



Published in final edited form as:

*Clin Cancer Res.* 2013 May 1; 19(9): 2493–2502. doi:10.1158/1078-0432.CCR-12-2117.

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

Assunta De Rienzo<sup>1</sup>, William G. Richards<sup>1</sup>, Beow Y. Yeap<sup>2</sup>, Melissa H. Coleman<sup>1</sup>, Peter Sugarbaker<sup>1</sup>, Lucian R. Chiriac<sup>3</sup>, Yaoyu E. Wang<sup>4</sup>, John Quackenbush<sup>5</sup>, Roderick V. Jensen<sup>6</sup>, and Raphael Bueno<sup>1</sup>

<sup>1</sup>The Thoracic Surgery Oncology Laboratory, and the International Mesothelioma Program ([www.impmeso.org](http://www.impmeso.org)), Division of Thoracic Surgery, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA 02115, USA

<sup>2</sup>Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

<sup>3</sup>Departments of Pathology, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA 02115, USA

<sup>4</sup>Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA

<sup>5</sup>Center for Cancer Computational Biology, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Harvard School of Public Health, Boston, Massachusetts 02215, USA

<sup>6</sup>Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA

### 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 MPM from among other potentially confounding diagnoses and differentiate the epithelioid from the sarcomatoid histological subtype of MPM. In addition, we searched for pathways relevant in MPM in comparison to other related cancers to identify unique molecular features in MPM.

**Experimental Design**—We performed 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 to other bioinformatic tools and validated this signature in an independent set of 170 samples. Functional enrichment analysis was performed 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.

---

Corresponding Author: Raphael Bueno, Division of Thoracic Surgery, Brigham and Women's Hospital 75 Francis Street, Boston, MA 02115, USA. Phone: (617) 732-6824; Fax: (617) 566-3441; [rbueno@partners.org](mailto:rbueno@partners.org).

The authors disclose no potential conflicts of interest

**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.

### Keywords

Mesothelioma; gene expression ratio test; microarray; diagnosis; gene enrichment analysis

---

### 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 histological 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 radiation therapy 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. Although histological examination is the gold standard for the diagnosis of MPM, 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 perform 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 able to classify types of tumors, to predict patient outcome, and sort groups of patients with different response to chemotherapy (7). We have previously described a gene expression ratio-based method used to translate comprehensive expression profiling data into simple clinical tests that are 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 between two clinically distinct conditions (such as diagnosis or prognosis) 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). However, to-date one limitation of the gene ratio technique is that an individual gene expression ratio determines a binary decision. Therefore, it is only capable of

distinguishing between two conditions 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 performed microarray analysis on 113 tumor specimens including MPMs, 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 differential diagnosis of MPM and to obtain insights via differential gene expression into pathways uniquely relevant in MPM in comparison to 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) and the Dana Farber Cancer Institute. 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 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 Triazol reagent (Invitrogen Corporation, Carlsbad, CA, USA) according to the manufacturer's instructions. DNase I (Invitrogen Corporation, Carlsbad, CA, USA) treatment was performed according to the manufacturer's instructions. RNA was quantified using an ND-1000 spectrophotometer (*NanoDrop* Fisher Thermo, Wilmington, DE, USA). The integrity of the RNA from the microarray set of samples was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

### Microarray experiments

To determine the levels of transcripts in each sample, 0.75 µg of total RNA was amplified using Illumina TotalPrep RNA amplification kit (Applied Biosystems, Foster City, CA, USA). cRNA was hybridized to Sentrix Human-6 Expression BeadChip (Illumina, San Diego, CA, USA), subsequently labeled with Cy3-streptavidin (Amersham Biosciences, Little Chalfont, United Kingdom), and scanned with a Bead Station (Illumina, San Diego, CA, USA). All the hybridization, washing, staining and scanner procedures were performed as recommended by the manufacturer. On a single BeadChip, six arrays were run in parallel. For quality control across platforms, two 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 (18). Benjamini-Hochberg method was used to correct p-value for multiple comparison tests. Probes with corrected values of  $<0.01$  and log fold change  $>1.5$  were included in the analysis. Hierarchical clustering was performed with Euclidean distance in R. All the analyses described in this section were performed using MultiExperiment Viewer (MeV), a Java application designed to allow the analysis of microarray data to identify patterns of gene expression and differentially expressed genes (19).

### 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 carcinoma (RC), and a fourth to distinguish MPM from thymoma. Our validated diagnostic test MPM vs. 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 two groups of samples in each test, we searched all of the genes represented on the Illumina microarray to identify those with highly significant difference ( $P < 1 \times 10^{-5}$ ) and with at least two-fold expression difference between matched training sets, using the same selection criteria as that published for MPM vs. 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  $\mu\text{g}$  of total RNA was reverse-transcribed using Taq-Man Reverse Transcription reagents (Applied Biosystems, Foster City, CA). Real-time quantitative PCR (RT-PCR) was performed using a SYBR-Green fluorometry-based detection system (Applied Biosystems, Foster City, CA), as previously described (10). Primer sequences (synthesized by Invitrogen Life Technologies) for all the tests are listed in Supplementary Table 1A and were used for RT-PCR as described in the Supplementary data and as previously published (9).

For each differential diagnosis test, samples were assigned to the diagnosis of MPM when the combined score was  $>1$  and of not-MPM when the combined score was  $<1$ . In the epithelioid vs. sarcomatoid test, the samples were assigned to the epithelioid histology when the combined score was  $>1$  and to sarcomatoid histology when the combined score was  $<1$ . A list of all the genes included in the tests is reported in Supplementary Table 1B.

### Cross validation

Random sub-sampling cross-validations were performed to evaluate the diagnostic power of our MPM signature using the ratio algorithm. In particular, we applied K-nearest neighbor (KNN) (20) and linear discriminator (LDA) analyses (21) using Bioconductor (17) on the 26 gene signatures used to distinguish the MPM samples. We randomly selected 70% of the samples as 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 three methods for comparison.

## Functional Enrichment Analysis

Functional enrichment analysis on both Gene Ontology (GO) (22) and KEGG (23) was performed using the DAVID web server (<http://david.abcc.ncifcrf.gov/>) (24, 25), using official gene symbols as identifiers to analyze GO biological process database and the KEGG database. Functional enrichment was determined by EASE score. Functional annotations with q-value of  $<0.2$ , corresponding to 20% False Discovery Rate (FDR), were considered to be significantly enriched, and the fold enrichments were then used to generate the functional heat map.

## Results

### Differential diagnosis of MPM using gene ratio-based tests

In order 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 carcinoma was excluded from analysis due to 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 able to discriminate MPM from all the other thoracic samples. We generated several tests and included in some analyses the normal pleura and the normal 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 due to 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 1B.

First, we applied our validated diagnostic test MPM vs. lung adenocarcinoma to all the microarray samples. We found that 80 of 112 samples were called 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.

Since all the NPs and normal lung samples were classified as MPM by the MPM vs. adenocarcinoma test using RT-PCR, we next developed a test to discriminate MPM from NP. The test was generated using only epithelioid MPMs because the epithelioid MPM subgroup is more similar to the normal pleura. When the MPM vs. NP test was applied to the remaining microarray cases called MPM by the MPM vs. adenocarcinoma test, 73 of 80 samples were called MPM. 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 vs. adenocarcinoma and MPM vs. NP tests were sarcomas, we next developed a test to distinguish MPM from sarcomas. This test was able to correctly classify all the 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 vs. sarcoma test was applied to RT-PCR expression data for the samples called MPM after the first two 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 seven samples that were incorrectly classified as MPM by RT-PCR were renal carcinomas (n = 2), thymoma (n = 2), melanoma (n = 2), and non-Hodgkins lymphoma (n = 1). Therefore, we next developed specific gene expression ratio tests to individually discriminate between MPM and those samples. When the MPM vs. RC and MPM vs. thymoma tests were performed 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 Figure 1 A, Band Supplementary Table 2.

We also developed several MPM vs. melanoma tests using different numbers of MPMs and all six 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 two groups of tumors.

### Microarray diagnosis-cross validation

To demonstrate that the gene ratio signature (the heat-map is showed in Supplementary Figure 1) has sensitivity and specificity similar to more complex algorithm, we compared the gene ratio approach to 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 three methods was the same (93%). However, the sequential gene ratio test had a sensitivity of 100%, whereas the other two methods had sensitivity 90%, indicating that gene ratio algorithm compared to the other two approaches is the most reliable to accurately identify all the MPM samples (Table 2).

### RT-PCR validation with a large independent set

To validate tests developed using the gene ratio algorithm, we analyzed by RT-PCR an independent test set of 100 MPM samples and 70 samples of tumors consisting the differential diagnosis of MPM (adenocarcinomas, sarcomas, and normal pleura). 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 which will need to be further validated in the future. We first applied the test MPM vs. adenocarcinomas. 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 MPM. When the MPM vs. 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 vs. sarcoma test was then applied to the samples called MPM, 90 of 98 MPM were properly classified as MPM. Eight MPM (2 epithelioid, 4 biphasic, and 2 sarcomatoid) were incorrectly classified as sarcomas. The lower specificity of the test MPM vs. sarcoma 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 two tests, MPM vs. RC and MPM vs. thymoma, appropriately classified all the MPM samples and included the miscalled sarcoma and NP samples in the MPM group (Supplementary Table 3). 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.

### **Diagnostic test epithelioid MPM vs. sarcomatoid MPM**

We next undertook to develop a test differentiating epithelioid from sarcomatoid types of MPM as this is clinically important in terms of staging and prognosis. Using the training set expression data, we developed a 4-gene 3-ratio test 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 independent test set of 100 MPM samples showing that 8 of 9 sarcomatoid samples (89%) and 62 of 63 (98%) epithelioid MPMs were correctly classified. One sarcomatoid sample was excluded from the analysis because the result of this test was non diagnostic (1.0). The biphasic MPMs were distributed to both MPM groups most likely according to their cellular heterogeneity.

### **Biological pathways differentially expressed between MPM and the other thoracic malignancies and between epithelioid MPM and sarcomatoid MPM**

To identify novel molecular pathways specific for MPM, we searched for differentially expressed genes for MPM vs. other tumor types. Linear model analysis was performed using the LIMMA package to detect differential expression between MPM and the other tumor types, and 167 probes, corresponding to 156 unique genes, were identified as differentially expressed ( $p$ -value  $< 0.01$ ) (Supplementary Table 4). These probes represent the minimal signature to distinguish MPM from all the other malignancies using the microarray expression data. We used the 167 probes to perform hierarchical clustering analysis and obtained a cluster dendrogram showing two major branches (Figure 2 A). A detailed description of the cluster dendrogram is reported in the Supplementary data. The heat-map of the 167 probes is shown in Figure 2B.

To determine the biological function of the 167 probes differentially expressed between MPM and all the other thoracic malignancies, we performed gene enrichment analysis to detect highly enriched functional terms and biological pathways 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 5). 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 up-regulated in MPM into at least four main groups: extracellular organization, development, response to endogenous, mechanical, or hormonal stimuli, and immune response.

### **Biological pathway differentially 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 two types (Supplementary Table 6). The dendrogram and the heat-map, displayed in Figure 3 A

and B showed that all the epithelioid and the sarcomatoid MPMs clustered into two distinct branches. When we searched for the biological function of the 183 probes, we found that the up-regulated pathways (Supplementary Table 7) in the epithelioid group were related to transmembrane receptor protein tyrosine kinase signaling, germ cell development, and regulation of cell proliferation. The down-regulated pathways (Supplementary Table 7) 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 genome-wide analytic technologies. Microarray based analysis has become an essential tool for the genetic profiling of biological samples due 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 for classifying 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 which class a new unknown tissue sample likely belongs to. However, methods for utilizing 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 cytological/cell block exam may lead to a diagnosis of some MPMs, the definitive diagnosis usually requires histopathologic examination of tumor tissue. Immunohistochemistry and other pathological 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 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 MPM patients. Both tests have been validated in several independent retrospective tissue biopsy sample sets as well as in an additional independent prospective cohort, utilizing 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 the 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 minimal signature identified by microarray analysis to distinguish MPM from all the other thoracic malignancies consisted of 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 a small number of genes are able to accurately distinguish MPMs from thoracic samples supporting our hypothesis that the gene ratio tests could provide a useful clinical adjunct in the diagnosis of MPM.

An important feature that underlies 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 in focus as compared to each single facet of the differential diagnosis. If a single expression based test is generated to identify one 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 demonstrated 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 histological 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 demonstrated 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 four main functions that are differentially regulated compare to 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, it has been shown a correlation between the presence of a lymphocyte infiltration and a better prognosis in MPM patients (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 having functions related to vasculature development, adhesion, and regulation of secretion were found differentially expressed between epithelioid and sarcomatoid types indicating that the two types have major molecular differences related to these pathways. Therefore, our findings suggest novel candidate genes and pathways 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 able to distinguish MPM from other common thoracic malignancies; we have generated a diagnostic gene ratio test able to identify the type of MPM; and we have provided novel molecular evidence to guide future investigations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The study sponsors played no role in the study design, in the collection, analysis, or interpretation of data, in the writing of the report, or in the decision to submit the paper for publication. We thank Angelina Lindsay and Stephen Addington for the technical support with the tissues; Dr. Chandrajit 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.

Support for this work was provided by National Cancer Institute (RO1-120528 to RB) as well as by grants from the International Mesothelioma Program at BWH (to RB) and the Maurice Favell Fund at the Vancouver Foundation (to RB).

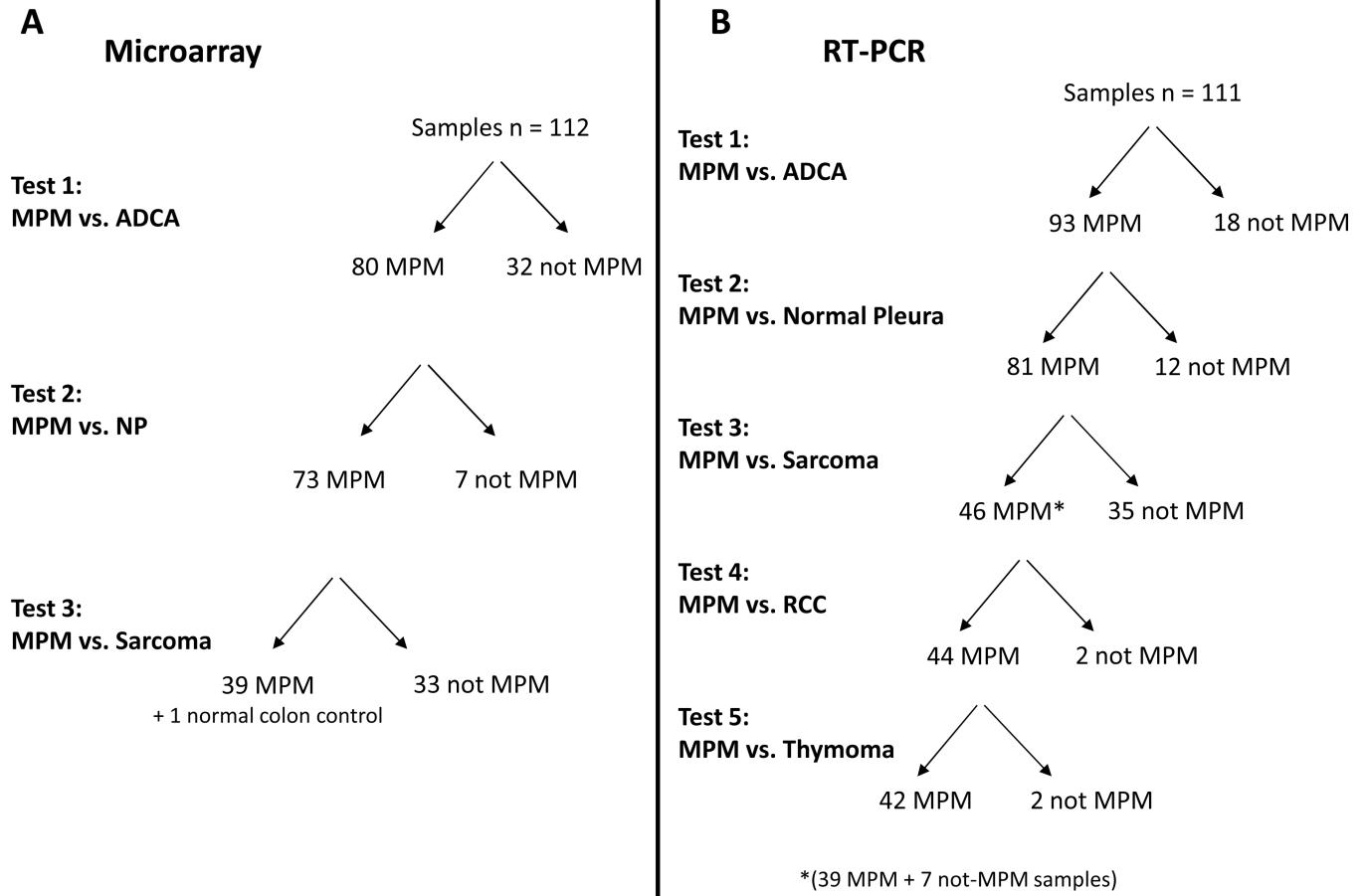
## References

1. Weder W. Mesothelioma. *Ann Oncol.* 2010; 21(Suppl 7):vii326–vii333. [PubMed: 20943637]
2. Sugarbaker DJ, Wolf AS, Chirieac LR, Godleski JJ, Tilleman TR, Jaklitsch MT, et al. Clinical and pathological features of three-year survivors of malignant pleural mesothelioma following extrapleural pneumonectomy. *Eur J Cardiothorac Surg.* 2011; 40:298–303. [PubMed: 21310625]
3. Wolf AS, Richards WG, Tilleman TR, Chirieac L, Hurwitz S, Bueno R, et al. Characteristics of malignant pleural mesothelioma in women. *Ann Thorac Surg.* 2010; 90:949–956. discussion 56. [PubMed: 20732523]
4. Ray M, Kindler HL. Malignant pleural mesothelioma: an update on biomarkers and treatment. *Chest.* 2009; 136:888–896. [PubMed: 19736192]
5. Stahel RA, Weder W, Felip E. Malignant pleural mesothelioma: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol.* 2009; 20(Suppl 4):73–75. [PubMed: 19454470]
6. Husain AN, Colby TV, Ordonez NG, Krausz T, Borczuk A, Cagle PT, et al. Guidelines for pathologic diagnosis of malignant mesothelioma: a consensus statement from the International Mesothelioma Interest Group. *Arch Pathol Lab Med.* 2009; 133:1317–1331. [PubMed: 19653732]
7. Quackenbush J. Microarray analysis and tumor classification. *N Engl J Med.* 2006; 354:2463–2472. [PubMed: 16760446]
8. Gordon GJ, Dong L, Yeap BY, Richards WG, Glickman JN, Edenfield H, et al. Four-gene expression ratio test for survival in patients undergoing surgery for mesothelioma. *J Natl Cancer Inst.* 2009; 101:678–686. [PubMed: 19401544]
9. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Ramaswamy S, et al. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res.* 2002; 62:4963–4967. [PubMed: 12208747]
10. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Richards WG, et al. Using gene expression ratios to predict outcome among patients with mesothelioma. *J Natl Cancer Inst.* 2003; 95:598–605. [PubMed: 12697852]
11. Gordon GJ, Rockwell GN, Godfrey PA, Jensen RV, Glickman JN, Yeap BY, et al. Validation of genomics-based prognostic tests in malignant pleural mesothelioma. *Clin Cancer Res.* 2005; 11:4406–4414. [PubMed: 15958625]
12. Bueno R, Loughlin KR, Powell MH, Gordon GJ. A diagnostic test for prostate cancer from gene expression profiling data. *J Urol.* 2004; 171:903–906. [PubMed: 14713850]

13. Dong L, Bard AJ, Richards WG, Nitz MD, Theodorescu D, Bueno R, et al. A gene ratio-based diagnostic test for bladder cancer. *Computational Biology and chemistry: Advances and Applications*. 2009; 2:17–22.
14. Gordon GJ, Richards WG, Sugarbaker DJ, Jaklitsch MT, Bueno R. A prognostic test for adenocarcinoma of the lung from gene expression profiling data. *Cancer Epidemiol Biomarkers Prev*. 2003; 12:905–910. [PubMed: 14504202]
15. Richards W, Van Oss S, Glickman J, Chirieac L, Yeap B, Dong L, et al. A microaliquoting technique for precise histological annotation and optimization of cell content in frozen tissue specimens. *Biotech Histochem*. 2007;1–9. [PubMed: 17917854]
16. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*. 2006; 24:1151–1161. [PubMed: 16964229]
17. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5:R80. [PubMed: 15461798]
18. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004; 3 Article3.
19. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. TM4 microarray software suite. *Methods Enzymol*. 2006; 411:134–193. [PubMed: 16939790]
20. Fisher R. The use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*. 1936; 7:179–188.
21. Duda, RO.; Hart, PE.; Stork, DG. *Pattern Classification*. Wiley-Interscience; 2001.
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–29. [PubMed: 10802651]
23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000; 28:27–30. [PubMed: 10592173]
24. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003; 4:P3. [PubMed: 12734009]
25. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]
26. Broadhead ML, Clark JC, Dass CR, Choong PF. Microarray: an instrument for cancer surgeons of the future? *ANZ J Surg*. 2010; 80:531–536. [PubMed: 20795968]
27. Gordon GJ. Transcriptional profiling of mesothelioma using microarrays. *Lung Cancer*. 2005; 49(Suppl 1):S99–S103. [PubMed: 15950812]
28. De Rienzo A, Dong L, Yeap BY, Jensen RV, Richards WG, Gordon GJ, et al. Fine-needle aspiration biopsies for gene expression ratio-based diagnostic and prognostic tests in malignant pleural mesothelioma. *Clin Cancer Res*. 2011; 17:310–316. [PubMed: 21088255]
29. Langer CJ, Besse B, Gualberto A, Brambilla E, Soria JC. The evolving role of histology in the management of advanced non-small-cell lung cancer. *J Clin Oncol*. 2010; 28:5311–5320. [PubMed: 21079145]
30. Leong AS, Zhuang Z. The changing role of pathology in breast cancer diagnosis and treatment. *Pathobiology*. 2011; 78:99–114. [PubMed: 21677473]
31. Leigh RA, Webster I. Lymphocytic infiltration of pleural mesothelioma and its significance for survival. *S Afr Med J*. 1982; 61:1007–1009. [PubMed: 7089768]
32. Robinson BW, Robinson C, Lake RA. Localised spontaneous regression in mesothelioma -- possible immunological mechanism. *Lung Cancer*. 2001; 32:197–201. [PubMed: 11325491]
33. Robinson C, Robinson BW, Lake RA. Sera from patients with malignant mesothelioma can contain autoantibodies. *Lung Cancer*. 1998; 20:175–184. [PubMed: 9733052]
34. Stamenkovic I, Yu Q. Merlin, a "magic" linker between extracellular cues and intracellular signaling pathways that regulate cell motility, proliferation, and survival. *Curr Protein Pept Sci*. 2010; 11:471–484. [PubMed: 20491622]

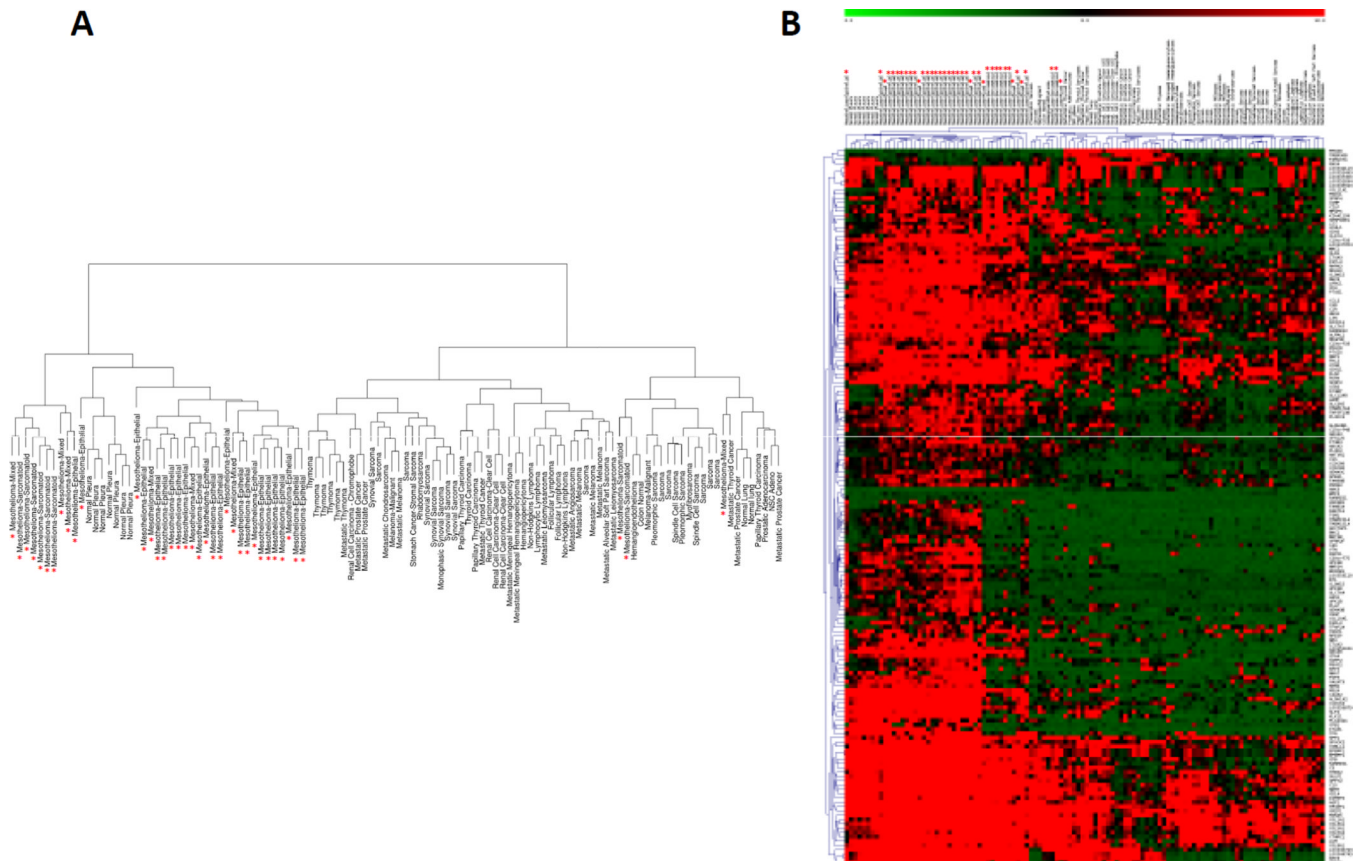
### Statement of Translational Relevance

Though long considered the gold-standard for diagnosing cancer, customary pathological 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 mesothelioma (MPM) as proof of principle of the applicability of this technique to cancer diagnosis. Since the gene ratio technique is binary, we used a sequential method similar to most clinical pathological diagnostic approaches. We developed several binary tests to differentiate MPM from all confounding diagnoses. Altogether, a 26-gene signature performed by the 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 able to differentiate the epithelioid from the sarcomatoid histological subtypes of MPM, a clinically important problem. Finally, we used the same dataset to discover molecular features unique to MPM.

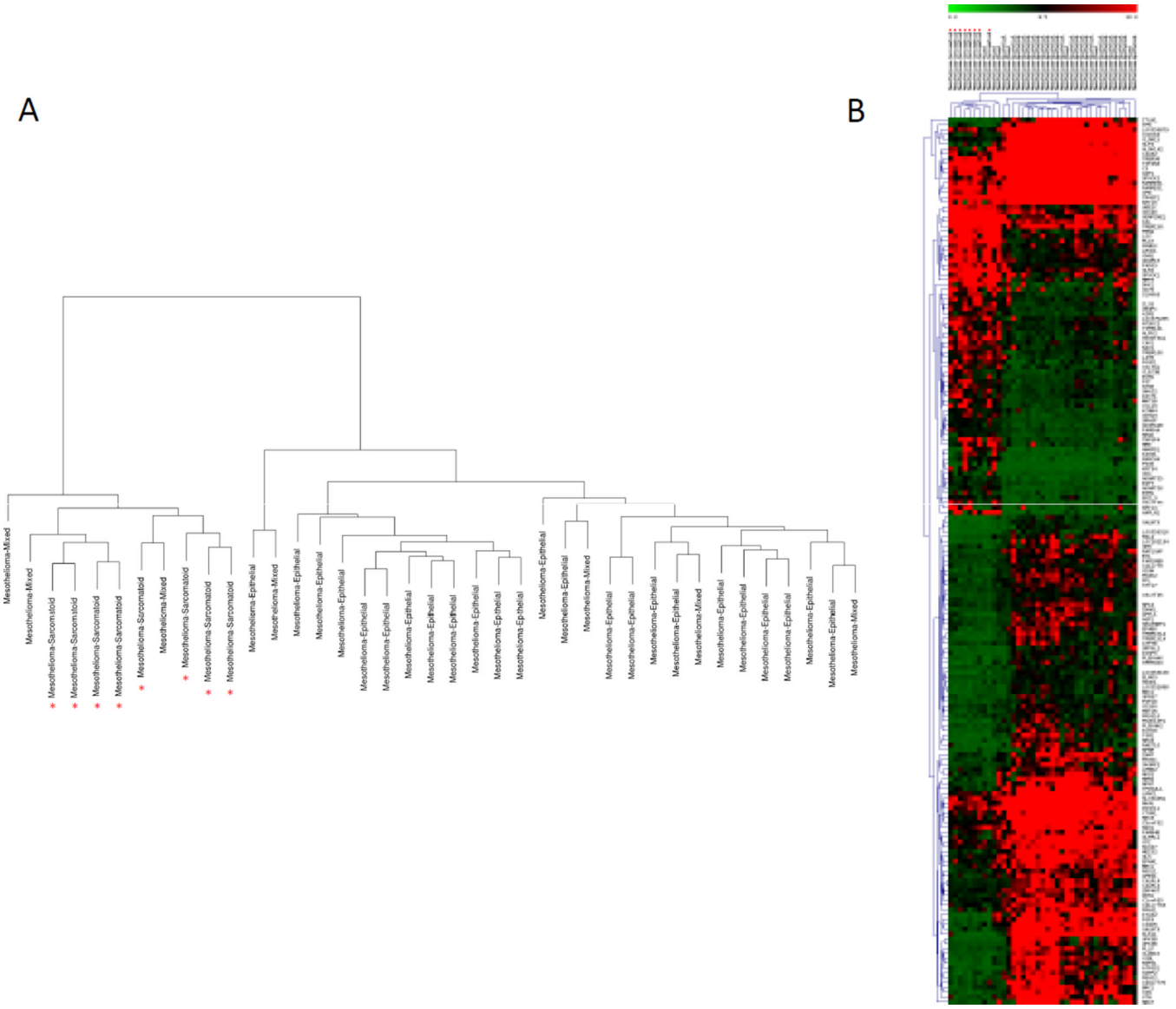


**Figure 1.** Schematic representation of the sequential diagnostic algorithms using the microarray data (A), and the RT-PCR (B).





**Figure 2.** Hierarchical clustering (A) and Two-way hierarchical clustering (B) of the samples using 167 probes differentially expressed between MPM and all the other thoracic malignancies. The red asterisk indicates the mesothelioma samples. In B, probes are annotated with gene symbol on the right. Relative gene expression levels are given by the scale at the top.



**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 biological samples. B) Two-way hierarchical clustering of the biological 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.

**Table 1**

Thoracic malignancies profiled using Illumina microarrays

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

**Table 2**

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

	<b>Specificity</b>	<b>Sensitivity</b>
Sequential Gene Ratio Test (hierarchical)	0.93	1
k-Nearest Neighbor Analysis	0.93	0.9
Linear Discriminate Analysis	0.93	0.86

n (iteration) =100

**Table 3**

Sensitivity and specificity of the 26-gene signature in the training and in the test set by the RT-PCR

	<b>Training Set (39 MPM + 72 DDX*)</b>	<b>Independent Test Set (100 MPM + 70 DDX*)</b>
Sensitivity	100%	92%
Specificity	90%	97%

\* Differential Diagnosis