

A multigene expression panel for the molecular diagnosis of Barrett's esophagus and Barrett's adenocarcinoma of the esophagus

Jan Brabender^{*1}, Paul Marjoram², Dennis Salonga³, Ralf Metzger¹, Paul M Schneider¹, Ji Min Park³, Sylke Schneider⁴, Arnulf H Hölscher¹, Jing Yin⁵, Stephen J Meltzer⁵, Kathleen D Danenberg³, Peter V Danenberg⁴ and Reginald V Lord⁶

¹Department of Visceral and Vascular Surgery, University of Cologne, 50931 Germany; ²Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90033, USA; ³Response Genetics Inc., Los Angeles, CA 90033, USA; ⁴Department of Biochemistry and Molecular Biology, University of Southern California Keck School of Medicine and USC/Kenneth Norris Comprehensive Cancer Center, Los Angeles, CA 90033, USA; ⁵University of Maryland School of Medicine and VA Medical Center, 22 S. Greene St., Rm. N3W62, Baltimore, MD 21201, USA; ⁶Department of Surgery, University of Southern California Keck School of Medicine and USC/Kenneth Norris Comprehensive Cancer Center, Los Angeles, CA 90033, USA

In order to identify genes or combination of genes that have the power to discriminate between premalignant Barrett's esophagus and Barrett's associated adenocarcinoma, we analysed a panel of 23 genes using quantitative real-time RT-PCR (qRT-PCR, Taqman[®]) and bioinformatic tools. The genes chosen were either known to be associated with Barrett's carcinogenesis or were filtered from a previous cDNA microarray study on Barrett's adenocarcinoma. A total of 98 tissues, obtained from 19 patients with Barrett's esophagus (BE group) and 20 patients with Barrett's associated esophageal adenocarcinoma (EA group), were studied. Triplicate analysis for the full 23 gene of interest panel, and analysis of an internal control gene, was performed for all samples, for a total of more than 9016 single PCR reactions. We found distinct classes of gene expression patterns in the different types of tissues. The most informative genes clustered in six different classes and had significantly different expression levels in Barrett's esophagus tissues compared to adenocarcinoma tissues. Linear discriminant analysis (LDA) distinguished four genetically different groups. The normal squamous esophagus tissues from patients with BE or EA were not distinguishable from one another, but Barrett's esophagus tissues could be distinguished from adenocarcinoma tissues. Using the most informative genes, obtained from a logistic regression analysis, we were able to completely distinguish between benign Barrett's and Barrett's adenocarcinomas. This study provides the first non-array parallel mRNA quantitation analysis of a panel of genes in the Barrett's esophagus model of multistage carcinogenesis. Our results suggest that mRNA expression quantitation of a panel of genes can discriminate between premalignant and malignant Barrett's disease. Logistic regression and LDAs can be used to further identify, from the complete panel, gene

subsets with the power to make these diagnostic distinctions. Expression analysis of a limited number of highly selected genes may have clinical usefulness for the treatment of patients with this disease.

Oncogene (2004) 23, 4780–4788. doi:10.1038/sj.onc.1207663
Published online 26 April 2004

Keywords: Barrett's esophagus; esophageal adenocarcinoma; gene expression; quantitative real-time RT-PCR

Introduction

Barrett's esophagus is a condition in which metaplastic columnar epithelium replaces the normal stratified squamous epithelium of the esophagus as a complication of chronic gastroesophageal reflux. Barrett's esophagus is the precursor of esophageal adenocarcinoma, the incidence of which has been increasing rapidly in the United States and other western countries (Devesa *et al.*, 1998). Esophageal adenocarcinoma usually presents at an advanced stage and undergoes a rapidly fatal course, with 5-year survival rates of approximately 25–30% (Jemal *et al.*, 2002). It is hoped that the identification of biomarkers associated with each Barrett's stage and with an increased cancer risk will lead to earlier detection and improved survival for patients with this disease.

Recent advances in biotechnology have revolutionized the large-scale analysis of gene expression. In particular, the development of high-throughput quantitative real-time PCR (Gibson *et al.*, 1996; Heid *et al.*, 1996) and cDNA microarray technologies has made it possible to analyse simultaneously the expression of a large panel of genes. These techniques, in combination with advances in bioinformatics, promise more accurate disease classification, earlier detection, and higher efficiency in the field of cancer diagnosis (Perou *et al.*,

*Correspondence: J Brabender, Department of Visceral and Vascular Surgery, University of Cologne, Joseph-Stelzmann Str. 9, 50931 Cologne, Germany; E-mail: jan.brabender@t-online.de
Received 24 August 2003; revised 19 February 2004; accepted 24 February 2004; Published online 26 April 2004

2000; Virmani *et al.*, 2002). cDNA microarray studies of different cancer types including esophageal and other gastrointestinal cancers (Selaru *et al.*, 2002a, b; Xu *et al.*, 2002; Zou *et al.*, 2002) have indicated the potential of a modern molecular taxonomy based on the statistical power of large phenotypic datasets. Direct analysis of cDNA microarrays generates precise ratios of gene expression, but it can be difficult to interpret this data. Moreover, certain bioinformatic methods, such as hierarchical clustering, have difficulty in recognizing subtle differences among predefined biological subgroups or identifying those genes which are most influential in these differences (Khan *et al.*, 2001).

We hypothesized that a valid molecular pathological classification of Barrett's premalignant and malignant tissues could be obtained by applying the new methods of expression analysis used for microarray data to the analysis of a relatively large number of highly selected genes whose expression had been measured by qRT-PCR. Underlying this proposal is the assumption that the expression of some genes is significantly and reliably different at different stages of the Barrett's metaplasia, dysplasia, adenocarcinoma, multistage carcinogenesis model. The genes chosen for our panel either had reported associations with Barrett's carcinogenesis or were filtered from a previously performed Affymetrix® cDNA microarray study (Scherf *et al.*, 2001). The genes selected from the microarray data (BFT, TSPAN, SPARC, TM4SF3) were, of the more than 50 000 genes or ESTs analysed, among those most differentially expressed in normal esophagus and esophageal adenocarcinoma tissues. This study is the first to use this approach to construct a molecular profile of Barrett's esophagus and esophageal adenocarcinoma.

Results

Gene expression and the progression to adenocarcinoma of the esophagus

We analysed the mRNA expression status of a panel of 23 genes and one internal reference gene by quantitative real-time RT-PCR (qRT-PCR, TaqMan®), performing more than 9016 separate PCR reactions. Table 1 shows the mRNA expression data for the 23 genes from our screen of 98 tissue specimens from 39 patients with different stages of Barrett's esophagus and/or adenocarcinoma. The alterations in mRNA expression of the genes are not uniform; they differ both in their pattern and their relative mRNA expression levels in various tissues. Accordingly, genes can be grouped by their expression behavior, as shown in Table 1. We provide a rationale for each of the gene classes in the following section.

Of the 23 genes, the most informative genes were those with significantly different expression levels between Barrett's esophagus tissues from patients with a maximum diagnosis of Barrett's esophagus (Barrett's esophagus but no cancer, BE group) and Barrett's associated adenocarcinoma tissues from the cancer

group (EA group). These genes were divided into three classes according to whether gene expression was progressively downregulated with progression from normal squamous esophagus through the metaplasia, dysplasia, adenocarcinoma stages of the Barrett's multi-step process (class A), upregulated (class C), or followed an 'on-off' expression pattern (class F). The remaining genes were less different in different Barrett's tissues and were grouped into classes B, D, and E, as detailed below.

Class A consists of the genes *BFT*, *GSTPI*, *RARγ*, and *RXRα* (Table 1). The relative mRNA expression levels for *BFT* ($P < 0.001$), *RXRα* ($P < 0.001$), *GSTPI* ($P = 0.003$), and *RARγ* ($P = 0.035$) were significantly lower in adenocarcinoma of the esophagus compared to Barrett's esophagus tissues from the BE group.

Class B consists of the genes *COX1* and *DPD*. These genes have the highest expression in normal squamous esophagus tissue and are downregulated towards progression to adenocarcinoma of the esophagus, but the expression levels in Barrett's esophagus (BE group) and adenocarcinoma tissues (EA group) were not significantly different.

The Class C genes (*COX2*, *DNMT3b*, *RARα*, *SPARC*) have very low expression levels in normal squamous esophagus tissues and are progressively upregulated during progression to esophageal adenocarcinoma. Furthermore, these genes have significantly higher mRNA expression (*COX2*: $P = 0.003$; *RARα*: $P = 0.009$; *DNMT3b*: $P = 0.021$; *SPARC*: $P < 0.001$) in adenocarcinoma tissues compared to Barrett's tissues from the group of patients with no cancer (BE group).

Class D consists of the genes *CDX2*, *C-myb*, *DNMT1*, *DNMT3a*, *ODC*, and *RXRγ*. These genes are generally upregulated with progression to adenocarcinoma but are not significantly differently expressed in Barrett's (BE group) and adenocarcinoma. Class F consists of the genes *BAX*, *DAPK*, *RXRβ*, *TM4SF3*, and *TSPAN*. These genes follow an 'on-off' regulation during progression to cancer. Following upregulation of gene expression from normal esophagus to Barrett's esophagus (BE group), these genes (*BAX*: $P = 0.009$; *DAPK*: $P = 0.005$; *RXRβ*: $P = 0.018$; *TM4SF3*: $P = 0.028$; *TSPAN*: $P < 0.001$) are significantly downregulated between the histologies Barrett's esophagus and adenocarcinoma of the esophagus. Class E consists of only two genes, *BCL2* and *TP*, with no significant alterations in gene expression between the different histologies of Barrett's disease.

Linear discriminant analysis (LDA)

To differentiate between histologies based on gene expression levels, a linear discriminant regression analysis was performed on the full data set. This analysis aims to find linear combinations of genes, the so-called *linear discriminant vectors*, which can discriminate between the different histologies. More formally, we let x_{ij} denote the expression value of gene j on sample i . Suppose we have n samples and p genes, then we can denote the data for each sample by a vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. LDA



Table 1

Class	Gene	mRNA expression (median and range)					Statistics				
		BE group		EA group			BE group	EA group		BE vs EA	
		N	BE	N	BE	EA	N vs BE ^a	N vs BE ^a	N vs EA ^a	BE vs EA ^a	BE vs EA ^b
A	BFT	50.81 (39.1–79.7)	8.61 (1.3–52.4)	48.45 (22.3–209.3)	5.43 (0.5–51.6)	2.21 (0.1–15.1)	<0.001	<0.001	0.001	<0.002	<0.001
	RXRa	2.37 (1.4–9.7)	2.5 (0.6–13.9)	2.44 (0.9–9.9)	1.67 (0.5–10.6)	0.99 (0.4–1.6)	0.658	0.057	<0.001	0.007	<0.001
	RARg	6.21 (2.8–8.8)	1.79 (0.2–5.9)	5.08 (1.1–8.4)	1.06 (0.4–4.6)	0.65 (0.1–2.2)	<0.001	<0.001	0.001	<0.057	0.035
	GSTPI	4.13 (1.9–7.1)	2.34 (0.5–5.4)	2.96 (0.3–4.9)	1.74 (0.2–4.6)	1.22 (0.2–2.8)	<0.001	0.15	0.002	0.247	0.003
B	COX1	2.14 (1.1–4.9)	1.01 (0.1–3.3)	2.48 (0.1–4.4)	1.77 (0.1–13.2)	0.71 (0.1–4.5)	0.007	0.681	0.002	0.005	0.396
	DPD	0.77 (0.1–2.2)	0.31 (0.1–3.2)	0.71 (0.1–2.3)	0.45 (0.1–1.4)	0.24 (0.1–0.5)	0.84	0.028	<0.001	0.015	0.224
C	COX2	0.14 (0.1–0.7)	0.71 (0.2–7.5)	0.33 (0.1–1.5)	1.43 (0.1–29.1)	3.41 (0.4–62.5)	0.001	0.012	<0.001	0.073	0.003
	RARa	0.13 (0.1–0.3)	0.33 (0.1–3.1)	0.22 (0.1–0.7)	0.71 (0.1–2.5)	0.84 (0.3–2.3)	0.001	<0.001	<0.001	0.455	0.009
	DNMT3b	0.07 (0.1–0.2)	0.21 (0.1–0.8)	0.06 (0–0.2)	0.15 (0.1–0.9)	0.43 (0.1–1.6)	0.001	0.001	<0.001	0.002	0.021
	SPARC	1.07 (0.3–2.9)	2.99 (0.1–7.2)	1.79 (0.1–16.6)	6.91 (0.1–53.6)	13.79 (1.8–105.2)	0.004	0.001	<0.001	0.008	<0.001
D	c-myb	0.36 (0–0.7)	0.82 (0.1–4.8)	0.71 (0–1.6)	0.59 (0.1–2.5)	1.11 (0–5.9)	0.003	0.575	0.057	0.079	0.247
	ODC	1.35 (0.5–8.9)	3.76 (1.2–48.1)	1.14 (0.1–2.5)	2.44 (0.3–7.3)	4.77 (1.9–28.7)	0.001	0.002	<0.001	<0.001	0.296
	CDX2	0.01 (0–0.1)	0.82 (0.1–3.6)	0.01 (0–0.7)	0.33 (0.1–4.3)	0.49 (0.1–2.5)	<0.001	<0.001	<0.001	0.911	0.309
	RXRg	0.27 (0–2.2)	2.42 (0.1–17.1)	0.48 (0–4.5)	2.52 (0–10.1)	1.91 (0.5–90.7)	<0.001	0.002	0.001	0.455	0.627
	DNMT3a	0.74 (0.3–1.2)	0.96 (0.2–1.8)	0.76 (0.2–1.8)	1.14 (0.3–4.1)	1.55 (0.4–4.7)	0.064	0.012	0.001	0.135	0.322
	DNMT1	0.34 (0.1–0.6)	0.32 (0.2–1.3)	0.31 (0.1–0.7)	0.41 (0.1–1.3)	0.46 (0.1–1.4)	0.243	0.601	0.044	0.073	0.607
E	Bcl2	0.99 (0.4–2.9)	0.95 (0.1–3.2)	1.35 (0.3–5.9)	2.46 (0.1–6.9)	1.43 (0.3–5.7)	0.501	0.411	0.654	0.184	0.134
	Tp	1.31 (0.5–5.6)	1.85 (0.4–6.2)	1.53 (0.2–7.8)	1.57 (0.1–8.6)	1.91 (0.2–6.8)	0.872	0.681	0.391	0.218	0.813
F	RXRb	0.81 (0.4–1.4)	1.04 (0.5–3.2)	1.05 (0.1–2.3)	0.98 (0.4–2.2)	0.75 (0.1–2.4)	0.011	0.654	0.709	0.126	0.018
	bax	0.61 (0.4–1.3)	2.26 (0.6–10.8)	0.59 (0.1–1.2)	1.15 (0.3–5.4)	1.15 (0.7–2.6)	<0.001	<0.001	<0.001	0.411	0.009
	DAPK	0.05 (0–0.1)	1.89 (0.1–26.4)	0.11 (0.1–1.2)	1.21 (0.1–7.2)	0.59 (0.1–3.9)	<0.001	<0.001	0.001	0.401	0.005
	TM4SF3	0.08 (0–1.3)	24.13 (0.1–65.9)	0.04 (0.1–0.2)	14.15 (0.1–65.9)	6.87 (0.2–35.7)	<0.001	0.001	<0.001	0.218	0.028
	TSPAN	0.01 (0.1–0.5)	16.97 (0.1–64.3)	0.04 (0–0.6)	11.38 (0.1–52.3)	1.76 (0.1–26.8)	<0.001	<0.001	0.001	0.007	<0.001

BE group: patient with maximum diagnosis of Barrett's esophagus without cancer; EA group: patients with Barrett's-associated adenocarcinoma of the esophagus; N: normal squamous esophagus; BE: Barrett's esophagus; EA: adenocarcinoma. ^aWilcoxon test. ^bMann-Whitney test. A rationale for each class of genes is provided in the Results section

looks for a linear function of the gene expression values, with the power to discriminate between the different classes of data (i.e. in our case, the different histologies). Such a function can be written as: $l(x_i) = c_1 \cdot x_{i1} + c_2 x_{i2} + \dots + c_p x_{ip}$. The number of linear discriminant functions is at most equal to $\min\{n, p\}$, and the functions are chosen in order to maximize the ratio of between-group to within-group sums of squares and cross-products. As an aid to intuition, one might consider a simple scenario in which we were attempting to discriminate between just two classes. Informally speaking, if a single discriminant function were used, it would be chosen so that it took a high (and roughly constant) value on samples from one class and a low (and roughly constant) value on samples from the other (as far as was possible).

The results are illustrated in Figure 1 and Table 2. These show that by using just three linear discriminant vectors based on the full panel of genes, we are able to differentiate normal squamous esophagus tissue obtained from the BE and EA group (histologies 1 and 3) from the other histologies, but not from each other. Importantly, using the full panel of genes, we were able to identify adenocarcinomas with an error rate of only 20%. This result is striking because by simply classifying samples into histologies at random we would expect an error rate of approximately 80%.

The issue of over-fitting is relevant here. Informally speaking, with so many genes, one might expect to classify anything with reasonable success. More formally, our linear discriminants may be unstable due to the relatively high number of variables (genes) we are using to categorize the samples. Consequently, as is common in such applications, we assess the success of the discriminant analysis by performing a *cross-validation* study (Dudoit *et al.*, 2002). To do this, we reanalyse the data set after having removed one of the samples. We find the linear discriminants for the reduced data set and then use these discriminants to predict the histology of the removed sample. We perform this procedure once for each sample we might remove.

The results are shown in Table 3 and verify that we are able to predict adenocarcinoma of the esophagus with an error rate of only 25% using the full panel of genes, again with an expected error rate of 80% if the histologies were assigned at random.

Owing to the potential clinical importance of expression markers that are reliably either present or absent in Barrett's adenocarcinoma but not in premalignant Barrett's tissues, we sought to differentiate between the histologies of Barrett's esophagus and adenocarcinoma. We started by performing a LDA using the full panel of genes. The results of the LDA and the corresponding cross-validation are shown in Tables 4 and 5. The possibility that some of the genes add only noise to our results inspired us to perform a sequential logistic regression analysis in order to detect the genes which most powerfully distinguish between Barrett's esophagus and adenocarcinoma of the esophagus. Interestingly, the logistic regression analysis revealed a combination of only three genes (*BFT*, *TSPAN*, *TP*)

to be the most powerful combination. We performed an LDA using just these three genes to discriminate between Barrett's esophagus and Barrett's associated adenocarcinoma. The cross-validation results for this study are listed in Table 6 and show that we were able to predict adenocarcinomas of the esophagus with an error rate of 0%.

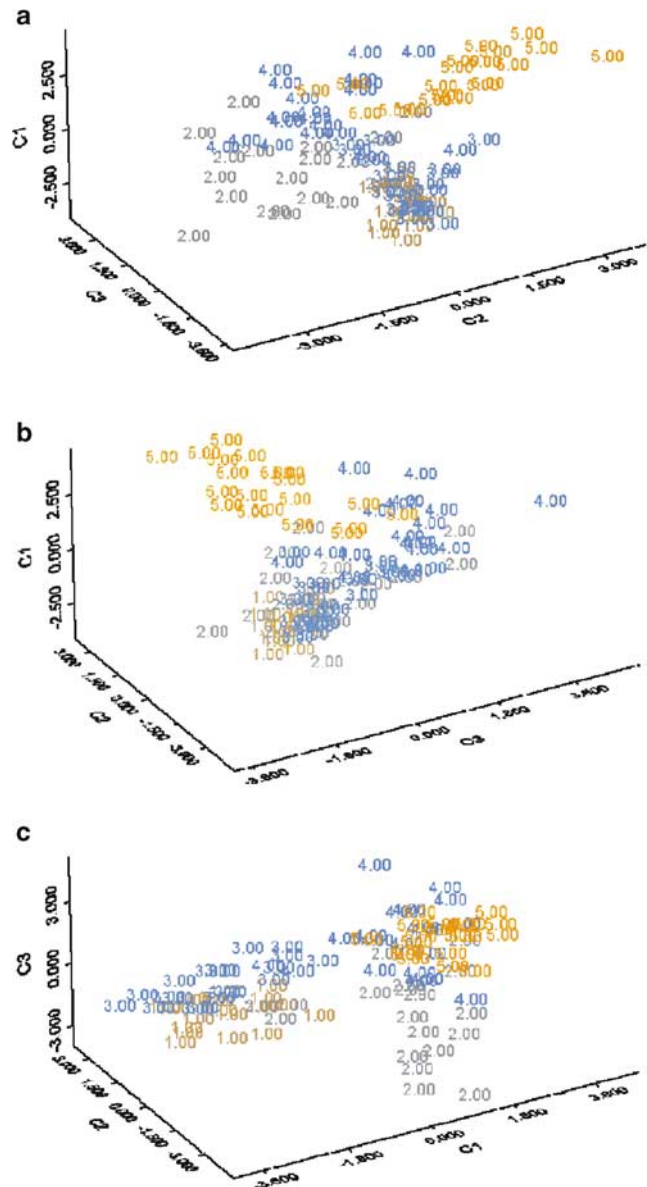


Figure 1 (a–c) Three two-dimensional representations of three-dimensional data from a LDA considering all histologies and using the full panel of genes. Points are labelled according to their histological type. The axes correspond to the first three linear discriminant vectors. Histologies: 1.00 = normal squamous esophagus from patients with Barrett's esophagus (BE group); 2.00 = Barrett's esophagus from patients with Barrett's esophagus (BE group); 3.00 = normal squamous esophagus from patients with esophageal adenocarcinoma (EA group); 4.00: Barrett's esophagus from patients with esophageal adenocarcinoma (EA group); 5.00 = Barrett's associated adenocarcinoma

Table 2 Linear discriminant analysis for all histologies using the full panel of genes

True histology	Predicted histologies					Error rate (%)
	NE (BE group)	BE (BE group)	NE (EA group)	BE (EA group)	EA	
NE (BE group)	14	0	5	0	0	26
BE (BE group)	3	12	0	3	1	37
NE (EA group)	5	0	14	1	0	30
BE (EA group)	0	2	2	15	1	25
EA	0	0	0	4	16	20
Overall						28 ^a

^aThe expected error rate is 80% if histologies were simply classified at random. The table shows the frequency with which each histology was predicted for each of the true histologies. NE: normal squamous esophagus; BE: Barrett's esophagus; EA: adenocarcinoma of the esophagus; BE group: patients with the maximum diagnosis of Barrett's esophagus; EA group: patients with Barrett's associated adenocarcinoma of the esophagus

Table 3 Cross-validation analysis for all histologies for the full panel of genes

True histology	Predicted histologies					Error rate (%)
	NE (BE group)	BE (BE group)	NE (EA group)	BE (EA group)	EA	
NE (BE group)	14	0	5	0	0	26
BE (BE group)	4	8	0	5	2	58
NE (EA group)	9	0	8	3	0	60
BE (EA group)	0	6	2	11	1	45
EA	0	0	0	5	15	25
Overall						43 ^a

^aThe expected error rate is 80% if histologies were simply classified at random. The table shows the frequency with which each histology was predicted for each of the true histologies. NE: normal squamous esophagus; BE: Barrett's esophagus; EA: adenocarcinoma of the esophagus; BE group: patients with the maximum diagnosis of Barrett's esophagus; EA group: patients with Barrett's associated adenocarcinoma of the esophagus

Table 4 Linear discriminant analysis for the two histologies, Barrett's esophagus and adenocarcinoma, using the full panel of genes

True histologies	Predicted histologies		Error rate (%)
	BE (BE group)	EA	
BE (BE group)	18	1	5
EA	0	20	0
Overall			3 ^a

^aThe expected error rate is 50% if histologies were simply classified at random. The table shows the frequency with which each histology was predicted for each of the true histologies. BE: Barrett's esophagus; EA: adenocarcinoma of the esophagus; BE group: patients with the maximum diagnosis of Barrett's esophagus; EA group: patients with Barrett's associated adenocarcinoma of the esophagus

Permutation test

We performed a permutation test to examine whether our results might be a consequence of over-fitting of the data. For example, the results of the permutation test corresponding to the data shown in Tables 2 and 3 are illustrated in Figure 2. Firstly, we created 50 data sets by using the old data on the full set of genes for all histologies and randomly permuting the histologies. Thus, each of the 50 new data sets has a different (randomly permuted) list of histologies, but the number of histologies of each type is constant across the data sets. We then performed an LDA on the permuted data sets and compared the results to that which we got when performing an identical analysis on the original data. If our original results were due to over-fitting, we would

Table 5 Cross-validation analysis for the two histologies, Barrett's esophagus and adenocarcinoma, using the full panel of genes

True histologies	Predicted histologies		Error rate (%)
	BE (BE group)	EA	
BE (BE group)	13	6	32
EA	8	12	40
Overall			36 ^a

^aThe expected error rate is 50% if histologies were simply classified at random. The table shows the frequency with which each histology was predicted for each of the true histologies. BE: Barrett's esophagus; EA: adenocarcinoma of the esophagus; BE group: patients with the maximum diagnosis of Barrett's esophagus; EA group: patients with Barrett's associated adenocarcinoma of the esophagus

Table 6 Cross-validation analysis for the two histologies, Barrett's esophagus and adenocarcinoma, using the three most informative genes (TSPAN, BFT, TP)

True histologies	Predicted histologies		Error rate (%)
	BE (BE group)	EA	
BE (BE group)	14	5	26
EA	0	20	0
Overall			13 ^a

^aThe expected error rate is 50% if histologies were simply classified at random. The table shows the frequency with which each histology was predicted for each of the true histologies. BE: Barrett's esophagus; EA: adenocarcinoma of the esophagus; BE group: patients with the maximum diagnosis of Barrett's esophagus; EA group: patients with Barrett's associated adenocarcinoma of the esophagus

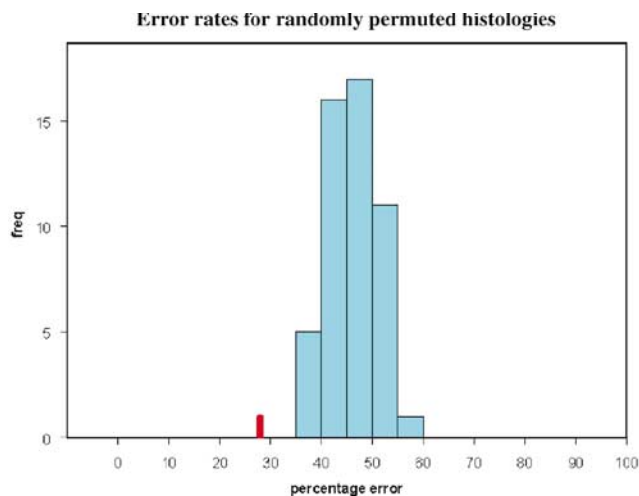


Figure 2 Histogram of the error rates for 50 randomly permuted data sets. The average classification error rate was 48% for the randomly permuted data performed for the complete set of genes and all five histologies compared to an error rate of 28% for the original data (shown as a red line and Table 2)

expect to do about as well when we performed the discriminant analysis on the permuted data sets as we did on the original data. As shown in Figure 2, the average classification error rate is 46% when fitting the original data and 81% for the subsequent cross-validation study. This compares to error rates of 28% (red line in Figure 2) for the corresponding analysis in Table 2, and 42% the cross-validation study in Table 3. None of the error rates in the permuted data were close to the values seen in the real data. The smallest error rate seen during the cross-validation tests for the permuted data was 68%, for example. These results fit with intuition. For randomly permuted data, there should be no relationship between histology and gene expression, and thus there should be no predictive power when attempting to assign a histology to the single, removed sample. Since we have five histologies, this essentially random prediction of histology for the removed sample results in an error rate of 80% in the cross-validation study, as per the above results. Consequently, we regard our observed cross-validation error rate of 42% as clear evidence of the existence of a strong signal in our data.

Since the effectiveness of any analysis is somewhat dependent on the assumptions made, we tried two other analysis techniques that have recently proven popular in the literature: Neural Nets and Multinomial regression. We conducted a cross-validation study for both analyses. In both cases, the degree of fit was comparable with, but slightly worse than that produced by the LDA (unpublished results).

Discussion

Most previous gene expression studies on Barrett's esophagus and esophageal adenocarcinoma have included only one or a few genes. Although these studies identified some important genes involved in this disease, the

potential translational use of the findings is limited because of the wide variation in expression levels in the same histopathologic stage tissues from different individuals, and even, in some cases, a wide variation in different areas of the esophagus in the same individual (Lord *et al.*, 2001). In an attempt to overcome these limitations, we measured the expression of a panel of 23 genes in different stages of Barrett's disease. Most of the genes studied have known association with gastrointestinal cancer, including esophageal adenocarcinoma in some cases: *RARs* (Lord *et al.*, 2001), *RXRs* (Brabender *et al.*, 2002a), *COX* (Wilson *et al.*, 1998), *DAPK* (Cohen and Kimchi, 2001), *ODC* (Brabender *et al.*, 2001b), *c-myb* (Brabender *et al.*, 2001a), *DPD* (Terashima *et al.*, 2002), *TP* (Suda *et al.*, 1999), *bcl2* (Woodward *et al.*, 2000), *Bax* (Woodward *et al.*, 2000), *DNMT's* (Eads *et al.*, 1999), and *GSTPI* (Brabender *et al.*, 2002b). The remaining genes were selected because they were highly differentially expressed in esophageal adenocarcinoma compared to normal esophagus in a previous microarray study (Scherf *et al.*, 2001).

The pattern of expression of many of the genes chosen was clearly different at different stages in the progression of Barrett's disease, with most of the differentially expressed genes having a pattern of either upregulation, downregulation, or an 'on-off' expression pattern with progression from normal squamous esophagus through the stages of premalignant Barrett's esophagus to esophageal adenocarcinoma. More than 9000 PCR reactions were performed, resulting in a quantity of data which, although small in comparison to that generated by many microarray studies, was nevertheless suitable for analysis using some of the bioinformatics methods employed for microarray data, such as LDA. LDA proceeds by finding linear combinations of the expression values for the genes, in such a way that the ratio of 'between histology' sum-of-squares to 'within-histology' sum-of-squares is as large as possible. In simple terms, it aims to find linear combinations of genes that can differentiate between the histologies. Using LDA on the full panel of genes, we were reassuringly able to differentiate normal squamous esophagus tissues from Barrett's and adenocarcinoma tissues, but normal esophagus specimens obtained from patients with Barrett's esophagus and from patients with esophageal adenocarcinoma could not be distinguished from each other. This result contrasts with previous findings that expression levels for some genes, for example, telomerase reverse transcriptase (hTERT) (Lord *et al.*, 2000), are significantly different in the normal esophagus of patients with cancer compared to the normal esophagus of patients with Barrett's esophagus without cancer, and suggests that the pathologic area, and not just the normal esophagus, must be biopsied in order to use expression quantitation for possible clinical decision-making.

A logistic regression analysis was performed in order to test the relative value of different genes in the panel and identify those with the most power to distinguish between the Barrett's esophagus and adenocarcinoma histopathologies. Interestingly, this analysis showed that some of the genes in the full panel add only noise to the

results. Indeed, a combination of only three very informative genes (*BFT*, *TSPAN*, *TP*) was found to be the most powerful combination for discriminating between Barrett's esophagus and Barrett's associated adenocarcinoma. Improving considerably on the results obtained using the full panel of 23 genes, the LDA and cross-validation study using these three most informative genes was able to perfectly predict adenocarcinomas of the esophagus (error rate of 0%, see Table 6). This result suggests that, in contrast to the perhaps natural assumption that medical applications of expression analysis will use the complete, or a very large, expression profile, an expression profile using relatively few genes may be both more informative and more suitable for clinical translation.

The histologic classification scheme of Barrett's lesions, which is based on somewhat subjective, non-quantitative evaluations of limited visual microscopic histologic data, may convey limited information regarding lesion biology. Thus, a more reliable and comprehensive method of distinguishing between these histologies would be clinically useful, as would a better way of discovering biomarkers of true cancer risk. To our knowledge, this is the first report applying bioinformatic tools combined with quantitative real-time PCR to discriminate between Barrett's lesions. The data herein suggest that subtle histologies can be distinguished using linear discriminant and logistic regression analyses based on a panel of gene expression data. In this respect, our approach may be complementary to histologic characterization. Previous studies on Barrett's esophagus disease have pointed out important molecular and histologic differences between different stage lesions (Brabender *et al.*, 2001a, b, 2002b, 2003; Lord *et al.*, 2000, 2001). Nevertheless, none of those differences has proven consistently capable of discriminating between different Barrett's histologies.

In conclusion, our results suggest that mRNA expression quantitation of a panel of highly selected genes has the potential to discriminate between pre-malignant and malignant Barrett's disease. Logistic regression and LDAs can be used to further identify, from the complete panel, gene subsets with the power to make these diagnostic distinctions. Prospective translational studies to determine the potential clinical value of this approach are warranted.

Materials and methods

Tissue samples for RT-PCR

In all, 98 tissue samples obtained at endoscopy or operation from 19 patients with Barrett's esophagus without adenocarcinoma (BE group) and 20 patients with Barrett's associated esophageal adenocarcinoma (EA group) were collected and immediately frozen in liquid nitrogen. Specimens were classified as intestinal metaplasia (IM) if IM but no dysplasia or cancer was present. Specimens were classified as dysplastic if either low-grade dysplasia (LGD) or high-grade dysplasia (HGD) was present. Dysplastic tissues were not divided into LGD or HGD groups because areas of LGD and HGD were

commonly present in the same specimen. Using these criteria, the following tissues were analysed: Barrett's IM ($n=16$), Barrett's dysplasia ($n=3$) and matching normal squamous tissue ($n=19$) in the BE group, and Barrett's adenocarcinoma of the esophagus ($n=20$), Barrett's IM ($n=5$), Barrett's dysplasia ($n=15$), and matching normal squamous esophagus tissues ($n=20$) in the EA group, for a total of 98 specimens. There were 30 men and 19 women, with a median age of 61.8 years (range 24–78 years). Endoscopic biopsies were obtained according to a protocol that required biopsy at 2 cm intervals from each quadrant (anterior, posterior, right, and left lateral positions) of the visible length of Barrett's mucosa and an additional biopsy from the normal appearing squamous mucosa of the esophagus. Normal esophagus biopsies were taken at least 4 cm proximal to the macroscopically abnormal epithelium. Part of the specimen or an adjacent specimen was fixed in formalin and paraffin for histopathological examination.

RNA extraction and cDNA synthesis

Total RNA was isolated by a single-step guanidinium isothiocyanate method using the QuickPrep™ *Micro* mRNA Purification Kit (Amersham Pharmacia Biotech Inc., Piscataway, NJ, USA) according to the manufacturer's instructions, and cDNAs were prepared as described previously (Chomczynski and Sacchi, 1987; Lord *et al.*, 2000).

PCR quantification of mRNA expression

cDNA quantitation of the 23 genes of interest and the internal reference gene (β -actin) was done using a fluorescence detection method (ABI PRISM 7700 Sequence Detection System, (TaqMan®) Applied Biosystems, Foster City, CA, USA), as described (Gibson *et al.*, 1996; Heid *et al.*, 1996; Lord *et al.*, 2000; Brabender *et al.*, 2001b).

The PCR reaction mixture consisted of 600 nM of each primer, 200 nM probe, 5 U AmpliTaq Gold Polymerase, 200 μ M each dATP, dCTP, dGTP, 400 μ M dUTP, 5.5 mM MgCl₂, and 1 \times Taqman Buffer A containing a reference dye, to a final volume of 25 μ l (all reagents Perkin-Elmer (PE) Applied Biosystems, Foster City, CA, USA). Thermal cycling was initiated with a denaturation step of 50°C for 10 s and 95°C for 10 min. The thermal profile for the PCR was 95°C for 15 s and 60°C for 1 min. Data obtained during 46 cycles of amplification were analysed. The primers and probes used are listed in Table 7. Parallel TaqMan® PCR reactions were independently performed for each gene and the β -actin reference gene for each sample. All samples were analysed with the full panel of genes and each reaction was at least repeated once. The ratio between the values obtained was used as a measure for the relative mRNA expression.

Statistical analysis

Taqman® analyses yield values that are expressed as ratios between two absolute measurements (gene of interest/internal reference gene). Gene expression levels in adenocarcinoma, Barrett's esophagus, and normal squamous esophagus tissues were compared using the Kruskal–Wallis test to identify significant differences in expressions among the histopathologic groups. The Kruskal–Wallis test was also used to compare the three groups of normal esophagus tissues. When the overall Kruskal–Wallis test (comparing three groups) was significant at the 0.05 level, pairwise comparisons were based on the Mann–Whitney test and the nominal *P*-value was reported. The Wilcoxon signed rank test was used for comparison of paired tissues. Statistical significance (with

Table 7 Expression primer and probe sequences

Gene	Forward primer (5'-3')	δ FAM-probe-TAMRA (5'-3')	Reverse primer (5'-3')
<i>β-Actin</i>	TGAGCGCGGCTACAGCTT	ACCACCACGGCCGAGCGG	TCCTTAATGTCACGCACGATT
<i>bax</i>	GACCCGGTGCCTCAGGAT	CCACCAAGAAGCTGAGCGAGTGTCTCA	CCTCTGCAGCTCCATGTTACTG
<i>Bcl-2</i>	TCGCCCTGTGGATGACTGA	TGAACCGGCACCTGCACACCTG	CAGAGACAGCCAGGAGAAATCAA
<i>BFT</i>	ACCAAGAGCTTCACCTTCTTCAA	TCCATGAGCCCGCTGTCGTCG	CTGGGTGCTGAGAACATGGA
<i>CDX2</i>	ACCAGGACGAAAGACAAATATCGA	TGTACACGGACCACCAGCGGCTG	TGTAGCGACTGTAGTAAAACCTCTCT
<i>c-myb</i>	ACGAGGATGATGAGGACTTTGAG	TGTGTGACCATGACTATGATGGGCTGCT	TTTTCCCAAGTGACGCTTT
<i>COX-1</i>	ATGATGGGCCTGCTGTTGA	CTGGCCTCAGCACTTGAATGACAA	CTCACCATGCCAAACCAGAA
<i>COX-2</i>	GCTCAAACATGATGTTTGCAATTC	TGCCCAGCACTTCACGCATCAGTT	GCTGGCCCTCGTTATGA
<i>DAPK</i>	GAATCCTAGACGTGGTCCGGTAT	TCAACGCTGGCTCCCATCAGACAGA	TGCTACGTGCTCGTGTCTGTT
<i>DNMT1</i>	GGTTCCTCCTCCTGGAGAATGTC	CCTTCAAGCGCTCCATGGTCTCTGAA	GGGCCACGCCGTAAGT
<i>DNMT3a</i>	CAATGACCTCTCCATCGTCAAC	AGCCGGCCAGTGCCCTCGTAG	CATGCAGGAGGCGGTAGAA
<i>DNMT3b</i>	CCATGAAGGTTGGCGACAA	CACTCCAGGAACCGTGAGATGTCCCT	TGGCATCAATCATCACTGGATT
<i>DPD</i>	CTGCCTTTGACTGTGCAACATC	CACACGGCGAGCTCCACAACGTAGA	ATTAACAAAGCCTTTTCTGAAGACGAT
<i>GSTP1</i>	CCTGTACCAGTCCAATACCATCCT	CTTCCCATAGAGCCCAAGGGTGCG	TGTAGATGAGGGAGATGTATTTGCA
<i>ODC</i>	TGTTGCTGCTGCCTCTACGTT	CATGAGTCCCACGCAGGCCCTG	GCTGGCATCCTGTTCTCTACTT
<i>RARα</i>	GAGCCGGTCCTTTGGTCAA	AGCTGGCCTTCAGGGCACCAAAA	CTGCGAGCATCACAGGACAT
<i>RARγ</i>	ACCGCGACAAAACTGTATCATC	TCATTTGACACAGCTTCTTGGACATG	CCTTCACCTCTTTCTTCTTCTGTTC
<i>RXRα</i>	AAGGACCGGAACGAGAATA	AGTCGACCAGCAGCGCCAACG	ATCCTCTCCACGGCATGT
<i>RXRβ</i>	CTCTGGATGATCAGGTCAATTTGCT	ACTCCTCATTGCCTCCTTTTACACCG	GCCATCTCGAACATCAATGGA
<i>RXRγ</i>	GGGAAGCTGTGCAAGAAGA	TCAGCTCGCTCCTGGCTCCTCTG	TGGTAGCACATTCTGCCTCACT
<i>SPARC</i>	TCTTCCCTGTACACTGGCAGTTC	CAGCTGGACCAGCACCCCATGAC	AGCTCGGTGTGGGAGAGGTA
<i>TM4SF3</i>	TTGATTGCTGTAGGTGCCATCA	CAGCATCCCAGGAAGCCCAGAATCAT	AACAGAAGCATGCAGCGACTTT
<i>TP</i>	CCTGCGGACGGAATCCT	CAGCCAGAGATGTGACAGCCACCG	TCCACGAGTTTCTTACTGAGAATGG
<i>TSPAN</i>	ACCACAATGGCTGAGCACTTC	TGGCAGGCACTACCAGCAACGTCA	TCCTGGGAACCATAATCTTTCTTG

Three primers are used in every reaction: two locus-specific PCR primers are flanking an oligonucleotide probe with a 5' fluorescent reporter dye (6-FAM) and a 3' quencher dye (TAMRA)

two-sided tests) was set at the 0.05 level. LDAs and logistic regression analyses were performed to find genes or combination of genes that have the power to discriminate between different histologies. Cross-validation studies were performed to assess the success of the LDA (Dudoit *et al.*, 2002). To further guard against the issue of over-fitting the data, we performed a permutation test in order to assess the significance of our findings.

The tissues were classified for the statistical analysis into the following groups: 1=normal squamous esophagus from patients with Barrett's esophagus (BE group); 2=Barrett's esophagus from patients with Barrett's esophagus (BE group); 3=normal squamous esophagus from patients with esophag-

geal adenocarcinoma (EA group); 4=Barrett's esophagus from patients with esophageal adenocarcinoma (EA group); 5=Barrett's associated adenocarcinoma. The laboratory and statistical analyses were both performed in a blinded fashion. The statistician was unaware, for example, that histologies 1 and 3 were both normal esophagus.

Acknowledgements

This work was supported in part by the Felix Burda Foundation for Cancer Research (JB), NIH grant CA 71716 (PVD), CA85069, CA01808, CA95323, CA098450, CA77057 (SJM).

References

- Brabender J, Danenberg KD, Metzger R, Schneider PM, Lord RV, Groshen S, Tsao-Wei DD, Park J, Salonga D, Holscher AH and Danenberg PV. (2002a). *Clin. Cancer Res.*, **8**, 438–443.
- Brabender J, Lord RV, Danenberg KD, Metzger R, Schneider PM, Park JM, Salonga D, Groshen S, Tsao-Wei DD, DeMeester TR, Holscher AH and Danenberg PV. (2001a). *J. Surg. Res.*, **99**, 301–306.
- Brabender J, Lord RV, Danenberg KD, Metzger R, Schneider PM, Uetake H, Kawakami K, Park JM, Salonga D, Peters JH, DeMeester TR and Danenberg PV. (2001b). *J. Gastrointest. Surg.*, **5**, 174–182.
- Brabender J, Lord RV, Metzger R, Park J, Salonga D, Danenberg KD, Danenberg PV, Holscher AH and Schneider PM. (2003). *Br. J. Cancer*, **89**, 1508–1512.
- Brabender J, Lord RV, Wickramasinghe K, Metzger R, Schneider PM, Park JM, Holscher AH, DeMeester TR, Danenberg KD and Danenberg PV. (2002b). *J. Gastrointest. Surg.*, **6**, 359–367.
- Chomczynski P and Sacchi N. (1987). *Anal. Biochem.*, **162**, 156–159.
- Cohen O and Kimchi A. (2001). *Cell. Death Differ.*, **8**, 6–15.
- Devesa SS, Blot WJ and Fraumeni JF. (1998). *Cancer*, **83**, 2049–2053.
- Dudoit S, Fridlyand J and Speed T. (2002). *J. Am. Stat. Assoc.*, **8**, 77–87.
- Eads CA, Danenberg KD, Kawakami K, Saltz LB, Danenberg PV and Laird PW. (1999). *Cancer Res.*, **59**, 2302–2306.
- Gibson UE, Heid CA and Williams PM. (1996). *Genome Res.*, **6**, 995–1001.
- Heid CA, Stevens J, Livak KL and Williams PM. (1996). *Genome Res.*, **6**, 986–994.
- Jemal A, Thomas A, Murray T and Thun M. (2002). *CA Cancer J. Clin.*, **52**, 23–47.
- Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C and Meltzer PS. (2001). *Nat. Med.*, **7**, 673–679.
- Lord RV, Salonga D, Danenberg KD, Peters JH, DeMeester TR, Park JM, Johansson J, Skinner KA, Chandrasoma P, DeMeester SR, Bremner CG, Tsai PI and Danenberg PV. (2000). *J. Gastrointest. Surg.*, **4**, 135–142.
- Lord RV, Tsai PI, Danenberg KD, Peters JH, DeMeester TR, Tsao-Wei DD, Groshen S, Salonga D, Park JM, Crookes PF, Kiyabu M, Chandrasoma P and Danenberg PV. (2001). *Surgery*, **129**, 267–276.
- Perou CM, Sorlie T, Eisen MB, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Aksten LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AI, Brown PO and Botstein D. (2000). *Nature*, **406**, 747–752.
- Scherf U, Williams AJ, Boland JF, Lord RV, Alvares C, Horne DT and Vockley JG. (2001). *Proc. Annu. Meet. Am. Assoc. Cancer Res.*, **42**, A3288.
- Selaru FM, Xu Y, Yin J, Zou T, Liu TC, Mori Y, Abraham JM, Sato F, Wang S, Twigg C, Olaru A, Shustova V, Leytin A, Hytiroglou P, Shibata D, Harpaz N and Meltzer SJ. (2002a). *Gastroenterology*, **122**, 606–613.
- Selaru FM, Zou T, Xu Y, Shustova V, Yin J, Mori Y, Sato F, Wang S, Olaru A, Shibata D, Greenwald BD, Krasna MJ, Abraham JM and Meltzer SJ. (2002b). *Oncogene*, **21**, 475–478.
- Suda Y, Kuwashima Y, Tanaka Y, Uchida K, Sakamoto H, Hashiguchi Y and Sekine T. (1999). *Gastric Cancer*, **2**, 165–172.
- Terashima M, Irinoda T, Fujiwara H, Nakaya T, Takagane A, Abe K, Yonezawa H, Oyama K, Inaba T, Saito K, Takechi T and Fukushima M. (2002). *Anticancer Res.*, **22**, 761–768.
- Virmani AK, Tsou J, Siegmund KD, Shen LY, Long TI, Laird PW, Gazdar AF and Laird-Offringa IA. (2002). *Cancer Epidemiol. Biomark. Prev.*, **11**, 291–297.
- Wilson KT, Fu S, Ramanujam KS and Meltzer SJ. (1998). *Cancer Res.*, **58**, 2929–2934.
- Woodward TA, Klingler PD, Genko PV and Wolfe JT. (2000). *Anticancer Res.*, **20**, 2427–2432.
- Xu Y, Selaru F, Yin J, Zou TT, Shustova V, Mori Y, Sato F, Liu TC, Olaru A, Wang S, Kimos MC, Perry K, Desai K, Greenwald BD, Krasna MJ, Shibata D, Abraham JM and Meltzer SJ. (2002). *Cancer Res.*, **62**, 3493–3497.
- Zou TT, Selaru F, Xu Y, Shustova V, Yin J, Mori Y, Shibata D, Sato F, Wang S, Olaru A, Deacu E, Liu TC, Abraham JM and Meltzer SJ. (2002). *Oncogene*, **21**, 4855–4862.