

ONCOGENOMICS

Transcriptional profiling suggests that Barrett's metaplasia is an early intermediate stage in esophageal adenocarcinogenesis

S Wang^{1,6}, M Zhan^{5,6}, J Yin¹, JM Abraham¹, Y Mori^{1,3}, F Sato^{1,2,3}, Y Xu¹, A Olaru¹, AT Berki¹, H Li², K Schulmann¹, T Kan¹, JP Hamilton¹, B Paun¹, MM Yu¹, Z Jin¹, Y Cheng¹, T Ito¹, C Mantzur¹, BD Greenwald¹ and SJ Meltzer^{1,3,4}

¹Division of Gastroenterology, Department of Medicine, Baltimore, USA; ²Department of Pathology, University of Maryland School of Medicine, Baltimore, USA; ³Marlene and Stewart Greenebaum Cancer Center, University of Maryland, Baltimore, USA; ⁴Baltimore VA Hospital, Baltimore, MD, USA and ⁵Bioinformatics Unit, National Institute on Aging, NIH, Baltimore, MD, USA

To investigate the relationship between Barrett's esophagus (BE) and esophageal adenocarcinoma (EAC), we determined gene expression profiles of discrete pathological stages of esophageal neoplasia using a sequence-verified human cDNA microarray. Fifty one RNAs, comprising 24 normal esophagi (NE), 18 BEs, and nine EACs were hybridized to cDNA microarrays. Five statistical analyses were used for the data analysis. Genes showing significantly different expression levels among the three sample groups were identified. Genes were grouped into functional categories based on the Gene Ontology Consortium. Surprisingly, the expression pattern of BE was significantly more similar to EAC than to NE, notwithstanding the known histopathologic differences between BE and EAC. The pattern of NE was clearly distinct from that of EAC. Thirty-six genes were the most differentially modulated, according to these microarray data, in BE-associated neoplastic progression. Twelve genes were significantly differentially expressed in cancer-associated BE's plus EAC (as a single combined tissue group) vs noncancer-associated BE's. These genes represent potential biomarkers to diagnose EAC at its early stages. Our results demonstrate that molecular events at the transcriptional level in BE are remarkably similar to BE's-associated adenocarcinoma of the esophagus. This finding alarmingly implies that BE is biologically closer to cancer than to normal esophagus, and that the cancer risk of BE is perhaps higher than we had imagined. These findings suggest that changes modulated at the molecular biologic level supervene earlier than histologic changes, and that BE is an early intermediate stage in the process of EAC.

Oncogene (2006) 25, 3346–3356. doi:10.1038/sj.onc.1209357; published online 30 January 2006

Keywords: Barrett's esophagus; esophageal adenocarcinoma; transcriptional profiling; bioinformatics; early detection

Correspondence: SJ Meltzer, Greenebaum Cancer Center, University of Maryland School of Medicine, 22 S. Greene St, Rm N3W62, Baltimore, MD 21201, USA.

E-mail: smeltzer@medicine.umaryland.edu

⁶These authors contributed equally to the work.

Received 16 August 2005; revised 12 October 2005; accepted 21 November 2005; published online 30 January 2006

Introduction

Barrett's esophagus (Barrett's metaplasia, BE) is characterized by specialized metaplastic intestinal epithelium replacing the normal squamous epithelium in the distal esophagus. It is considered to be a consequence of long-lasting and severe gastroesophageal reflux disease (GERD). A metaplasia-dysplasia-carcinoma sequence links Barrett's esophagus with adenocarcinoma of the distal esophagus (EAC), which is one of the fastest-increasing cancers in the Western world (Powell and McConkey, 1992; Blot and McLaughlin, 1999; Burdiles *et al.*, 2003; Cossentino and Wong, 2003; von Rahden *et al.*, 2003). BE progresses toward EAC without causing obvious symptoms. Therefore, EAC are usually detected at advanced stages, when available treatments are not effective. The only available means of early detection of EAC is regular endoscopic surveillance for patients with BE. However, the clinical benefit of endoscopic surveillance remains to be established because the development of EAC from BE takes many years and occurs in only a limited percentage of BE subjects (Stoltzing *et al.*, 1998; Umansky *et al.*, 2001; Ferguson and Durkin, 2002; Spechler, 2002). Risk stratification of patients with BE for the progression to EAC has been attempted, with a relative paucity of biomarkers characteristic to the precancerous lesions BE, low- and high-grade dysplasia (Abraham *et al.*, 1996; Going *et al.*, 2002; Wang *et al.*, 2003; Kimos *et al.*, 2004; McManus *et al.*, 2004). Abnormalities of the cell cycle regulatory genes *p53*, *p16* and *cyclin D1*, as well as loss of heterozygosity at chromosomes 3p, 5q, 9p and 17p are associated with EAC (Sanz-Ortega *et al.*, 2003; Suspiro *et al.*, 2003). Despite these findings, the molecular basis of the relationship between BE and EAC remains largely unclear.

Gene expression profiling studies have been used in the examination of cancer progression, diagnosis, drug target discovery, and gene therapy evaluation (Brabender *et al.*, 2004; Chang *et al.*, 2004; Tamoto *et al.*, 2004). Our previous cDNA microarray-based expression profiling studies global genetic signatures in esophageal cancer (Selaru *et al.*, 2002; Xu *et al.*, 2002). In the current study, we conducted cDNA microarray-based

expression profiling of esophageal tissues at various stages of carcinogenesis from patients with and without EAC in order to clarify the relationship between the stages of neoplastic progression in EAC and pre-EAC. Functional classification of differentially expressed genes was then performed to discover molecular pathways and subgroupings associated with each carcinogenic transition. Furthermore, we examined global gene signatures as a potential means of risk stratification at precancerous stages.

Results

Gene expression profile search

Principal Component Analysis (PCA) and clustering analysis not only provide sample grouping patterns but also the relationship among those patterns. In this study, PCA and cluster analysis were based on correlation coefficients and Euclidean distances as the similarity metric, respectively, gathered into three major groups based on pathologic features. As seen in Figure 1, the normal esophagi (NE) group was clearly distinct from EAC. BE was located intermediately between EAC and NE, with significantly more similarity to EAC. Distinct expression profiles of EAC and BE could also be separated statistically by PCA analysis at components 5 and 6. BEs were subclassified into two categories: BEs with (BE) or without (B) concurrent EAC. Distinct expression profiles were also observed between Bt vs B or Bt plus EAC vs B at component 20 ($P < 0.01$). These observations suggest that there is a unique gene expression pattern in NE, BE and EAC, at each stage of neoplastic progression; but that the biological phenotype of BE is distinct from, but somewhat similar to, EAC. These findings suggest that BE may constitute a more advanced intermediate stage in esophageal carcinogenesis, rather than a premalignant stage *per se*.

Identification of differentially expressed genes

Genes differentially expressed among the histological groups NE, BE, and EAC were identified jointly by ANOVA and SAM. These analyses gave rise to similar results when the P -value was < 0.001 between EAC or BE vs NE (or 0.005 between EAC and BE) in ANOVA. The FDR was < 0.001 (or 0.005 between EAC and BE) in SAM with more than two-fold changes. Such differentially expressed genes were subjected to further hierarchical clustering analysis. These cross-validation results showed clear separation among the three pathological groups and coherent clustering of samples within each group. Four hundred and fifty-seven genes were significantly differentially expressed in EAC vs NE, of which 242 and 215 genes were up- and down-regulated, respectively. Two hundred and ninety-five genes were differentially expressed between BE and NE, of which 162 genes were upregulated and the remainder were downregulated. However, only 36 genes were significantly differentially expressed between EAC and BE. This finding reveals that there are more similarities between EAC and BE in global gene expression profiles

than between BE and NE, and supports the observation from PCA and cluster analysis that BE may be an early intermediate stage in esophageal adenocarcinogenesis.

Surprisingly, EAC and BE shared 212 genes that were differentially expressed from NE (Table 1 and Figure 2). The fold change of significant differential expression between EAC and NE was greater than the fold change in BE vs NE, as well as in EAC vs BE. These in-common genes may have occurred as late events during the transition from BE to EAC. The other non-shared genes (245 in EAC and 83 in BE) are characteristic of the expression profiles of EAC-specific and, more importantly, precancer BE stage-specific genes, respectively.

In order to identify biomarkers specific to BE, we further explored differentially expressed genes in the subgroup (Bt plus EAC) vs B at $FDR < 2.7$ ($P < 0.01$ in PCA), as well as the group of Bt vs B. These selected genes were jointly analysed by PAM for additional cross-validation of biomarker selection in the BE stage.

Gene functional classification

To clarify distinctive biological functional characteristics of BE and EAC, 245 EAC-specific and 83 BE-specific genes were annotated for their biological processes. These annotations are briefly summarized in Table 2. As shown, the selected BE- and EAC-specific genes were both involved in metabolism, cellular physiologic processes, cell-cell communication, and responses to stimuli (EASE score ≤ 0.05). Compared with NE, 38.8% of EAC-specific genes and 33.1% of BE-specific genes were classifiable into cell growth and/or maintenance. In total, 37.2 and 17.3% of these genes were enhanced in the function of signal transduction in EAC and BE, respectively. EAC-specific genes were, however, mostly different in response to biologic stimuli and stress, cell mobility and cell-cell signal transduction, whereas BE-specific genes were different in cell adhesion and regulation of cell proliferation, catabolism and lipid metabolism. Notably, of the 11 BE-specific genes classified as being involved in the metabolic processes, eight were upregulated (Supplemental Materials Tables 1 and 2), which included *Fatty acid binding protein (FABP)*, *carbonic anhydrase II (CA2)*, *Clusterin (CLU)*, *acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain (ACADS)* and *fucosidase, alpha-L-1 (FUCA1)*. Of the 32 EAC-specific genes from the same category, 25 were downregulated. Differentially expressed genes between EAC and NE in the processes of cell growth and maintenance, and signal transduction included upregulated *matrix metalloproteinase 7 (MMP7)*, *Insulin-like growth factor binding protein 7 (IGFBP7)*, *Insulin-like growth factor binding protein 3 (IGFBP3)*, Chemokine (C-X-C motif) ligand 3 (*CXCL3*), *Tumor necrosis factor receptor superfamily, member 12A (TNFRSF12)*, *BCL2-antagonist/killer 1 (BAK1)* and downregulated *programmed cell death 4 (neoplastic transformation inhibitor)*.

Diagnostic markers and sample prediction

In order to identify potential biomarkers for the diagnosis of EAC at an early stage, we performed

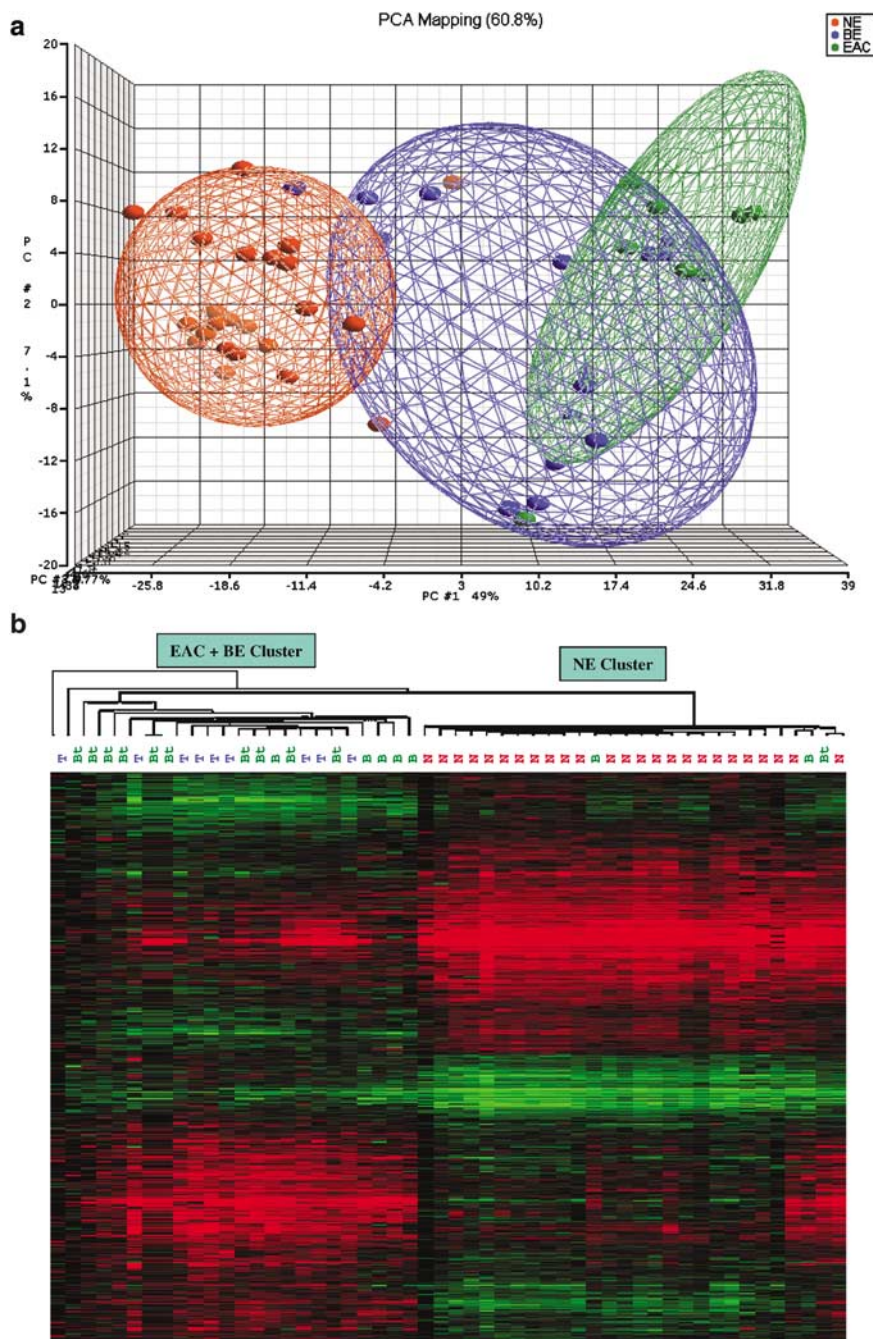


Figure 1 Distinct histopathological tissue type gene expression groupings. **(a)** PCA analysis of esophageal tissue-derived microarray data. Different tissue types are color-coded and show a clear grouping pattern, while EAC and BE partially overlap each other. The wire meshes were constructed based on a standard deviation of 1 to outline the boundaries of each tissue subgroup. The colors of the wire meshes were the same as for the tissue types. **(b)** Average linkage hierarchical clustering of all 51 specimens based on data from 6153 genes. Samples labeled red are from NE, green are from BE, and blue are from EAC (T). BEs are subclassified into two groups: BE with concurrent EAC (BE) and BE without concurrent cancer (B). Each column represents a specimen and each row represents a gene. Genes with the most pronounced differences in expression between the two main clusters are displayed. Twenty-four NEs clustered with only 4 BEs (the NE cluster), while 14 BEs clustered with 9 EACs (the EAC cluster).

Prediction Analysis of Microarrays (PAM) for each pair of sample groups to cross-validate the identification of differentially expressed genes. More importantly, PAM was performed to develop possible strategies of molecular diagnosis. We first separated BE and EAC from NE, and then distinguished between BE and EAC. The first separation required only 15 genes with the error

rate at 0.1. The second separation required 50 genes, with the error rate at less than 0.25. Based on these results, we identified 36 genes that were differentially expressed in EAC vs BE in both SAM and ANOVA. Of these 36, 12 genes were identified by SAM from the (Bt plus EAC) vs B comparison that were also significantly differentially expressed in EAC vs BE (Table 3 and

Table 1 Functional categories of significantly differentially expressed genes in EAC and BE^a relative to NE

| Gene symbol | Gene description | Fold change | |
|--|---|-------------|----------|
| | | EAC vs NE | BE vs NE |
| <i>Antiapoptosis or apoptosis^b</i> | | | |
| TNFRSF10C | Tumor necrosis factor receptor superfamily, member 10c | 4.70 | 3.12 |
| LTA | Lymphotoxin alpha (TNF superfamily, member 1) | 4.58 | 4.04 |
| MAL | Mal, T-cell differentiation protein | 0.21 | 0.43 |
| LGALS7 | Lectin, galactoside-binding, soluble 7 | 0.07 | 0.33 |
| <i>Cell adhesion, motility and proliferation</i> | | | |
| ARPC3 | Actin related protein 2/3 complex, subunit 3 | 9.34 | 20.59 |
| GAL5 | Galectin-5 | 9.23 | 11.90 |
| TSPAN-1 | Tetraspan TM4SF | 8.23 | 7.30 |
| LAMC2 | Laminin, gamma 2 | 6.27 | 2.90 |
| TS4SF8 | Transmembrane 4 superfamily member 9 | 4.51 | 3.64 |
| RELN | Reelin | 0.16 | 0.42 |
| ANXA1 | Annexin A1 | 0.14 | 0.30 |
| ARS | ARS component B | 0.03 | 0.28 |
| <i>Cell cycle, cell growth and proliferation</i> | | | |
| CRIP1 | Cysteine-rich protein 1 | 5.33 | 3.70 |
| EMP1 | Epithelial membrane protein 1 | 0.20 | 0.28 |
| VAT1 | Vesicle amine transport protein 1 homolog | 0.18 | 0.27 |
| ALOX12 | Arachidonate 12-lipoxygenase | 0.09 | 0.30 |
| KRT6A | Keratin 6A | 0.04 | 0.22 |
| <i>Cellular physiological process</i> | | | |
| TM4SF3 | Transmembrane 4 superfamily member 4 | 8.26 | 4.54 |
| KRT8 | Keratin 8 | 7.73 | 5.19 |
| COL4A2 | Collagen, type IV, alpha 2 | 5.38 | 2.02 |
| SERPINH1 | Serine proteinase inhibitor, clade H | 4.70 | 2.36 |
| ST3GAL4 | ST3 beta-galactoside alpha-2,3-sialyltransferase 4 | 0.24 | 0.32 |
| KRT4 | Keratin 4 | 0.17 | 0.23 |
| PGD | Phosphogluconate dehydrogenase | 0.15 | 0.24 |
| KRT1 | Keratin 1 (epidermolytic hyperkeratosis) | 0.06 | 0.15 |
| <i>Metabolism and catabolism</i> | | | |
| LYZ | Lysozyme (renal amyloidosis) | 7.50 | 6.99 |
| CYBA | Cytochrome b-245, alpha polypeptide | 7.32 | 4.32 |
| HGD | Homogentisate 1,2-dioxygenase | 6.38 | 4.15 |
| PSMB8 | Proteasome subunit, beta type 8 | 5.37 | 2.13 |
| AZGP1 | Alpha-2-glycoprotein 1, zinc | 4.98 | 4.68 |
| FDXR | Ferredoxin reductase | 0.24 | 0.47 |
| CDA | Cytidine deaminase | 0.17 | 0.47 |
| GPX3 | Glutathione peroxidase 3 | 0.17 | 0.31 |
| CBR3 | Carbonyl reductase 3 | 0.16 | 0.29 |
| SULT2B1 | Sulfotransferase family, cytosolic 2B, member 1 | 0.12 | 0.33 |
| <i>Regulation of transcription</i> | | | |
| CREB3L1 | CAMP responsive element binding protein 3-like 1 | 9.40 | 9.08 |
| FOXA3 | Forkhead box A3 | 9.16 | 5.31 |
| TCEAL1 | Transcription elongation factor A(SII)-like 1 | 5.80 | 3.24 |
| NR0B2 | Nuclear receptor subfamily 0, group B, member 2 | 4.67 | 4.95 |
| PAX9 | Paired box gene 9 | 0.20 | 0.35 |
| RARG | Retinoic acid receptor, gamma | 0.20 | 0.24 |
| TRIM29 | Tripartite motif-containing 29 | 0.14 | 0.21 |
| CRABP2 | Cellular retinoic acid binding protein 3 | 0.13 | 0.18 |
| MAFG | V-maf musculoaponeurotic fibrosarcoma oncogene homologG | 0.11 | 0.21 |
| NFRKB | Nuclear factor related to kappa B binding protein | 0.09 | 0.33 |
| ECM1 | Extracellular matrix protein 1 | 0.08 | 0.47 |
| <i>Signal transduction</i> | | | |
| CXCL3 | Chemokine (C-X-C motif) ligand 3 | 9.18 | 3.53 |
| GDF15 | Growth differentiation factor 15 | 6.07 | 3.82 |
| PDGFA | Platelet-derived growth factor alpha polypeptide | 5.18 | 2.35 |
| GJB1 | Gap junction protein, beta 2 | 4.83 | 5.58 |
| INSR | Insulin receptor | 4.68 | 2.92 |
| INI1A | Integrase interactor 1a protein | 0.25 | 0.36 |
| ARF4L | ADP-ribosylation factor 4-like | 0.16 | 0.39 |
| NEDD9 | Neural precursor cell expressed, developmentally down-regulated 9 | 0.13 | 0.32 |

Table 1 (continued)

| Gene symbol | Gene description | Fold change | |
|--------------------------------------|--|-------------|----------|
| | | EAC vs NE | BE vs NE |
| <i>Biological processing unknown</i> | | | |
| PNPLA2 | Patatin-like phospholipase domain containing 2 | 16.33 | 6.46 |
| AGR2 | Anterior gradient 2 homolog | 14.63 | 7.87 |
| AADAT | Aminoacidase aminotransferase | 4.60 | 2.38 |
| ANXA8 | Annexin A8 | 0.22 | 0.39 |
| SERPINB3 | Serine proteinase inhibitor, clade B member 3 | 0.21 | 0.26 |
| SPRR2C | Small proline-rich protein 2C | 0.19 | 0.45 |
| CSTA | Cystatin A (stefin A) | 0.17 | 0.27 |
| LY6G6C | Lymphocyte antigen 6 complex, locus G6C | 0.15 | 0.20 |
| S100A2 | S100 calcium-binding protein A2 | 0.14 | 0.25 |
| TPD52L2 | Tumor protein D52-like 2 | 0.14 | 0.27 |
| CSTB | Cystatin B (stefin B) | 0.11 | 0.29 |
| KIAA0657 | KIAA0657 Protein | 0.11 | 0.20 |
| PPL | Periplakin | 0.06 | 0.27 |

^aList of genes that are over 4.5-fold differentially expressed in esophageal adenocarcinoma (EAC) vs normal esophageal mucosa (NE), or differentially expressed in Barrett's esophagus (BE) vs NE. The complete list of these genes may be obtained from the author on request. ^bThis functional classification in biological processes does not include the categories of immune response, transport, development, or proteolysis and peptidolysis.

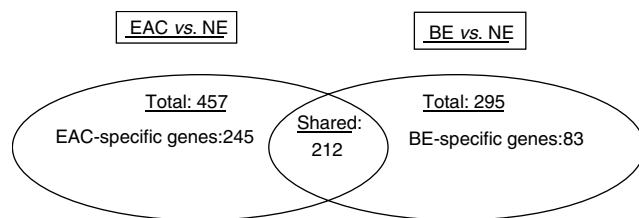


Figure 2 Paired analysis of significant gene expression in adenocarcinoma vs normal esophageal mucosa and Barrett's metaplasia vs normal esophageal mucosa. The left Venn diagram represents the total number of genes that were significantly differentially expressed between EAC and NE, whereas the right Venn diagram represent the total number of genes significantly differentially expressed between BE and NE. The overlap between these two differentially expressed sets contained 212 genes. The remaining 245 genes in EAC vs NE and 83 genes in BE vs NE were designated as EAC-specific or BE-specific for the purposes of functional classification.

Figure 3a). Three genes, *TNFRSF12A*, *GXCL3* and *Myeloid-associated differentiation marker* were differentially expressed not only in both compared subgroups, but also in Bt vs B.

Experimental validation of candidate biomarkers

Among all the differentially expressed genes identified, we were particularly interested in finding genes responsible for two progressive transitions: from BE to EAC, and from nontumor-risk-associated BE to tumor-associated Barrett's esophagus (i.e., in this study, Bt). Based on PCA, SAM and PAM, we selected four genes from the 12-gene set (see above), *TNFRSF12A*, *MMP7*, *CXCL3* and *C10orf116*, which showed up- or down-regulation in BE vs EAC or B vs Bt, for further validation by quantitative RT-PCR (Figure 3b). Three of these genes (*TNFRSF12A*, *MMP7*, and *CXCL3*) exhibited an increased mRNA level in Bt and EAC

Table 2 Classification of biologic process in transcriptional profiling of EAC- or BE-specific genes

| Biological process | EAC vs NE (%) (n = 251) | BE vs NE (%) (n = 83) |
|--|----------------------------|--------------------------|
| Metabolism | 33.56 ^a | 33.33 |
| Cell growth and/or maintenance | 38.77 | 33.14 |
| Signal transduction | 37.21 | 17.39 |
| Nucleoside, nucleotide and nucleic acid metabolism | 16.78 | 9.80 |
| Response to biotic stimulus | 15.75 | 7.84 |
| Immune response | 13.70 | 7.84 |
| Response to stress | 11.64 | 7.84 |
| Catabolism | 12.67 | 15.69 |
| Regulation of metabolism | 12.33 | 5.88 |
| Organogenesis | 10.96 | 5.88 |
| Phosphorus metabolism | 9.93 | 7.84 |
| Response to external stimulus | 8.22 | 3.92 |
| Cell adhesion | 6.85 | 11.76 |
| Cell motility | 6.85 | 3.92 |
| Lipid metabolism | 5.51 | 11.76 |
| Cell death | 6.16 | 4.84 |
| Cell-cell signaling | 5.14 | 3.92 |
| Organic acid metabolism | 4.45 | 3.92 |
| Neurophysiological process | 3.77 | 1.96 |
| Regulation of cell proliferation | 3.42 | 13.73 |
| Regulation of programmed cell death | 3.42 | 1.08 |
| Energy pathways | 2.40 | 3.92 |
| Regulation of signal transduction | 2.05 | 1.96 |
| Cell growth | 4.05 | 3.92 |
| Regulation of cell growth | 2.05 | 1.96 |

Gene Ontology terms in biologic process that are identified from the differentially expressed genes between EAC and NE and between BE and NE. ^aThe number in the columns represents the percentage of genes in each group that involve each particular category of biologic process. n, number of population genes; NE, normal esophageal mucosa; BE, Barrett's esophagus; EAC, esophageal adenocarcinoma.

samples compared with noncancer BE samples, while only *C10orf116* showed a decreased mRNA level in Bt and EAC compared to BE, matching our cDNA microarray results.

Table 3 Fold change of candidate progression biomarkers in esophageal neoplasia

| Gene symbol | Gene description | Fold change ^a | | |
|-------------|--|--------------------------|---------|-----------|
| | | EAC & Bt vs B | Bt vs B | EAC vs BE |
| MMP7 | Matrix metalloproteinase 7 | 42.13 | NS | 41.56 |
| CXCL3 | Chemokine (C-X-C motif) ligand 3 | 6.46 | 2.50 | 5.88 |
| TNFRSF12A | Tumor necrosis factor receptor superfamily, member 12A | 4.07 | 3.04 | 2.37 |
| MYADM | Myeloid-associated differentiation marker | 3.94 | 2.11 | 3.67 |
| DUSP2 | Dual specificity phosphatase 2 | 3.03 | NS | 3.27 |
| SC65 | Synaptonemal complex protein SC65 | 2.85 | NS | 2.45 |
| PLAUR | Plasminogen activator, urokinase receptor | 2.78 | NS | 2.55 |
| PRG1 | Proteoglycan 1, secretory granule | 2.66 | NS | 2.69 |
| HSPE1 | Heat shock 10kDa protein 1 | 2.29 | NS | 2.75 |
| C10orf116 | Chromosome 10 open reading frame 116 | 0.22 | NS | 0.23 |
| EREG | Epiregulin | 0.31 | NS | 0.27 |
| CNN3 | Calponin 3, acidic | 0.25 | NS | 0.30 |

^aFold Change base in SAM. NE, normal esophageal esophagus from patients without concomitant EAC; B, Barrett's esophagus from patients without concomitant EAC; Bt, Barrett's mucosa from patients with concomitant EAC; EAC, esophageal adenocarcinoma. NS, not selected by SAM.

Immunohistochemical analysis demonstrated differential expression of *MMP7* at the protein level in different tissue types (Figure 3c). *MMP7* was highly expressed in EAC, compared to NE or BE in our cDNA microarray data. By immunohistochemical staining, the *MMP7* protein was highly expressed in five EAC samples (5/9), highly or intermediately expressed in four Bt samples (4/5), and negatively expressed in all three Barrett's mucosa and normal esophageal squamous mucosa. The positively stained with *MMP7* in EAC sections were mostly in well-differentiated tubular adenocarcinoma. This protein expression pattern corresponded to the results in both of cDNA microarray and TaqMan RT-PCR.

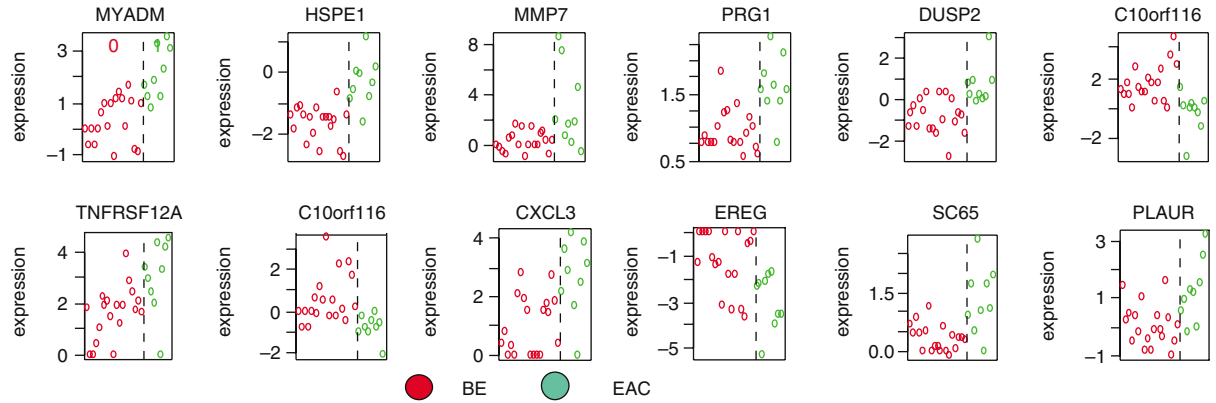
Discussion

Barrett's esophagus has been long recognized as a key precursor lesion of EAC, which is derived from GERD (Offner *et al.*, 1996; Chen and Yang, 2001). The progression from BE to EAC takes a number of years, and the rate of progression to cancer among BE cohorts is only 0.4–5% per year (Cameron, 2002). Molecular investigation has provided evidence that multiple genetic alterations are involved in the development and progression of EAC, but it is still lacking information on the similarities and differences in gene expression and biological functional features between BE and EAC, as well as NE. Through careful sample collection and multiple bioinformatics approaches, we located a set of genes that were expressed in EAC vs NE, a set differentially expressed in BE vs NE, and a set common to both comparisons. In addition, we found genes that distinguished BE with concurrent EAC (Bt) from BE without concurrent EAC (B). These sets of genes can provide important clues for functional study. Using an annotation based on functional classification, we found that individual genes differentially expressed in EAC vs NE and BE vs NE function in the immune response, DNA repair, regulation of cell growth, apoptosis,

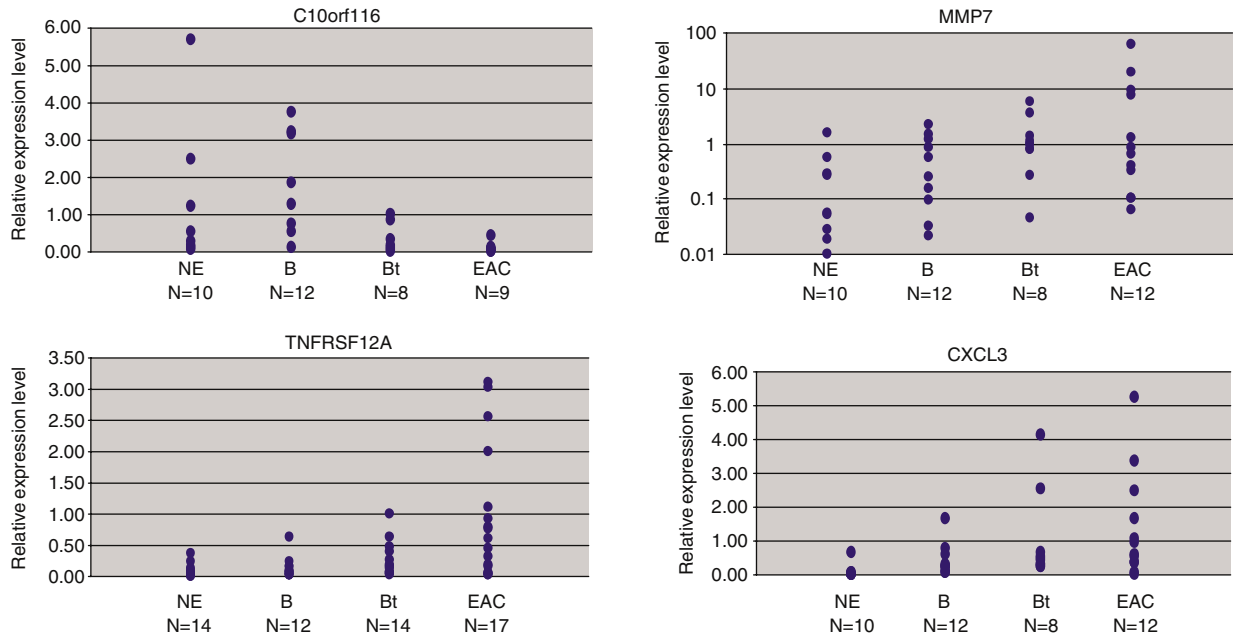
regulation of the cell cycle, and organic acid metabolism. In contrast, biological processes associated with the BE-specific gene set were mostly comparable to those of the set of EAC-specific genes: both were characterized by cellular physiological processes, cell communication, regulation of cellular processes, and metabolism. These findings also applied to the set of shared genes between EAC and BE, which further demonstrate that the gene transcriptional profiles in EAC and BE are exceptionally similar. Since BE has been commonly believed to be very benign, very close in its biology to normal esophagus. BE arises from a stem cell present in the basal layer of NE. The vast majority of BE never progresses to EAC. BE would be expected to more closely resemble NE than EAC. It was very surprising, and turned out to be the main finding of the current report, that the expression pattern of BE more closely resembled EAC than NE. Several findings supporting the advanced neoplastic nature of BE have appeared in studies of BE and EAC. For example, aneuploidy and loss of heterozygosity have been observed in metaplastic mucosa from Barrett's patients with dysplasia or EAC (Reid *et al.*, 1987; Meltzer *et al.*, 1994). *p53* tumor suppressor gene mutation has been reported in BE (Boynton *et al.*, 1991; Raskind *et al.*, 1992). One IHC study found that the mean positive cell rate (PR) of Ki-67 was 4% in normal squamous epithelium (NE) and 25% in BE and 42% in EAC (Fujii *et al.*, 2003). The mean PR of PCNA was 6, 30 and 55% in the NE, BE and EAC respectively (Fujii *et al.*, 2003).

The environment of GERD has been suggested as a possible explanation for the similarity of molecular features between BE and EAC. It has been proposed as a factor that triggers the molecular process of progression, which starts from chronic inflammation, to genetic and epigenetic changes, and finally to cancer. Chen *et al.* has observed that pathological progression starts from GERD to BE, BE with dysplasia, and finally EAC in animal models (Chen *et al.*, 2000; Chen and Yang, 2001). Gastric acid, bile acid and digestive enzymes induce irritation and inflammation in the esophagus.

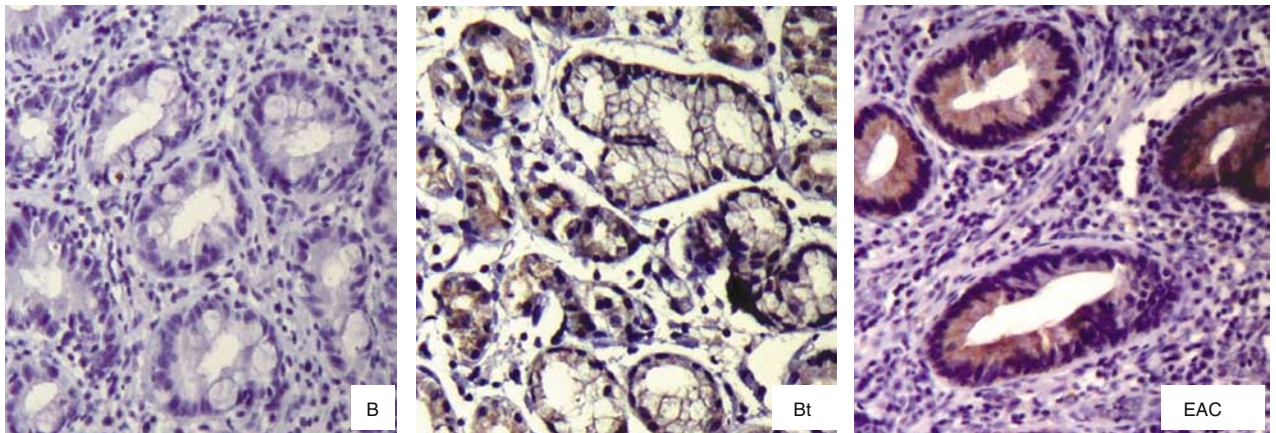
a Differentially expressed candidate biomarkers in BE and EAC by PAM



b Biomarkers differentially expressed in various esophageal tissues



c Differential expression of MMP7 protein in various Barrett's esophageal tissues and esophageal adenocarcinoma



Since the squamous epithelium of esophagus is more vulnerable to the refluxate and BE epithelium is more resistant, replacement by BE epithelium more or less relieves the symptoms of GERD. BE has been proposed as a metaplasia of pluripotent stem cells in the basal cell layer upon repeated stimulation from refluxate (Li *et al.*, 1994; Chen and Yang, 2001). Persistent stimulation of BE epithelial cells by inflammatory growth factors may result in a series of genetic and epigenetic changes, DNA repair and cellular modulation to damage and the environment. These modulatory effects may appear in corresponding morphological and molecular phenotypes in cells. However, molecular phenotype may supervene much earlier than morphological phenotype. For example, using fluorescence *in situ* hybridization, Cesar *et al.* (2004) found aneuploidy of chromosomes 3, 7, 8, 9, and 17 and deletion and overexpression of the *TP53* gene in intestinal metaplastic gastric tissue from non-gastric cancer patients.

Interestingly, in the current study, most differences in gene expression between EAC and BE were greater in both up- or downregulation in cell mobility, signal transduction, and regulation of cell proliferation and programmed cell death, whereas individual genes differentially expressed in BE vs NE were more involved in the regulation of lipid, alcohol, carbohydrate and organic acid metabolism. Arachidonic acid metabolism and reactive oxygen species have been proposed as inflammatory mediators produced by inflammatory cells in the esophagus (Wilson *et al.*, 1998; Shirvani *et al.*, 2000). Reactive oxygen species may cause DNA strand breaks, DNA base modification, lipid peroxidation and protein oxidation (Hyun *et al.*, 2004). Arachidonic acid metabolism and reactive oxygen species together stimulate the growth of BE epithelium, provoke growth and disordered metabolism, and alter gene expression and cell cycle control (Buttar *et al.*, 2002; Chen *et al.*, 2002). These biologic tendencies are in agreement with our current microarray findings in these epithelia.

Twelve genes were significantly differentially expressed both between EAC and BE, and between (Bt plus EAC) vs (B). As a group, these 12 genes may be considered as candidate biomarkers for early diagnosis or risk stratification particularly worthy of further study. These genes were also highly significantly differentially expressed between EAC and NE. The biological process of *CXCL3* involves the response to inflammation and G-protein coupled receptor protein signaling. It has been reported as a member of the *growth regulated oncogene (gro)* family in human colon

carcinoma cells (Li *et al.*, 2004). *TNFRSF12A* is a precursor of tumor necrosis factor receptor superfamily member *FN14*, which involves the process of apoptosis, cell adhesion, and cell motility. Elevated *FN14* expression was found in human liver cancer cell lines and hepatocellular carcinoma specimens (Meighan-Mantha *et al.*, 1999; Wiley and Winkles, 2003). Both genes, *CXCL3* and *TNFRSF12A*, are involved in cytokine-cytokine receptor interaction, a cancer-related pathway. These two genes, *CXCL3* and *TNFRSF12A*, were overexpressed in both Bt and EAC samples relative to NE and B samples in our present study.

MMP7 has been considered a target in the Wnt signal pathway (Zhai *et al.*, 2002; Schwartz *et al.*, 2003). Wnt signaling regulates various developmental processes and can lead to cancer formation. *MMP7* has been reported overexpressed in human gastric cancers (Mori *et al.*, 2002; Yamamoto *et al.*, 2004) and colorectal cancers (Matsushima *et al.*, 1998; Hovanes *et al.*, 2001). Recently, Chung *et al.* conducted an IHC analysis of regulation of the β -catenin signaling pathway on a breast cancer tissue microarray. They found that *MMP7* was expressed in 75% of 346 lymph node-negative breast carcinomas, while nuclear expression of p53 was noted in only 31% of the tumors (Chung *et al.*, 2004). This information signifies that *MMP7* may represent a novel target in the process of esophageal adenocarcinogenesis.

In summary, this cDNA microarray analysis identified three major transcriptional profiles. We obtained 212 genes whose differential expression vs NE was shared between EAC and BE, and two profiles unique to EAC vs NE and BE vs NE. Based on these last two specific profiles, and using ANOVA, SAM and PAM, we selected 36 genes accounting for the most differentially modulated events in Barrett's-associated neoplastic progression. Twelve genes were significantly differentially expressed both between EAC and BE, and between Bt plus EAC vs B. These genes are suggested as potential biomarkers to diagnose EAC at its earlier stages. Our results demonstrate that molecular events elements at the transcriptional level in BE are remarkably similar to Barrett's-associated adenocarcinoma of the esophagus, notwithstanding the notable histopathologic distinctions between BE and EAC. This finding alarmingly implies that BE is biologically closer to cancer than to normal esophagus, and that the cancer risk of BE is perhaps higher than we had imagined. These findings suggest that changes modulated at the molecular biologic level supervene earlier than histologic changes. The current results suggest that BE represents

Figure 3 Biomarker prediction and data validation. (a) Biomarker predicted by PAM. This Figure shows gene expression (Y-axis) of each biomarker in samples. Each green dot represents an EAC sample and each red dot stand for a BE sample. When all biomarkers were considered together, the EAC and BE samples could be separated from each other at an error rate of less than 0.25 based on gene expression levels. (b) Real-time quantitative RT-PCR validation. Four individual genes were tested for cDNA data validation. Each panel represents a tested gene. Each dot represents an individual tissue sample. The Y-axis of each figure indicates the relative mRNA expression level of the tested gene, which that was normalized against β -actin. (c) Immunohistochemical (IHC) analysis in Barrett's mucosa and EAC. Tissue sections were stained with polyclonal anti-MMP7 antibody. B, BE from patients without concomitant EAC; Bt, BE from patients with concomitant EAC; EAC, esophageal adenocarcinoma. This figure clearly shows negative staining in Barrett's sections but positive staining in both Bt and EAC sections, which match the cDNA microarray results.

an early intermediate step in the process of esophageal adenocarcinogenesis.

Materials and methods

Specimens and RNA extraction

For cDNA microarray analysis, 51 esophageal specimens were obtained from 32 individuals during endoscopy at the University of Maryland Medical System. Total RNAs were extracted from freshly frozen specimens using RNeasy kit (Qiagen, Valencia, CA, USA). The samples consisted of 24 normal esophageal mucosa (NE), 18 BE tissues and nine EAC tissues. Of these 24 NE specimens, nine were obtained from patients with both BE and EAC, six differentially modulated, according to these microarray data, in BE-associated neoplastic progression were from patients with BE alone, and nine were from individuals with no BE or EAC. The nine individuals without BE or EAC underwent endoscopic examination during which the esophagus and gastroesophageal junction were histologically normal. Of 18 BE specimens, 11 were from patients with both BE and EAC (Bt) and seven were from patients with BE alone (B).

For real-time quantitative RT-PCR, 57 specimens were used consisting of 14 NEs, 26 BEs, and 17 EACs. Of these 57 specimens, 21 were also used for cDNA microarray analysis.

For immunohistochemical analysis, 16 specimens consisting of seven NEs, eight BEs, and nine EACs were used for the validation of differentially expressed genes.

The histology of each specimen was examined using hematoxylin and eosin staining by an expert gastrointestinal pathologist at the University of Maryland. All specimens were collected from patients prior to chemotherapy or radiation. All protocols were approved by the Institutional Review Board at the University of Maryland, Baltimore. Informed consent was obtained prior to endoscopy from all patients.

cDNA microarrays

Amplified RNAs (aRNA) were prepared from 3–20 μ g of each total RNA using AmpliScript T7-flash transcription kit (Epicentre, Madison, WI, USA) as described previously (Xu *et al.*, 2002). Reference RNA was prepared from an equimolar mixture containing aRNAs from eight human malignant cell lines as described previously (Xu *et al.*, 2002). Six micrograms of sample and reference aRNA were labeled with Cy3 and Cy5, respectively, and purified with a Microcon YM-30 microcentrifuge filter (Millipore Corporation, Billerica, MA, USA) as described previously (Xu *et al.*, 2002). Sample and reference probes were then co-hybridized to an in-house cDNA microarray containing 8064 sequence-verified human cDNAs (Xu *et al.*, 2002).

Real-time quantitative RT-PCR

Template cDNAs were synthesized from 500 ng of total RNA using a SuperScript™ II kit (Invitrogen Life Technologies, Carlsbad, CA, USA) and random hexamers. Twenty microliter of PCR reaction mixture contained template cDNA, 100 nM of both forward and reverse primers, 100 nM of TaqMan[®] probe in 1 \times TaqMan Universal PCR Master Mix (Applied Biosystems, Foster City, CA, USA). PCR amplification was performed on an ABI 7700 sequence analyzer (Applied Biosystems, Foster City, CA, USA) as follows: 10 min of initial denaturation at 95°C, followed by 40 cycles of 15 s of denaturation at 95°C, and 1 min of annealing/extension at

60°C. HeLa cells were used as the quantification standard. Human β -actin mRNA was used for normalization of template loading quantity. Each specimen was analysed in duplicate. The expression index was calculated according to the following formula for the relative expression of target mRNA:

$$\text{expression index} = (\text{TarS}/\text{TarR})/(\beta\text{-actinS}/\beta\text{-actinR})$$

where TarS and TarR represent levels of mRNA expression for the target gene in the samples and reference cDNA, respectively, and β -actinS and β -actinR correspond to the amplified levels of β -actin in the sample and reference cDNA, respectively. The sequences of primers and probes are listed in supplemental Table 3).

Immunohistochemistry (IHC)

To confirm *MMP7* expression at protein level, IHC was performed on 11 formalin-fixed paraffin-embedded tissues histologically characterized into four distinct tissue groups: seven normal esophageal mucosae, three Barrett's esophageal mucosae, five Barrett's with high-grade dysplasia, and nine esophageal adenocarcinoma. Immunohistochemical staining was performed using the standard labeled streptavidin-biotin-peroxidase complex method (DAKO Corporation, Carpinteria, CA, USA). Tissue sections were cut at 5 μ m, deparaffinized in xylene, and rehydrated gradually through graded ethanols. After a heated antigen unmasking treatment (Vectors Lab., Burlingame, CA, USA), sections were pretreated with 0.3% hydrogen peroxide (DAKO Corporation, Carpinteria, CA, USA), and DAKO protein block serum (DAKO Corporation, Carpinteria, CA, USA), and incubated with 1:100 monoclonal anti-human Pro-MMP7 antibody (R&D System, Minneapolis, MN, USA) at 4°C overnight. Next, immunohistochemical staining and hematoxylin nuclear counterstaining were performed with the standard protocol from the Kit of DAKO LSAB + System, HRP (DAKO Corporation, Carpinteria, CA, USA).

Data analysis

Within-slide and inter-slide normalization of signal intensities from each microarray were conducted using the LOWESS curve-fitting method (Mori *et al.*, 2003). After normalization, 6153 genes showing expression values in more than 38 of 51 (75%) samples were used for further analyses. With normalized log ratios of gene expression levels, we first identified genes that were differentially expressed among the three major sample groups using analysis of variance (ANOVA) and significance analysis of microarray (SAM). Differentially expressed genes were determined based on the false negative rate (*P*-value) in ANOVA, then by the false discovery rate (FDR) via permutations of repeated measurements in SAM (Tusher *et al.*, 2001). In SAM, relative differences in gene expression in two classes were defined as $d(i) = (\text{ave}_1(i) - \text{ave}_2(i)) / (s(i) - s_0)$, where $\text{ave}_1(i)$ and $\text{ave}_2(i)$ are the average levels of gene expression (*i*) in classes 1 and 2, and $s(i)$ is the gene-specific scatter which is the standard deviation of repeated expression measurements. The functional categories in which differentially expressed genes were highly frequent and over-represented were discovered based on the Gene Ontology database using the software FatiGO (Harris *et al.*, 2004) and EASA (Hosack *et al.*, 2003) (EASE score ≤ 0.05).

We then conducted principal component analysis (PCA), average linkage hierarchical clustering analysis, and prediction analysis of microarrays (PAM), which, in addition to cross-validating the ANOVA and SAM-based detection of differentially expressed genes, uncovered variation patterns and provided diagnostic markers for sample prediction, respec-

tively. The PCA and clustering analyses were based on correlation coefficients and Euclidean distances as the similarity metric, respectively, using differentially expressed genes identified as well as all genes. In PAM analysis, a list of significant genes whose expression characterizes each diagnostic class, in this case NE, BE, and EAC, was obtained. The average gene expression level in each class was divided by the within-class standard deviation. The nearest centroid classification computed takes the gene expression profile from a new sample, and compares it to each of these class centroids (Tibshirani *et al.*, 2002). The resulting graphs show the shrunken class centroids for genes that have at least one nonzero difference each the diagnostic class. The Genes with nonzero components in each class are almost mutually exclusive and represent candidate biomarkers for the diagnosis of each class. For cross-validation of prediction results,

multiple classification processes were performed on two data sets randomly constructed each time from the entire gene expression data set. The first data set, consisting of 70% of the total data, was used as the training data set, and the other data set, containing the remaining 30% of data, was used for the data prediction and verification process. The final biomarkers were determined in such a way that the misclassification error rate was minimal. All data analyses were performed using the software program Partek™ and the bioconductor package (Gentleman *et al.*, 2004).

Acknowledgements

This work was supported by the grants CA85069, CA01808, CA95323, DK67872, CA10676.

References

- Abraham JM, Wang S, Suzuki H, Jiang HY, Rosenblum-Vos LS, Yin J *et al.* (1996). *Cell Growth Differ* **7**: 855–860.
- Blot WJ, McLaughlin JK. (1999). *Semin Oncol* **26**(Suppl 15): 2–8.
- Boynton RF, Huang Y, Blount PL, Reid BJ, Raskind WH, Haggitt RC *et al.* (1991). *Cancer Res* **51**: 5766–5769.
- Brabender J, Marjoram P, Salonga D, Metzger R, Schneider PM, Park JM *et al.* (2004). *Oncogene* **23**: 4780–4788.
- Burdiles P, Csendes A, Smok G, Braghetto I, Korn O. (2003). *Rev Med Chil* **131**: 587–596.
- Buttar NS, Wang KK, Leontovich O, Westcott JY, Pacifico RJ, Anderson MA *et al.* (2002). *Gastroenterology* **122**: 1101–1112.
- Cameron AJ. (2002). *Dis Esophagus* **15**: 106–108.
- Cesar AC, Borim AA, Caetano A, Cury PM, Silva AE. (2004). *Cancer Genet Cytogenet* **153**: 127–132.
- Chang Y, Gong J, Liu B, Zhang J, Dai F. (2004). *World J Gastroenterol* **10**: 3194–3196.
- Chen X, Yang CS. (2001). *Carcinogenesis* **22**: 1119–1129.
- Chen X, Ding YW, Yang G, Bondoc F, Lee MJ, Yan CS. (2000). *Carcinogenesis* **21**: 257–263.
- Chen X, Li N, Wang S, Hong J, Fang M, Youselfson J *et al.* (2002). *Carcinogenesis* **23**: 2095–2102.
- Chung GG, Zerkowski MP, Ocal IT, Dolled-Filhart M, Kang JY, Psyrri A *et al.* (2004). *Cancer* **100**: 2084–2092.
- Cossentino MJ, Wong RK. (2003). *Semin Gastrointest Dis* **14**: 128–135.
- Ferguson MK, Durkin A. (2002). *J Gastrointest Surg* **6**: 29–35; discussion 36.
- Fujii T, Nakagawa S, Hanzawa M, Sueyoshi S, Fujita H, Shirouzu K *et al.* (2003). *Oncol Rep* **10**: 427–431.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S *et al.* (2004). *Genome Biol* **5**: R80.
- Going JJ, Keith WN, Neilson L, Stoeber K, Stuart RC, Williams GH. (2002). *Gut* **50**: 373–377.
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R *et al.* (2004). *Nucleic Acids Res* **32**(Database issue): D258–D261.
- Hosack DA, Dennis Jr G, Sherman BT, Lane HC, Lempicki RA. (2003). *Genome Biol* **4**: R70.
- Hovanes K, Li TW, Munguia JE, Truong T, Milovanovic T, Lawrence MJ *et al.* (2001). *Nat Genet* **28**: 53–57.
- Hyun DH, Gray DA, Halliwell B, Jenner P. (2004). *J Neurochem* **90**: 422–430.
- Kimos MC, Wang S, Borkowski A, Yang GY, Yang CS, Perry K *et al.* (2004). *Int J Cancer* **111**: 415–417.
- Li A, Varney ML, Singh RK. (2004). *Clin Exp Metastasis* **21**: 571–579.
- Li H, Walsh TN, O'Dowd G, Gillen P, Byrne PJ, Hennessy TP. (1994). *Surgery* **115**: 176–181.
- Matsushima T, Mori M, Kido A, Adachi Y, Sugimachi K. (1998). *Oncol Rep* **5**: 73–76.
- McManus DT, Olaru A, Meltzer SJ. (2004). *Cancer Res* **64**: 1561–1569.
- Meighan-Mantha RL, Hsu DK, Guo Y, Brown SA, Feng SL, Peifley KA *et al.* (1999). *J Biol Chem* **274**: 33166–33176.
- Meltzer SJ, Yin J, Manin B, Rhyu MG, Cottrell J, Hudson E *et al.* (1994). *Cancer Res* **54**: 3379–3382.
- Mori M, Mimori K, Yoshikawa Y, Shibuta K, Utsunomiya T, Sadanaga N *et al.* (2002). *Surgery* **131**(Suppl): S39–S47.
- Mori Y, Selaru FM, Sato F, Yin J, Simms LA, Xu Y *et al.* (2003). *Cancer Res* **63**: 4577–4582.
- Offner FA, Lewin KJ, Weinstein WM. (1996). *Hum Pathol* **27**: 885–889.
- Powell J, McConkey CC. (1992). *Eur J Cancer Prev* **1**: 265–269.
- Raskind WH, Norwood T, Levine DS, Haggitt RC, Rabinovitch PS, Reid BJ. (1992). *Cancer Res* **52**: 2946–2950.
- Reid BJ, Haggitt RC, Rubin CE, Rabinovitch PS. (1987). *Gastroenterology* **93**: 1–11.
- Sanz-Ortega J, Hernandez S, Saez MC, Sierra E, Sanz-Ortega G, Torres A *et al.* (2003). *Hepatogastroenterology* **50**: 404–407.
- Schwartz DR, Wu R, Kardias SL, Levin AM, Huang CC, Shedden KA *et al.* (2003). *Cancer Res* **63**: 2913–2922.
- Selaru FM, Zou T, Xu Y, Shustova V, Yin J, Mori Y *et al.* (2002). *Oncogene* **21**: 475–478.
- Shirvani VN, Ouatu-Lascar R, Kau BS, Omary MB, Triadafilopoulos G. (2000). *Gastroenterology* **118**: 487–496.
- Spechler SJ. (2002). *Med Clin N Am* **86**: 1423–1445, vii.
- Stoltzing O, Schneider PM, Becker K, Wegerer S, Siewert JR, Holscher AH. (1998). *Langenbecks Arch Chir Suppl Kongressbd* **115**(Suppl I): 485–489.
- Suspiro A, Pereira AD, Afonso A, Albuquerque C, Chaves P, Soares J *et al.* (2003). *Am J Gastroenterol* **98**: 728–734.
- Tamoto E, Tada M, Murakawa K, Takada M, Shindo G, Teramoto K *et al.* (2004). *Clin Cancer Res* **10**: 3629–3638.
- Tibshirani R, Hastie T, Narasimhan B, Chu G. (2002). *Proc Natl Acad Sci USA* **99**: 6567–6572.
- Tusher VG, Tibshirani R, Chu G. (2001). *Proc Natl Acad Sci USA* **98**: 5116–5121.

- Umansky M, Yasui W, Hallak A, Brill S, Shapira I, Halpern Z *et al.* (2001). *Oncogene* **20**: 7987–7991.
- von Rahden BH, Stein HJ, Siewert JR. (2003). *Curr Oncol Rep* **5**: 203–209.
- Wang S, Mori Y, Sato F, Yin J, Xu Y, Zou TT *et al.* (2003). *Oncogene* **22**: 467–470.
- Wiley SR, Winkles JA. (2003). *Cytokine Growth Factor Rev* **14**: 241–249.
- Wilson KT, Fu S, Ramanujam KS, Meltzer SJ. (1998). *Cancer Res* **58**: 2929–2934.
- Xu Y, Selaru FM, Yin J, Zou TT, Shustova V, Mori Y *et al.* (2002). *Cancer Res* **62**: 3493–3497.
- Yamamoto H, Horiuchi S, Adachi Y, Taniguchi H, Nosho K, Min Y *et al.* (2004). *Carcinogenesis* **25**: 325–332.
- Zhai Y, Wu R, Schwartz DR, Darrah D, Reed H, Kolligs FT *et al.* (2002). *Am J Pathol* **160**: 1229–1238.

Supplementary Information accompanies the paper on the Oncogene website (<http://www.nature.com/onc>)