCR analysis, we validated our results for some of genes of interest from the 75-gene set.

Discussion

In this study, we could identify a 75-gene signature discriminating between node-positive and node-negative primary lung adenocarcinomas. Hierarchical clustering using this gene set generated a distinct independent group composed of node-positive cases, including the metastatic tumors and the peripheral well-differentiated components, separate from node-negative tumors and normal lung tissues. Striking transcriptional similarities were observed between samples from the same case and unsupervised hierarchical clustering showed tight clustering. Hierarchical clustering using the 75-gene set also showed tight clustering, reflecting similarity also in the metastatic gene expression signature. Indeed, statistical comparison of gene expression levels between pairs of primary tumors and their lymph node metastases revealed no differences responsible for the metastasis, implying that metastatic potential might be established early in the pathogenesis of tumors. This result is in keeping with recent array findings suggesting that metastatic potential is encoded in the bulk of a primary tumor [7]. More recently, D'Arrigo et al. [23] similarly showed striking transcriptional similarity using 10 pairs of matching primary colorectal cancers and distant metastases.

Several studies have resulted in lists of metastasis-related or malignancy-related genes in lung cancers [5–7, 22, 24]. However, most authors did not use microdissection but rather RNAs isolated from tumor masses. As their gene lists differed widely and had only few genes in common, the 75-gene list we identified also differed widely from theirs. Ein-Dor et al. [25] reported that thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Kikuchi et al. [22] used microdissected samples of 22 primary lung adenocarcinoma cases and identified 40 genes whose expression levels could separate cases according to their lymph node status. One of the 40 genes was *IQGAP1*, also included in our 75-gene metastatic signature. In our hospital, survival for p-Stage I lung cancer patients is good as compared with the literature. Whereas the reported 5-year survival for p-Stage I patients with non-small cell lung cancer is about 60% [1], it is 90% in our hospital [2]. We usually perform thorough dissection of lymph nodes and make a detailed histopathological examination. This accurate assessment of lymph node status is clearly advantageous for comparison of node-positive and node-negative tumors.

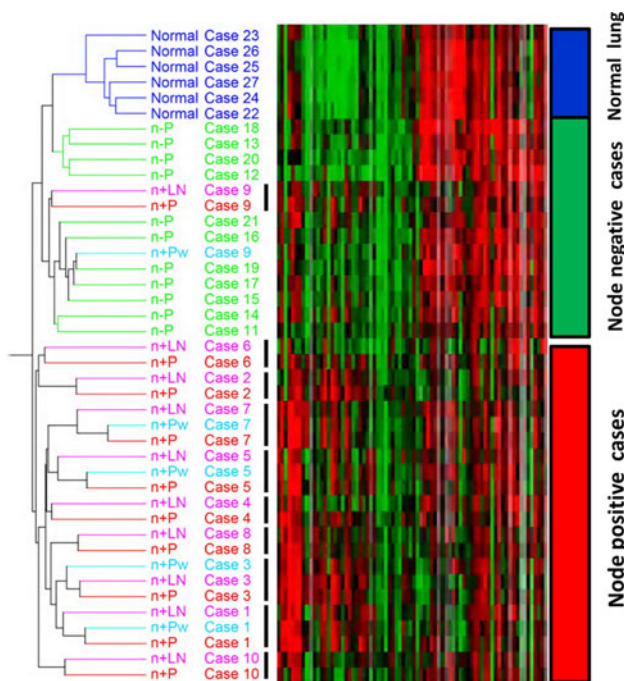


Fig. 7 Hierarchical clustering for the 75 genes against 42 samples comprising 10 pairs of node-positive primary tumors (n + P) and node-positive lymph nodes (n + LN), 11 node-negative primary tumors (n-P), 5 well-differentiated components of the node-positive primary tumors (n + Pw), and 6 normal tissues (Normal). Remarkably, nine of the 10 node-positive cases formed a distinct independent group. Pairs of primary and metastatic tumors clustered next to each other (bars)

In the 75-gene set we identified, *HOXB2* was the top discriminating gene between node-positive and node-negative primary lung adenocarcinomas. Aberrant expression of *HOX* genes has been implicated in leukemias and various solid cancers, including lung cancers, with likely involvement of the gene products in features of malignant progression, such as invasion and metastasis. Overexpression of *HOXD3* is known to induce coordinate expression of metastasis-related genes in lung cancer cells [26]. A recent study further revealed ectopic *HOXB2* expression in pancreatic cancers and some proportion of precursor lesions, pancreatic intraepithelial neoplasias, possibly associated with a poor prognosis [21]. In our current study, we observed *HOXB2* overexpression not only in the central but also in the peripheral zones of node-positive primary tumors. Malignant potential associated with *HOXB2* expression might thus be acquired early in tumorigenesis.

Our results imply that the metastatic potential might be encoded in the entirety of each primary lung tumor including the morphologically less malignant component. This has profound clinical implications. It offers a rationale for therapeutic applications based on the expression profile of the primary tumor. Even if the peripheral component of the tumor is sampled by bronchoscopic or needle biopsy, the metastatic potential of the tumor could be predicted with accuracy. Such evaluation of metastatic potential would help spare unnecessary lymph node dissection for low-risk patients. However, the 75-gene signature needs to be confirmed using independent samples and further research is required to clarify the included molecular functions.

Very recently, using real time RT-PCR analysis, we investigated the transcriptional levels of the top metastasis-related genes using 96 independent test lung adenocarcinoma samples and investigated their correlations with prognosis [27]. We could document evidence that p-Stage I patients with *HOXB2* up-regulation have a worse prognosis than those with *HOXB2* down-regulation ($P = 0.0065$). Comparing tumors and corresponding normal lung tissue, we confirmed *HOXB2* up-regulated lesions to have much higher *HOXB2* expression than the corresponding normal tissue.

In conclusion

Recent studies support the hypothesis that metastatic potential is acquired early in tumorigenesis and that the majority of tumor cells have the potential to metastasize. Although gene expression profiles that can classify cancer patients according to the risk of recurrence have been found in many studies, most of these were retrospective. Very recently, by gene expression profiling, Potti et al. [9] proposed a “lung metagene model” that could identify individuals at increased risk for disease recurrence with

stage IA NSCLC. They are now planning to use this for a prospective randomized clinical trial. In the future, more emphasis needs to be placed on translational research to enable basic research findings to be applied in the clinic.

References

1. Brundage MD, Davies D, Mackillop WJ (2002) Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest* 122(3):1037–1057. doi:10.1378/chest.122.3.1037
2. Miyoshi T, Satoh Y, Okumura S et al (2003) Early-stage lung adenocarcinomas with a micropapillary pattern, a distinct pathologic marker for a significantly poor prognosis. *Am J Surg Pathol* 27(1):101–109. doi:10.1097/00000478-200301000-00011
3. Inamura K, Shimoji T, Ninomiya H et al (2007) A metastatic signature in entire lung adenocarcinomas irrespective of morphological heterogeneity. *Hum Pathol* 38(5):702–709. doi:10.1016/j.humpath.2006.11.019
4. Garber ME, Troyanskaya OG, Schluens K et al (2001) Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 98(24):13784–13789. doi:10.1073/pnas.241500798
5. Bhattacharjee A, Richards WG, Staunton J et al (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98(24):13790–13795. doi:10.1073/pnas.191502998
6. Beer DG, Kardia SL, Huang CC et al (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8(8):816–824
7. Ramaswamy S, Ross KN, Lander ES et al (2003) A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33(1):49–54. doi:10.1038/ng1060
8. Liotta LA, Kohn EC (2003) Cancer's deadly signature. *Nat Genet* 33(1):10–11. doi:10.1038/ng0103-10
9. Potti A, Mukherjee S, Petersen R et al (2006) A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* 355(6):570–580. doi:10.1056/NEJMoa060467
10. Sun Z, Aubry MC, Deschamps C et al (2006) Histologic grade is an independent prognostic factor for survival in non-small cell lung cancer: an analysis of 5018 hospital- and 712 population-based cases. *J Thoracic Cardiovasc Surg* 131(5):1014–1020. doi:10.1016/j.jtcvs.2005.12.057
11. Inamura K, Fujiwara T, Hoshida Y et al (2005) Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. *Oncogene* 24(47):7105–7113. doi:10.1038/sj.onc.1208858
12. Brunet JP, Tamayo P, Golub TR et al (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci USA* 101(12):4164–4169. doi:10.1073/pnas.0308531101
13. Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC (2004) World health organization classification of tumours: pathology and genetics of tumours of the lung, pleural, thymus and heart. Springer, Berlin
14. Jones MH, Virtanen C, Honjoh D et al (2004) Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *Lancet* 363(9411):775–781. doi:10.1016/S0140-6736(04)15693-6
15. Virtanen C, Ishikawa Y, Honjoh D et al (2002) Integrated classification of lung tumors and cell lines by expression profiling.

- Proc Natl Acad Sci USA 99(19):12357–12362. doi:[10.1073/pnas.192240599](https://doi.org/10.1073/pnas.192240599)
16. Sasatomi E, Finkelstein SD, Woods JD et al (2002) Comparison of accumulated allele loss between primary tumor and lymph node metastasis in stage II non-small cell lung carcinoma: implications for the timing of lymph node metastasis and prognostic value. *Cancer Res* 62(9):2681–2689
 17. Kang Y, Siegel PM, Shu W et al (2003) A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 3(6):537–549. doi:[10.1016/S1535-6108\(03\)00132-6](https://doi.org/10.1016/S1535-6108(03)00132-6)
 18. Kakiuchi S, Daigo Y, Tsunoda T et al (2003) Genome-wide analysis of organ-preferential metastasis of human small cell lung cancer in mice. *Mol Cancer Res* 1(7):485–499
 19. van 't Veer LJ, Dai H, van de Vijver MJ et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871):530–536. doi:[10.1038/415530a](https://doi.org/10.1038/415530a)
 20. Terasaki H, Niki T, Matsuno Y et al (2003) Lung adenocarcinoma with mixed bronchioloalveolar and invasive components: clinicopathological features, subclassification by extent of invasive foci, and immunohistochemical characterization. *Am J Surg Pathol* 27(7):937–951. doi:[10.1097/00000478-200307000-00009](https://doi.org/10.1097/00000478-200307000-00009)
 21. Segara D, Biankin AV, Kench JG et al (2005) Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia. *Clin Cancer Res* 11(9):3587–3596. doi:[10.1158/1078-0432.CCR-04-1813](https://doi.org/10.1158/1078-0432.CCR-04-1813)
 22. Kikuchi T, Daigo Y, Katagiri T et al (2003) Expression profiles of non-small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. *Oncogene* 22(14):2192–2205. doi:[10.1038/sj.onc.1206288](https://doi.org/10.1038/sj.onc.1206288)
 23. D'Arrigo A, Belluco C, Ambrosi A et al (2005) Metastatic transcriptional pattern revealed by gene expression profiling in primary colorectal carcinoma. *Int J Cancer* 115(2):256–262. doi:[10.1002/ijc.20883](https://doi.org/10.1002/ijc.20883)
 24. Xi L, Lyons-Weiler J, Coello MC et al (2005) Prediction of lymph node metastasis by analysis of gene expression profiles in primary lung adenocarcinomas. *Clin Cancer Res* 11(11):4128–4135. doi:[10.1158/1078-0432.CCR-04-2525](https://doi.org/10.1158/1078-0432.CCR-04-2525)
 25. Ein-Dor L, Zuk O, Domany E (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci USA* 103(15):5923–5928. doi:[10.1073/pnas.0601231103](https://doi.org/10.1073/pnas.0601231103)
 26. Miyazaki YJ, Hamada J, Tada M et al (2002) HOXD3 enhances motility and invasiveness through the TGF-beta-dependent and -independent pathways in A549 cells. *Oncogene* 21(5):798–808. doi:[10.1038/sj.onc.1205126](https://doi.org/10.1038/sj.onc.1205126)
 27. Inamura K, Togashi Y, Okui M et al (2007) HOXB2 as a novel prognostic indicator for stage I lung adenocarcinomas. *J Thoracic Oncol* 2(9):802–807. doi:[10.1097/JTO.0b013e3181461987](https://doi.org/10.1097/JTO.0b013e3181461987)