Sequential Binary Gene Ratio Tests Define a Novel Molecular Diagnostic Strategy for Malignant Pleural Mesothelioma

Assunta De Rienzo, William G. Richards, Beow Y. Yeap, Melissa H. Coleman, Peter E. Sugarbaker, Lucian R. Chirieac, Yaoyu E. Wang, John Quackenbush, Roderick V. Jensen, and Raphael Bueno

Abstract

**Purpose:** To develop a standardized approach for molecular diagnostics, we used the gene expression ratio bioinformatic technique to design a molecular signature to diagnose malignant pleural mesothelioma (MPM) from among other potentially confounding diagnoses and differentiate the epithelioid from the sarcomatoid histologic subtype of MPM. In addition, we searched for pathways relevant in MPM in comparison with other related cancers to identify unique molecular features in MPM.

**Experimental Design:** We conducted microarray analysis on 113 specimens including MPMs and a spectrum of tumors and benign tissues comprising the differential diagnosis of MPM. We generated a sequential combination of binary gene expression ratio tests able to discriminate MPM from other thoracic malignancies. We compared this method with other bioinformatic tools and validated this signature in an independent set of 170 samples. Functional enrichment analysis was conducted to identify differentially expressed probes.

**Results:** A sequential combination of gene expression ratio tests was the best molecular approach to distinguish MPM from all the other samples. Bioinformatic and molecular validations showed that the sequential gene ratio tests were able to identify the MPM samples with high sensitivity and specificity. In addition, the gene ratio technique was able to differentiate the epithelioid from the sarcomatoid type of MPM. Novel genes and pathways specifically activated in MPM were identified.

**Conclusions:** New clinically relevant molecular tests have been generated using a small number of genes to accurately distinguish MPMs from other thoracic samples, supporting our hypothesis that the gene expression ratio approach could be a useful tool in the differential diagnosis of cancers. *Clin Cancer Res; 19(9); 1–10. ©2013 AACR.*

Introduction

Malignant pleural mesothelioma (MPM) is an aggressive malignancy arising from the mesothelial cells of the pleura. This cancer is by and large associated with asbestos exposure, and its worldwide incidence continues to increase even though the commercial use of asbestos has been banned in many Western countries (1). Morphologically, MPM is subclassified into 3 histologic types: epithelioid, sarcomatoid, and biphasic (mixed epithelioid and sarcomatoid). The epithelioid histologic subtype of MPM is the most common and is associated with a more favorable prognosis. However, the overall survival even for patients with epithelioid tumors is dismal. The expected median survival of the average patient diagnosed with MPM is between 4 and 12 months (2). Aggressive cytoreductive therapy followed by combination chemo- and radiotherapy has been associated with prolonged survival in selected patients with early MPM as well as in a number of long-term survivors (2, 3).

Making the correct diagnosis of MPM can be challenging in some cases. The epithelioid type may be difficult to distinguish from adenocarcinoma or thymoma metastatic to the pleura, and the sarcomatoid type of MPM from some sarcomas or other tumors with sarcomatoid histologies (4). Other malignancies in the differential diagnosis of pleural tumors include hemangioendothelioma, thyroid cancer, renal cell carcinoma, lymphoma, undifferentiated carcinomas, and prostate cancer. Benign pleuritis or mesothelial cell proliferation can also confound the diagnosis of MPM.
Translational Relevance

Although long considered the gold standard for diagnosing cancer, customary pathologic approaches are not always successful. Therefore, we applied the bioinformatic technique of gene expression ratio tests to develop and validate molecular signatures for the differential diagnosis of malignant pleural mesothelioma (MPM) as proof-of-principle of the applicability of this technique to cancer diagnosis. As the gene ratio technique is binary, we used a sequential method similar to most clinical pathologic diagnostic approaches. We developed several binary tests to differentiate MPM from all confounding diagnoses. Ultimately, a 26-gene signature derived from sequential gene ratio tests diagnosed MPM with high sensitivity and specificity. This signature required fewer genes than the one identified by standard bioinformatics. We used the same technique to develop a test capable of differentiating the epithelioid from the sarcomatoid histologic subtypes of MPM, a clinically important problem. Finally, we used the same dataset to discover molecular features unique to MPM.

Although long considered the gold standard for diagnosing cancer, immunohistochemical analysis using a panel of both positive and negative stains is customarily required for making a diagnosis of MPM (5). However, no single immunohistochemical stain is diagnostic in all cases. In challenging cases, it is occasionally necessary to carry out electron microscopy to definitively determine the diagnosis (6).

Microarray profiling technology uses gene-specific probes that represent thousands of individual genes allowing simultaneous measurement of their levels of expression in a single experiment. Microarrays have been successfully applied to cancer research for the discovery of novel biomarkers. In particular, many studies have shown that this technology can be used to identify specific signatures capable of classifying types of tumors, predicting patient outcome, and sorting groups of patients with different response to chemotherapy (7). We have previously described a gene expression ratio-based method that translates comprehensive expression profiling data into simple clinical tests based on the expression levels of a relatively small number of genes (8–11). This method identifies genes that are differentially expressed in a statistically significant manner in a pair of distinct clinical conditions and specifies ratios of expression levels for gene pairs that can alone or in combination predict the condition. In particular, we have reported that combinations of a small number of carefully chosen and validated gene expression ratios can be used to develop diagnostic and prognostic tests for several types of cancer (8, 9, 11–14). One limitation of the gene ratio technique, however, is that an individual gene expression ratio determines a binary decision. Therefore, it is only capable of distinguishing between two conditions, thereby limiting its ability to predict one condition among more than two alternatives.

In this study, we investigate whether this limitation may be overcome by iterative application of the gene ratio tests. We use this approach to discriminate MPM from all the other potentially confounding diagnoses as proof-of-principle of the applicability of this method to differential diagnosis. We conducted microarray analysis on 113 tumor specimens, including MPM, and a spectrum of other common thoracic malignancies and benign tissues using Illumina whole genome microarrays. Our goals were to develop methodology using the gene ratio technique to define molecular signatures relevant to the differential diagnosis of MPM and to obtain insights via differential gene expression into pathways uniquely relevant in MPM in comparison with other related cancers.

Materials and Methods

Tumor samples and RNA extraction

Studies using human tissues were approved by and conducted in accordance with the policies of the Institutional Review Boards at the Brigham and Women’s Hospital (BWH; Boston, MA) and the Dana-Farber Cancer Institute (Boston, MA). All tumor samples were collected at surgery as discarded specimens, fresh frozen, stored, and annotated by the institutional tumor bank. For the microarray experiments, 113 tumor and normal samples were used. The histologic distribution and the number of the samples are displayed in Table 1. All the samples included in the microarray experiments had at least 70% tumor cell content, as previously determined (15). For validation of the novel gene expression ratio tests, independent test sets of MPM (n = 100; 63 epithelioid, 27 biphasic, 10

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesothelioma</td>
<td>39</td>
</tr>
<tr>
<td>Epithelioid</td>
<td>24</td>
</tr>
<tr>
<td>Biphasic</td>
<td>7</td>
</tr>
<tr>
<td>Sarcomatoid</td>
<td>8</td>
</tr>
<tr>
<td>Sarcomas</td>
<td>26</td>
</tr>
<tr>
<td>Melanoma</td>
<td>6</td>
</tr>
<tr>
<td>Metastatic thyroid cancer</td>
<td>6</td>
</tr>
<tr>
<td>Lymphomas</td>
<td>5</td>
</tr>
<tr>
<td>Prostate carcinomas</td>
<td>5</td>
</tr>
<tr>
<td>Renal carcinomas</td>
<td>5</td>
</tr>
<tr>
<td>Thymoma</td>
<td>6</td>
</tr>
<tr>
<td>Hemangioendothelioma/pericytoma</td>
<td>4</td>
</tr>
<tr>
<td>Pleura from patients with benign</td>
<td>7</td>
</tr>
<tr>
<td>pulmonary diseases</td>
<td></td>
</tr>
<tr>
<td>Normal Colon</td>
<td>1</td>
</tr>
<tr>
<td>Normal Lung</td>
<td>2</td>
</tr>
<tr>
<td>Lung Adenocarcinoma</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. Thoracic malignancies profiled using Illumina microarrays
sarcomatoid), sarcoma (n = 38); a kind gift from Dr. C.P. Raut (on behalf of the BWH Sarcoma Tumor Bank), adenocarcinoma (n = 20), and normal pleurae in undisturbed and inflamed states from patients without malignancies (n = 12) were used. The visceral pleural surface consists of connective tissue (70%–80%) and a single layer of flattened mesothelial cells (20%–30%), and these cellular contents were represented in all the normal specimens. RNA was extracted using Trizol reagent (Invitrogen Corporation) according to the manufacturer’s instructions. DNase I (Invitrogen Corporation) treatment was conducted according to the manufacturer’s instructions. RNA was quantified using an ND-1000 spectrophotometer (NanoDrop, Fisher Thermo). The integrity of the RNA from the microarray set of samples was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies).

Microarray experiments
To determine the levels of transcripts in each sample, 0.75 μg of total RNA was amplified using the Illumina TotalPrep RNA amplification kit (Applied Biosystems). cRNA was hybridized to Sentrix Human-6 Expression BeadChip (Illumina), subsequently labeled with Cy3-streptavidin (Amersham Biosciences), and scanned with a Bead Station (Illumina). All the hybridization, washing, staining, and scanner procedures were conducted as recommended by the manufacturer. On a single BeadChip, 6 arrays were run in parallel. For quality control across platforms, 2 MicroArray Quality Control (MAQC) samples (MAQCa and MAQCb) were included in the analysis (16). A blind control was also added to check the variability of the expression values across the chips. The probe intensity distribution was examined for quality control, and outliers were removed. Expression profile raw data are available at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42977).

For clustering and functional enrichment analysis, the arrays were normalized by quantile normalization using Bioconductor (17), and differentially expressed probes were identified by linear model using the LIMMA package (17).折叠模型 for multiple comparison tests. Probes with corrected values of less than 0.01 and log fold-change more than 0.75 were identified as differentially expressed probes. One microgram of total RNA was reverse-transcribed using the TaqMan Reverse Transcription Reagents (Applied Biosystems). Real-time quantitative PCR (RT-PCR) was conducted using a SYBR-Green fluorometry-based detection system (Applied Biosystems), as previously described (10). Primer sequences (synthesized by Invitrogen Life Technologies) for all the tests are listed in Supplementary Table S1A and were used for RT-PCR as described in the Supplementary data and as previously published (9).

For each differential diagnostic test, samples were assigned to the diagnosis of MPM when the combined score was more than 1 and not-MPM when the combined score was less than 1. In the epithelioid versus sarcomatoid test, the samples were assigned to the epithelioid histology when the combined score was more than 1 and to sarcomatoid histology when the combined score was less than 1. A list of all the genes included in the tests is reported in Supplementary Table S1B.

Cross validation
Random subsampling cross validations were conducted to evaluate the diagnostic power of our MPM signature using the gene-ratio algorithm. In particular, we applied K-nearest neighbor (KNN; ref. 20) and linear discriminant analyses (LDA; ref. 21) using Bioconductor (17) on the 26-gene signatures used to distinguish the MPM samples. We randomly selected 70% of the samples as the training set to train each classifier, and used the remaining 30% of samples as the test set to evaluate the performance of each classifier. This process was repeated for 10,000 independent iterations to determine the average sensitivities and specificities for each of these 3 methods for comparison.

Identification of diagnostic molecular markers and data analysis
Four different gene ratio–based tests were developed to individually distinguish MPM from other specific diagnoses: one to distinguish MPM from normal pleura (NP), a second to distinguish MPM from all sarcomas, a third to distinguish MPM from renal cell carcinoma (RCC), and a fourth to distinguish MPM from thymoma. Our validated diagnostic test for MPM versus adenocarcinoma was also added to this analysis (9). In addition, we developed a gene ratio–based diagnostic test to discriminate the epithelioid MPM from the sarcomatoid MPM subtype. Detailed information about the training sets chosen for each ratio test is included in the Supplementary data. To find genes differentially expressed between 2 groups of samples in each test, we searched all of the genes represented on the Illumina microarray to identify those with a highly significant difference (P < 1 × 10⁻⁵) and with at least a 2-fold expression difference between matched training sets, using the same selection criteria as that published for MPM versus adenocarcinoma (9). For each test, we chose from 5 to 15 genes meeting these criteria for further analysis. We determined the diagnostic accuracy of candidate gene expression ratio tests as previously described (9).

Real-time quantitative PCR
One microgram of total RNA was reverse-transcribed using the TaqMan Reverse Transcription Reagents (Applied Biosystems). Real-time quantitative PCR (RT-PCR) was conducted using a SYBR-Green fluorometry-based detection system (Applied Biosystems), as previously described (10). Primer sequences (synthesized by Invitrogen Life Technologies) for all the tests are listed in Supplementary Table S1A and were used for RT-PCR as described in the Supplementary data and as previously published (9).

For each differential diagnostic test, samples were assigned to the diagnosis of MPM when the combined score was more than 1 and not-MPM when the combined score was less than 1. In the epithelioid versus sarcomatoid test, the samples were assigned to the epithelioid histology when the combined score was more than 1 and to sarcomatoid histology when the combined score was less than 1. A list of all the genes included in the tests is reported in Supplementary Table S1B.

Functional enrichment analysis
Functional enrichment analysis on both Gene Ontology (GO; ref. 22) and KEGG (kyoto encyclopedia of genes and genomes; ref. 23) was conducted using the DAVID (data-base for annotation, visualization, and integrated discovery) web server (http://david.abcc.ncifcrf.gov/; refs. 24, 25),
using official gene symbols as identifiers to analyze the GO biologic process database, and the KEGG database. Functional enrichment was determined by the expression analysis systematic explorer score. Functional annotations with q value less than 0.2, corresponding to a 20% false discovery rate (FDR), were considered to be significantly enriched, and the fold enrichments were then used to generate the functional heatmap.

Results
Differential diagnosis of MPM using gene ratio-based tests
To discover molecular signatures that would determine the differential diagnosis of MPM, we profiled 113 thoracic malignancies and control tissues using Illumina whole genome microarrays (Table 1). One sample of renal cell carcinoma was excluded from analysis because of poor array quality. To define a diagnostic algorithm to distinguish MPM from all the other samples, we initially explored the possibility of generating a single diagnostic gene expression ratio test capable of distinguishing MPM from all the other thoracic samples. We generated several tests and included in some analyses the normal pleura and lung samples and in others only the tumors. Using the microarray expression values, a few tests were able to correctly discriminate MPM from all the other samples in the training set. We next attempted to use RT-PCR to examine the best expression-based tests using the same specimens analyzed in the microarray. RT-PCR is widely considered the gold standard for gene expression measurement because of its high assay specificity, high detection sensitivity, and wide linear dynamic range. We found that the accuracy of the MPM diagnosis in the training sample set was 80% indicating that the single ratio-based tests were not sufficiently accurate for this application.

The next strategy was to develop a sequential combination of binary diagnostic tests able to distinguish MPM from all the other thoracic samples. Our aim was to mimic the clinical diagnostic practice of pathologists using immunohistochemistry for the diagnosis of MPM, where it has become a standard to use panels of positive and negative antibodies that are sequentially applied and that can vary depending on the differential diagnosis (6). The genes selected for each diagnostic test are reported in Supplementary Table S1B.

First, we applied our validated diagnostic test for MPM versus lung adenocarcinoma to all the microarray samples. In 80 of 112 samples tested, the call was made for MPM. Those included all 39 known MPMs. We repeated the analysis on the same samples using the RT-PCR and found that 93 of 111 were called MPM including all 39 known MPM samples. One metastatic melanoma was excluded from the RT-PCR analysis because its RNA was degraded and no additional specimen from the same sample was available. Because all the normal pleura and lung samples were classified as MPM by the MPM versus adenocarcinoma test using RT-PCR, we next developed a test to discriminate MPM from NP. The test was generated using epithelioid MPM samples only because the epithelioid MPM subgroup is more similar to the NP. When the MPM versus NP test was applied to the remaining microarray cases determined to be MPM by the MPM versus adenocarcinoma test, the call was made for MPM in 73 of 80 samples. Using RT-PCR, 81 of 93 were called MPM. In both platform applications, all the MPMs and the normal samples were correctly classified.

Because most of the samples erroneously called MPM by both the MPM versus adenocarcinoma and MPM versus NP tests were sarcomas, we next developed a test to distinguish MPM from sarcoma. This test correctly classified all 39 known MPM samples and all the remaining “not-MPM” samples using the microarray expression data. The only misclassified sample still classified as MPM was the normal colon control. When the MPM versus sarcoma test was applied to RT-PCR expression data for the samples called MPM after the first 2 tests, 46 of 81 samples, including all known MPM, were called MPM. The overall sensitivity for the sequential diagnosis of MPM was 100% with both microarray and RT-PCR applications of the test, whereas the overall specificity was 99% using the microarray platform and 90% using the RT-PCR platform. The 7 samples that were incorrectly classified as MPM by RT-PCR were renal carcinomas (n = 2), thymoma (n = 2), melanoma (n = 2), and non-Hodgkin’s lymphoma (n = 1). Therefore, we next developed specific gene expression ratio tests to individually discriminate between MPM and those samples. When the MPM versus RC and MPM versus thymoma tests were conducted using either microarray or RT-PCR expression values, all the samples were correctly classified. A schematic representation of the sequential application of the binary tests for both microarray and RT-PCR and the related test data are presented in Fig. 1A and B, and Supplementary Table S2.

We also developed several MPM versus melanoma tests using different numbers of MPMs and all 6 melanoma samples. Those successfully classified all the samples using the microarray data, but called some MPM samples as not-MPM when analyzed by RT-PCR. We assume that the number of the melanoma samples analyzed herein was inadequate to discriminate genes differentially expressed between the 2 groups of tumors.

Microarray diagnosis cross validation
To show that the gene ratio signature (the heat-map is shown in Supplementary Fig. S1) has sensitivity and specificity similar to the more complex algorithm, we compared the gene ratio approach with other machine-learning–based approaches. In particular, we applied the KNN and the LDA analyses to the 26-gene microarray signatures used to distinguish the MPM samples from all the other thoracic malignancies. We found that in the training set the specificity for the 3 methods was the same (93%). However, the sequential gene ratio test had a sensitivity of 100%, whereas the other 2 methods had sensitivities of 90% or less, indicating that gene ratio algorithm more reliably and accurately identified all of the MPM samples (Table 2).
RT-PCR validation with a large independent set

To validate tests developed by the gene ratio algorithm, we used RT-PCR to analyze an independent test set of 100 MPM samples and 70 samples of tumors comprising the differential diagnosis of MPM (adenocarcinomas, sarcomas, and NP). A sufficient number of thymoma and RCC specimens was not available to be included in the test set because of the rarity of such metastatic diseases to the pleura. We included these tests as a proof-of-principle that will need to be further validated in the future. We first applied the MPM versus adenocarcinoma test. All the MPMs and adenocarcinomas were correctly classified. One sarcoma and 2 NP samples were classified as adenocarcinoma, whereas all the remaining sarcoma and NP samples were classified as MPM. When the MPM versus NP test was applied to the samples called MPM on the previous test, 98 MPM and 7 NP samples were correctly classified. Two MPM samples were called not-MPM, but additional review showed that the actual specimens used for the RT-PCR had 0% tumor content. One NP and all the sarcomas were classified as MPM using that test. The MPM versus sarcoma test was then applied to the samples determined to be MPM, and 90 of 98 MPM were properly classified as MPM. Eight MPM cases (2 epithelioid, 4 biphasic, and 2 sarcomatoid) were incorrectly classified as sarcomas. The lower specificity of the MPM versus sarcoma test is most likely due to a subgroup of MPM that have expression profiles similar to the sarcoma samples. Also, 2 samples (1 miscalled NP sample and 1 sarcoma) were incorrectly classified as MPM. The remaining 2 tests, MPM versus RCC and MPM versus thymoma, appropriately classified all the MPM samples and included the miscalled sarcoma and NP samples in the MPM group (Supplementary Table S3). Even though we used a specific sequential order for applying the tests, the same results were obtained in all possible sequences.

The overall diagnostic sensitivity of the 26-gene signatures in the test set analyzed by RT-PCR was 92%, whereas the specificity was 97%. The results are schematically represented in Table 3.

Table 2. Comparison of cross-validation results for the sequential gene ratio test and the KNN ad LDA analyses

<table>
<thead>
<tr>
<th></th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential gene ratio test</td>
<td>0.93</td>
<td>1</td>
</tr>
<tr>
<td>(hierarchical)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k-Nearest neighbor analysis</td>
<td>0.93</td>
<td>0.9</td>
</tr>
<tr>
<td>Linear discriminate analysis</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td>n (iteration) = 100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Diagnostic test epithelioid MPM versus sarcomatoid MPM

Next we developed a test to differentiate the epithelioid from sarcomatoid subtypes, as this is clinically important for staging and prognosis in MPM. Using the training set expression data, we developed a 4-gene 3-ratio test that is able to distinguish all the epithelioid MPM from all the sarcomatoid MPM samples. The test was then validated by RT-PCR in the same 39 training set MPM samples. All the epithelioid and sarcomatoid MPM were correctly classified. The same test was then applied using RT-PCR to an
Table 3. Sensitivity and specificity of the 26-gene signature in the training and in the test set by the RT-PCR

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Independent Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>(39 MPM + 72 DDX&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>(100 MPM + 70 DDX&lt;sup&gt;a&lt;/sup&gt;)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPM</td>
<td>100%</td>
<td>90%</td>
</tr>
<tr>
<td>sarcomatoid</td>
<td>92%</td>
<td>97%</td>
</tr>
</tbody>
</table>

<sup>a</sup>Differential diagnosis.

Biological pathways differentially expressed between MPM and other thoracic malignancies

To identify novel molecular pathways specific for MPM, we searched for differentially expressed genes for MPM versus other tumor types. Linear model analysis was conducted using the LIMMA package to detect differential expression between MPM and other tumor types, and 167 probes, corresponding to 156 unique genes, were identified as differentially expressed (P < 0.01; Supplementary Table S4). These probes represent the minimum signature required to distinguish MPM from all the other malignancies using the microarray expression data. We used the 167 probes to carry out hierarchical clustering analysis and obtained a cluster dendrogram showing 2 major branches (Fig. 2A). A detailed description of the cluster dendrogram is reported in the Supplementary data. The heatmap of the 167 probes is shown in Fig. 2B.

To determine the biologic function of the 167 probes differentially expressed between MPM and all the other thoracic malignancies, we conducted gene enrichment analysis to detect highly enriched functional terms and biologic pathway definitions according to the Gene Ontology Biological Process and the KEGG databases, respectively, using the DAVID web server (23, 24). Forty-five pathways were significantly enriched (q value < 0.2) in the MPM group (Supplementary Table S5). Because of the experimental design and the heterogeneity of the thoracic tumors in the comparison, the analysis was not able to identify pathways specifically enriched in the other thoracic malignancy group.

We classified the pathways upregulated in MPM into at least 4 main groups: extracellular organization, development, response to endogenous, mechanical, or hormonal stimuli, and immune response.

Biological pathway differently expressed between epithelioid MPM and sarcomatoid MPM

When the same analysis was applied to the MPM subtype expression data, we found 183 significant probes corresponding to 172 genes differentially expressed between the 2 types (Supplementary Table S6). The dendogram and the heatmap, displayed in Fig. 3A and B, showed that all the epithelioid and the sarcomatoid MPMs clustered into 2 distinct branches. When we searched for the biologic function of the 183 probes, we found that the upregulated pathways (Supplementary Table S7) in the epithelioid group were related to transmembrane receptor protein tyrosine kinase signaling, germ cell development, and regulation of cell proliferation. The downregulated pathways (Supplementary Table S7) in the epithelioid group were related to response to external stimulus, blood vessel development, cell adhesion, and regulation of secretion.

Discussion

A major recent focus of medical science has been to characterize the molecular basis of cancer using genomewide analytic technologies. Microarray-based analysis has become an essential tool for the genetic profiling of biologic samples owing to its ability to assess the expression of thousands of genes simultaneously (26). The practical applications of this technology include the identification of biomarkers associated with the disease and of expression patterns of genes that distinguish subclasses of samples. In cancer, molecular profiling can be used to distinguish subclasses among tumors that appear identical under the microscope but may have distinct clinical features and therapeutic considerations. A key step in the practical implementation of this technology is the development of tools that classify tumor samples according to their gene expression levels in a reproducible manner that can be used clinically. Specifically, given a collection of gene expression data, grouped into classes, the goal is to determine to which class a new unknown tissue sample likely belongs. However, methods for using gene expression profiling with microarrays or other platforms are not yet sufficiently established or in widespread clinical use, and further optimizations are necessary before reliable and accessible techniques for such clinical applications are available for most tumors. The work described herein aims to create an algorithm that can be applied to facilitate the application of gene expression data to cancer diagnostics.

The use of diagnostic algorithms is well established in surgical pathology practice and is based on developing a differential diagnosis related to the tissue of origin and the disease process. Although examination of a pleural effusion by cytologic/cell block exam may lead to a diagnosis of some MPMs, the definitive diagnosis usually requires histopathologic examination of tumor tissue. Immunohistochemistry and other pathologic tests may not always provide an optimal answer, and determining the specific gene expression pattern of a tumor, perhaps by profiling with microarrays or RT-PCR, may represent an adjunct tool to resolve diagnostic dilemmas and increase the diagnostic
accuracy. It may provide additional support for a diagnosis by identifying tumor-specific genetic signatures, help in prognostication, and even aid in determining the best therapeutic options (26).

In recent years, we have focused on using gene expression measurements to predict clinical parameters in cancer. Specifically, we have investigated the feasibility of using ratios of gene expression levels and rationally chosen thresholds to accurately distinguish between genetically different tissues. We have developed gene expression ratio–based tests to discriminate MPM from lung adenocarcinoma, and to predict the outcome of patients with...
MPM. Both tests have been validated in several independent retrospective tissue biopsy sample sets as well as in an additional independent prospective cohort using both fresh-frozen tumor biopsies and ex vivo fine-needle aspiration biopsies (8–11, 27, 28).

In this study, we explored the possibility of using gene expression ratio tests to distinguish MPM from all other potentially confounding thoracic malignancies and normal tissues that represent the actual spectrum of differential diagnosis of MPM as a proof-of-principle. Our hypothesis was that the gene ratio method can be used to translate the genomic signature into diagnostic tests that may aid in the diagnosis of MPM. We profiled 112 thoracic samples consisting of the differential diagnostic spectrum of MPM using Illumina whole genome microarrays. We discovered that a single signature by any methodology was not sufficiently accurate, but that the application of a sequential combination of binary gene expression ratio tests was able to reliably distinguish MPM from all the other thoracic samples. This sequential approach is quite similar to most clinical diagnostic approaches used in the routine clinical laboratory. Most interestingly we found that even in the training set the sequential gene ratio approach was more accurate than a single gene ratio test method as well as any other bioinformatic algorithm. In addition, the sequential combination of binary gene ratio tests required the analysis of a signature of only 26 genes, whereas the minimum signature identified by microarray analysis to distinguish MPM from all the other thoracic malignancies consisted of

Figure 3. Graphic display of hierarchical cluster analysis using 183 probes corresponding to 172 genes differentially expressed between epithelioid and sarcomatoid MPM. The red asterisk indicates the sarcomatoid mesothelioma samples. A, hierarchical clustering of the biologic samples. B, two-way hierarchical clustering of the biologic samples and the extracted genes. Probes are annotated with gene symbol on the right. Relative gene expression levels are given by the scale at the top.
167 probes. Furthermore, the latter approach would require an array platform with its inherent limitations.

The findings described herein indicate that tests generated from a relatively small number of genes are able to accurately distinguish MPMs from thoracic samples supporting our hypothesis that gene ratio tests could provide a useful clinical adjunct in the diagnosis of MPM.

An important underlying feature of this new methodology is that a comprehensive diagnostic molecular test may be assembled one component at a time according to the molecular characteristic of the tumor under scrutiny as compared with each single facet of the differential diagnosis. If a single expression-based test is generated to identify 1 tumor from all the other samples, the selection of the number and the types of samples may influence the choice of the genes and the algorithms selected and, consequently, introduce bias. The method proposed herein allows for an independent development of each of the components to maximize its accuracy. Thus, we believe that this approach will be broadly applicable to other tumor types.

Recent clinical trials have shown that determining the accurate histology is an important factor for individualizing treatment, based on either safety or efficacy outcomes in several types of cancers (29, 30). Here, we show for the first time that the gene ratio technique is also able to distinguish between histologic subtypes of MPM with very high sensitivity and specificity supporting the hypothesis that this technique may be useful for other applications in cancer.

It is currently accepted that different cancers require specific somatic alterations of genes and pathways. We undertook pathway analysis of diverse thoracic malignancies and subtypes of MPM and have shown that there is a significant enrichment in gene expression related to specific functions in different tumors. We showed that in MPM, there are several pathways related to 4 main functions that are differentially regulated compared with all the other thoracic malignancies. Some of these functions have already been explored in MPM. Several investigations have studied the role of the immune system in MPM; in particular, a correlation has been shown between the presence of lymphocyte infiltration and better prognosis in patients with MPM (31–33). Some genes involved in the extracellular organization, such as MERLIN (34), have also been shown to play a key role in MPM.

Interestingly, several genes that have functions related to vasculature development, adhesion, and regulation of secretion were found to be differentially expressed between epithelioid and sarcomatoid types indicating that the 2 types have major molecular differences related to these pathways. Therefore, our findings suggest novel candidate genes and pathways that are preferentially activated or inactivated in each type of MPM.

In summary, using expression profiles we have identified a sequential combination of binary gene expression ratio tests that can distinguish MPM from other common thoracic malignancies; we have generated a diagnostic gene ratio test able to identify the subtypes of MPM; and we have provided novel molecular evidence to guide future investigations.

Disclosure of Potential Conflicts of Interest

R. Bueno has a commercial research grant from Myriad, Genentech, Novartis, Pam Gen, Siemens, Exosome Diagnostics, and Castle Biosciences; is a consultant/advisory board member of Exosome and CollaboRX, and has expert testimony in Defense chart reviews. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: A. De Rienzo, B.Y. Yeap, J. Quackenbush, R. Bueno Development of methodology: A. De Rienzo, R. Bueno Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. De Rienzo, W.G. Richards, M.H. Coleman, I.R. Chirieac Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. De Rienzo, W.G. Richards, B.Y. Yeap, L. R. Chirieac, Y.E. Wang, J. Quackenbush, R.V. Jensen, R. Bueno Writing, review, and/or revision of the manuscript: A. De Rienzo, W.G. Richards, B.Y. Yeap, J. Quackenbush, R.V. Jensen, R. Bueno Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): W.G. Richards, M.H. Coleman, P.E. Sugarbaker, R. Bueno Study supervision: R. Bueno

Acknowledgments

The authors would like to thank Angelina Lindsay and Stephen Addington for the technical support with the tissues; Dr. Chandrjajit P. Raut, M.D., M.Sc., and the Center for Sarcoma and Bone Oncology at the Dana-Farber/Brigham and Women’s Cancer Center for the banked sarcoma tissues.

Grant Support

This work was supported by National Cancer Institute (RO1-120528 to R. Bueno) as well as by grants from the International Mesothelioma Program at BWH (to R. Bueno) and the Maurice Favell Fund at the Vancouver Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 5, 2012; revised February 26, 2013; accepted March 1, 2013; published OnlineFirst March 14, 2013.

References

De Rienzo et al.

Sequential Binary Gene Ratio Tests Define a Novel Molecular Diagnostic Strategy for Malignant Pleural Mesothelioma


Clin Cancer Res  Published OnlineFirst March 14, 2013.