Published in final edited form as:
*J Gastroenterol*. 2009 ; 44(9): . doi:10.1007/s00535-009-0082-2.

# Differential Gene Expression in Normal Esophagus and Barrett's Esophagus

**Jacob Wang**[1,2], **Rong Qin**[1], **Yan Ma**[1], **Huiyun Wu**[3], **Heiko Peters**[4], **Matthew Tyska**[5], **Nicholas J. Shaheen**[2], and **Xiaoxin Chen**[1,2,*]

[1]Cancer Research Program, Julius L. Chambers Biomedical/Biotechnology Research Institute, North Carolina Central University, 700 George Street, Durham, NC 27707

[2]Center for Gastrointestinal Biology and Disease, Division of Gastroenterology and Hepatology, Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

[3]Department of Biostatistics, Vanderbilt University School of Medicine, Vanderbilt-Ingram Cancer Center/Biostatistics Shared Resource, Nashville, TN 37232-2158

[4]Institute of Human Genetics, International Centre for Life, University of Newcastle upon Tyne, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK

[5]Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN 37232

## Abstract

**Purpose—**As the premalignant lesion of human esophageal adenocarcinoma (EAC), Barrett's esophagus (BE) is characterized by intestinal metaplasia in the normal esophagus (NE). Gene expression profiling may help us understand the potential molecular mechanism of human BE.

**Methods—**We analyzed three microarray datasets (two cDNA arrays and one oligonucleotide array) and one SAGE dataset with SAM and SAGE(Poisson) to identify individual genes differentially expressed in BE. GSEA was used to identify *a priori* defined sets of genes that were differentially expressed. These gene sets were either grouped according to certain signaling pathways (GSEA curated), or the presence of consensus binding sequences of known transcription factors (GSEA motif). Immunohistochemical staining (IHC) was used to validate differential gene expression.

**Results—**Both SAM and SAGE(Poisson) identified 68 differentially expressed genes (55 BE genes and 13 NE genes) with an arbitrary cutoff ratio (≥4 fold). With IHC on matched pairs of NE and BE tissues from 6 patients, these genes were grouped into 6 categories: Category I (25 genes only expressed in BE), Category II (5 genes only expressed in NE), Category III (8 genes expressed more in BE than in NE), and Category IV (2 genes expressed more in NE than in BE). Differential expression of the remaining genes was not confirmed by IHC either due to false

*Correspondence: Cancer Research Program, Julius L. Chambers Biomedical/Biotechnology Research Institute, North Carolina Central University, 700 George Street, Durham, NC 27707. Tel: 919-530-6425; Fax: 919-530-7780; lchen@nccu.edu.

**Authors' contributions** JW performed statistical analysis, participated in immunostaining, and drafted the manuscript.
RQ participated in immunostaining and read the slides.
YM carried out the sectioning and participated in immunostaining.
HW instructed the statistical analysis and data interpretation
HP and MT helped with immunostaining
NJS participated in study design, collected the human samples, and participated in data interpretation.
XC designed the study, reviewed the slides, participated in the data interpretation, and drafted the manuscript with JW.
All authors read and approved the final manuscript.

**Competing interests** The authors declare that they have no competing interests.

discovery (Category V), or lack of proper antibodies (Category VI). Besides individual genes, the TGFβ pathway and several transcription factors (CDX2, HNF1, and HNF4) were identified by GSEA as enriched pathways and motifs in BE. Apart from 9 target genes known to be up-regulated in BE, IHC staining confirmed up-regulation of 19 additional CDX1 and CDX2 target genes in BE.

**Conclusion—**Our data suggested an important role of CDX1 and CDX2 in the development of BE. The IHC-confirmed gene list may lead to future studies on the molecular mechanism of BE.

### Keywords

Barrett's esophagus; intestinal metaplasia; expression profile; SAM; GSEA

## Introduction

Barrett's esophagus (BE) is characterized by replacement of normal squamous epithelium (NE) of the esophagus by intestinalized columnar epithelium, due to chronic gastroesophageal reflux disease. Patients with BE are at an increased risk of esophageal adenocarcinoma (EAC), which is now the most rapidly increasing type of cancer in Western countries (1).

The pathogenesis of intestinal metaplasia (IM) in BE is poorly understood. Stem cells in the esophageal squamous epithelium or the submucosal glands have been proposed as the celluar origins of IM (2). Regardless of the site of origin, however, it is generally believed that inflammatory mediators and/or gastroesophageal refluxate trigger the pluripotent stem cells to differentiate into intestinalized columnar epithelium. Recent studies have focused on the potential roles of acid and bile, inflammatory mediators, and intestinal transcription factors in the pathogenesis of human BE (2-4). Despite these efforts, the mechanism of IM in BE remains unclear.

The current literature contains 14 studies concerning differential gene expression in NE and BE (5-18). Several different technologies have been used in these investigations, including oligonucleotide microarrays (5, 6, 10-13), cDNA microarrays (7-9, 15, 17, 18), serial analysis of gene expression (SAGE) (16), and proteomics (14). These studies have generated a wealth of information regarding the differential expression of individual genes. Knowledge of individual genes involved in the development of BE, however, is inadequate for mechanistic understanding. Distinct genes do not detect biological processes, and it is unlikely that the focus on disparate genes will present a cohesive view of the mechanism of BE. Therefore it is necessary to improve data analysis of gene expression profiles in order to identify useful links (both individual genes and biological processes) for mechanistic studies.

Recently, there has been a shift toward pathway-based analysis of microarray data. Gene Ontology (GO) has been the most popularly used annotation. The GO annotations are vague, however, and do not provide specific mechanistic information. Furthermore, an individual gene may be involved in multiple cellular processes. An alternative to GO is gene set enrichment analysis (GSEA), a computational method that determines whether the difference in expression of an *a priori* defined set of genes in two different phenotypes is statistically significant. This pathway-based approach has proved successful in discovering molecular pathways involved in human diseases.

The aim of this study was to subject the publicly available gene expression data to the same set of statistical tools in order to discover differential gene expression patterns in NE and

BE. The purpose was to discover not only individual genes, but also potential molecular pathways leading to BE.

## Materials and Methods

### Gene array datasets

Of the 14 published studies involving gene expression profiling in human BE, six studies made their raw data publicly available (7-9, 11, 12, 16). Three sets of these raw data were accessible through the GEO Datasets available on PubMed (GSE2769, GSE1420, and GSE2444). Others were available from the Stanford MicroArray Database (http://genomewww5.stanford.edu/), EMBL-EBI (http://www.ebi.ac.uk/aerep/dataselection?expid=956697506), and the author's website (http://gastrolab.coi.waw.pl/barrett/). One dataset was generated using SAGE (16), two using oligonucleotide microarrays (11, 12), and three using cDNA microarrays (7-9). Various numbers of tissue samples were used in these six studies:

1. The study by van Baal *et al* utilized SAGE to analyze tissue samples from 21 patients with non-dysplastic BE. NE and BE tissues were obtained from each patient with confirmed histology (16). For simplicity, the data generated by this study was referred to as the "van Baal dataset."
2. The study by Kimchi *et al* utilized the U133A Affymetrix GeneChip to analyze 8 matched samples of NE, BE, and EAC. Histological diagnosis was "obtained from experienced gastrointestinal pathologists" (11). Here, the data from this study was referred to as the "Kimchi dataset."
3. The study by Gomes *et al* analyzed 9 samples of NE, 10 samples of BE, and 5 samples of EAC with an in-house cDNA array consisting of 4,800 sequences (7). Here, the expression data generated from this study was referred to as the "Gomes dataset."
4. The study by Hao *et al* analyzed 15 samples of NE, 14 samples of BE, and 5 samples of EAC with an in-house cDNA array. NE, BE, and EAC (if present) tissue samples were obtained from unselected patients. BE and EAC were confirmed by histopathology. Duplicate microarrays were performed for one of the samples (677N) (9). Here, the data from this study was referred to as the "Hao dataset."
5. The study by Greenawalt *et al* analyzed 39 samples of NE, 25 samples of BE, and 38 samples of EAC with an in-house cDNA array of 10,500 elements representing ~9,400 unique cDNA clones. Each tissue sample was halved, with one piece undergoing histopathological review and the other prepared for RNA extraction (8). Here, the data from this study was referred to as the "Greenawalt dataset."
6. The study by Ostrowski *et al* utilized the U133A Affymetrix 2.0 GeneChip to analyze tissue samples obtained from 29 patients with long-segment BE. Histology of normal squamous epithelium and BE was histologically confirmed (12). Here, the data of 23 paired samples were used and was referred to as the "Ostrowski dataset."

### Initial analysis and exclusion of data

During the initial significance analysis of microarrays (SAM), the Kimchi dataset only had 95 differentially expressed genes of statistical significance at a false discovery rate (FDR) of 1%. Of these, none of the known markers of BE were present. The Gomes dataset had no genes with statistical significance using SAM analysis. Furthermore, a simple *t*-test also

failed to detect any differential gene expression of statistical significance ($p < 0.05$). Therefore, these two datasets were excluded from this study.

Three datasets were then used for analysis: the Hao dataset and the Greenawalt dataset (cDNA microarrays), and the van Baal dataset (SAGE). For analysis of the Hao dataset, three BE samples (673, B1, and B6) were eliminated, since Sample 673 exhibited dysplasia, and Samples B1 and B6 were obtained from BE adjacent to tumor mass. The fourth dataset, the Ostrowski dataset (oligonucleotide microarray), was used to validate our findings based on the first three datasets.

### Statistical analysis for identification of differentially expressed genes

Our data analysis followed the flowchart presented in Figure 1. In order to determine individual genes which were differentially expressed between NE and BE, SAM was used to analyze the original, normalized Hao dataset and Greenawalt dataset (http://www stat.stanford.edu/~tibs/SAM/) (19). Delta values were adjusted until the FDR was less than 1%. The van Baal dataset was analyzed using a Poisson approach (http:// genome.dfci.harvard.edu/sager/) at a significance level of 0.05 (20). An arbitrary 4-fold cut-off threshold was then applied to the lists of significant genes from the SAM analysis and the Poisson-based analysis. Significant genes picked up by at least two of the three datasets were selected as individual genes differentially expressed between NE and BE. These genes were designated the SAM/SAGE overlap gene list. Differential expression of some of these genes was confirmed by immunohistochemical staining (IHC) to create the IHC-confirmed gene list.

### Statistical analysis for identification of differentially expressed gene sets

In order to explore potential involvement of signaling pathways or transcription factors, GSEA was performed on the Hao dataset and the Greenawalt dataset. GSEA is a computational method that determines whether a set of genes shows statistically significant differences between two biological states (http://www.broad.mit.edu/gsea). Two databases were utilized in our analysis, GSEA motif databases which contain genes that share *cis* regulatory motifs that are conserved across the human, mouse, rat, and dog genomes, and GSEA curated databases which contain genes on certain molecular pathways.

Selected gene sets identified from GSEA were then visualized with Gene Map Annotator and Pathway Profiler (GenMapp, http://www.genmapp.org) using pathways modified from the GenMapp database. A GenMapp was constructed to visualize the regulation of downstream target genes of CDX1 and CDX2 by manually checking the literature.

The individual genes and pathways detected in the Hao and Greenawalt datasets were then confirmed using the Ostrowski dataset. This dataset was subjected to the same analysis (SAM, GSEA, and GenMapp), and the results were compared to those obtained from the analysis of the Hao and Greenawalt datasets.

### Human tissue samples

Six matched pairs of tissue samples of NE and BE were obtained from the Department of Medicine, University of North Carolina at Chapel Hill. These 6 patients were mid-age Caucasian American men on proton pump inhibitors for various durations. Biopsy tissue samples of NE and BE were harvested during the first endoscopic examination. None of these patients had dysplasia or cancer. Clinical diagnosis was confirmed by histopathology. These samples were used for validation of the SAM/SAGE overlap gene list. Use of human samples has been approved by the IRB Committee. All human samples were coded with

patient identifiers removed. Diagnosis of BE was made according to the established criteria (21).

### Immunohistochemical staining

Expression of the SAM/SAGE overlap gene list, diagnostic gene set, and selected target genes of CDX1 and CDX2 were examined with IHC (Table 1 and Table 2). In brief, paraffin-embedded tissue sections were deparaffinized, rehydrated, and pretreated by heating the slides for 5–10 min in 10mM citrate buffer. IHC was performed with the ABC kit (Vector Labs, Capenteria, CA) according to the manufacturer's instructions. The sources of the primary antibodies, catalogue numbers, and working concentrations are listed in Table 1. Normal serum or phosphate buffered saline were used as negative controls, instead of the primary antibodies. Both positive and negative control slides were processed in parallel. Alcian blue staining (1% in 3% acetic acid, pH2.5, for 10 min) was performed after IHC on some slides to visualize cells that produce acidic mucin.

IHC for each gene was scored by one pathologist (RQ) as either positive or negative in the epithelium of interest. Areas of NE and BE were evaluated separately. In NE, expression in basal, suprabasal, and superficial cell layers was described separately. In BE, expression in columnar epithelial cells, goblet cells, and Paneth cells were also described separately. Localization in the nuclei, cytoplasm, or membrane was indicated.

## Results

### Individual genes differentially expressed in BE

The result of overlapping the lists of significant genes generated from the cDNA microarrays (Hao dataset and Greenawalt dataset) and SAGE (Van Baal dataset) studies was a short list of 68 genes, the SAM/SAGE overlap gene list (Table 1). Output files were provided in Supplementary File 1. In this list, 55 up-regulated genes in BE were regarded as BE genes, and 13 down-regulated genes as NE genes.

Differential expression of these 68 genes was examined on six matched pairs of human NE and BE samples with IHC (Figure 2). Following IHC, the SAM/SAGE overlap gene list was furthered classified into six categories based on their expression patterns in NE and BE: (1) Category I: genes expressed in BE only (25 genes); (2) Category II: genes expressed in NE only (5 genes); (3) Category III: genes expressed at a higher level in BE than NE (8 genes); (4) Category IV: genes expressed at a higher level in NE than BE (2 genes); (5) Category V: genes expressed at roughly the same level in NE and BE (6 genes); and (6) Category VI: genes for which no quality antibodies were available for IHC (22 genes). Since expression of many genes were examined in a small number of samples, we applied strict criteria when the IHC staining was interpreted. A higher level of expression referred to increased expression when all 6 pairs of samples showed the same tendency. Cellular sources and localization were also characterized and shown in Table 1.

An IHC-confirmed gene list was then generated after exclusion of Category V and Category VI. This list consisted of 40 genes, 33 BE genes (25 genes of Category I and 8 genes of Category III) and 7 NE genes (5 genes of Category III and 2 genes of Category IV).

### Transcription factors identified by GSEA using motif gene sets

GSEA performed with the motif databases identified several gene sets that contain genes sharing transcription factor binding sites defined in the TRANSFAC database (version 7.4, http:///www.gene-regulation.com/). In the Hao dataset, 24 gene sets were enriched for NE

and 9 for BE ($p < 0.05$). In the Greenawalt dataset, GSEA detected 1 enriched gene set in NE and 18 in BE ($p < 0.05$).

Of the transcription factors identified from these two datasets, CDX2, HNF1, and HNF4 are known to be critical for gastrointestinal development. Since previous studies had suggested an important role of CDX1 and CDX2 in intestinal metaplasia *in vitro* (22, 23), we manually checked the literature for target genes of CDX1 and CDX2. In total, we found 14 CDX1 target genes and 71 CDX2 target genes (Supplementary Table 1). Most of these target genes are direct transcriptional targets, and some are genes indirectly regulated and identified by microarray screening. In addition, there are 22 genes downstream to ATOH1, a CDX2 target gene and a transcription factor essential for goblet cell development (24).

Using GenMapp, we found that among 71 CDX2 target genes, 31 were up-regulated in BE, 7 were down-regulated and 33 were not significantly changed or not included in the arrays. Among 14 CDX1 target genes, 6 were up-regulated in BE, 1 was down-regulated and 7 were not significantly changed or not included (Figure 3A). Since previous studies have clearly demonstrated overexpression of both CDX1 and CDX2 in BE (25, 26), overexpression of CDX1 and CDX2 target genes in BE suggested that both CDX1 and CDX2 might play causative roles in the development of BE.

Seven target genes of CDX1 and CDX2 (FABP1, CDH17, HEPH, TFF3, AGR2, LYZ, and MUC5AC) were also present in the SAM/SAGE Overlap Gene List and were only tested once. Two target genes (VIL1 and MUC2) have been examined by us previously (27), leaving 28 target genes of CDX1 and CDX2 for further examination by IHC. Differential expression of 19 genes (11 genes of Category I and 8 genes of Category III) was confirmed (Table 2; Figure 4). We failed to detect 6 genes with IHC (CA1, CLU, HOXA9, MMP7, TM4SF4, and UGT2B7) due to lack of proper antibodies. GFI1, HOXA10, and NR0B2 were found to be more highly expressed in NE than BE (data not shown).

## Molecular pathways identified by GSEA using curated gene sets

GSEA was also performed using the curated databases to discover molecular pathways involved in the development of BE (Supplementary File 1). At asignificance level of 0.05, GSEA detected 48 enriched gene sets in NE and 43 enriched gene sets in BE using the Hao dataset. GSEA also detected 6 enriched gene sets in NE and 81 enriched gene sets in BE using the Greenawalt dataset.

Among these pathways, the TGFβ pathway was identified as an activated pathway in BE in both datasets. The expression pattern of BMP/TGFβ pathway genes was visualized using GenMapp (Figure 3B). Consistent with the recent findings that BMP4 induced metaplastic changes in esophageal cells *in vitro* (28), the widespread up-regulation of BMP/TGFβ target genes in BE suggested an important role of BMP/TGFβ signaling pathway in the development of BE.

## Confirmation of above findings using oligonucleotide microarray data (the Ostrowski dataset)

In order to confirm the above results using data obtained from a different array platform, we also analyzed the Ostrowski dataset for differential gene expression with SAM and GSEA. Of the 68 genes in the above-mentioned SAM/SAGE overlap gene list, 57 were also found to be significantly up-regulated or down-regulated in BE. Of the remaining 11 genes which were not detected in the Ostrowski dataset (BCMP11, FOXA3, GDDR, ITLN1, KCNE3, OCIAD2, ORF1-FL49, REG4, TM4SF8, GJB2, and MALAT1), none were originally included in the U133A Affymetrix 2.0 gene chip used by Ostrowski *et al.*

GSEA performed on the Ostrowski dataset using the motif database detected 4 gene sets in NE and 99 gene sets in BE ($p<0.05$). CDX2, HNF1, and HNF4 were among the significant gene sets of BE. Besides these, GSEA also identified SOX9, TTF1, and GATA6 as enriched gene sets in BE.

Using the curated database, GSEA detected 15 gene sets in NE samples and 292 gene sets in BE ($p<0.05$). The TGFβ signaling pathway was among the enriched gene sets of BE. This was consistent with the discovery by Ostrowski *et al.* (12).

GenMapp showed that many target genes of CDX1 and CDX2 and the BMP/TGFβ signaling pathway were up-regulated in BE (Supplementary Figure 1A, B). This was consistent with the above results in Figure 3 using the data from cDNA microarrays.

## Discussion

In this study, we identified 40 genes differentially expressed in NE and BE, which were validated by IHC staining. Our data also suggested that multiple genetic pathways (e.g., BMP/TGFβ pathway) and transcription factors (e.g., CDX1, CDX2) might participate in the development of BE. This observation is consistent with a well-accepted theory that transdifferentiation or metaplasia is mediated by activation or inactivation of transcription factors, which further regulate their target genes specific for certain cellular functions (29).

Among the 40 genes differentially expressed in NE and BE, there were 33 genes expressed only in BE (Category I, 25 genes) or expressed in BE more than in NE (Category III, 9 genes). Some of these genes (e.g., AGR2, ANXA10, CLDN18, KRT8, MUC3B, MUC5AC, TACSTD1/EpCAM, TFF1, TFF2, TFF3, TSPAN1) have already been reported to be up-regulated in human BE (14, 16, 30-37). Some genes are known differentiation markers of certain cell lineages in BE. For example, LYZ is a known marker of Paneth cells, KRT8 a marker of columnar epithelial cells, and TFF3 a marker of goblet cells. None of the markers of enteroendocrine cells appeared in our gene list. It was probably due to a small population of such cells in BE. Many genes (e.g., CDH17, CTSE, FABP1, FOXA3, HEPH, IQGAP2, ITLN1, LGALS2, LIPF, MEP1A, MYO1A, PIGR, SELENBP1) are known to be highly expressed in the stomach and/or the intestine. Function-wise, some genes (e.g., CA2, CLDN18) were up-regulated in BE as a defensive mechanism against reflux (32, 38). Besides these 40 genes whose differential expression was validated by IHC, seven genes of Category VI (BCMP11, GDDR, GKN1, ITLN1, MUC13, REG4, TM4SF4) were very likely to increase in BE versus NE because of their expression patterns in the gastrointestinal epithelium and gastric IM. As a TFF2-binding protein, GDDR was shown to be up-regulated in human BE (Dr. William Otto, Cancer Research UK, London, UK; personal communication). These data clearly showed that BE is a phenotype of multi-directional differentiation, suggesting columnar differentiation during the development of BE.

Seven genes were expressed only in NE (Category II, 5 genes) or expressed more in NE than in BE (Category IV, 2 genes). These 7 genes are known to be expressed in stratified squamous epithelium, including the esophagus (39-45). Some of these genes were found to be down-regulated in esophageal squamous cell carcinoma, and associated with clinical prognosis (39, 42, 43, 45). Down-regulation of these squamous marker genes suggested squamous dedifferentiation in BE. These data were consistent with our previous study showing loss of squamous markers and gain of columnar markers as two facets of BE (27).

Several transcription factors were identified by GSEA as potential players in BE, including CDX2, HNF1, HNF4, SOX9, TTF1, and GATA6. Among these factors, Sox9 is required for differentiation of Paneth cells (46). TTF1 regulates expression of CLDN18, one of the BE genes identified in this study (Table 1; Figure 2IC) (47). GATA6 has been identified as a

gene differentially expressed in human BE (48). However, CDX2 might be a pivotal switch between intestinal columnar epithelium and squamous epithelium in the gastrointestinal tract.

In normal intestinal epithelium, CDX2 and CDX1 are expressed in most cell lineages with Paneth cells having a lower level of expression than other cells. Squamous epithelial cells of normal human esophagus do not express CDX2 or CDX1, while submucosal glands weakly express CDX2 in the cytoplasm. In human BE, CDX2 is expressed in both goblet and non-goblet cells. Dysplasia and EAC may have decreased levels of or even absence of expression of CDX2 and CDX1. In EAC, a high level of CDX2 expression was usually associated with well or moderately differentiated tumors. A low level of CDX2 mRNA was detectable by RT-PCR in biopsy samples of squamous epithelium of GERD patients, even before the appearance of CDX2 protein and other marker genes of intestinal metaplasia and histological metaplasia. CDX2 expression also precedes CDX1 expression. Many “marker” genes of BE, such as VIL, GCC, SI, MUC2, TFF3, FABP1, are known to be regulated by CDX2 and/or CDX1 at the transcriptional level. Treatment of human and rodent esophageal squamous epithelial cells with either acid or bile acids, which mimics gastroesophageal reflux, induced expression of CDX2. Transfection of CDX2 into human esophageal squamous epithelial cells induced metaplastic changes in morphology and gene expression. CDX2 regulates expression of critical genes of goblet cell differentiation, such as ATOH1, KLF4, MUC2, MUC4, RETNLB and TFF3 (22). It should be noted that the presence of goblet cells is diagnostic of BE.

We manually picked up 14 target genes of CDX1 and 71 target genes of CDX2 from the literature (Figure 4A). Using IHC, differential expression of 19 genes (11 genes of Category I and 8 genes of Category III) was confirmed (Table 2; Figure 5). In addition, 9 target genes of CDX1 and CDX2 (VIL, MUC2, FABP1, CDH17, HEPH, TFF3, AGR2, LYZ, and MUC5AC) have been described above (Table 1 and Figure 2) or previously by us (27). These data highly suggested a causative role of CDX1 and CDX2 in the development of human BE.

GSEA analysis identified TGFβ pathway as one of the signaling pathways activated in BE (Figure 4B). This is consistent with a recent study showing increased expression of BMP4 and activation of its signaling pathway in BE. Treatment of primary squamous cells with BMP4 induced squamous dedifferentiation and columnar differentiation (4, 28). Recently, CDX2 was found to interact with SMAD3 independent of SMAD4, resulting in stimulation of SMAD3 transcriptional activity. CDX1 also interacted with SMAD3 by inhibiting SMAD3/SMAD4-dependent transcription (49). Treatment of human gastric cancer cells with BMP2 and BMP4 induced expression of CDX2, as well as that of MUC2 and LI-Cadherin (50). These data suggested that multiple pathways and factors might interact with each other in driving the development of BE. However, caution should be excised when GSEA data are interpreted. Functional analysis *in vivo* will determine whether and how these signaling pathways and transcription factors may participate in the development of BE.

In summary, we identified and validated genes differentially expressed in human BE. CDX1/CDX2 and the BMP/TGFβ signaling pathway may participate in the development of BE. Further studies may be warranted to determine how these differentially expressed genes are regulated in BE, and how these signaling pathways (e.g., BMP/TGFβ pathway) and transcription factors (e.g., CDX1, CDX2) may interact with each other to mediate the development of BE.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Dr. Li Cai, Department of Biomedical Engineering, Rutgers University, Piscataway, NJ, in helping in SAGE(Poisson) analysis; Dr. Vasilis Vasiliou, Molecular Toxicology & Environmental Health Sciences Program, Department of Pharmaceutical Sciences, University of Colorado Health Sciences Center, Denver, CO, for providing the anti-ALDH3A1 used in this study; Drs. Danielle M. Greenwalt and Wayne A. Phillips, Centre for Cancer Genomics and Predictive Medicine, Peter MacCallum Cancer Centre, Victoria, Australia, for providing original data and information regarding pathology; Dr. Jerzy Ostrowski, Department of Gastroenterology, Medical Center for Postgraduate Education, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland, for providing original data and information regarding pathology; Dr. William R. Otto, Histopathology Unit, London Research Institute, Cancer Research UK, London, UK, for staining our tissue sections for GDDR.

This study was supported by NIH grants U56 CA092077 and P20 MD000175.

## The abbreviations used are

| | |
|---|---|
| **BE** | Barrett's esophagus |
| **EAC** | esophageal adenocarcinoma |
| **FDR** | false discovery rate |
| **GenMapp** | Gene Map Annotator and Pathway Profiler |
| **GO** | gene ontology |
| **GSEA** | gene set enrichment analysis |
| **IHC** | immunohistochemical staining |
| **IM** | intestinal metaplasia |
| **NE** | normal esophagus |
| **SAGE** | serial analysis of gene expression |
| **SAM** | significance analysis of microarrays |

## References

1. Chen X, Yang CS. Esophageal adenocarcinoma: a review and perspectives on the mechanism of carcinogenesis and chemoprevention. Carcinogenesis. 2001; 22:1119–29. [PubMed: 11470739]
2. Jankowski JA, Harrison RF, Perry I, Balkwill F, Tselepis C. Barrett's metaplasia. Lancet. 2000; 356:2079–85. [PubMed: 11145505]
3. Fitzgerald RC. Barrett's oesophagus and oesophageal adenocarcinoma: how does acid interfere with cell proliferation and differentiation? Gut. 2005; 54(Suppl 1):i21–6. [PubMed: 15711004]
4. Krishnadath KK. Novel findings in the pathogenesis of esophageal columnar metaplasia or Barrett's esophagus. Curr Opin Gastroenterol. 2007; 23:440–5. [PubMed: 17545783]
5. Barrett MT, Yeung KY, Ruzzo WL, et al. Transcriptional analyses of Barrett's metaplasia and normal upper GI mucosae. Neoplasia. 2002; 4:121–8. [PubMed: 11896567]
6. Fox CA, Sapinoso LM, Zhang H, et al. Altered expression of TFF-1 and CES-2 in Barrett's Esophagus and associated adenocarcinomas. Neoplasia. 2005; 7:407–16. [PubMed: 15967118]
7. Gomes LI, Esteves GH, Carvalho AF, et al. Expression profile of malignant and nonmalignant lesions of esophagus and stomach: differential activity of functional modules related to inflammation and lipid metabolism. Cancer Res. 2005; 65:7127–36. [PubMed: 16103062]

8. Greenawalt DM, Duong C, Smyth GK, et al. Gene expression profiling of esophageal cancer: comparative analysis of Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma. Int J Cancer. 2007; 120:1914–21. [PubMed: 17236199]
9. Hao Y, Triadafilopoulos G, Sahbaie P, Young HS, Omary MB, Lowe AW. Gene expression profiling reveals stromal genes expressed in common between Barrett's esophagus and adenocarcinoma. Gastroenterology. 2006; 131:925–33. [PubMed: 16952561]
10. Helm J, Enkemann SA, Coppola D, Barthel JS, Kelley ST, Yeatman TJ. Dedifferentiation precedes invasion in the progression from Barrett's metaplasia to esophageal adenocarcinoma. Clin Cancer Res. 2005; 11:2478–85. [PubMed: 15814623]
11. Kimchi ET, Posner MC, Park JO, et al. Progression of Barrett's metaplasia to adenocarcinoma is associated with the suppression of the transcriptional programs of epidermal differentiation. Cancer Res. 2005; 65:3146–54. [PubMed: 15833844]
12. Ostrowski J, Mikula M, Karczmarski J, et al. Molecular defense mechanisms of Barrett's metaplasia estimated by an integrative genomics. J Mol Med. 2007; 85:733–43. [PubMed: 17415542]
13. Ostrowski J, Rubel T, Wyrwicz LS, et al. Three clinical variants of gastroesophageal reflux disease form two distinct gene expression signatures. J Mol Med. 2006; 84:872–82. [PubMed: 16924468]
14. Pohler E, Craig AL, Cotton J, et al. The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. Mol Cell Proteomics. 2004; 3:534–47. [PubMed: 14967811]
15. Selaru FM, Zou T, Xu Y, et al. Global gene expression profiling in Barrett's esophagus and esophageal cancer: a comparative analysis using cDNA microarrays. Oncogene. 2002; 21:475–8. [PubMed: 11821959]
16. van Baal JW, Milano F, Rygiel AM, et al. A comparative analysis by SAGE of gene expression profiles of Barrett's esophagus, normal squamous esophagus, and gastric cardia. Gastroenterology. 2005; 129:1274–81. [PubMed: 16230080]
17. Wang S, Zhan M, Yin J, et al. Transcriptional profiling suggests that Barrett's metaplasia is an early intermediate stage in esophageal adenocarcinogenesis. Oncogene. 2006; 25:3346–56. [PubMed: 16449976]
18. Xu Y, Selaru FM, Yin J, et al. Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. Cancer Res. 2002; 62:3493–7. [PubMed: 12067993]
19. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001; 98:5116–21. [PubMed: 11309499]
20. Cai L, Huang H, Blackshaw S, Liu JS, Cepko C, Wong WH. Clustering analysis of SAGE data using a Poisson approach. Genome Biol. 2004; 5:R51. [PubMed: 15239836]
21. Lewin, K.; Appelman, H. Barrett's esophagus, columnar dysplasia, and adenocarcinoma of the esophagus. In: Appelman, KLaH, editor. Tumors of the Esophagus and Stomach. AFIP; Washington D.C.: 1995. p. 99-144.
22. Liu T, Zhang X, So CK, et al. Regulation of Cdx2 expression by promoter methylation, and effects of Cdx2 transfection on morphology and gene expression of human esophageal epithelial cells. Carcinogenesis. 2007; 28:488–96. [PubMed: 16990345]
23. Wong NA, Wilding J, Bartlett S, et al. CDX1 is an important molecular mediator of Barrett's metaplasia. Proc Natl Acad Sci U S A. 2005; 102:7565–70. [PubMed: 15894614]
24. Yang Q, Bermingham NA, Finegold MJ, Zoghbi HY. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. Science. 2001; 294:2155–8. [PubMed: 11739954]
25. Eda A, Osawa H, Satoh K, et al. Aberrant expression of CDX2 in Barrett's epithelium and inflammatory esophageal mucosa. J Gastroenterol. 2003; 38:14–22. [PubMed: 12560917]
26. Silberg DG, Furth EE, Taylor JK, Schuck T, Chiou T, Traber PG. CDX1 protein expression in normal, metaplastic, and neoplastic human alimentary tract epithelium. Gastroenterology. 1997; 113:478–86. [PubMed: 9247467]
27. Chen X, Qin R, Liu B, et al. Multilayered epithelium in a rat model and human Barrett's esophagus: similar expression patterns of transcription factors and differentiation markers. BMC Gastroenterol. 2008; 8:1. [PubMed: 18190713]

28. Milano F, van Baal JW, Buttar NS, et al. Bone morphogenetic protein 4 expressed in esophagitis induces a columnar phenotype in esophageal squamous cells. Gastroenterology. 2007; 132:2412–21. [PubMed: 17570215]
29. Slack JM, Tosh D. Transdifferentiation and metaplasia--switching cell types. Curr Opin Genet Dev. 2001; 11:581–6. [PubMed: 11532402]
30. Brabender J, Marjoram P, Salonga D, et al. A multigene expression panel for the molecular diagnosis of Barrett's esophagus and Barrett's adenocarcinoma of the esophagus. Oncogene. 2004; 23:4780–8. [PubMed: 15107828]
31. Glickman JN, Blount PL, Sanchez CA, et al. Mucin core polypeptide expression in the progression of neoplasia in Barrett's esophagus. Hum Pathol. 2006; 37:1304–15. [PubMed: 16949933]
32. Jovov B, Van Itallie CM, Shaheen NJ, et al. Claudin-18: a dominant tight junction protein in Barrett's esophagus and likely contributor to its acid resistance. Am J Physiol Gastrointest Liver Physiol. 2007; 293:G1106–13. [PubMed: 17932229]
33. Kumble S, Omary MB, Fajardo LF, Triadafilopoulos G. Multifocal heterogeneity in villin and Ep-CAM expression in Barrett's esophagus. Int J Cancer. 1996; 66:48–54. [PubMed: 8608965]
34. Madsen J, Nielsen O, Tornoe I, Thim L, Holmskov U. Tissue localization of human trefoil factors 1, 2, and 3. J Histochem Cytochem. 2007; 55:505–13. [PubMed: 17242463]
35. Mitas M, Almeida JS, Mikhitarian K, et al. Accurate discrimination of Barrett's esophagus and esophageal adenocarcinoma using a quantitative three-tiered algorithm and multimarker real-time reverse transcription-PCR. Clin Cancer Res. 2005; 11:2205–14. [PubMed: 15788668]
36. van Baal JW, Bozikas A, Pronk R, et al. Cytokeratin and CDX-2 expression in Barrett's esophagus. Scand J Gastroenterol. 2008; 43:132–40. [PubMed: 18224560]
37. Wong NA, Warren BF, Piris J, Maynard N, Marshall R, Bodmer WF. EpCAM and gpA33 are markers of Barrett's metaplasia. J Clin Pathol. 2006; 59:260–3. [PubMed: 16473928]
38. Christie KN, Thomson C. The distribution of carbonic anhydrase II in human, pig and rat oesophageal epithelium. Histochem J. 2000; 32:753–7. [PubMed: 11254091]
39. Xia SH, Hu LP, Hu H, et al. Three isoforms of annexin I are preferentially expressed in normal esophageal epithelia but down-regulated in esophageal squamous cell carcinomas. Oncogene. 2002; 21:6641–8. [PubMed: 12242662]
40. Mobasheri A, Wray S, Marples D. Distribution of AQP2 and AQP3 water channels in human tissue microarrays. J Mol Histol. 2005; 36:1–14. [PubMed: 15703994]
41. Chu PG, Weiss LM. Keratin expression in human tissues and neoplasms. Histopathology. 2002; 40:403–39. [PubMed: 12010363]
42. Gerber JK, Richter T, Kremmer E, et al. Progressive loss of PAX9 expression correlates with increasing malignancy of dysplastic and cancerous epithelium of the human oesophagus. J Pathol. 2002; 197:293–7. [PubMed: 12115874]
43. Ihrie RA, Marques MR, Nguyen BT, et al. Perp is a p63-regulated gene essential for epithelial integrity. Cell. 2005; 120:843–56. [PubMed: 15797384]
44. South AP. Plakophilin 1: an important stabilizer of desmosomes. Clin Exp Dermatol. 2004; 29:161–7. [PubMed: 14987275]
45. Nishimori T, Tomonaga T, Matsushita K, et al. Proteomic analysis of primary esophageal squamous cell carcinoma reveals downregulation of a cell adhesion protein, periplakin. Proteomics. 2006; 6:1011–8. [PubMed: 16400690]
46. Mori-Akiyama Y, van den Born M, van Es JH, et al. SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. Gastroenterology. 2007; 133:539–46. [PubMed: 17681175]
47. Niimi T, Nagashima K, Ward JM, et al. claudin-18, a novel downstream target gene for the T/EBP/NKX2.1 homeodomain transcription factor, encodes lung- and stomach-specific isoforms through alternative splicing. Mol Cell Biol. 2001; 21:7380–90. [PubMed: 11585919]
48. Haveri H, Westerholm-Ormio M, Lindfors K, et al. Transcription factors GATA-4 and GATA-6 in normal and neoplastic human gastrointestinal mucosa. BMC Gastroenterol. 2008; 8:9. [PubMed: 18405344]
49. Calon A, Gross I, Lhermitte B, et al. Different effects of the Cdx1 and Cdx2 homeobox genes in a murine model of intestinal inflammation. Gut. 2007; 56:1688–95. [PubMed: 17595234]

50. Barros R, Pereira B, Duluc I, et al. Key elements of the BMP/SMAD pathway co-localize with CDX2 in intestinal metaplasia and regulate CDX2 expression in human gastric cell lines. J Pathol. 2008

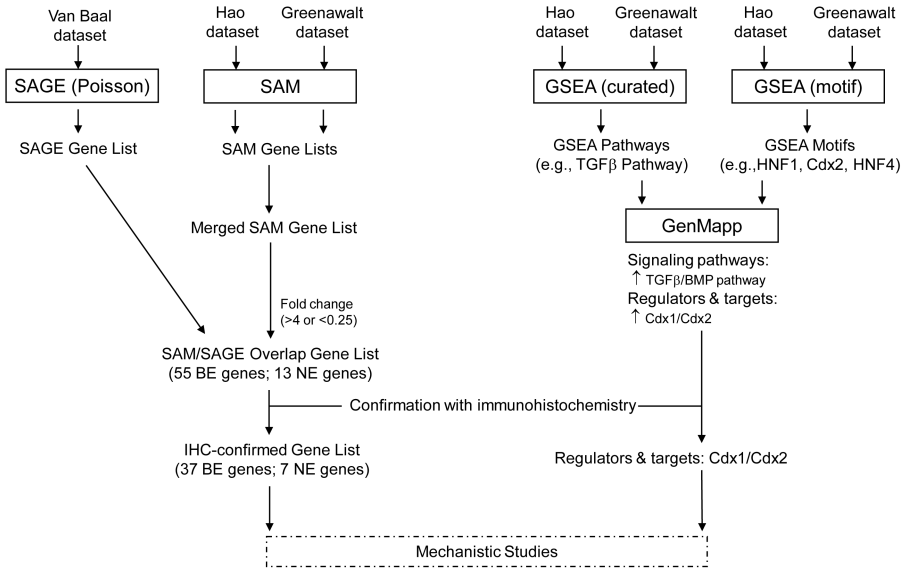


**Figure 1.**

The overall design of this study. The right branch represented the process of generating a list of individual genes differentially expressed in NE and BE, while the left branch indicated the methodology of pathway analysis. IHC was used to validate differential gene expression in NE and BE.

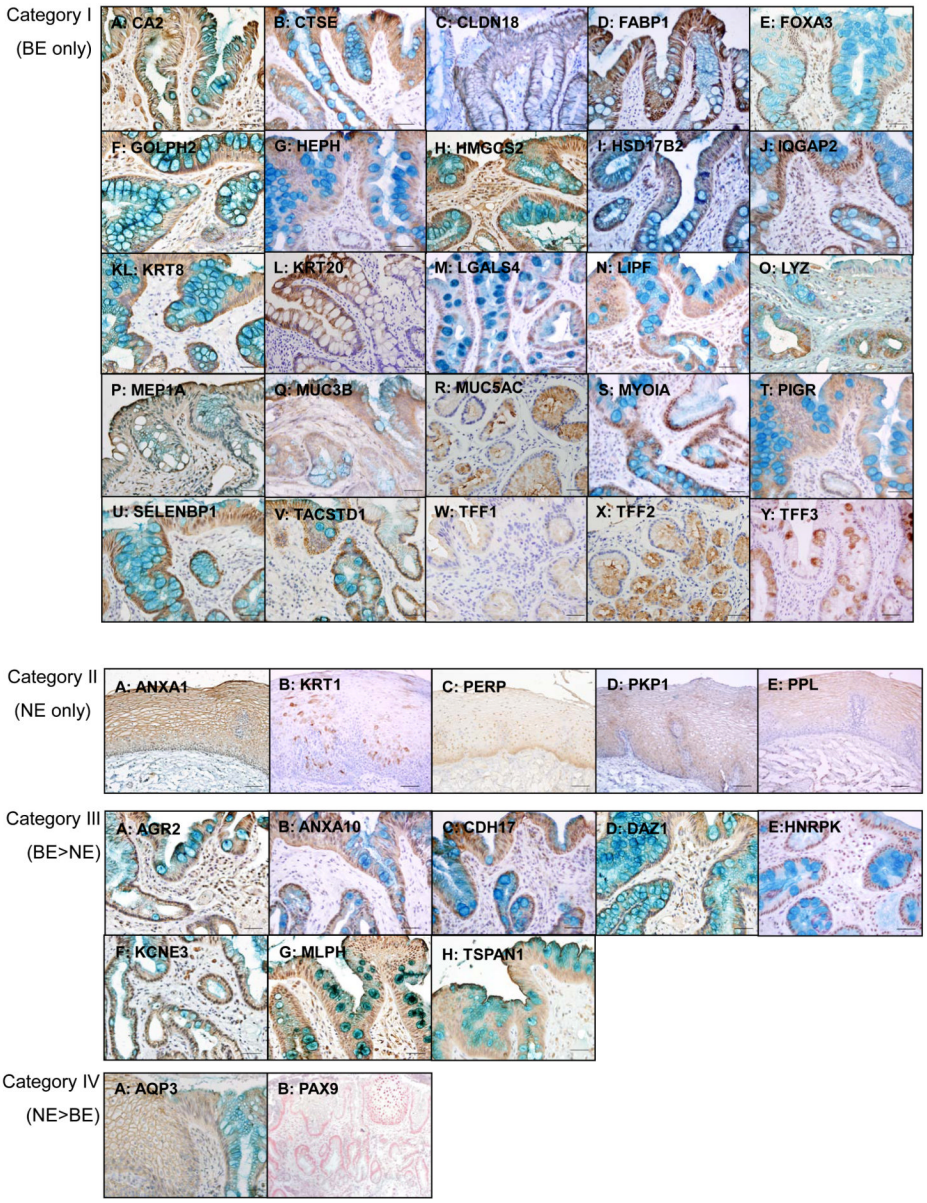


**Figure 2.**

Confirmation of differential gene expression in human BE by IHC. Expression of four categories of genes are shown here: (1) Category I: genes expressed in BE only (25 genes); (2) Category II: genes expressed in NE only (5 genes); (3) Category III: genes expressed higher in BE than NE (8 genes); (4) Category IV: genes expressed higher in NE than BE (2 genes). Due to the space limitations, staining of the NE slides for Category III genes were not shown. Scale bar = 50μm.

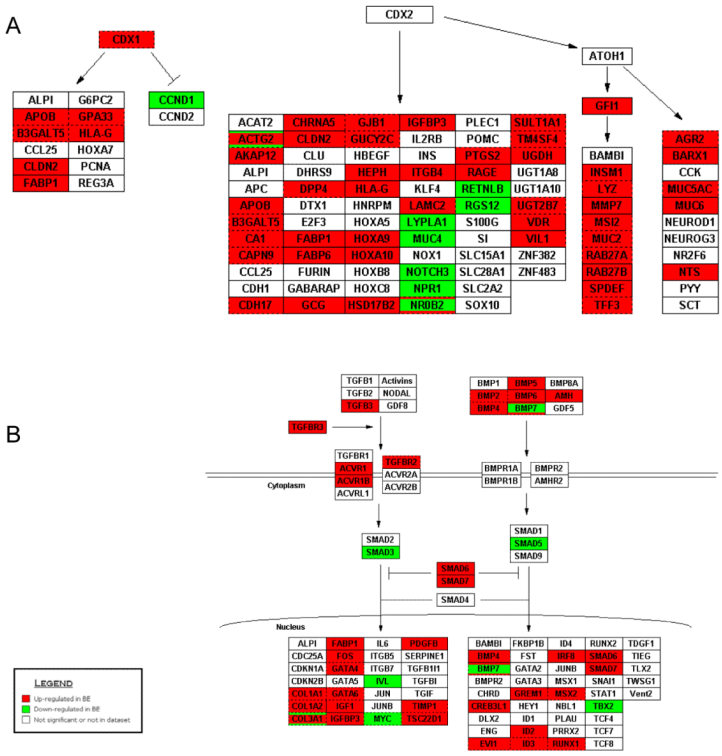


**Figure 3.**

Visualization of target genes of CDX1 and CDX2, and the BMP/TGFβ signaling pathway with GenMapp, using cDNA array data (the Hao dataset and the Greenawalt dataset). Red boxes indicated genes that were up-regulated in BE, while green boxes indicated genes that were down-regulated in BE. Uncolored boxes denoted genes that were not found to be differentially expressed or were not present in the dataset. Dashed boxes represented genes that had multiple spots on the microarray showing differential expression. The central color of dashed boxes with multiple colors denoted the expression pattern for the mode of spots, while the rim color denoted the expression pattern of the minority of spots. Figure 4A showed differential expression of target genes of CDX1 and CDX2. Figure 4B showed differential expression of genes of the BMP/TGFβ signaling pathway.

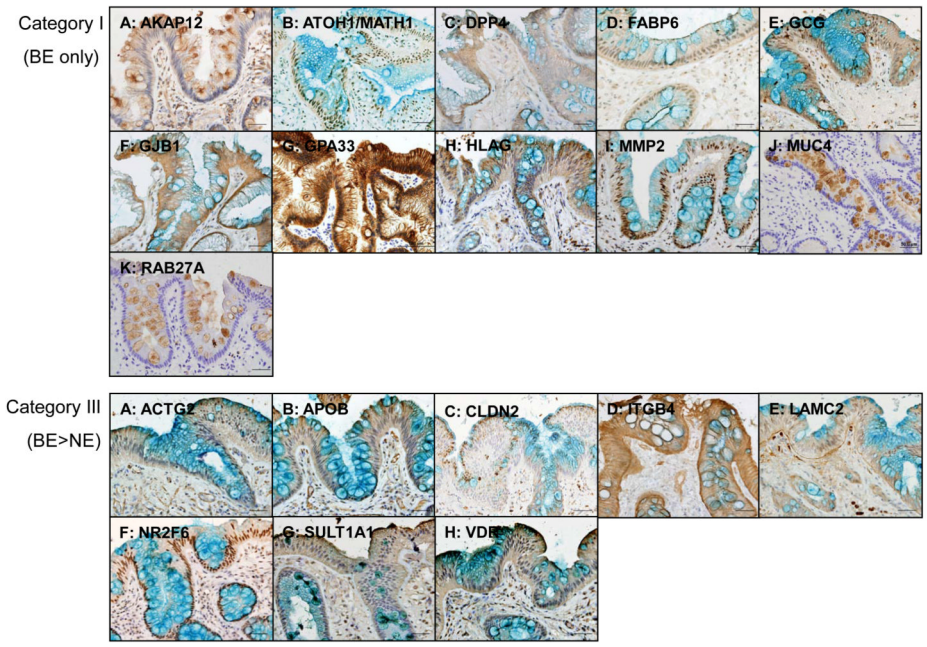


**Figure 4.**

IHC staining of BE tissue for selected target genes of CDX1 and CDX2. Differential expression of 19 genes (11 genes of Category I and 8 genes of Category III) were validated. Scale bar = 50μm.

**Table 1**

Differentially expressed genes in human normal and Barrett's esophagus: SAM/SAGE overlap gene list [1]

| Gene Symbol | Gene Name | Antibody for IHC [2] | Category [3] | Cells (Localization) [4] |
|---|---|---|---|---|
| AGR2 | Anterior gradient 2 homolog (Xenopus laevis) | 1:50, NB110-17780, Novus | III | Columnar, Goblet<br>Basal, Parabasal, Superficial (C,M) |
| ANXA10 | Annexin A10 | 1:1000, AF3544, R&D | III | Columnar, Goblet<br>Basal, Parabasal (C,M) |
| BCMP11 | Breast cancer membrane protein 11; anterior gradient 3 homolog | - | VI | - |
| CA2 | Carbonic anhydrase II | 1:100, AF2184, R&D | I | Columnar, Goblet (C,M) |
| CDH17 | Cadherin 17, LI cadherin (liver-intestine) | 1:100, MAB1032, R&D | III | Columnar, Goblet<br>Basal, Parabasal (C,M) |
| CLDN18 | Claudin 18 | 1:250, 38-8100, Invitrogen | I | Columnar, Goblet (M) |
| CTSE | Cathepsin E | 1:100, AF1294, R&D | I | Columnar, Goblet (C) |
| DAZ4 | Deleted in azoospermia 1 | 1:500, H00001617-A01, Novus | III | Columnar, Goblet<br>Basal, Parabasal (C,N) |
| FABP1 | Fatty acid binding protein 1, liver | 1:100, GTX27807, Genetex | I | Columnar, Goblet (C) |
| FOXA3 | Forkhead box A3; HNF3g | 1:100, PAI-17038, ABR | I | Columnar, Goblet (N) |
| GDDR | Blottin; Down-regulated in gastric cancer; gastrokine 2 | - | VI | - |
| GKN1 | Gastrokine 1; CA11 | - | VI | - |
| GOLPH2 | Golgi phosphoprotein 2; GP73 | 1:400, LS-C682, Lifespan | I | Columnar, Goblet (C,M) |
| HEPH | Hephaestin | 1:500, sc-49970, Santa Cruz | I | Columnar, Goblet (C) |
| HMGCS2 | 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial) | 1:100, NB120-14302, Novus | I | Columnar, Goblet (C,N) |
| HNRPK | Heterogeneous nuclear ribonucleoprotein K | 1:25, Auvation | III | Columnar, Goblet<br>Basal, Parabasal (N) |
| HSD17B2 | Hydroxysteroid (17-beta) dehydrogenase 2 | 1:250, 10978-1-AP, Proteintech | I | Columnar, Goblet (C) |
| IGHG1 | Immunoglobulin heavy constant gamma 1 (G1m marker) | - | VI | - |
| IGJ | Immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides | - | VI | - |

| Gene Symbol | Gene Name | Antibody for IHC [2] | Category [3] | Cells (Localization) [4] |
|---|---|---|---|---|
| IQGAP2 | IQ motif containing GTPase activating protein 2 | 1:250, 18448076, USB | I | Columnar, Goblet (C,N) |
| ITLN1 | Intelectin 1 (galactofuranose binding) | - | VI | - |
| KCNE3 | Potassium voltage-gated channel, Isk-related family, member 3 | 1:200, LS-C863, Lifespan | III | Columnar, Goblet<br>Basal, Parabasal, Superficial (C,N) |
| KRT20 | Keratin 20 | - | I | Columnar, Goblet (C,M) |
| KRT8 | Keratin 8 | 1:100, NCL-L-CK8, Novocastra | I | Columnar, Goblet (C,M) |
| LGALS4 | Lectin, galactoside-binding, soluble, 4 (galectin 4) | 1:250, AF1227, R&D | I | Columnar, Goblet (C) |
| LIPF | Lipase, gastric | 1:500, sc50938, Santa Cruz | I | Columnar, Goblet (C) |
| LYZ | Lysozyme (renal amyloidosis) | 1:800, A0099, Dako | I | Paneth (C) |
| MEP1A | Meprin A, alpha (PABA peptide hydrolase) | 1:50, M2930, USB | I | Columnar, Goblet (C,M) |
| MGC27165 | Hypothetical protein MGC27165 | - | VI | - |
| MLPH | Melanophilin | 1:200, 10338-1-AP, Proteintech | III | Columnar, Goblet, Parabasal, Superficial (C,N,M) |
| MUC13 | Mucin 13, cell surface associated | - | VI | - |
| MUC3B | Mucin 3B, cell surface associated | MA1-21217, ABR | I | Columnar, Goblet (C) |
| MUC5AC | Mucin 5, subtypes A and C, tracheobronchial/gastric | - | I | Columnar, Goblet (C) |
| MYO1A | Myosin IA | 1:250, Provided by Dr. Matthew J. Tyska | I | Columnar, Goblet (C,N) |
| OCIAD2 | OCIA domain containing 2 | - | VI | - |
| ORF1-FL49 | Putative nuclear protein ORF1-FL49 | - | VI | - |
| PIGR | Polymeric immunoglobulin receptor | 1:250, AF2800, R&D | I | Columnar, Goblet (C) |
| PROM1 | Prominin 1; CD133 | - | VI | Columnar, Goblet, Basal (C) |
| REG4 | Regenerating islet-derived family, member 4 | - | VI | - |
| RGS2 | Regulator of G-protein signalling 2, 24kDa | - | VI | Columnar, Goblet<br>Basal, Parabasal, Superficial (C,N) |
| RNASE1 | Ribonuclease, RNase A family, 1 (pancreatic) | - | VI | - |
| RNASE4 | Ribonuclease, RNase A family, 4 | 1:500, LS-C11241, Lifespan | V | Columnar, Goblet |

| Gene Symbol | Gene Name | Antibody for IHC [2] | Category [3] | Cells (Localization) [4] |
|---|---|---|---|---|
| | | | | Basal, Parabasal, Superficial (N) |
| SELENBP1 | Selenium binding protein 1 | 1:500, M061-3, MBI | I | Columnar, Goblet (C,N) |
| SEMA4G | Sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4G | 1:500, H00057715-A01, Novus | V | Columnar, Goblet<br>Basal, Parabasal, Superficial (C,N) |
| SLC39A7 | Solute carrier family 39 (zinc transporter), member 7 | - | VI | - |
| SULT1C1 | **Sulfotransferase family, cytosolic, 1C, member 1 | - | VI | - |
| TACSTD1 | Tumor-associated calcium signal transducer 1; EpCAM | 1:50, MA1-06502, ABR | I | Columnar, Goblet (C,M) |
| TFF1 | Trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in) | | I | Columnar, Goblet (C) |
| TFF2 | Trefoil factor 2 (spasmolytic protein 1) | - | I | Columnar, Goblet (C) |
| TFF3 | Trefoil factor 3 (intestinal) | 1:250, sc28927, Santa Cruz | I | Goblet (C) |
| TM4SF4 | Transmembrane 4 L six family member 4; intestinal and liver tetraspan membrane protein (il-TMP) | - | VI | - |
| TM4SF8 | Tetraspanin 3 | - | VI | - |
| TPD52 | Tumor protein D52 | 1:200, sc33842, Santa Cruz | V | Columnar, Goblet<br>Basal, Parabasal, Superficial (C) |
| TSPAN1 | Tetraspanin 1 | 1:500, PAI-57842, ABR | III | Columnar, Goblet<br>Basal, Parabasal, Superficial (C,M) |
| VILL | Villin-like | - | VI | - |
| ADH7 | Alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide | 1:500, PAI-50112, ABR | V | Columnar, Goblet<br>Basal, Parabasal, Superficial (C,N) |
| ALDH3A1 | Aldehyde dehydrogenase 3 family, memberA1 | Provided by Dr. Vasilis Vasiliou | V | Columnar, Goblet<br>Basal, Parabasal, Superficial (C) |
| ANXA1 | Annexin A1 | 1:25, RB-10659-P0, Thermo | II | Parabasal, Superficial (M) |
| AQP3 | Aquaporin 3 | 1:500, NB110-13175, Novus | IV | Columnar, Goblet Basal, Parabasal, Superficial (C,M) |
| CITED2 | Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2 | 1:800, MA3-16523, ABR | V | Columnar, Goblet<br>Basal, Parabasal, Superficial (N) |
| GJB2 | Gap junction protein, beta 2, 26kDa | - | VI | - |

| Gene Symbol | Gene Name | Antibody for IHC [2] | Category [3] | Cells (Localization) [4] |
|---|---|---|---|---|
| | (connexin 26) | | | |
| KRT1 | Keratin 1 (epidermolytic hyperkeratosis) | 1:40, NCL-CK1, Novocastra | II | Parabasal, Superficial (C) |
| MALAT1 | Metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA) | - | VI | - |
| PAX9 | Paired box gene 9 | Provided by Dr. Heiko Peters | IV | Columnar, Goblet<br>Basal, Parabasal, Superficial (N) |
| PERP | PERP, TP53 apoptosis effector | 1:500, LS-C400, Lifespan | II | Basal, Parabasal, Superficial (C) |
| PKP1 | Plakophilin 1 (ectodermal dysplasia/skin fragility syndrome) | PP1-5C2, Progen | II | Basal, Parabasal, Superficial (C,M) |
| PPL | Periplakin | 1:250, sc16754, Santa Cruz | II | Parabasal, Superficial (M) |
| RAB38 | RAB38, RAS oncogene family | - | VI | - |

[1]Differential expression of these genes were also reported in other microarray and proteomics studies, which were excluded in this study (5, 6, 11, 14, 17): 1) genes up-regulated in BE: AGR2, ANXA10, CA2, CDH17, CTSE, DAZ1, FABP1, FOXA3, GCG, HSD17B2, KRT8, LGALS4, LYZ, MEP1A, MUC5AC, RGS2, TFF1, TM4SF4, TM4SF8, TSPAN1; 2) genes down-regulated in BE: ALDH3A1, ANXA1, AQP3, KRT1, PAX9, PPL.

[2]Information of antibody dilution, catalogue number, and manufacturer are provided.

[3]Genes were categorized according to their expression pattern: 1) Category I: BE only (25 genes); 2) Category II: NE only (5 genes); 3) Category III: BE>NE (8 genes); 4) Category IV: NE>BE (2 genes); 5) Category V: NE=BE (6 genes); 6) Category VI: others (not tested or good antibodies unavailable; 22 genes).

[4]Columnar stands for “columnar epithelial cells”; Goblet for “goblet cells”; Paneth for “Paneth cells”; Basal for “basal cells in squamous epithelium”; Parabasal for “parabasal cells in squamous epithelium” and Superficial for “superficial cells in squamous epithelium”. C: cytoplasmic localization; M: membranous localization; and N: nuclear localization.

*J Gastroenterol*. Author manuscript; available in PMC 2013 October 29.

**Table 2**

Expression of selected target genes of CDX1 and CDX2 in human normal and Barrett's esophagus [1]

| Gene symbol | Gene name | Antibody for IHC [2] | Category[3] | Cells (Localization) [3] |
|---|---|---|---|---|
| ACTG2 | actin, gamma 2, smooth muscle, enteric | 1:100, PPI-50055, ABR | III | Columnar, goblet, basal, parabasal, superficial (C) |
| AKAP12 | A-kinase anchor protein 12 | 1:1000, G3795, Sigma | I | Columnar, goblet (C) |
| APOB | apolipoprotein B | 1:40, MAB4124, R&D | III | Columnar, goblet, basal, parabasal, superficial (C) |
| ATOH1 | atonal homolog 1 (Drosophila) | 1:500, ab27667, ABcam | I | Columnar, goblet (N) |
| CLDN2 | Claudin 2 | 1:250, 32-5600, Invitrogen | III | Columnar, goblet, basal, parabasal, superficial (C) |
| DPP4 | dipeptidyl-peptidase 4, CD26 | 1:100, ab22613, ABcam | I | Columnar, goblet (C) |
| FABP6 | fatty acid binding protein 6, ileal | 1:40, AF3880, R&D | I | Columnar, goblet (C,M) |
| GCG | Glucagon | Ab18459, ABcam | I | Columnar, goblet (C) |
| GJB1 | gap junction protein, beta 1, 32kDa | 1:100, MA1-25094, ABR | I | Columnar, goblet (C) |
| GPA33 | Glycoprotein A33 | 1:40, AF3080, R&D | I | Columnar, goblet (C,M) |
| HLAG | major histocompatibility complex, class I, G | 1:100, ab26090, ABcam | I | Columnar, goblet (C) |
| ITGB4 | integrin beta 4 | 1:40, AF4060, R&D | III | Columnar, goblet, basal, parabasal, superficial (C,M) |
| LAMC2 | laminin, gamma 2 | 1:50, sc28330, Santa Cruz | III | Columnar, goblet, basal, parabasal (C,M) |
| MMP2 | matrix metalloproteinase 2 | 1:50, ab2462, ABcam | I | Columnar, goblet (N) |
| MUC4 | Mucin 4 | 1:200, GTX93066, GeneTex | I | Goblet (C) |
| NR2F6 | nuclear receptor subfamily 2, group F, member 6 | 1:100, RB-10471-P0, Thermo | III | Columnar, goblet, basal, parabasal, superficial (N) |
| RAB27A | RAB27A, member RAS oncogene family | 1:100, GTX92731, GeneTex | I | Goblet (C) |
| SULT1A1 | sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1 | 1:100, 10911-2-AP, Proteintech | III | Columnar, goblet, basal, parabasal, superficial (C,M) |
| VDR | vitamin D receptor | 1:50, PP-H4537-00, R&D | III | Columnar, goblet, basal, parabasal, superficial (C,M) |

[1] In addition, 9 target genes of CDX1 and CDX2 (VIL, MUC2, FABP1, CDH17, HEPH, TFF3, AGR2, LYZ, and MUC5AC) have been listed in Table 1 or studied in our previous study (27). GPA33, ITGB4, and Muc6 were also reported by other array and proteomics studies which were excluded from this study (5, 6, 11, 14, 17).

[2] Information of antibody dilution, catalogue number, and manufacturer are provided.

[3] see Footnotes of Table 1 for detailed explanation.