Genome-Wide Analysis of Barrett’s Adenocarcinoma. A First Step Towards Identifying Patients at Risk and Developing Therapeutic Paths1,2

Yiyang Dai*, 3, 4, Qing Wang*, 3, Adrian Gonzalez Lopez†, Mario Anders‡, 5, Peter Malfertheiner§, Michael Vieth¶ and Wolfgang Kemmner*

*Translational Oncology, Experimental Clinical Research Center (ECRC), Charité Campus Buch, 13125 Berlin, Germany; †Department of Anesthesiology and Operative Intensive Care Medicine, Charité Campus Virchow-Klinikum, 13353 Berlin, Germany; ‡Gastroenterologie und Interdisziplinäre Endoskopie, Vivantes Wenckebach-Klinikum, 12099 Berlin, Germany; §Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University Magdeburg, 39120 Magdeburg, Germany; ¶Institute of Pathology, Klinikum Bayreuth, 95445 Bayreuth, Germany

Abstract

BACKGROUND: Barrett’s esophagus metaplasia is the key precursor lesion of esophageal adenocarcinoma. The aim of this study was to find a subset of markers that may allow the identification of patients at risk for esophageal adenocarcinoma, and to determine genes differentially expressed in esophageal squamous cell carcinoma.

METHODS: Laser capture microdissection technique was applied to procure cells from defined regions. Genome-wide RNA profiling was performed on esophageal adenocarcinoma (n = 21), Barrett’s esophagus (n = 20), esophageal squamous carcinoma (n = 9) and healthy esophageal biopsies (n = 18) using the Affymetrix Human Genome U133plus 2.0 array. Microarray results were validated by quantitative real-time polymerase chain reaction in a second and independent cohort and by immunohistochemistry of two putative markers in a third independent cohort.

RESULTS: Through unsupervised hierarchical clustering and principal component analysis, samples were separated into four distinct groups that match perfectly with histology. Many genes down-regulated in esophageal cancers belong to the epidermal differentiation complex or the related GO-group “cornified envelope” of terminally differentiated keratinocytes. Similarly, retinol metabolism was strongly down-regulated. Genes showing strong overexpression in esophageal carcinomas belong to the GO groups extracellular region /matrix such as MMP1, CTHRC1, and INHBA. According to an analysis of genes strongly up-regulated in both esophageal adenocarcinoma and Barrett’s esophagus, REG4 might be of particular interest as an early marker for esophageal adenocarcinoma.

CONCLUSIONS: Our study provides high quality data, which could serve for identification of potential biomarkers of Barrett’s esophagus at risk of esophageal adenocarcinoma progression.

Translational Oncology (2018) 11, 116–124

Address all correspondence to: Wolfgang Kemmner, Translational Oncology, Experimental Clinical Research Center (ECRC), Charité Campus Buch, Lindenberger Weg 80, 13125 Berlin, Germany.
E-mail: wolfgang.kemmner@charite.de

1 Funding: This work was supported by grants from Charité University Hospital and Otto-von-Guericke University Magdeburg, Germany.
2 Disclaimers: The authors indicated no potential conflicts of interest.
3 Both authors contributed equally to this work.
4 Present address: Department of Gastroenterology, The Fourth Affiliated Hospital of Zhejiang University School of Medicine, 310,013 Hangzhou, China.
5 Present address: Zentrum für Radiologie und Endoskopie Universitätsklinikum Hamburg-Eppendorf, Martinistraße 52, 20,251 Hamburg, Germany.
Received 1 August 2017; Revised 12 October 2017; Accepted 23 October 2017
© 2017 The Authors. Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1936-5233
https://doi.org/10.1016/j.tranon.2017.10.003
Introduction
Barrett’s esophagus, a condition in which a metaplastic columnar mucosa that confers a predisposition to cancer replaces an esophageal squamous mucosa, is a major risk factor for esophageal adenocarcinoma (EAC). The frequency of this deadly tumor has increased significantly during the past four decades [1]. Barrett’s esophagus (BE) is associated with a 40-fold increased risk for the development of EAC, which gradually progresses to carcinoma through a metaplasia–dysplasia–carcinoma sequence [2]. Thus, a better understanding of the pathogenesis of Barrett’s associated cancer is required for risk prediction and the development of effective therapeutic strategies. Esophageal squamous cell carcinoma (ESCC) is another histological subtype which is detected mostly at advanced stages leading to a low survival of the patients [3].

In many types of carcinomas, biomarkers can enhance our ability for diagnosis, prognosis, and for therapy prediction. At present there are no clinical or histological parameters that allow the identification of patients bearing a high risk of progression of BE to EAC. One of the reasons might be that esophageal Barrett’s adenocarcinomas are quite heterogeneous. A number of possible markers have been identified, e.g. aneuploidy [4], copy number alterations [5] and deletion of the p16 tumor suppressor gene [6]. Similarly, miRNAs might be potential biomarkers [7]. EGFR was found to be overexpressed in high-grade dysplasia and adenocarcinoma of the esophagus [8]. However, none of the currently known molecular biomarkers has sufficient predictive power to identify BE progression in a clinically useful manner [9].

Gene expression profiling by microarrays is a powerful technique for delineating biomarkers and pathways associated with cancer progression [10]. Here, also laser capture microdissection (LMD) was applied to collect cells from defined regions of histologic tissue. We focused on neoplastic progression and found a set of genes representing potential biomarkers which might help to create a more standardized, reproducible and reliable diagnostic procedure in order to stratify patients with BE in terms of their risk to develop cancer. Expression of these genes can be determined in a quantitative way, and thus may be able to complement conventional semi-quantitative histological analysis.

Materials and Methods

Patients and Tissue Specimens
Unselected patients seeking treatment for Barrett’s esophagus or esophageal cancer at Charité-Universitätsmedizin (Berlin, Germany) were enrolled in the study. Gene expression profiling using whole-human-genome microarrays (Affymetrix U133plus2) was performed on 21 EAC, 20 BE, 9 ESCC and 18 NE biopsies (biopsies from patients without esophageal pathology), collected from 68 individuals. Patients’ age ranged from 22 to 79 years, with a median age of 58 years. The validation cohort for RT-PCR verification consisted of 49 chemotherapy-naive patients with known EAC of histological grading above 7.0 was processed further. The Affymetrix Human Genome U133plus 2.0 array (Affymetrix, Santa Clara, CA) was used for this study. Preparation of labeled cRNA and hybridization were done using the gene chip hybridization, wash, and stain kit (Affymetrix, Santa Clara, CA), as described previously [11–13]. Two cycle labeling was applied to all samples. Chip data quality was strictly controlled. Besides the common quality control provided by gene chip operating software, a scientific workflow for chip quality control was set up in our lab including pseudo image plots, scatter MA plot and relative log expression plot. Additionally, correlation analysis across gene chips was also performed by Agilent GeneSpring software GX 10.2.

Data Analysis
Probe set data were generated with GeneChip Operation Software (GCOS, Affymetrix). Raw data were further processed using tools available from Bioconductor version 2.6 [14]. Background correction and normalization were carried out using the Affymetrix package [15]. Results were filtered by the detection P-value, considering probes with P-value <0.01 as ‘present’ and all others ‘absent’. Probes ‘present’ in at least 50% of the samples of one group were examined for further analysis. Differential expression with respect to control samples was determined using the limma package in Bioconductor [16]. The False Discovery Rate (FDR) was employed using Benjamini-Hochberg procedure for multiple testing of the resulting P-value significance. Principal component analysis (PCA) and unsupervised hierarchical clustering were used to evaluate the relationships between patients’ groups. Gene annotations were based on Affymetrix Human Genome U133 Plus 2.0 Array annotation data (hg-u133-plus2 package). Gene Ontology enrichment was carried out using the DAVID Functional
Results and Discussion

was rabbit polyclonal anti-human INHBA antibody (Sigma-Aldrich (Abcam ab72527, 1:1000). Primary antibody for INHBA detection
fixed paraffin sections using the biotin blocking system (Dako, Gene Expression Profiles of Samples are Consistent with
Annotation Bioinformatics Microarray Analysis [17]. A list of all the
genes included in these microarrays and the normalized data have been
deposited in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/info/linking.html) under GEO accession number
GSE26886.

RNA Extraction and qRT-PCR

RNA extractions were carried out using the RNasy Mini Kit (Qiagen, Valencia, CA). Total RNA quality and yield were assessed using a
bioanalyzer system (Agilent 2100 Bioanalyzer; Agilent Technologies, Palo Alto, CA) and a spectrophotometer (NanoDrop ND-1000; NanoDrop Technologies, Wilmington, DE). Only RNA with an
RNA integrity number >7.0 was used for RT-PCR analysis. For
quantification, the expression levels of genes found by microarray
analyses, RT-PCR was performed using pre-designed gene-specific TaqMan probes and primer sets purchased from Applied Biosystems.
RT-PCR amplification was carried out using RNA UltraSense One-Step Quantitative RT-PCR System (Invitrogen) on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems), according to the manufacturer’s instructions. All samples were amplified in triplicate
reactions. Gene expression was quantified relative to the expression of internal control glyceraldehyde-3-phosphate dehydrogenase (GAPDH) using Sequence Detector Software (SDS 2.2, Applied Biosystems) and the relative quantification method for quantitative RT-PCR. In addition, GAPDH levels were also used to determine whether a given sample contained sufficient mRNA to be included in the study. Samples for
which the Ct value for GAPDH above 22 were excluded from analysis. The relative expression of each individual gene was calculated based on the
determined ΔCT values. The ΔCT value is the difference between the Ct value for an individual gene and the internal reference control gene
GAPDH [18]. ΔCT values correspond to a log-2 scale. The result of quantitative gene expression was calculated as 2^-ΔCT relative to
GAPDH expression in the corresponding samples.

Immunohistochemistry

Immunohistochemistry was performed on 5 μm thin formalin
fixed paraffin sections using the biotin blocking system (Dako, Glostrup, Denmark) and the ImmPRESS universal reagent (Vector Laboratories Inc., Burlingame, CA). Primary antibody for CTHRC1
detection was mouse monoclonal anti-human CTHRC1 antibody (Abcam ab72527, 1:1000). Primary antibody for INHBA detection
was rabbit polyclonal anti-human INHBA antibody (Sigma-Aldrich Corporation, HPA020031, 1:200).

Results and Discussion

Gene Expression Profiles of Samples are Consistent with
Histological Typing

Gene expression profiling using whole-human-genome microar-
rays (Affymetrix U133plus2) was performed on 21 EAC, 20 BE, 9
ESCC and 18 NE biopsies. By unsupervised hierarchical clustering
the tissue samples were separated into 4 distinct groups consistent
with their histological subtype. Clustering of the gene expression
profiles is depicted in the dendogram (Figure 1). All 18 samples
taken from normal esophageal epithelium cluster together on the left
side group called CONTROL. Esophageal carcinoma samples belong
to the group in the middle, but they fall into two distinctly separated
clusters. On the middle left, esophageal squamous cell carcinomas
(ESCC) cluster together, and on the middle right esophageal
adenocarcinomas (EAC) are found. Samples taken from Barrett’s
esophagus (BE) cluster together on the right side. Please note that for
each group all samples cluster together with not one mismatch.
Another approach, Principal component analysis (Supplement Figure 1)
demonstrated a similar grouping. The excellent agreement of gene
expression profiles with the histological subtypes might be a result of laser
capture microdissection, which is likely to be the superior approach in
terms of controlling tissue purity.

Several gene expression profiling studies of Barrett’s esophagus and
esophageal carcinomas have been published to date, e.g., [19]. Cellular contamination within patient samples, or a mixture of tissue
types within a biopsy is a strong challenge for efforts to determine an
accurate cancer–related gene expression profile because gene
expression measurements can become skewed by this heterogeneity.
Here, we have applied laser microdissection (LMD) in order to get
access to almost “pure” samples of Barrett’s epithelium, esophageal
adenocarcinoma or squamous carcinoma tissue. LMD enables the
isolation of specific cell populations from complex tissues under
morphological control, which is likely to be the superior approach in
terms of controlling tissue purity. In our study, LMD was applied not
only for all esophageal tumor samples, but also for BE samples
because BE biopsies often contain only few BE glands besides
squamous epithelium or stromal cells. LMD enabled us to obtain
purified columnar cells from each BE biopsy and use them as a source
for RNA extraction. Unsupervised hierarchical clustering and
principal component analysis of microarray data obtained distin-
guished samples into four distinct groups according to their
histopathology, demonstrating the quality of the LMD approach
and the microarray analysis.

Validation in an Independent Cohort

To verify our observed gene expression results, we investigated
expression levels of 12 randomly chosen genes with P < .01 by
real-time PCR and confirmed conclusively the microarray results
generated from the training cohort. Comparison of the obtained
fold-changes by array analysis versus real-time PCR confirmed our
approach as conservative, revealing higher expression differences in
the former (Supplement Table 2).

Examination of the Marker Expression on the Protein Level

CTHRC1 and INHBA antibodies detected a specific protein band
on western blots and thus were used for another validation of gene
expression results. Immunohistochemical analysis demonstrated
differential abundance of CTHRC1 and INHBA on the protein
level in paraffin tissue specimens of a third independent patient
cohort of 12 EAC, 5 intraepithelial neoplasia (IEN) specimens, and 7
BE from patients without tumor formation in the first five years after
biopsy sampling representing low risk BE, and 12 NE specimens.
Immunohistochemical staining showed that CTHRC1 was abundant
in 10 of 12 EAC specimens and in the 5 IEN specimens, but staining
was low in 7 BE specimens from patients without tumor formation.
In NE biopsies no CTHRC1 staining was detected. INHBA was
highly abundant in all EAC samples, highly abundant in the IEN
specimens, but staining was low in the BE samples and not detected at
all in NE specimens (Figure 2).

Analysis of Differentially Expressed Genes

The aim of this study was the generation of a database, which
allows the identification of genes or groups of genes with strong
differential expression in esophageal carcinomas. To get an overview, gene ontology analysis was used to determine significantly over- or underrepresented biological processes, cellular components, and molecular functions within the entire set of differentially expressed genes. Entrez IDs were assigned to gene ontology categories according to the annotations in the Affymetrix Human Genome U133 Plus 2.0 Array annotation data (hgu133plus2 package). Then these Entrez IDs were subjected to DAVID functional annotation bioinformatic microarray analysis, using the default feature listings and algorithm settings, with the whole human genome as background. Ontology

Figure 1. Unsupervised hierarchical clustering of samples obtained from four different tissue types. Clustering of the gene expression profiles of the 68 cases under study produced four main clusters. Depicted is a full-length view of the cluster dendrogram with cases orientated along the horizontal axis. Hierarchical clustering analysis is called unsupervised if the relatedness of cases is determined only by the similarity of their whole human genome expression profiles, independently of clinical or pathological parameters. Here, hierarchical cluster analysis showed that the four tissue types are clearly distinct. All 18 samples taken from healthy esophageal epithelium (NE) cluster together in the group on the left side. Carcinoma samples belong to the cluster group in the middle, encompassing two strictly separated clusters of esophageal squamous cell carcinomas (ESCC) and of esophageal adenocarcinomas (EAC). Samples taken from Barrett’s esophagus (BE) form another cluster group on the right side. Please note that all samples of the different tissue types cluster together with not one mismatch.

Figure 2. Immunohistochemical detection of CTHRC1 (A-D) and INHBA (E-H) expression in healthy epithelium, BE, intraepithelial neoplasia and esophageal adenocarcinoma. Immunohistochemical detection of CTHRC1 and INHBA (red) was performed on 5 μm thin formalin fixed paraffin sections. Cell nuclei were counterstained using hematoxylin (blue). No specific CTHRC1 / INHBA expression could be found in healthy esophageal epithelium (A + E), low CTHRC1 / INHBA expression was detected in Barrett’s glands (B + F). In adenocarcinomas, high CTHRC1 and INHBA expression was predominantly observed in mucosa glands (D + H), moreover, high expression of CTHRC1 / INHBA was observed in the intraepithelial neoplasia (C + G). Representative sections are shown. Negative controls were obtained by omission of primary antibody (data not shown). Pictures are 200× magnified.
Table 1. Peak Genes Up-Regulated in Esophageal Carcinomas

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Description</th>
<th>Entrez</th>
<th>BE logFC</th>
<th>BE adj.P</th>
<th>EAC logFC</th>
<th>EAC adj.P</th>
<th>ESCC logFC</th>
<th>ESCC adj.P</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGN</td>
<td>biglycan</td>
<td>633</td>
<td>0.18</td>
<td>5.56E-01</td>
<td>2.56</td>
<td>4.84E-13</td>
<td>2.70</td>
<td>6.16E-10</td>
</tr>
<tr>
<td>C11orf96</td>
<td>chromosome 11 open reading frame 96</td>
<td>387763</td>
<td>2.12</td>
<td>3.20E-08</td>
<td>4.61</td>
<td>9.83E-21</td>
<td>3.76</td>
<td>2.93E-12</td>
</tr>
<tr>
<td>COCH</td>
<td>coagulation factor C homolog, coelhin (Limulus polyphemus)</td>
<td>1690</td>
<td>0.76</td>
<td>1.03E-01</td>
<td>2.95</td>
<td>8.35E-09</td>
<td>2.77</td>
<td>1.33E-05</td>
</tr>
<tr>
<td>COL6A1</td>
<td>collagen, type VIII, alpha 1</td>
<td>1295</td>
<td>-0.67</td>
<td>1.45E-01</td>
<td>2.56</td>
<td>1.59E-07</td>
<td>2.72</td>
<td>1.23E-05</td>
</tr>
<tr>
<td>CTHRC1</td>
<td>collagen triple helix repeat containing 1</td>
<td>115908</td>
<td>3.00</td>
<td>7.35E-10</td>
<td>5.31</td>
<td>1.50E-19</td>
<td>5.50</td>
<td>5.21E-15</td>
</tr>
<tr>
<td>FOXC1</td>
<td>forkhead box C1</td>
<td>2296</td>
<td>-1.35</td>
<td>2.50E-03</td>
<td>1.23</td>
<td>4.79E-03</td>
<td>2.57</td>
<td>1.68E-05</td>
</tr>
<tr>
<td>IL6</td>
<td>interleukin 8</td>
<td>3576</td>
<td>0.68</td>
<td>2.48E-01</td>
<td>2.73</td>
<td>6.77E-06</td>
<td>2.58</td>
<td>9.25E-04</td>
</tr>
<tr>
<td>INHBA</td>
<td>inhibin, beta A</td>
<td>3624</td>
<td>0.33</td>
<td>3.88E-01</td>
<td>4.26</td>
<td>1.23E-17</td>
<td>5.13</td>
<td>8.01E-16</td>
</tr>
<tr>
<td>LEPRE4</td>
<td>leprecan-like 4</td>
<td>10609</td>
<td>0.99</td>
<td>5.00E-04</td>
<td>3.03</td>
<td>6.17E-17</td>
<td>3.15</td>
<td>6.56E-13</td>
</tr>
<tr>
<td>MMP1</td>
<td>matrix metalloproteinase 1 (intestinal collagenase)</td>
<td>4312</td>
<td>2.51</td>
<td>3.07E-04</td>
<td>5.49</td>
<td>3.87E-12</td>
<td>7.12</td>
<td>6.53E-12</td>
</tr>
<tr>
<td>MTHFD1L</td>
<td>methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like</td>
<td>25902</td>
<td>0.30</td>
<td>3.51E-01</td>
<td>2.55</td>
<td>3.11E-07</td>
<td>2.94</td>
<td>2.54E-10</td>
</tr>
<tr>
<td>NNNMT</td>
<td>nitrinaminid N-methyltransferase</td>
<td>4837</td>
<td>2.10</td>
<td>1.08E-05</td>
<td>4.18</td>
<td>4.33E-14</td>
<td>4.06</td>
<td>1.02E-09</td>
</tr>
<tr>
<td>PEMPA1</td>
<td>prostate transmembrane protein, androgen induced 1</td>
<td>56937</td>
<td>1.27</td>
<td>7.58E-04</td>
<td>4.37</td>
<td>1.38E-18</td>
<td>3.94</td>
<td>4.07E-12</td>
</tr>
<tr>
<td>RGS1</td>
<td>regulator of G-protein signaling 1</td>
<td>5996</td>
<td>0.05</td>
<td>9.16E-01</td>
<td>2.89</td>
<td>3.07E-09</td>
<td>2.77</td>
<td>4.90E-06</td>
</tr>
<tr>
<td>SOCS3</td>
<td>suppressor of cytokine signaling 3</td>
<td>9021</td>
<td>1.02</td>
<td>1.70E-02</td>
<td>3.27</td>
<td>2.75E-11</td>
<td>3.89</td>
<td>5.91E-10</td>
</tr>
<tr>
<td>SPP1</td>
<td>secreted phosphoprotein 1 / osteopontin</td>
<td>6606</td>
<td>0.02</td>
<td>9.73E-01</td>
<td>2.89</td>
<td>1.20E-06</td>
<td>5.57</td>
<td>5.96E-11</td>
</tr>
<tr>
<td>SULF1</td>
<td>sulfatase 1</td>
<td>23213</td>
<td>-0.16</td>
<td>7.48E-01</td>
<td>3.05</td>
<td>1.51E-08</td>
<td>3.13</td>
<td>4.29E-06</td>
</tr>
<tr>
<td>USP42</td>
<td>ubiquitin specific peptidase 42</td>
<td>84132</td>
<td>0.02</td>
<td>9.23E-01</td>
<td>2.25</td>
<td>6.15E-18</td>
<td>2.91</td>
<td>2.70E-06</td>
</tr>
<tr>
<td>WDR72</td>
<td>WD repeat domain 72</td>
<td>256764</td>
<td>1.93</td>
<td>7.52E-03</td>
<td>4.68</td>
<td>3.28E-09</td>
<td>5.76</td>
<td>1.97E-08</td>
</tr>
<tr>
<td>WISP1</td>
<td>WNT1 inducible signaling pathway protein 1</td>
<td>8840</td>
<td>0.17</td>
<td>6.57E-01</td>
<td>2.26</td>
<td>4.30E-08</td>
<td>3.10</td>
<td>1.48E-08</td>
</tr>
</tbody>
</table>

Depicted are genes whose expression is strongly up-regulated in esophageal carcinomas in comparison to NE. From left to right: Gene symbol and description; Entrez-ID; BE fold change, fold increase of gene expression in BE vs. healthy samples; BE adjusted P-value; EAC fold change, fold increase of gene expression in esophageal adenocarcinomas vs. healthy samples; EAC adjusted P-value; ESCC fold change, fold increase of gene expression in esophageal squamous carcinoma vs. healthy samples; ESCC adjusted P-value. Note that a log2-fold change is used. P-value, significance of this difference in gene expression.

categories with FDR adjusted (Benjamini) P values <.05 were recorded (Suppl. Table 1).

This database can be queried in different ways - pursuing different aims. In order to detect cancer-associated changes but to rule out epithelium-specific changes, esophageal adenocarcinomas were compared with their precursor tissue type, BE, and squamous carcinomas were compared with healthy normal esophageal tissue, NE. In an approach to identify the most consistently changed individual genes, we searched for so-called peak genes with a fold change $>6$-fold increase/decrease between any two tissue groups.

On the one hand, BE but also EAC and ESCC samples showed an up-regulated representation of genes coding for proteins of the extracellular environment (Table 1) such as MMP1, SPP1, CTHRC1, INHBA and SULF1. Matrix metalloproteinases (MMPs) are a family of zinc-dependent proteases that target the degradation and proteolytic process of components of the extracellular matrix [20]. An elevated expression of MMP1 in esophageal cancers has been described [21]. Another gene, strongly up-regulated in both types of esophageal carcinomas, Activin, the disulfide-linked homodimer of inhibin-ßA (INHBA), is a ligand of the transforming growth factor-beta (TGF-ß) superfamily. INHBA overexpression promotes cell proliferation and may be epigenetically regulated in esophageal adenocarcinoma [22]. Overexpression of activin A in esophageal squamous cell carcinoma has been associated with advanced nodal status, clinical stage, and a worse overall prognosis [23]. Similarly, collagen triple helix repeat-containing 1 (CTHRC1) is known to be aberrantly up-regulated in most human solid tumors [24]. CTHRC1 drives cancer progression perhaps by increasing cancer cell migration [25]. High levels of PMEPA1/TMEPAI expression have been found in renal cell carcinoma, colon cancer, breast cancer, and ovarian cancer as well as in several cancer cell lines [26,27]. Genome-wide studies, which compared the gene expression levels of invasive cancer tissues with normal counterpart tissues or pre-invasive cancers, suggested that TMEPAI is one of the most highly inducible genes in invasive cancers that converts TGF-ß signaling from a tumor suppressor to a tumor promoter [28]. These lines of evidence suggest an oncogenic function of TMEPAI in many cancers. Genes strongly overexpressed in esophageal carcinomas might be helpful for diagnosis and future therapeutic strategies.

On the other hand, genes belonging to the GO group epithelium development were down-regulated in BE but also in EAC and ESCC samples when compared to healthy esophageal epithelium. Interestingly, many of the down-regulated peak genes in esophageal cancers (Table 2) belong to the so-called epidermal differentiation complex (EDC) [29]. Other genes strongly down-regulated in EAC and ESCC such as CRTC1 and SCEL belong to the GO-group cornified envelope of terminally differentiated keratinocytes. Expression of these genes is necessary for differentiation frequently turned down in carcinomas. For instance, CRNN, which was found to be significantly down-regulated in esophageal carcinomas, is a member of the EDC. The loss of CRNN expression correlates with longer tumor length, deeper tumor invasion depth, more advanced lymph node metastasis, and poorer survival [30]. Another gene belonging to the EDC is S100A9, which was also strongly down-regulated in both EAC and ESCC. In a proteomic approach, S100A9 was found to be strongly down-regulated in ESCC [31]. Interestingly, also the transglutaminase 3, which is responsible for crosslinking EDC matrix proteins was down-regulated. This means that the formation of the epidermal layer is disturbed in the carcinomas. Similarly SCEL, responsible for the assembly of proteins in the cornified envelope was strongly down-regulated. Recent results that SCEL is necessary for metastatic colorectal cancer tumor growth in the liver [32].

Interestingly, also the expression of MAL (T-cell differentiation-related gene) was significantly reduced in both EAC and ESCC (Table 2). MAL promoter hypermethylation correlating with decreased mRNA expression was reported [33] MAL is thought to be a novel suppressor gene [34] with diagnostic value in colorectal and esophageal cancers [35]. The protein has been localized to the endoplasmic reticulum of T-cells and is a candidate linker protein in T-cell signal transduction. The role of MAL for cancer immunity has not been determined yet.

Retinol metabolism was also down-regulated in both types of carcinomas, which is reflected by a strong loss of the expression of
Table 2. Peak Genes Down-Regulated in Esophageal Carcinomas

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Description</th>
<th>Entrez</th>
<th>BE logFC</th>
<th>BE adj.P</th>
<th>EAC logFC</th>
<th>EAC adj.P</th>
<th>ESCC logFC</th>
<th>ESCC adj.P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH7</td>
<td>alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide</td>
<td>131</td>
<td>-4.03</td>
<td>-6.12</td>
<td>2.28E-21</td>
<td>-4.24</td>
<td>1.98E-10</td>
<td></td>
</tr>
<tr>
<td>CRNN</td>
<td>Cornulin</td>
<td>49,860</td>
<td>-6.01</td>
<td>-9.17</td>
<td>5.23E-20</td>
<td>-9.22</td>
<td>6.37E-15</td>
<td></td>
</tr>
<tr>
<td>DHR59</td>
<td>dehydrogenase/reductase (SDR family) member 9</td>
<td>10,170</td>
<td>-1.51</td>
<td>-3.84</td>
<td>4.12E-11</td>
<td>-4.45</td>
<td>1.86E-09</td>
<td></td>
</tr>
<tr>
<td>DSC2</td>
<td>desmocollin 2</td>
<td>1824</td>
<td>-1.48</td>
<td>-3.74</td>
<td>3.27E-13</td>
<td>-4.09</td>
<td>1.51E-10</td>
<td></td>
</tr>
<tr>
<td>DSG3</td>
<td>desmoglein 3</td>
<td>1830</td>
<td>-5.04</td>
<td>-7.04</td>
<td>2.27E-19</td>
<td>-3.47</td>
<td>8.82E-06</td>
<td></td>
</tr>
<tr>
<td>EMP1</td>
<td>epithelial membrane protein 1</td>
<td>2012</td>
<td>-3.32</td>
<td>-5.68</td>
<td>8.34E-18</td>
<td>-4.67</td>
<td>2.91E-10</td>
<td></td>
</tr>
<tr>
<td>FLG</td>
<td>filaggrin</td>
<td>2312</td>
<td>-3.94</td>
<td>-5.99</td>
<td>7.96E-16</td>
<td>-5.72</td>
<td>8.26E-11</td>
<td></td>
</tr>
<tr>
<td>GRHL3</td>
<td>grainyhead-like 3 (Drosophila)</td>
<td>57,822</td>
<td>-4.56</td>
<td>-5.33</td>
<td>1.77E-20</td>
<td>-4.17</td>
<td>1.81E-11</td>
<td></td>
</tr>
<tr>
<td>IL1RN</td>
<td>interleukin 1 receptor antagonist</td>
<td>3557</td>
<td>-2.47</td>
<td>-5.51</td>
<td>5.58E-17</td>
<td>-5.20</td>
<td>1.64E-11</td>
<td></td>
</tr>
<tr>
<td>KRT4</td>
<td>keratin 4</td>
<td>3851</td>
<td>-5.36</td>
<td>-8.57</td>
<td>1.16E-15</td>
<td>-8.22</td>
<td>9.77E-11</td>
<td></td>
</tr>
<tr>
<td>MALL</td>
<td>mal, T-cell differentiation protein</td>
<td>4118</td>
<td>-4.50</td>
<td>-8.95</td>
<td>1.01E-21</td>
<td>-9.01</td>
<td>2.42E-16</td>
<td></td>
</tr>
<tr>
<td>MAL</td>
<td>mal, T-cell differentiation protein-like</td>
<td>7851</td>
<td>-1.70</td>
<td>-3.91</td>
<td>3.81E-11</td>
<td>-4.17</td>
<td>1.70E-08</td>
<td></td>
</tr>
<tr>
<td>S100A8</td>
<td>S100 calcium binding protein A8</td>
<td>6279</td>
<td>-5.18</td>
<td>-8.87</td>
<td>9.59E-15</td>
<td>-3.98</td>
<td>1.11E-03</td>
<td></td>
</tr>
<tr>
<td>S100A9</td>
<td>S100 calcium binding protein A9</td>
<td>6280</td>
<td>-5.13</td>
<td>-7.82</td>
<td>5.04E-17</td>
<td>-4.88</td>
<td>9.69E-07</td>
<td></td>
</tr>
<tr>
<td>SCEL</td>
<td>sccelin</td>
<td>8796</td>
<td>-5.39</td>
<td>-8.30</td>
<td>2.71E-19</td>
<td>-7.72</td>
<td>4.35E-13</td>
<td></td>
</tr>
<tr>
<td>SERPINB11</td>
<td>serpin peptidase inhibitor, clade B (ovalbumin), member 11</td>
<td>89,778</td>
<td>-2.69</td>
<td>-3.03</td>
<td>2.51E-17</td>
<td>-2.89</td>
<td>5.76E-12</td>
<td></td>
</tr>
<tr>
<td>SERPINB13</td>
<td>serpin peptidase inhibitor, clade B (ovalbumin), member 13</td>
<td>5275</td>
<td>-6.23</td>
<td>-8.59</td>
<td>9.12E-18</td>
<td>-5.04</td>
<td>1.51E-06</td>
<td></td>
</tr>
<tr>
<td>SERPINB2</td>
<td>serpin peptidase inhibitor, clade B (ovalbumin), member 2</td>
<td>5055</td>
<td>-5.29</td>
<td>-7.68</td>
<td>5.23E-20</td>
<td>-6.32</td>
<td>7.07E-12</td>
<td></td>
</tr>
<tr>
<td>SERPINB4</td>
<td>serpin peptidase inhibitor, clade B (ovalbumin), member 4</td>
<td>6318</td>
<td>-4.76</td>
<td>-6.95</td>
<td>5.53E-13</td>
<td>-6.15</td>
<td>7.70E-08</td>
<td></td>
</tr>
<tr>
<td>SPINK7</td>
<td>serine peptidase inhibitor, Kazal type 7 (purative)</td>
<td>84,651</td>
<td>-4.18</td>
<td>-7.40</td>
<td>2.24E-19</td>
<td>-7.23</td>
<td>6.51E-14</td>
<td></td>
</tr>
<tr>
<td>SPRR1A</td>
<td>small proline-rich protein 1A</td>
<td>6698</td>
<td>-5.76</td>
<td>-9.65</td>
<td>6.86E-16</td>
<td>-5.21</td>
<td>4.41E-05</td>
<td></td>
</tr>
<tr>
<td>SPRR1B</td>
<td>small proline-rich protein 1B</td>
<td>6699</td>
<td>-5.34</td>
<td>-8.60</td>
<td>3.64E-15</td>
<td>-3.78</td>
<td>1.19E-03</td>
<td></td>
</tr>
<tr>
<td>SPRR3</td>
<td>small proline-rich protein 3</td>
<td>6707</td>
<td>-5.33</td>
<td>-9.49</td>
<td>3.09E-17</td>
<td>-8.46</td>
<td>6.13E-11</td>
<td></td>
</tr>
<tr>
<td>TGM3</td>
<td>transglutaminase 3</td>
<td>7053</td>
<td>-5.73</td>
<td>-8.39</td>
<td>8.01E-22</td>
<td>-7.90</td>
<td>2.55E-15</td>
<td></td>
</tr>
</tbody>
</table>

Depicted are genes whose expression is strongly down-regulated in esophageal carcinomas in comparison to NE. From left to right: Gene symbol and description; Entrez-ID; BE fold change, fold decrease of gene expression in BE vs. healthy samples; EAC fold change, fold decrease of gene expression in EAC vs. healthy samples; ESCC adjusted P-value; ESCC fold change, fold decrease of gene expression in esophageal squamous carcinoma vs. healthy samples; ESCC adjusted P-value; Note that a log2-fold change is used. P-value, significance of this difference in gene expression.

ADH7 (Table 2, Figure 3). The enzyme encoded by ADH7 is a member of the alcohol dehydrogenase family. However, it is inefficient in ethanol oxidation, but active as a retinol dehydrogenase. Since retinoic acid regulates the proliferation, differentiation, and apoptosis of premalignant and malignant cells during carcinogenesis [36], it is conceivable that a disturbance of its formation leads to various disorders including malignancy. Moreover, we noted a reduced activity in genes that are related to retinol metabolism (UGT2B15, ADH7, ADH1C, DHR59, CYP4B1) within EAC compared to BE. These changes may signify a kind of dedifferentiation and indicate that cancer cells are more susceptible to vitamin A insufficiency. Indeed, recent studies indicated that loss of retinoic acid signaling is an early event during esophageal carcinogenesis [37] and that administration of retinoic acid reduces growth of esophageal squamous carcinoma cell lines [38].

In an approach to identify genes, which might be helpful for early diagnosis of esophageal adenocarcinomas, we searched for genes which were strongly up-regulated in EAC but also in BE. Genes like...
this (Table 3) often have a role for transcriptional activation of other genes such as GATA6, TOX3, PROM or are components of the cell membrane. These genes could be helpful for early diagnosis of cancer in Barrett tissues but their expression could also be increased due to the replacement of the squamous mucosa by columnar epithelium in BE and EAC. Therefore, the expression of the encoded proteins in gastrointestinal carcinomas was checked using the protein atlas database [39]. For instance, GREM1 (gremlin 1, DNA family BMP antagonist), ITTLN1 (intelectin 1) and TM4SF4 (transmembrane 4 L six family member 4) were significantly up-regulated in BE and EAC. Expression of these genes was highly homogeneous in all samples of a given tissue and therefore might be of particular value for early diagnosis. Interestingly, the encoded proteins were detected in stomach carcinomas, but not in the respective non-malignant tissues. Similarly, REG4 (Regenerating islet-derived family, member 4) was significantly up-regulated in both BE and EAC and TOX3. Colorectal cancers along with few stomach, pancreatic and ovarian cancers showed moderate to strong cytoplasmic immunoreactivity for REG4. Aberrant expression of REG4 was associated with the growth, survival, adhesion and resistance to apoptosis of tumor cells (e.g. [40]). Moreover, REG4 is considered a novel prognostic factor in some gastroenterological carcinomas [41] and was identified as a transcriptional target of GATA6 [42]. Patients suffering from esophageal cancer with an amplification of GATA6 have a poorer survival [43]. In our study, GATA6 was found to be significantly up-regulated in both BE and EAC (Table 3).

Several researchers have investigated gene expression in Barrett’s esophagus. Greenawalt et al. [44] found that SOCS3, MMP3 and MMP10 were up-regulated in ESCC and EAC, and that SERPINB2, SEPINB3 and SPINK5 were down-regulated in esophageal carcinomas, which is in accordance with our data. Interestingly, they also searched for genes up-regulated in BE and found an increased expression of TFF1, LGALS4 and GATA6 (Table 3). Botelho et al. [45] searched for gene expression using paraffin-embedded Barrett’s esophagus tissue samples. They also found down-regulation of ADH7 in esophageal carcinomas and an up-regulation of TSPAN8 in BE (Tables 2,3). Hyland et al. [46] performed gene expression analysis of Barrett’s esophagus and matched normal mucosa. Ten of the top 50 genes in their list of differentially expressed genes in BE vs. NE belong to the 25 peak genes depicted in Table 3. Warnecke-Eberz et al. [19] proposed a diagnostic signature of 19 genes for esophageal carcinomas of both subtypes that contained at least 3 genes of Table 1.

Recently, the Cancer Genome Atlas Research Network performed a comprehensive molecular analysis of esophageal carcinomas derived from Western and Eastern populations [47]. The previously detected amplifications of GATA4 and GATA6 genes were confirmed. Interestingly, an up-regulation of the p63 transcription network only in ESCC was reported. Accordingly, we found an up-regulation of TP63 in ESCC (0.67 log2fold vs. NE) in comparison to EAC (4.6 log2fold) and BE (−4.6 log2fold). However, we noted a down-regulation of the EDC genes in esophageal carcinomas in comparison to NE (Table 2). EDC genes are thought to be target genes of p63 [44].

In summary, a database of gene expression in BE and esophageal carcinomas was generated which can be queried in different ways, e.g. in order to find carcinoma-associated markers with differential expression in both esophageal squamous cell carcinomas and adenocarcinomas. Another aim is to identify a subset of markers that may allow detecting patients at risk for esophageal adenocarcinoma at a preneoplastic stage. Of course, there are many other kinds of possible queries, e.g. related to networks of genes, which are affected by the changes in differentiation during cancer development. The examples given here are intended to show that by using this database genes or groups of related genes can be identified which might be helpful for early diagnosis of esophageal carcinomas or for treatment and monitoring of this disease. The next step is to kick off a retrospective study with more patients and long follow-up in order to evaluate the changes in gene expression identified in our experimental approach.
Conclusions
Our study provides high quality data, e.g. in order to determine a subset of markers for the identification of patients at risk for esophageal adenocarcinoma at a preneoplastic stage or markers which may lead to a better disease management.

Acknowledgements
The authors thank Mrs. Sabine Grigull, Mrs. Gudrun Koch and Mrs Claudia Roefzaad for excellent technical assistance.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.tranon.2017.10.003.

References


