Characterization of cancer stroma markers: In silico analysis of an mRNA expression database for fibroblast activation protein and endosialin

Helmut Dolznig¹, Norbert Schweifer², Christina Puri¹, Norbert Kraut¹, Wolfgang J. Rettig², Dortscho Kerjaschki¹, and Pilar Garin-Chesa¹,²

¹Institute of Clinical Pathology, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria
²Boehringer Ingelheim Austria GmbH, Dr. Boehringergasse 5-10, 1120 Vienna, Austria

Keywords: human, stromal cells, tumor markers, fibroblast activation protein, endosialin protein, mRNA, immunohistochemistry, tissue distribution

Standardized, high-throughput RNA detection with microarray chips allows for the construction of genome-wide databases for tissue specimens suitable for in silico electronic Northern blot (eNorthern) analysis of marker genes. We used the BioExpress™ database, which contains transcriptional profiles of normal and cancer samples, to examine two putative markers of cancer stroma: fibroblast activation protein-alpha (FAP-alpha) and endosialin. Analyses for FAP-alpha showed that normal tissues generally lack RNA signals, with the exception of endometrium. Typing of tumors revealed prominent FAP-alpha signals in cancer types marked by desmoplasia, and localization of FAP-alpha in reactive cancer stroma was confirmed by immunohistochemistry. A subset of sarcomas displayed prominent FAP-alpha signals localizing to the malignant cells. For endosialin, eNorthern analyses showed low to moderate RNA signals in many normal organs, whereas immunohistochemistry revealed endosialin in only some tissues, such as endometrium. Endosialin was detected at the RNA and protein level in sarcomas, notably malignant fibrous histiocytomas. Low to moderate endosialin RNA signals were found in epithelial cancer types for which immunostaining identifies expression in subsets of tumor capillaries or fibroblasts. These findings extend the FAP-alpha and endosialin profiling in silico to an unbiased tumor database and place both molecules in a novel context of endometrial biology and sarcoma subtyping. Our findings suggest that BioExpress™ can be searched directly for tumor stroma markers but may need prior enrichment for markers with narrow cellular representation, such as endosialin. Constructing databases from microdissected cancer tissues may be an essential step for tumor stroma-targeted therapies.

Introduction

With the advances in human genome sequencing and bioinformatic annotation, and with new technologies for standardized, high-throughput DNA and RNA detection in hand, novel opportunities abound for investigating the molecular pathology of human cancer. While much initial work focused on the characterization of cell lines cultured in vitro or on cancer xenograft models, which are more readily handled in an experimental setting, attention is now shifting to biopsy and surgical specimens of human cancer that have not been manipulated in vitro. With more mature technologies available, these efforts are contributing to the large-scale analysis of germline and somatic mutations, loss of heterozygosity, and DNA amplifications during cancer progression, as well as accompanying RNA expression studies to define comprehensive oncogene signature profiles (1, 2, 3, 4).

To fully exploit the results of such studies, especially RNA expression profiles, it is desirable to standardize sample collection, processing, and testing in order to facilitate repeated in silico database mining of a well-defined tissue collection for unlimited sets of marker genes and for data-mining across multiple independent databases. In the present study, we have used one such database, the BioExpress™ database containing Affymetrix U133 GeneChip expression profiles of thousands of tissues (5, 6), to examine two putative cancer markers, FAP-alpha and endosialin.

Neither FAP-alpha nor endosialin was identified as a classical cancer cell marker, and their typing on in vitro cancer cell lines has not been very informative. Instead, both are markers of distinct cell types within the reactive stromal compartment of cancer tissues, which has emerged as an important factor in cancer angiogenesis, locally invasive growth, and metastatic spread of malignant cells (7, 8, 9). Specifically, FAP-alpha was discovered with a monoclonal antibody, mAb F19, which in immunohistochemical tests binds to the reactive stromal fibroblasts in many epithelial cancers, but not to the malignant epithelial cells, and also not to most normal human tissues examined (10, 11). The biological function of FAP-alpha is unclear in light of the inconspicuous phenotype of fap(-/-) knockout mice (12), but genetic and biochemical studies have identified a type II cell surface serine protease capable of degrading gelatin and collagen (13, 14), and cell line studies suggest a link to melanocyte transformation independent of this peptidase function (15). Based on the selective expression of FAP-alpha in normal and cancer tissues, radiolabeled anti-FAP-alpha antibodies have been shown to target breast cancers in a xenograft model (16) and in metastases from colorectal cancers in patients (17, 18).

Endosialin was also discovered with a monoclonal antibody, mAb FB5, which has been shown by immunohistochemistry to bind to subsets of capillaries in human cancers, but not to capillary endothelium or other cell types in most normal human tissues (19). The physiological role of endosialin is unknown. Biochemically, it is a member of the C-type lectin superfamily, a highly glycosylated and sialylated class I transmembrane...
Cancer Immunity, Vol. 5, p. 10 (3 August 2005)

The present study was designed to further our understanding of the FAP-alpha and endosialin expression profiles in the context of cancer pathology, and to validate the use of global RNA expression databases for standardized tumor stromal marker analysis.

Figure 1

Expression profiles of FAP-alpha in normal and malignant human epithelial tissues. (A) Whisker-box plots for normal (green), malignant (red), and malignant metastatic lesions (orange); the number of samples for each tissue is given. The bold lines in each box indicate median values, and box limits represent the first and third quartiles; the whiskers extend to 1.5 times this interquartile range. The highest value in each cohort outside the whiskers is marked by the letter “x.” For significant upregulation of FAP-alpha in cancers versus matched normal tissues, the \( P \)-values in Student’s \( t \)-tests are shown. (B) Patterns of FAP-alpha expression, as determined by immunohistochemistry with mAb F19, are shown. The avidin biotin complex immunoperoxidase method was used, followed by hematoxilin counterstaining. Appropriate negative controls with unrelated antibodies were run in all assays and did not yield any immunoperoxidase staining. SCC, squamous cell carcinoma; LN, lymph node; (P), primary; (M), metastases; Pap ca, papillary carcinoma; AdSq Ca, adenosquamous carcinoma; LCC, large cell carcinoma; SCLC, small cell lung carcinoma; NEC, neuroendocrine cancer; Inflamm, inflammatory; MAC, mucinous adenocarcinoma; HCC, hepatocellular carcinoma; IDC, intraductal carcinoma; DC, ductal carcinoma; LC, lobular carcinoma; MG, medullary carcinoma; MMT, Mullerian mixed tumor; CICC, clear cell carcinoma; GCC, granular cell carcinoma; OC, oncocytoma; TCC, transitional cell carcinoma; WT, Wilms’ tumor. Normal head and neck tissues consisted of seven normal parotid gland tissues and six normal larynx samples.
Results

In silico analysis of the BioExpress™ database

The eNorthern gene expression profiles for FAP-alpha and endosialin were derived for close to 3200 distinct human tissue samples in the database, including more than 1700 human cancer specimens. No a priori effort was made to adjust sample numbers for any given normal tissue or cancer type due to the prevalence or expected statistical variations in marker gene expression, and sample diagnoses were taken at face value from the database. As one consequence, some cancer types were amply represented in the collection (for example, 236 samples of ductal carcinoma of the breast), while others were less abundantly represented (for example, 4 samples for endometrial papillary adenocarcinoma). Moreover, in this initial search, no detailed subclassification of cancer types was attempted.

Upregulation of FAP-alpha in cancer tissues versus normal organs

The summary of the in silico expression analysis for FAP-alpha, as illustrated in the whisker-box plots in Figure 1, shows that the FAP-alpha RNA signal is well represented in a range of cancer types while being restricted in the panel of normal tissues. Among the cancers with readily detectable FAP-alpha expression, squamous cell carcinomas of head and neck, lung, colorectal, pancreas, and breast carcinomas are noteworthy, as is the marked variability in RNA levels (interquartile range) as seen, for instance, in lung and colorectal cancers. Among the normal tissues tested, the uterus, cervix, skin, and breast were among those showing a positive RNA signal. From this sample set, it is not apparent whether noncancerous, reactive processes such as fibrosis or inflammation may have contributed to FAP-alpha expression. Median RNA signals were generally higher for the malignant tissues than for their matched, normal counterparts, as illustrated by the statistical analysis in Figure 1. It is notable, however, that certain epithelial cancer types, such as prostate cancer and ovarian carcinomas, as well as neuroendocrine tumors, showed no or only minimal gene expression. In Figure 2, additional tissue profiles are presented for skin and skin cancers, lymphoid tissues, brain, bone tumors, and soft tissue sarcomas. FAP-alpha expression is comparable in skin and skin tumors, but is markedly elevated in subsets of bone and soft tissue sarcomas, notably malignant fibrous histiocytomas and spindle cell sarcomas.

In order to relate RNA expression within complex cancer tissues to actual cellular expression, we extended some of the previous immunohistochemical studies aimed at FAP-alpha-positive cancers. In representative examples shown in Figure 1B, it is apparent that gene expression in epithelial cancers is sharply restricted to the reactive stromal fibroblasts of the cancer stroma, which traverses the malignant epithelial components and contains tumor vascular endothelium. In contrast, FAP-alpha expression in sarcomas can be localized to the malignant cell components (Figure 2B).

Comparison of FAP-alpha expression in normal tissues

The FAP-alpha mRNA expression profile in normal human tissues (n = 1455; 45% of 3200 cases in total) is summarized in Figure 3 to better allow a direct comparison. Nonmalignant breast tissues, cervix uteri, endometrium, pancreas, placenta,
and skin show definite signals, even if not as high as some of the cancer tissues in Figures 1 and 2. Figure 3B shows that FAP-alpha expression is localized to the stroma rather than to the epithelial component, whereas a comparison with colonic mucosa shows a complete absence of FAP-alpha immunostaining. For the endometrium, we have found that expression decreases during the mid-secretory phase. No expression was found in samples of atrophic endometrium.

Expression of endosialin in cancers compared to normal organs
The whisker-box plots in Figure 4 show that endosialin mRNA expression signals do not differ generally between normal and cancer tissues of the corresponding organs, and that the expression values in normal tissues are mostly higher than those seen in Figure 1 for FAP-alpha. Whereas the lack of a distinction between normal and cancer tissues is informative for endosialin, less may be deduced from the overall higher expression values for endosialin compared to FAP-alpha, since these depend to some extent on the hybridization properties of selected probe sets. The interpretation of the endosialin mRNA profile is further complicated by the patterns illustrated in Figure 4B, which shows endosialin immunostaining in highly variable subsets of cancer stromal cells, commonly, but not always, showing features of tumor capillary endothelium. Congruent with the FAP-alpha analysis, endosialin expression was also profiled in the skin, lymphoid, brain, bone, and soft tissue panel depicted in Figure 5. In this panel of tissues, marked endosialin mRNA expression in a subset of soft tissue sarcomas was most notable, including malignant fibrous histiocytomas, spindle-cell sarcomas, and liposarcomas. As shown in Figure 5B, endosialin expression in sarcomas corresponds to a classical tumor antigen on malignant cells, rather than to a tumor stroma-restricted marker.

Comparison of endosialin expression in normal tissues
The summary of endosialin mRNA signals in normal tissues in Figure 6 shows low to moderate expression across all tissues tested and highlights the increased signals in tissues such as the mammary gland, cervix uteri, endometrium, and skin. At least for the three examples shown in Figure 6B, the higher mRNA levels for the endometrium appear to translate into endosialin expression, which is also confirmed by immunohistochemistry in the stromal compartment of the endometrium; whereas tissues such as normal liver and normal prostate, which have lower mRNA signals, appear endosialin-negative by immunohistochemistry.

Discussion
The present study, along with other investigations using BioExpress™ or similar databases with global mRNA expression profiles (1, 2, 3, 4, 5, 6), strengthens a new paradigm in cancer
biology whereby large collections of pathologically well-defined diseased tissues, matched with normal controls, are converted into versatile resources for cancer bioinformatics. To be most useful, sampling methods and database annotations for clinico-pathological features, including detailed diagnostic assessment and oncogene mutation status, should follow standardized principles in order to facilitate cross-referencing among these databases. In the past, access to fresh tissue samples suitable for RNA analysis was limited, and studies with insufficient sample numbers or reliance on cultured cell line data were common. The electronic, tissue-based expression profiles by far outweigh the information gleaned from cultured cell lines and allow multiple marker genes to be studied, at any time, in the original sample set.

With these improvements in RNA expression studies, the focus shifts to another challenging dimension of the molecular pathology of human cancer, namely genetic and phenotypic variability within a given cancer type. It is obvious to the pathologist that cancer biopsies or surgical specimens used for RNA extractions represent complex mixtures of malignant cells, residual normal tissues, and reactive stromal tissue associated with cancer cell proliferation, invasion, and metastasis. These stromal elements include newly formed blood vessels, desmoplastic reactions with reactive stromal fibroblasts, and variable extracellular matrix deposition and inflammatory infiltrates. Far from being mere bystanders, these adaptive changes in the tissue stroma may support cancer progression or limit cancer cell spread (7, 8, 9).

Expression profiles of endosialin in normal and malignant human epithelial tissues. (A) Whisker-box plots showing expression profiles determined by eNorthern analysis. SCC = squamous cell carcinoma; LN = lymph node; (P) = primary; (M), metastases; Pap ca = papillary carcinoma; AdSq Ca = adenosquamous carcinoma, LCC = large cell carcinoma; SCLC = small cell lung carcinoma; NEC = neuroendocrine cancer, Inflamm = inflammatory; MAC = mucinous adenocarcinoma; HCC = hepatocellular carcinoma; IDC = intraductal carcinoma; DC = ductal carcinoma; LC = lobular carcinoma; MC = medullary carcinoma; MMT = Miederian mixed tumor; ClCC = clear cell carcinoma; GCC = granular cell carcinoma; OC = oncocytoma; TCC = transitional cell carcinoma; WT = Wilms’ tumor. Normal head and neck tissues consisted of seven normal parotid gland tissues and six normal larynx samples. Patterns of endosialin expression, as determined by immunohistochemistry with mAb FB5, are shown in (B). The avidin biotin complex method was used, followed by hematoxylin counterstaining.
The FAP-alpha serine protease appears to be a characteristic marker of reactive stromal fibroblasts of several types of human epithelial cancers, a conclusion supported by the present analysis. Since the stromal compartment in cancers with desmoplastic reactions, such as pancreatic, breast, lung, or squamous cell carcinomas of the head and neck, may comprise from 10 to 90% of the corresponding tissue mass, it is not surprising that FAP-alpha RNA signals are readily picked up in our study. It is noteworthy that a subset of the sarcomas studied is also FAP-alpha positive, and only the combined use of RNA profiling and immunohistochemistry with FAP-alpha-specific antibodies makes it possible to distinguish between stromal marker expression in carcinomas and malignant cell expression in mesenchymal tumors. The upregulation of FAP-alpha RNA in pancreas cancers was noted previously with the help of the BioExpress™ database (6), and FAP-alpha RNA upregulation in samples of aggressive fibromatosis was described with the use of the Affymetrix GeneChip U133 arrays incorporated in BioExpress™ (22). Functional data suggesting that FAP-alpha may act as a tumor suppressor in malignant melanomas (15, 23) seem to contrast with the results of this study; however, there is a report in which FAP-alpha appears to localize to nonmelanoma cells (24). We now show FAP-alpha RNA and protein expression in the stroma of cycling endometrium, and it is tempting to speculate about FAP-alpha function in tissue remodeling, both in this context and in other known sites of human FAP-alpha expression, including rheumatoid arthritis, wound healing (11, 25), liver cirrhosis (26, 27), and skin carcinogenesis (15, 24), or in *Xenopus laevis* morphogenesis (28). The generation and initial characterization of a *fap(-/-)* knockout mouse (12) has not yet provided any specific clues to this proposed function.

For endosialin, also known as TEM1 (21), the RNA profile from the present study and the results of prior immunohistochemical analyses (19) and RNA in situ hybridization experiments (21) are more divergent. Presumably, the low to moderate RNA expression in most normal organs, as reported here for human tissues and elsewhere for selected mouse and human tissues (29), is not strictly reflected by methods that address cellular localization, namely immunostaining or in situ RNA hybridization. Whether this is due to different detection limits or to differences in total RNA versus polysome-bound RNA (30), discrepancies between mRNA and protein levels (31), or protein detection with mAb FB5, remains to be explored. It is apparent, however, that in some normal tissues, such as the stroma of normal endometrium, and in subsets of sarcomas, notably malignant fibrous histiocytomas, consistent endosialin expression is seen with RNA methods and immunostaining and, in these cases, is not due to vascular endothelial expression. The fact that tumor capillary endothelial expression of endosialin in carcinomas, which sparked the initial interest in and discovery of these molecules, is not clearly apparent from the database mining (with the possible exception of kidney cancers), may have an even simpler explanation. The proportion of endosialin-expressing stromal cells, namely tumor endothelium, pericytes

![Figure 5](image)

**Expression profiles of endosialin in neuroectodermal, hematopoietic, and mesenchymal cancers, and in related normal tissues.** (A) Whisker-box plots showing expression profiles determined by eNorthern analysis. SCC = squamous cell carcinoma, BCC = basal cell carcinoma, Mel (P) = primary melanoma, Mel (M) = metastatic melanoma, NHD = Non-Hodgkin’s disease; HD = Hodgkin’s disease; GBM = glioblastoma; GCT = giant cell tumor of bone; OS = osteosarcoma; CHS = chordosarcoma; MFH = malignant fibrous histiocytoma; SpCSa = spindle cell sarcoma; LS = liposarcoma; LMS = leiomyosarcoma; SyCS = synovial cell sarcoma. Normal soft tissue consisted of normal skeletal muscle, adipose tissue from the thigh and uterine ligament soft tissues. Endosialin expression in (B) was determined by immunohistochemistry with mAb FB5.
and activated fibroblasts (19, 32, 33, 34) in these tissues may be so low that the corresponding signal is hidden within the relatively high signal for normal tissues.

The relative ease and speed of performing eNorthern analyses in the BioExpress™ database will allow us, as a next step, to search for putative regulators of cancer stroma induction and function, such as cytokines, growth factors, matrix proteins, and proteases, that may show distinct patterns of coregulation. A comparable analysis for remodeling processes in normal developing or cycling tissues and inflammatory lesions seems feasible.

Due to the current interest in cancer stroma function as a potential target for novel cancer therapies, it may be worthwhile to consider supplemental RNA expression databases constructed from microdissected cancer and normal tissues (35), in which minor tissue elements or cell types are more prominently represented. This approach may be too cumbersome for single-marker gene studies, but ultimately may be rewarding if the resulting data cover the human genome with all possible sets of marker genes and are made broadly available to cancer researchers.

**Abbreviations**
eNorthern, electronic Northern; FAP-alpha, fibroblast activation protein-alpha

**Acknowledgements**
We are grateful to Michael Seewald, Christian Stratowa, and Christian Haslinger for their assistance with bioinformatics. This work was funded, in part, by the GEN-AU program of the Austrian Ministry of Education, Science and Arts.

**References**


29. Opavsky R, Haviernik P, Jurkovicova D, Garin MT, Copeland NG (Affymetrix, Inc., Santa Clara, CA, USA): Briefly, these chips performed on Affymetrix HG-U133A/B oligonucleotide chips BioExpress™ database (36). The respective hybridizations were the upper whisker limits, are marked by an “x.” The human interquartile range; the positions of extreme values, if above the first and third quartiles. Whiskers extend to 1.5 times the line indicates the median; the box (green, normal; red, tumor; orange, metastasis) represents the interquartile range between the upper and lower quartiles. The scaling factor is calculated by removing the highest and the lowest 2% of the nonnormalized expression values and calculating the mean for the remaining values, as a trimmed mean. One hundred divided by the trimmed mean gives the scaling factor, where one hundred is the standard value used by GeneLogic.


Materials and methods

In silico expression analysis

For expression analysis, box- and whisker-plots were generated with the eNorthern tool of the GeneExpress Genesis 2.5 software, which is based on normalized gene expression data extracted from the BioExpress™ database (Gene Logic, Inc., Gaithersburg, MD, USA) as reported (36). The bold center line indicates the median; the box (green, normal; red, tumor; orange, metastasis) represents the interquartile range between the first and third quartiles. Whiskers extend to 1.5 times the interquartile range; the positions of extreme values, if above the upper whisker limits, are marked by an “x.” The human sample collection has been described by the originator of the BioExpress™ database (36). The respective hybridizations were performed on Affymetrix HG-U133A/B oligonucleotide chips (Affymetrix, Inc., Santa Clara, CA, USA): Briefly, these chips are based on 25-mer oligonucleotides and allow the detection of more than 33,000 well-substantiated human genes, with probe sets of 11 oligonucleotides used per transcript.

Statistical analysis

Chip analysis data were normalized with the statistical algorithm implemented in the Microarray Suite version 5.0 (Affymetrix, Inc.). According to this algorithm, the raw expression intensity for a given chip experiment is multiplied by a global scaling factor to allow comparisons among chips. The scaling factor is calculated by removing the highest and the lowest 2% of the nonnormalized expression values and calculating the mean for the remaining values, as a trimmed mean. One hundred divided by the trimmed mean gives the scaling factor, where one hundred is the standard value used by GeneLogic.

Statistical significance of the differences between tumor and normal samples was verified by calculating Student’s t-test P-values using the Comparative Analysis tool of GeneExpress. The P-values (displayed in gray to the right of the box plots in Figures 1, 2, 4, and 5) are given only if significant upregulation (P < 0.01) of the gene of interest was detected. The Affymetrix identification code for FAP-alpha is 209955_s_at and that for endosialin 219025_at.

Comparison of eNorthern and immunohistochemistry results

Although the tissue samples used to construct the BioExpress™ database were not available for side-by-side RNA and immunohistochemistry analysis in this study, we used matched tissue types, both normal and malignant, to illustrate certain patterns of gene expression based on staining with mAb F19 and mAb FB5, respectively. Human tissues were collected at the Institute of Pathology, Medical University of Vienna, Austria, following institutional guidelines, and processed for immunohistochemistry following published procedures (10, 11, 19). Briefly, tissues were embedded in OCT compound (Miles, Naperville, IL, USA), snap-frozen in isopentane precooled in liquid N2, and stored at -70°C. The avidin biotin complex immunoperoxidase method was used, and adjacent serial sections were stained with hematoxylin–eosin for morphological evaluation. For selected sarcoma cases, expression patterns are illustrated using antibody-stained tissue sections from previously reported studies (10, 11, 19, 23, 25).

Contact

Address correspondence to:

Dr. Pilar Garin-Chesa
Institute of Clinical Pathology
Medical University of Vienna
Waehringer Guertel 18-20
1090 Vienna
Austria
Tel.: +43 1 40400-4980
Fax: +43 1 40400-5179
E-mail: pilar.garin-chesa@meduniwien.ac.at