J Gastroenterol (2021) 56:791–807
https://doi.org/10.1007/s00535-021-01802-2

The Japanese Society of Gastroenterology

ORIGINAL ARTICLE—ALIMENTARY TRACT

Alteration of protein expression and spliceosome pathway activity during Barrett's carcinogenesis

Christoph Stingl[1] · Angela Bureo Gonzalez[2] · Coşkun Güzel[1] · Kai Yi Nadine Phoa[2] · Michail Doukas[3] · Gerben Eise Breimer[4,5] · Sybren Lodewijk Meijer[4] · Jacques Johannes Bergman[2] · Theo Marten Luider[1]

Received: 6 November 2020 / Accepted: 18 June 2021 / Published online: 5 July 2021
© The Author(s) 2021

Abstract

Background Barrett's esophagus (BE) is a known precursor lesion and the strongest risk factor for esophageal adenocarcinoma (EAC), a common and lethal type of cancer. Prediction of risk, the basis for efficient intervention, is commonly solely based on histologic examination. This approach is challenged by problems such as inter-observer variability in the face of the high heterogeneity of dysplastic tissue. Molecular markers might offer an additional way to understand the carcinogenesis and improve the diagnosis—and eventually treatment. In this study, we probed significant proteomic changes during dysplastic progression from BE into EAC.

Methods During endoscopic mucosa resection, epithelial and stromal tissue samples were collected by laser capture microdissection from 10 patients with normal BE and 13 patients with high-grade dysplastic/EAC. Samples were analyzed by mass spectrometry-based proteomic analysis. Expressed proteins were determined by label-free quantitation, and gene set enrichment was used to find differentially expressed pathways. The results were validated by immunohistochemistry for two selected key proteins (MSH6 and XPO5).

Results Comparing dysplastic/EAC to non-dysplastic BE, we found in equal volumes of epithelial tissue an overall up-regulation in terms of protein abundance and diversity, and determined a set of 226 differentially expressed proteins. Significantly higher expressions of MSH6 and XPO5 were validated orthogonally and confirmed by immunohistochemistry.

Conclusions Our results demonstrate that disease-related proteomic alterations can be determined by analyzing minute amounts of cell-type-specific collected tissue. Further analysis indicated that alterations of certain pathways associated with carcinogenesis, such as micro-RNA trafficking, DNA damage repair, and spliceosome activity, exist in dysplastic/EAC.

Keywords Barrett's esophagus · Adenocarcinoma · Laser capture microdissection · Mass spectrometry · Proteomics

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00535-021-01802-2.

✉ Christoph Stingl
c.stingl@erasmusmc.nl

[1] Department of Neurology, Erasmus University Medical Center, PO Box 20440, 3000 CA Rotterdam, The Netherlands

[2] Department of Gastroenterology and Hepatology, Amsterdam University Medical Centers, Amsterdam, The Netherlands

[3] Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands

[4] Department of Pathology, Amsterdam University Medical Centers, Amsterdam, The Netherlands

[5] Present Address: Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands

Abbreviations

AGC	Automatic gain control
APA	Alternative polyadenylation
BE	Barrett's esophagus
DSB	Double-strand breakage
EAC	Esophageal adenocarcinoma
ELISA	Enzyme-linked immunosorbent assay
EMR	Endoscopic mucosa resection
ER-cap	Endoscopic resection cap technique
ESI	Electrospray ionization
FDR	False discovery rate
FFPE	Formalin fixed and paraffin embedded

GERD	Gastroesophageal reflux disease
GO	Gene ontology
HE	Hematoxylin and eosin
HGD	High-grade dysplasia
iBAQ	Intensity-based absolute quantification
IHC	Immunohistochemistry
LC–MS	Liquid chromatography coupled to mass spectrometry
LC	Liquid chromatography
LCM	Laser capture microdissection
LFQ	Label-free quantification
LGD	Low-grade dysplasia
MALDI	Matrix-assisted laser desorption/ionization
MMR	Mismatch repair genes
MS/MS	Tandem (or fragment) mass spectrum
NDBE	Non-dysplastic Barrett's epithelium
NHEJ	Non-homologous end joining
SNP	Single-nucleotide polymorphism
TMA	Tissue micro-array
u	Atomic mass unit
UTR	Untranslated region

Introduction

In Barrett's esophagus (BE), the normal squamous lining of the lower esophagus is replaced by gastric type columnar epithelium [1]. This condition is considered a consequence of chronic gastroesophageal reflux disease (GERD). Because BE is asymptomatic, it is most commonly diagnosed by endoscopy in patients with GERD symptoms [2]. It is, therefore, difficult to assess the prevalence for the general population, and a biased group of patients undergo endoscopy because of symptoms that are not necessarily related to BE [3]. Dependent on the scope and population of a study, the reported average prevalence of histologically confirmed BE is around 1.5% (0.1–9.0%) [3–5]. BE is considered a premalignant precursor for esophageal adenocarcinoma (EAC), which might progress continuously through the sequence of low-grade dysplasia (LGD), high-grade dysplasia (HGD) and ultimately adenocarcinoma. It follows that both non-dysplastic BE and dysplastic BE are important risk factors for EAC [6]. The prognosis of EAC is poor; the 5-year survival rate is low, at 10–18% dependent on sex and ethnicity [7, 8]. EAC occurs predominately in males, with the highest rates in Western and Central Asia regions [9, 10], and is currently the sixth most frequent cancer, with the highest increase of incidence rate in the past 3 decades [11].

BE is diagnosed by the presence of endoscopically visible and histopathologically confirmed metaplasia [12]. The grade of dysplasia is strongly related to the risk of carcinogenesis [13] and defines the intensity of the required surveillance and treatment [14]. However, distinguishing between different grades of dysplasia is challenging and in the past resulted in low inter-observer agreement and variation in the assessment of risk of progression between studies [15]. As a consequence, the risk prediction of EAC solely on basis of the dysplastic grade is of limited reliability, potentially may lead to overtreatment [16]. The pathological progression from BE into EAC is associated with biological processes such as proliferation, tumor suppression, cell adhesion and inflammation. Molecules involved in these pathways might predict the development of EAC. A wide range of molecular markers have been studied, such as genomic alterations, epigenetic markers and proteins expression[17–19]: DNA copy number variations and aneuploidy have been found to be altered in EAC [20], and regions of loss of heterozygosity have been identified as promising predictive markers for EAC [21]. Gains of chromosomes 7 and 17 determined by FISH have been found correlated with the grade of oncogenic progression; the detection rate of dysplasia improved when the assessment of these gains was added to cytology [22]. EAC is characterized by a high mutational burden due to genomic micro-satellite instability compared to other cancers [23]. The predictive power of mutational load is limited for EAC, because in non-dysplastic BE (NDBE), the mutational load is also already elevated [24]. Correspondingly, gene expression studies have shown that the transcription profile of BE is more similar to that of EAC than that of normal esophagus [25]. Alterations of driver genes and frequency of genetic events have been found associated with EAC development [26]. Overexpression of p53, determined by immunohistochemistry (IHC), to predict development of EAC has been intensively studied [27]. Loss of heterozygosity in chromosome 17p is linked to inactivation of the p53 tumor suppressor gene. This inactivation of p53 was found in a higher frequency in HGD patients compared to NDBE patients, and is associated with a higher risk of progression to EAC [28]. Consequentially, p53 immunostaining has been suggested as an adjunct molecule marker for the diagnosis of dysplasia in BE [29].

Far fewer studies comparing the proteomes of BE and EAC tissue have been conducted. Zhao and co-workers compared premalignant Barrett metaplasia tissues with esophageal adenocarcinoma tissues taken from the same six patients. 2D liquid chromatography protein separation and time-of-flight mass spectrometry (MS) identified 38 differentially expressed proteins, of which 20 correlated with mRNA expression levels; and validated by IHC (3 of 3 proteins positive) [30]. Elsner and co-workers used imaging MS to determine *m*/*z* profiles of metaplastic and carcinogen tissue areas in a set of fresh-frozen samples

taken from 38 Barrett's adenocarcinoma patients. They found 22 *m/z* species that were differentially expressed and identified six of these as proteins potentially involved in tumor development and metastasis [31]. Through an LC–MS analysis of NDBE, HGD, and EAC epithelium, Zaidi and co-workers determined a diagnostic 4-protein biomarker panel that was successfully evaluated in serum by an ELISA assay on an independent cohort to discriminate between GERD and EAC patients with an accuracy of 87% [32]. O'Neill and co-workers acquired by MS-based proteomics a set of more than 6000 proteins from EAC, normal esophagus and gastric tissue samples of seven patients. Around half of the proteins quantified in tumor samples were differentially expressed, and quantification was successfully validated by IHC staining of seven proteins [33]. Despite these efforts, so far, none of the potential markers has been further developed for application in clinical practice.

In this study, we conducted an analysis specifically on the epithelial cell compartment and the surrounding stroma to determine proteomic alterations related to Barrett's carcinogenesis. Because the proportion of epithelial cells relative to all cells of a specimen as well as the proportion of dysplastic/EAC epithelial cells relative to all epithelial cells vary widely, laser capture microdissection (LCM) was chosen as an appropriate method to collect samples that are, from a microscopical perspective, sufficiently uniform in tissue volume and stage of disease [34]. Proteins were identified and quantified by label-free bottom-up proteomics using high-resolution LC–MS. Results were validated by IHC for two selected proteins. Knowledge about these proteins and the underlying functions and pathways might add another puzzle piece to the molecular mechanisms of Barrett's carcinogenesis. This addition could ultimately help to accurately predict the risk of carcinogenic progression, and thus decide on the most effective treatment and disease management.

Methods

Patient materials and characteristics

Patients were included between March 2011 and June 2015 at the Amsterdam University Medical Centers (AMC) and divided into two groups according to their histopathological diagnosis: HGD/EAC versus non-dysplastic BE. The study was approved by the medical ethics review board of the AMC (Dutch trial registration number NTR3249, https://www.trialregister.nl). Patients scheduled for EMR of BE containing HGD or early cancer were assessed for eligibility during endoscopy. Patients were excluded when the whole EMR specimen was needed for clinical decision making, when there were no visible abnormalities to target for resection, when *en bloc* resection was preferred, or when EMR was finally not performed. Eligible for inclusion in the non-dysplastic BE group were those patients with a scheduled surveillance endoscopy when no dysplasia had been found during endoscopies for at least two years previously, if no visible abnormalities in the Barrett's esophagus had been detected in the two most recent surveillance endoscopies, and when the prior biopsies had been reviewed and diagnosed as NDBE by an expert pathologist ($N = 11$). Patients were excluded if dysplasia was found in the EMR specimen taken. For all samples used in this study, written informed consent was given from the patients prior to EMR endoscopy.

Endoscopic mucosa resection

High-resolution endoscopy for the visualization of the Barrett segment was performed. The extent of columnar lined esophagus was documented according to the Prague C&M classification [35]. Lesions were described by the Paris classification and were resected piecemeal using the ER-cap-based technique, as described before [36]. In short, in this technique, the mucosa is lifted with saline after demarcation with coagulation. Using a transparent distal attachment placed on the tip of the endoscope, the lesion is pulled by suction into the cap. The trapped lesion is then enclosed by a snare loop and cut with electrocautery. In this study, lesions were lifted with saline without adrenaline to prevent any interaction with the proteomic analysis. We used the ER-cap technique and not the more commonly used multi-band mucosectomy technique since the latter causes venous congestion of the resection specimen which might affect proteomic analysis. Dysplastic patients were included only if the lesion was resected piecemeal, which is usually the case when the lesion has a > 2 cm diameter. The endoscopists assessed whether all specimens were needed for clinical evaluation, e.g., for the identification of the infiltration depth. If enough material was collected for clinical evaluation, the remaining material was used for this study. In the non-dysplastic group, a random portion of the Barrett segment was chosen for resection.

ER specimens were retrieved from the patient after resection and immediately pinned down, snap-frozen in liquid nitrogen, and stored at -80 °C. Specimens for clinical evaluation were further collected according to the hospital protocol and sent for pathological review. The snap-frozen specimens were transferred on dry ice to the Erasmus University Medical Center, Rotterdam and stored at -80 °C for proteomic analysis.

Sample preparation

Fresh-frozen EMR specimens were cut in 8 µm thick sections that were placed on PEN membrane slides (Zeiss, Göttingen, Germany), fixated with 70% ice-cold ethanol and stored at − 80 °C until further processing. Before LCM, mounted EMR sections were thawed, hematoxylin and eosin stained, and air dried. Immediately thereafter, tissue areas of 600,000 μm^2 were collected by LCM, yielding a tissue volume of approximately 4,800,000 μm^3 (corresponding to an estimated number of 4800 cells under the simplified assumption that the size of a cell corresponds to a cube with an edge length of 10 µm). Micro-sections, collected in the cap of the collection tube (Zeiss *AdhesiveCap*), were then transferred using 20 µL 0.2% aqueous Rapigest (Waters, Milford, MA, USA) into a sample tube (*Eppendorf LoBind*) and stored at − 80 °C until digestion. Next, LCM microsections were thawed, heated to 95 °C for 2 min and lysed in a sonification cell disruptor (Branson sonifier, 70% intensity). Ammonium bicarbonate was added to 50 mM (final) and the tissue lysate was reduced at 5 mM dithiothreitol (1 h at 57 °C) and afterwards alkylated at 15 mM iodoacetamide (1 h in darkness at room temperature). Samples were digested by addition of 50 ng trypsin (trypsin gold, Promega, Madison, WI, USA) and over-night incubated at 37 °C. Digests were stopped and detergent (Rapigest) hydrolyzed simultaneously by addition of 0.5% trifluoroacetic acid (final), incubation at 37 °C for 1 h and subsequently centrifuged (14,000*g*, 10 min). Finally, digests were transferred to LC vials and stored at + 4 °C until measurements. Unless otherwise noted, all reagents were purchased from Sigma Aldrich.

LC–MS acquisition

LC–MS analysis was conducted on a nano-LC system coupled to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Twenty µL (entire volume) of digest was loaded onto a trap column (C18 PepMap, 300 µm ID × 5 mm, 5 µm, 100 Å; Thermo Fisher Scientific) and desalted for 10 min using 0.1% trifluoroacetic acid at a flow rate of 20 µL/min. Subsequently, the trap column was switched in-line with the analytical column (PepMap C18, 75 µm ID × 500 mm, 3 µm, 100 Å) and peptides were eluted using a binary 90′ gradient increasing solvent B from 4 to 38%, whereby solvent A was 0.1% formic acid, solvent B 80% acetonitrile and 0.08% formic acid, flow rate 300 nL/min and column temperature 40 °C. For electrospray ionization, nano ESI emitters (New Objective, Woburn, MA, USA) were used and a spray voltage of 1.7 kV applied. A data-dependent acquisition MS method was used with an Orbitrap survey scan (range 375–1500 *m/z*, resolution of 120,000, AGC target 400,000), followed by consecutively isolation, fragmentation (HCD, 35% NCE) and detection (ion trap, AGC 10,000) of the peptide precursors detected in the survey scan until a duty cycle time of 3 s was exceeded ('Top Speed' method). Precursor masses that were selected once for MS/MS were excluded for subsequent fragmentation for 60 s.

Samples of each cell type and EMR specimen were prepared and analyzed in duplicate ($N = 92$ runs, of 23 samples × 2 cell types × 2 replicates). A total of 91 measurements were successfully completed (1 failed for a technical reason). The sample set was split in two parts according to the cell type (epithelial or stromal) and both sets were subsequently analyzed independently of each other. Acquired data have been made publicly available through the ProteomeXchange Consortium using the PRIDE identifier PXD020903 [37].

Protein identification and quantification

Protein identification and label-free quantification (LFQ) was carried out, separately for epithelial and stromal samples, by the quantitative proteomics software package MaxQuant [38, 39] (version 1.6.1), using the internal search engine Andromeda [40] applying the following settings: human (*Homo Sapiens*) subset of the uniprot swissprot database (20,194 entries; version: 12. November 2015), carbamidomethylation (+ 57.021 u) of cysteine as fixed modification, oxidation (+ 15.995 u) of methionine, proline and lysine and protein N-terminal acetylation (+ 42.0106 u) as variable modification, tryptic cleavage allowing two miscleavages, 10 ppm precursor tolerance, 0.5 u fragment tolerance, and ESI-trap as instrument type. For the label-free quantification, the parameter multiplicity was set to 1, label-free quantification set to LFQ, and calculation of iBAQ values activated; otherwise the default settings were used. Next, we combined results of all samples by cell type, applied filtering of identification (local protein false discovery rate, $FDR < 1\%$, local peptide $FDR < 0.1\%$, minimum 2 peptide/protein identified) and conducted protein grouping using the software package Scaffold (Proteome Software, version 4.10, batch Q +). Protein-sample table containing protein abundances (iBAQ values) and spectrum reports were exported and used for further data analysis.

Statistical calculations and analysis of differentially expressed proteins were carried out with the statistical software package R [41]. First, iBAQ values of the individual runs were aggregated by sample, then ^{2}log -transformed and the missing values replaced, by a sample-specific zero imputation value calculated as abundance mean minus 4 standard deviations [42]. Distributions of protein abundances were tested for normality with the

Shapiro–Wilk test (normality indicated by a $P > 0.1$). Because abundances were normally distributed for just 24% of the proteins (786 of 3226), we concluded that in general the requirements for parametric tests were not fulfilled. Therefore, we used the non-parametric Wilcoxon rank-sum test in combination with Benjamini–Hochberg correction for multiple hypothesis testing to find potentially significant different protein abundances between dysplastic/EAC and non-dysplastic specimen. All proteins with an $FDR < 5\%$ were reported as differentially expressed proteins.

Gene set enrichment analysis

Gene set enrichment analysis was performed on the set of differentially expressed proteins ($FDR < 5\%$) queried against the protein–protein interaction (PPI) database *STRING* (https://string-db.org). This database contains pathway annotations from *KEGG* (https://www.genome.jp/kegg/) and *Reactome* (https://reactome.org/). Analysis was performed using the software *Cytoscape* (v. 3.7.2) [43]. To query the PPI network and to conduct gene set enrichment, we used the add-in Cytoscape StringApp (v 1.5.1.) [44] applying a confidence cut-off of 0.4, no additional interactors and the set of all identified and quantified proteins served as reference gene set to assess the statistical background. Prior to conducting functional enrichment, the PPI network was clustered by the interaction strength applying MCL clustering with granularity set to 2.0 using the add-in ClusterMaker [45]; subsequently functional enrichment was carried out on the four largest clusters. Pathways (*KEGG* and *Reactome*) with an $FDR < 0.05\%$ were exported and used for interpretation of the data. For analysis of functional similar proteins and PPI of the spliceosome-related proteins we used the software and database of *GeneMANIA* [46, 47] (through the Cytoscape App *GeneMANIA* [48], version 3.5.2; H. Sapiens data set, version 2021-04-29-core). In a first analysis, a PPI search was conducted using the 19 significantly up-regulated and spliceosome-related gene products (Table 1: genes of

Table 1 Specimen characteristics and results of pathological diagnosis of EMR specimen at different phases of the study and parts of the specimen

Specimen ID	Sex	Age	Diagnosis of patient	Diagnosis EMR, FF half	Diagnosis EMR, FFPE half	Category for statistics
ER081	M	67.0	EAC	EAC	LGD	Dysplasia/EAC
ER084	M	67.8	EAC	HGD	LGD	Dysplasia/EAC
ER086	M	62.1	EAC	LGD	LGD	Dysplasia/EAC
ER090	M	66.8	EAC	LGD	EAC	Dysplasia/EAC
ER096	M	51.6	EAC	EAC	EAC	Dysplasia/EAC
ER097	F	67.6	EAC	HGD	EAC	Dysplasia/EAC
ER102	M	67.0	EAC	LGD	n.a	Dysplasia/EAC
ER108	M	67.0	EAC	EAC	EAC	Dysplasia/EAC
ER082	M	84.7	HGD	HGD	HGD	Dysplasia/EAC
ER093	M	53.2	HGD	NDBE	n.a	Dysplasia/EAC
ER094	M	71.1	HGD	LGD	HGD	Dysplasia/EAC
ER103	M	65.0	HGD	LGD	LGD	Dysplasia/EAC
ER106	M	82.0	HGD	LGD	LGD	Dysplasia/EAC
ER083	M	71.7	NDBE	NDBE	n.a	NDBE
ER085	M	54.0	NDBE	NDBE	n.a	NDBE
ER087	M	62.2	NDBE	NDBE	n.a	NDBE
ER088	M	69.3	NDBE	NDBE	n.a	NDBE
ER089	M	80.8	NDBE	NDBE	n.a	NDBE
ER095	M	74.2	NDBE	NDBE	n.a	NDBE
ER098	M	62.2	NDBE	NDBE	n.a	NDBE
ER099	M	60.6	NDBE	NDBE	n.a	NDBE
ER104	M	58.6	NDBE	NDBE	n.a	NDBE
ER105	M	59.7	NDBE	NDBE	n.a	NDBE

Sex: *F* female, *M* male; age: age at day the sample was resected; diagnosis of patient: stage on the basis of worst pathological diagnosis; *diagnosis EMR, FF half* stage on the basis of fresh-frozen half of EMR specimen, *diagnosis EMR, FFPE half*: stage on the basis of formalin-fixed paraffin-embedded half of EMR specimen

pathway HSA-72163) to determine the top 20 related genes. For the second analysis, the set of all significantly up-regulated proteins (Supplementary Table S1) was used without allowing inclusion of related genes.

IHC validation

To evaluate the results of the discovery proteomics study, we performed IHC in a set of 23 formalin-fixed and paraffin-embedded (FFPE) tissue samples obtained by EMR with antibodies specific for MSH6 (1:100 diluted; AC-0047EUA, Epitomics) and XPO5 (1:400 diluted; HPA018402, Atlas Antibodies). A tissue micro-array (TMA) with 2 mm cores was prepared for 17 EMR specimens and tissue sections from 6 additional biopsy samples were mounted individually on glass slides. Stained slides were scanned and images acquired were loaded into the digital pathology software pathXL (Philips) for review and scoring by three expert pathologists. Intensity and frequency of nuclear staining of MSH6 and XPO5 and cytoplasmic staining of XPO5 were scored, and the IHC score was computed as the sum of the products of intensity and frequency of each intensity level as follows:

$$\text{score (IHC)} = \sum_{I=0}^{3} F \times I,$$

whereby the intensity (I) rated the staining intensity from 0 to 3 (0 = negative, 1 = weak, 2 = moderate, and 3 = intensive) and frequency (F) described the proportion of epithelial cells for each intensity level (0–3). Significances between IHC scores of NDBE and dysplastic/EAC specimen were calculated by Wilcoxon rank-sum test.

Results

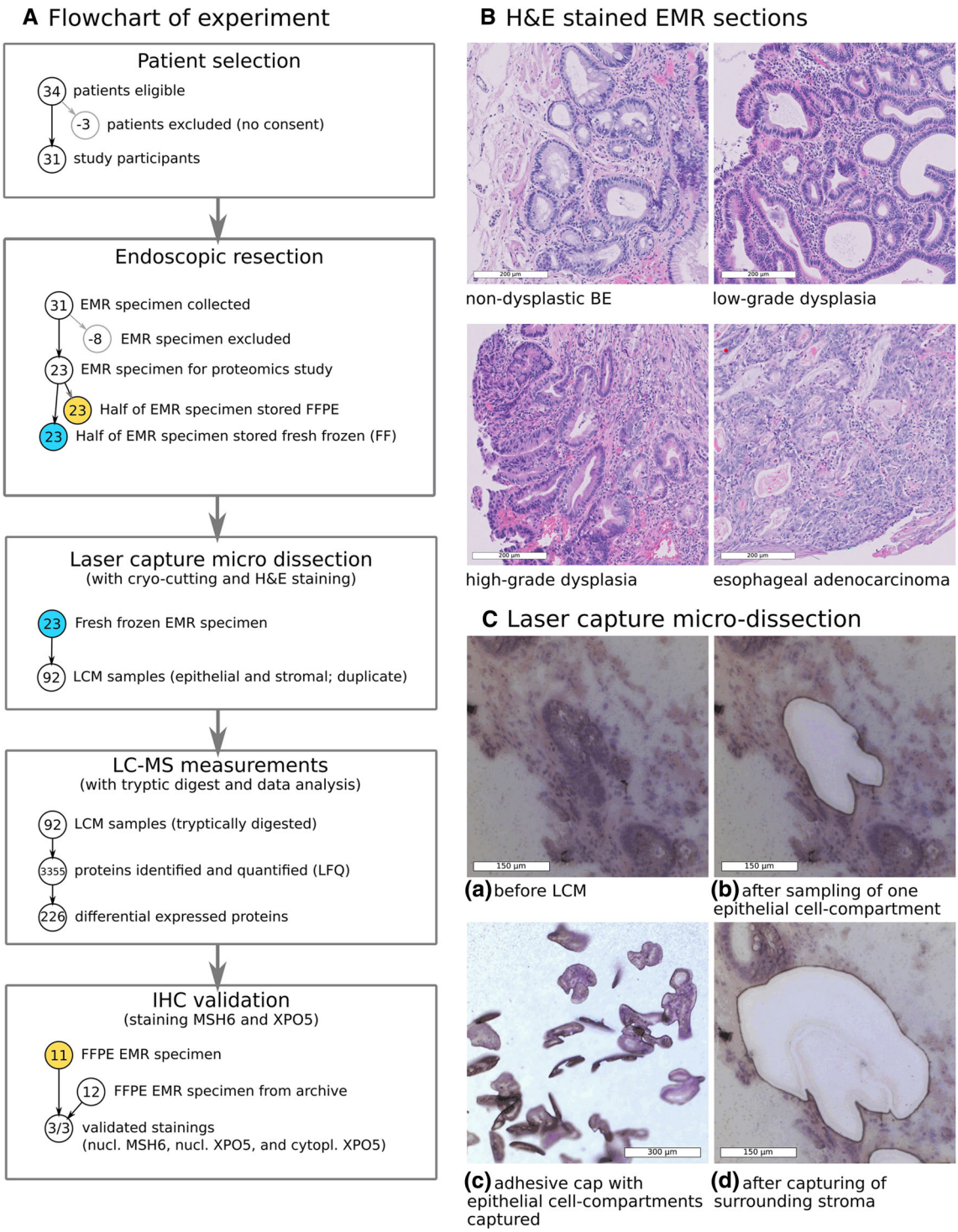
Characteristics of the sample collection and specimen

Specimen from 11 of the initial 34 patients were excluded for the following reasons: no consent given (3), no endoscopic resection due to submucosal growth (2), small lesions for which the whole specimen was needed for clinical evaluation (3), another endoscopic treatment (radio-frequency ablation) was used instead of EMR (1), absence of dysplasia in a specimen taken from a dysplastic esophagus (1), or presence of dysplasia in a specimen taken without prior analysis of dysplasia (1). Hence, a total of 23 specimens—13 dysplasia/EAC and 10 NDBE—were used for analysis (Fig. 1). Among the patients with dysplasia/EAC, HGD was diagnosed in five cases and EAC in eight cases as the most advanced stage. Most patients were male, with only one female in the dysplastic/EAC group. Median age was comparable between groups (NDBE = 62.2 years; HGD/EAC = 67.0 years; $P = 0.64$). The median Barrett length was C3M4 for the non-dysplastic group and C3M6 for the dysplastic/EAC group.

In the course of the discovery experiment, grade of dysplasia/EAC was determined on basis of the section that was cut from the fresh-frozen EMR specimen and used for LCM. Thereby, three specimens were diagnosed as EAC, three as HGD, and six as LGD. In one sample from a patient with dysplastic BE no dysplastic tissue was found. This sample was kept in the study, but was excluded from statistical comparison between dysplastic and non-dysplastic samples. Sections of the FFPE halves of the EMR specimen used for the IHC validation experiment were reviewed and graded as well by an expert pathologist. Four EMR sections were graded as EAC, two sections as HGD and five sections as LGD (Table 1). Twelve samples of the set of FFPE halves of the initial set of EMR specimen were not available for IHC validation, and were replaced by an additional 10 non-dysplastic and two dysplastic/EAC specimens.

Differential protein quantification

In epithelial samples, we quantified a total of 4059 proteins. In dysplastic/EAC we quantified 13% ($P = 0.01$) more proteins with a 48% higher total abundance ($P = 0.04$) than in non-dysplastic samples. In stromal cells, a total of 2409 proteins were quantified; the total protein abundances of dysplastic/EAC samples was not higher than that of the non-dysplastic samples. The numbers of quantified proteins and protein groups and the related responsibilities are detailed in the Supplemental Figures S1 and S2. For statistical analysis of epithelial samples, we used 3226 proteins out of these 4059 quantified proteins that were present (quantified) in at least seven samples (> 30% of samples). By unsupervised principal component analysis (PCA, Fig. 2) and unsupervised hierarchical clustering (Supplemental Figure S4A), protein profiles clustered primarily by the disease stage (non-dysplastic samples and dysplastic/EAC samples), except one LGD sample and the sample from a NDBE EMR specimen of the patient with dysplastic BE that clustered closer to non-dysplastic samples (Supplemental Figures S4A and S4B). As a result, we found 226 differentially expressed proteins ($FDR < 5\%$), of which 209 were up-regulated and 17 were down-regulated in dysplastic/EAC samples (Supplemental Tables S1 and S2). In stromal samples, ratio of fold-change and significance of change between non-dysplastic and dysplastic/EAC stromal samples were calculated for 1778 proteins with a minimum occurrence of 7 samples out of the total of 2409 proteins quantified. Unsupervised hierarchical


A Flowchart of experiment
Patient selection
34 patients eligible
-3 patients excluded (no consent)
31 study participants
Endoscopic resection
31 EMR specimen collected
-8 EMR specimen excluded
23 EMR specimen for proteomics study
23 Half of EMR specimen stored FFPE
23 Half of EMR specimen stored fresh frozen (FF)
Laser capture micro dissection
(with cryo-cutting and H&E staining)
23 Fresh frozen EMR specimen
92 LCM samples (epithelial and stromal; duplicate)
LC-MS measurements
(with tryptic digest and data analysis)
92 LCM samples (tryptically digested)
3355 proteins identified and quantified (LFQ)
226 differential expressed proteins
IHC validation
(staining MSH6 and XPO5)
11 FFPE EMR specimen
12 FFPE EMR specimen from archive
3/3 validated stainings
(nucl. MSH6, nucl. XPO5, and cytopl. XPO5)
B H&E stained EMR sections
non-dysplastic BE
low-grade dysplasia
high-grade dysplasia
esophageal adenocarcinoma
C Laser capture micro-dissection
(a) before LCM
(b) after sampling of one epithelial cell-compartment
(c) adhesive cap with epithelial cell-compartments captured
(d) after capturing of surrounding stroma

◀**Fig. 1** **A** Flowchart of sample collection, discovery and validation experiment. Thirty-one out of 34 initially eligible patients participated and underwent the surgical procedure, and EMR specimens of finally 23 patients could be included in the proteomics discovery experiment (samples were excluded when, e.g., the EMR was needed for clinical validation, when no EMR was taken because of submucosal growth or other endoscopic treatment). The specimens (13 dysplastic/EAC and 10 non-dysplastic) were immediately split into two halves. One half was snap-frozen (FF) for the proteomics discovery experiment, while the other half was formalin fixed and paraffin embedded (FFPE) and used for clinical evaluation. Next, in the discovery experiment, volumes of around 4.8 million μm^3 epithelial and stromal tissue were captured by LCM of each sample in duplicate. Samples were tryptically digested, measured by LC–MS, and quantitative protein profiles were determined and compared. For the validation experiment, 11 dysplastic samples of the discovery experiment and 12 additional samples (2 dysplastic and 10 non-dysplastic) were used to score the abundances of MSH6 and XPO5 by IHC staining. **B** Representative scans of Hematoxylin–Eosin-stained EMR sections of NDBE, LGD, HGD and EAC tissue (scale bar corresponds to 200 µm). **C** Images taken during LCM showing tissue before LCM, after sampling of one epithelial compartment (micro section), all microsections of one sample collected in the adhesive cap of a sampling vial, and the tissue section after capturing of the surrounding stroma

clustering and unsupervised PCA did not show formation of any distinct clusters, and no significant fold-change of protein abundance passed the *FDR* filter criteria of < 5% (Supplemental Figure S4C and Figure S5). Because of the high FDR of quantitative differences of proteins in stroma, we did not conduct further analysis on that part of the dataset. A list of all proteins identified in epithelial and stromal samples is available as supplementary data (Table S3).

Gene set enrichment analysis

Gene set enrichment was carried out by querying a protein-interaction network on the basis of the 226 differential abundant proteins, further clustered by the functional interaction subnetworks which were subjected to gene set enrichment analysis using all quantified proteins ($N = 3226$) as a background reference set. We found 12 *Reactome* pathways and 6 KEGG pathways that passed the filter criteria of 5% FDR, at least five matching proteins and a minimum 5% pathway coverage (Table 2). The most significant enrichment was found for the *Reactome* pathways *mRNA Splicing—Major Pathway* (HSA-72163, 19 significant genes overlap) and the superordinated pathways *Processing of Capped Intron-Containing Pre-mRNA* (HSA-72203, 20 genes; Fig. 3 and supplemental Figure S6) and *Metabolism of RNA* (HSA-8953854, 29 genes) as well as for the KEGG pathway *Spliceosome* (map03040, 11 genes). Next, for the 19 up-regulated proteins of the splicing pathway, we determine a set of another 20 proteins based on their known and expected protein–protein interactions (*GeneMANIA* search). Interestingly, 19 of these 20 proteins were identified but were not included in the set of differentially expressed proteins because the significance of these proteins did not meet the threshold of FDR < 5%. However, when these proteins were examined individually with less stringent filtering criteria ($P < 0.05$; corresponding to an FDR < 15%), 14 of the 19 proteins passed this reduced confidence threshold (Supplementary Table S4). This accumulation of less significant proteins is nevertheless highly significant ($P < 0.0001$) and is supporting our findings. Also consistent with results of the pathway enrichment analysis, the analysis of GO term enrichment based on all up-regulated proteins also revealed that the most significantly enriched terms were associated with splicing and spliceosome-related processes (Supplementary Table S5).

Validation by immunohistochemistry

For technical orthogonal validation of the discovery study, we performed an IHC staining for MSH6 (discovery study: *FDR* = 0.03, ^{2}log fold-change = 4.72 up-regulated; Supplemental Table S1) and XPO5 (*FDR* = 0.03, ^{2}log fold-change = 5.71 up-regulated, Supplemental Table S1) in a set of 23 FFPE samples. Staining of nuclear MSH6, nuclear XPO5 and cytoplasmic XPO5 was present in all NDBE samples (median IHC scores: nuclear MSH6 = 2.0, nuclear XPO5 = 1.6, and cytoplasmic XPO5 = 1.0) and was increased about 0.44–0.85 score points in dysplastic/EAC samples (nuclear MSH6 = 2.5, nuclear XPO5: 2.2, and cytoplasmic XPO5 = 1.9; Figs. 4 and 5; Supplemental Figure S7). Overall, the increase of IHC staining was significant when mean scores of all three pathologists were used, but also, with one exception (cytoplasmic XPO5 by one reviewer, $P = 0.058$), on the basis of the three individual reviews (Supplemental Figure S8A). Moderate to mainly strong correlations were found between the reviewers, with correlation coefficients ranging from 0.66 to 0.91 for the final review (Supplemental Figure S8B).

Discussion

Tissue samples with different pathologic grades of the progression sequence from non-dysplastic Barrett's epithelium, dysplasia and ultimately Barrett's adenocarcinoma are characterized by high heterogeneity, and apparently non-dysplastic cells can be found in close spatial proximity to dysplastic and cancerous cells. In an earlier study, we used biopsies taken from non-dysplastic tissue to successfully analyze the proteomes of epithelial and surrounding stromal cells [34]. In the present study, however,

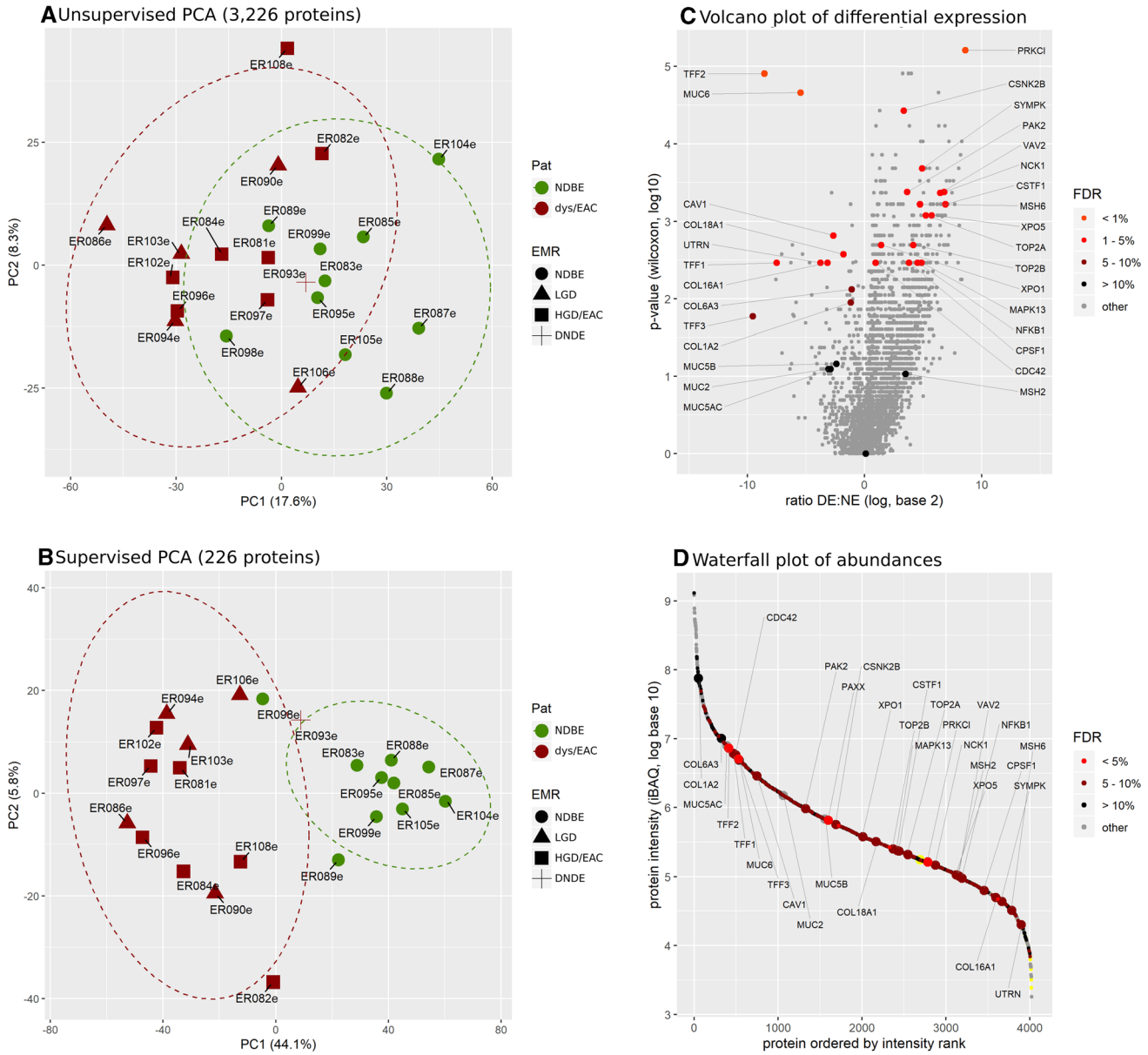


Fig. 2 Results of protein quantification and differential quantitative analysis. Unsupervised principal component analysis (PCA) on 23 epithelial samples on the basis of 3226 quantified proteins (**A**) and supervised PCA of the set of 226 significant differentially expressed proteins (**B**). Icon colors label for dysplastic (red) and non-dysplastic (green) patients, and the icon shape refer to the stage of dysplasia assessed during the LCM experiment. Volcano plot (scatter plot of fold-change between dysplastic and non-dysplastic samples versus significance of fold-change) of differential quantitative analysis comparing dysplastic and non-dysplastic samples; colors of dots indicate the FDR of the hit (Benjamini–Hochberg correction); in total 226 proteins were found significantly up-regulated with an *FDR* < 5%, of which 209 in dysplastic samples (**C**). Scatter plot of protein abundance (iBAQ) vs intensity-based rank of protein (*waterfall plot*); red dots indicate significantly differentially expressed proteins. An abundance range of about 6 magnitudes of order is covered, and significant hits were found over almost the entire abundance range (**D**)

using biopsies from dysplastic/EAC tissue did not prove to be a viable option for this analysis, because in the majority of cases not enough cells of the selected type and disease stage could be found. Moreover, making a clear histopathological diagnosis on the basis of a single fresh-frozen biopsy was often not possible. Therefore, for the sake of pathohistological confidence, we used fresh-frozen EMR specimen obtained by ER-cap resection as sample type to obtain biological specimen with high biological and clinical fidelity. As a consequence, sample sizes were small, but on the other hand, fewer samples had to be excluded for the reason of insufficient tissue with the targeted cell type or uncertain stage of dysplasia. Nevertheless, we still found different stages of dysplasia/EAC

Table 2 List of significantly enriched pathways (Reactome and KEGG) determined by String gene set enrichment analysis

Nr.	Pathway name	Src	ID	*n*	*N*	*S*	FDR	Csig%	Cid%	Genes
1	mRNA splicing—major pathway	R	HSA-72163	19	151	180	9.05e-09	12.6	83.9	CSTF1, PRPF19, SYMPK, DNAJC8, HNRNPU, SF3B3, HNRNPD, HNRNPM, DHX15, HNRNPA1, HNRNPH1, CTNNBL1, DHX9, HNRNPR, CRNKL1, PUF60, RBM8A, SRRT, CPSF1
2	Processing of capped Intron-containing Pre-mRNA	R	HSA-72203	20	179	244	9.05e-09	11.2	73.4	CSTF1, PRPF19, SYMPK, DNAJC8, HNRNPU, SF3B3, HNRNPD, HNRNPM, DHX15, HNRNPA1, HNRNPH1, CTNNBL1, DHX9, HNRNPR, CRNKL1, PUF60, ZC3H11A, RBM8A, SRRT, CPSF1
3	Metabolism of RNA	R	HSA-8953854	29	397	721	9.05e-09	7.3	55.1	CSTF1, PRPF19, WDR77, SYMPK, DNAJC8, NSUN2, HNRNPU, SF3B3, HNRNPD, NCL, HNRNPM, DHX15, HNRNPA1, HNRNPH1, CTNNBL1, DHX9, ADAR, HNRNPR, XRN2, CRNKL1, XPO1, ANP32A, PUF60, ZC3H11A, RBM8A, SUPT5H, DDX6, SRRT, CPSF1
4	T cell receptor signaling pathway	K	map04660	6	22	86	3.71e-06	27.3	25.6	MAPK13, NFKB1, PAK2, VAV2, CDC42, NCK1
5	VEGFA–VEGFR2 Pathway	R	HSA-4420097	6	37	95	0.00012	16.2	38.9	MAPK13, PAK2, CAV1, VAV2, CDC42, NCK1
6	Spliceosome	K	map03040	11	103	122	0.00014	10.7	84.4	PRPF19, HNRNPU, TCERG1, SF3B3, HNRNPM, DHX15, HNRNPA1, CTNNBL1, CRNKL1, PUF60, RBM8A
7	Proteoglycans in cancer	K	map05205	6	65	165	0.00057	9.2	39.4	MAPK13, STAT3, ARHGEF1, CAV1, VAV2, CDC42
8	TCR signaling	R	HSA-202403	6	64	126	0.00078	9.4	50.8	NFKB1, PAK2, PSMF1, PSMB10, NCK1, PSMB3
9	Fc epsilon receptor (FCERI) signaling	R	HSA-2454202	6	63	210	0.00078	9.5	30.0	NFKB1, PAK2, PSMF1, PSMB10, VAV2, PSMB3
10	Signaling by Interleukins	R	HSA-449147	8	142	452	0.00078	5.6	31.4	NFKB1, STAT3, PAK2, PSMF1, PSMB10, LCP1, CDC42, PSMB3
11	Leukocyte transendothelial migration	K	map04670	5	43	75	0.00081	11.6	57.3	MAPK13, F11R, VAV2, MLLT4, CDC42
12	Rap1 signaling pathway	K	map04015	5	46	162	0.00083	10.9	28.4	MAPK13, PRKCI, VAV2, MLLT4, CDC42
13	Tight junction	K	map04530	5	62	101	0.0018	8.1	61.4	CGN, PRKCI, F11R, MLLT4, CDC42
14	MAPK6/MAPK4 signaling	R	HSA-5687128	5	55	94	0.0026	9.1	58.5	PAK2, PSMF1, PSMB10, CDC42, PSMB3
15	Apoptosis	R	HSA-109581	6	99	179	0.0035	6.1	55.3	STAT3, PAK2, PSMF1, PSMB10, DBNL, PSMB3
16	Interleukin-1 family signaling	R	HSA-446652	5	64	139	0.0035	7.8	46.0	NFKB1, STAT3, PSMF1, PSMB10, PSMB3
17	Signaling by the B Cell Receptor	R	HSA-983705	5	60	175	0.0035	8.3	34.3	NFKB1, PSMF1, PSMB10, NCK1, PSMB3
18	C-type lectin receptors (CLRs)	R	HSA-5621481	5	71	144	0.0037	7.0	49.3	NFKB1, PAK2, PSMF1, PSMB10, PSMB3

Source = *Reactome* (R) or KEGG (K); *n* = number of significantly differentially expressed genes matching to the pathway; *N* = number of pathway-related genes products identified (used as background set); *S* size of pathway in terms of total number of genes linked to pathway (*Reactome* or KEGG). *FDR* false discovery rate of enrichment, *Csig%* coverage of identified set of pathway genes by significantly expressed genes (*n*/*N*), *Cid%* coverage of all pathway genes by identified genes (*N*/*S*)

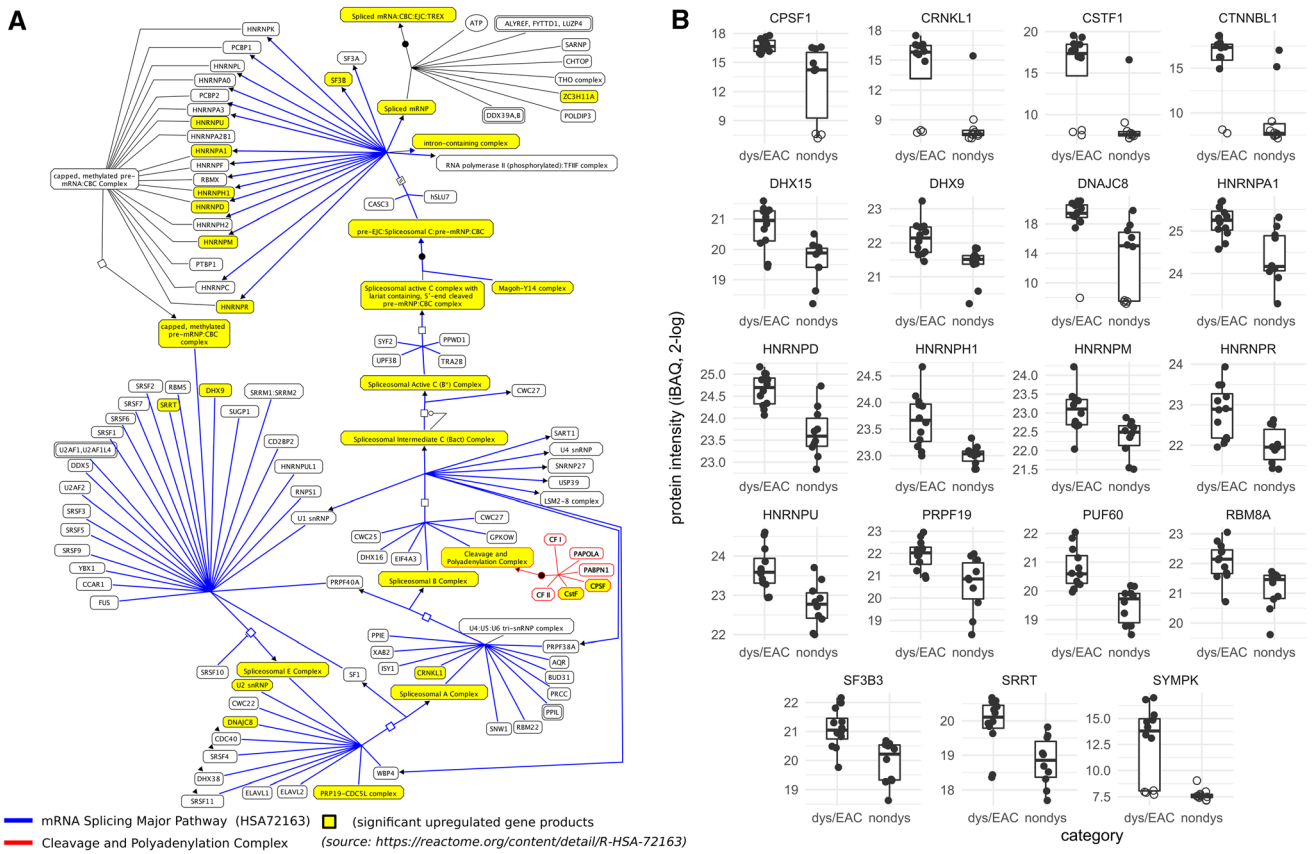


Fig. 3 **A** Significantly up-regulated proteins in the mRNA Splicing Major Pathway (pathway source: Reactome https://reactome.org/content/detail/R-HSA-72163, pathway plot generated by Cytoscape using the ReactomeFI plugin and manually simplified and annotated; pathways with full details in Figure S6) and **B** box-plots of protein intensity grouped by disease stage of nineteen differentially expressed proteins of the mRNA Splicing Major Pathway

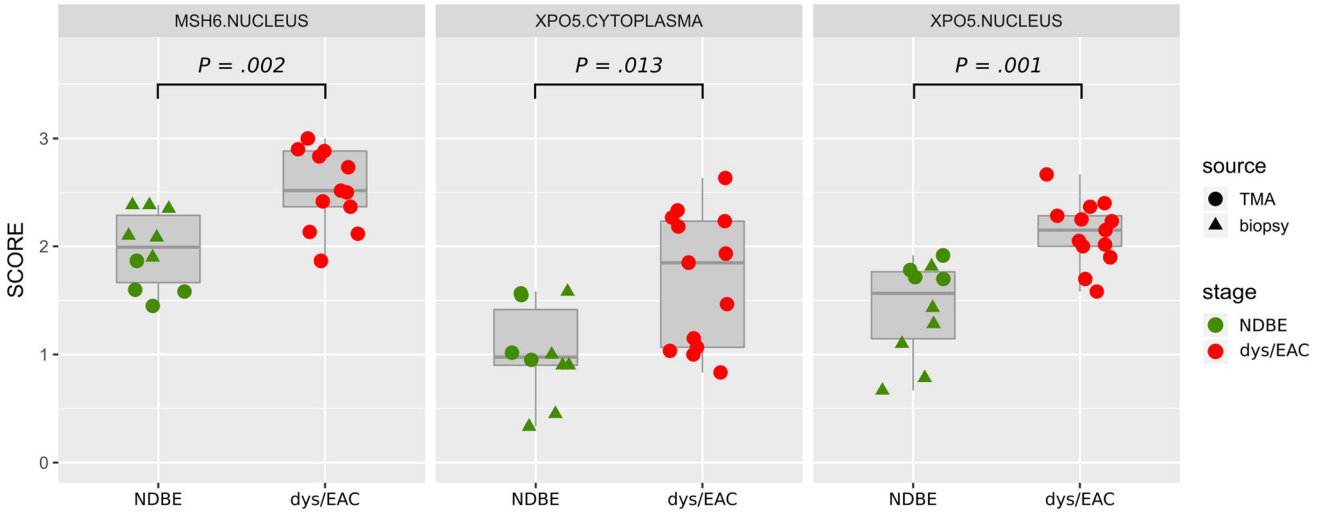


Fig. 4 Results of technical IHC validation. IHC scores of nuclear MSH6, nuclear XPO5 and cytoplasmic XPO5 in esophageal tissue from patients diagnosed for NDBE (non-dysplastic) and dysplasia/EAC. IHC scores are the mean scores of all three reviewers; see supplemental Fig. 8 for individual scores. Significant differences (determined by Wilcoxon rank-sum test) were found in all three cases (nuclear MSH6: NDBE = 2.0 vs dys/EAC = 2.6; $P = .016$; cytoplasmic XPO5: 1.0 vs 1.8; $P = .046$; nuclear XPO5: 1.6 vs 2.0; $P = .010$, respectively) and confirmed the results of the proteomics discovery study. Icon shape indicates whether specimens were TMA cores or whole biopsies

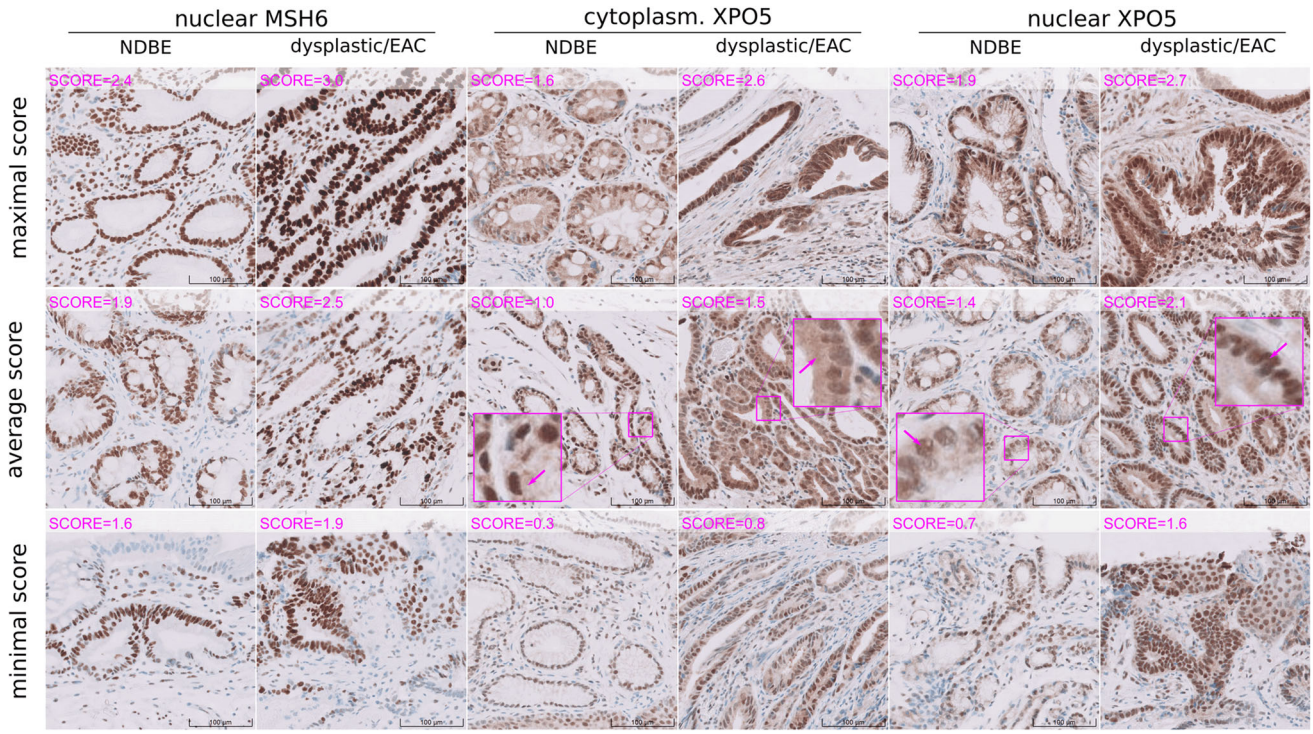


Fig. 5 IHC staining. IHC of non-dysplastic (non-dys.) and dysplastic/EAC (dys.) samples stained for MSH6 and XPO5. Tissues were scored for nuclear MSH6, nuclear XPO5 and cytoplasmic XPO5. The sample with, respectively, the highest, middle and lowest IHC score (mean value of three reviews) is shown in the various panels. Close-ups indicating cytoplasmic and nuclear XPO5 staining (arrows). (20× magnification, scale bar corresponds to 100 µm)

between different specimens of an individual but also within the same specimens, and staging coincided just partially between samples. This heterogeneity of dysplastic tissue in samples that are spatially very close to each other is probably a substantial factor for the common disagreement on grading of dysplasia reported in literature [49–52]. For the above reasons, during the statistical analysis, we were constrained to categorize solely NDBE and dysplastic/EAC tissue, without further differentiating between the grades of dysplasia.

When comparing the protein abundances between non-dysplastic and dysplastic/EAC samples, we found more different protein identifications and a higher total protein abundance in dysplastic/EAC epithelial tissue; interestingly, this observation did not hold for stromal tissue. Because same volumes of epithelial tissue were collected for both non-dysplastic and dysplastic/EAC cells, higher protein abundance is likely to be a result of the higher density of the dysplastic epithelial compartment, probably due to the absence of goblet cells and abnormal cellular organization such as atypic and crowded nuclei, and irregular compacted growth. Because for statistically analyses normalized abundances were used that corrected for variation due to overall differences, the higher number of differentially up-regulated proteins could not be exclusively explained by an overall difference of total protein abundance. The increased number and abundance of proteins are therefore probably mainly related to the higher heterogeneity of dysplastic/EAC tissue. An overall up-regulation in number of proteins was also shown in earlier studies that used LCM sampling to investigate malignant epithelial cells in breast cancer [53] and cervix carcinoma [54]. The up-regulation in these two studies as well might be explained by the possibility of a higher density of tumor cells.

In the present study, up-regulation of proteins in dysplastic samples did not apply uniformly to all types of proteins. Secreted mucins and mucin-associated trefoil factors had lower abundance in dysplastic/EAC tissue compared to non-dysplastic tissue. This group of proteins has characteristic expression patterns that in general decline during the progression from BE into EAC. The group includes MUC2, characteristically secreted by goblet cells; gastric MUC5AC, expressed at the surface epithelium and the submucosal glands; and MUC6 and MUC5B, found inside the glands [55–57]. Associated to mucins, too, is the family of trefoil factors, including the gastric tumor suppressors TFF1 and TTF2, which are co-localized with MUC5AC and MUC6, respectively; and TFF3, which is typically not secreted by gastric mucosa but, like MUC2, by goblet cells. Trefoil factors are essential in mucosal protection and repair, and decreased expression is

associated with increased risk of dysplastic progression [58–60]. The presence of TFF3 in samples taken by *Cytosponges* [61] further allows specific and sensitive diagnosis of BE [62, 63]. Our results are consistent with the expected expression profile, because we found MUC6, TFF1 and TFF2 significantly down-regulated in dysplastic samples. In contrast, significant abundance differences between dysplastic and non-dysplastic samples were not found for MUC2, MUC5AC, MUC5B and TFF3. The probable reason for this latter finding is that LGD samples, from the group of dysplastic samples, had abundancies comparable to those of the non-dysplastic group. These abundance patterns of mucins and TFFs reflect the origin of secretion [59]. TFF2 and MUC6, secreted from mucous neck cells of fundic glands, show reduced expression already in LGD tissue. MUC5AC and TFF1, which are expressed in gastric mucosa cells, were similarly expressed in LGD and non-dysplastic tissue. The extent of expressions of MUC2 and TFF3, which are expressed by goblet cells, in LGD samples was between that of non-dysplastic and dysplastic/EAC tissue. However, the detection of mucins and TFFs in dysplastic/EAC samples indicates moreover that the samples collected were heterogeneous in terms of dysplasia and EAC. Collagens as well had in general a lower expression in dysplastic/EAC epithelium. Still, two of these, the endostatin precursor and angiogenesis inhibitor COL18A1 [64] and COL16A1, were significantly lower expressed in dysplastic/EAC tissue. This latter finding may be related to the proportional reduction of extra-cellular matrix surrounding epithelial cells in response to the dysplastic growth of the epithelial compartment.

A group of up-regulated proteins—PAXX, TOP2A, TOP2B, and MSH6—are involved in the stimulation of cellular response to DNA damage. PAXX is executing ligation in damage repair as response to double-strand breakage (DSB), mediated by TOP2A [65, 66]. There is further evidence that mismatch repair (MMR) genes are involved in DSB repair [66, 67] and that MSH6 regulates NHEJ activity by interaction with Ku70 [68]. Up-regulation of MSH6 and other MMR genes, such as MSH2 and MLH, have also been reported in various types of cancer [67, 69]. Mutations of MMR genes causes micro-satellite instability, which in turn leads to increased mutation rates that can ultimately lead to cancer. Still, microsatellite instability is less common in BE-associated EAC [70, 71]. In the present study, the elevated levels of MSH6 determined in the discovery experiment were confirmed by significantly increased MSH6 IHC staining in dysplasia/EAC.

Exportin-5 (XPO5) transports micro-RNAs and proteins from the nucleus to the cytoplasm [72]. MicroRNAs are small non-coding RNAs that regulate gene expression by binding to mRNA during translation, which process is frequently dysregulated in cancer [73]. In normal, dysplastic and cancerous Barrett's epithelium, discriminating micro-RNA signatures have been found for different stages of dysplasia of esophageal tissue [74]. In colorectal cancer, elevated expression of XPO5 is primarily found in the nucleus and correlates with advanced disease stage and poor prognosis. XPO5 has an oncogenic role because its down-regulation reduces the invasive capacities and cell proliferation [75]. In prostate carcinoma, a DNA microarray analysis revealed that XPO5 was 1.6-fold up-regulated [76]. Another exportin, XPO1, showed distinct nucleic and cytoplasmic staining patterns associated to the Gleason score [77]. In the present study, we validated the elevated expression of XPO5 in dysplastic/EAC tissue by IHC staining. This revealed a significant increase in cytoplasmic and nuclear XPO5 staining in dysplastic/EAC tissue.

The most significantly up-regulated protein in our study, PRKCI, is a known genetic driver and genomic EAC biomarker [26, 79]. PRKCI is an oncogene that shows increased copy numbers in invasive tumors and has a locus in a commonly amplified region due to 9p loss of heterozygosity during progression of BE to EAC [78, 79]. CSNK2A and CSNK2B, which both were found significantly up-regulated, are subunits of the protein kinase CK2 (CSNK2), which has been associated with various cancer types, such as breast, lung, colon, and prostate cancer. CSNK2 is an emerging candidate for targeted therapy [80]. Deregulation of the regulatory subunits is suggested to promote various cancer types, and are considered potential biomarkers and therapeutic targets [81]. In a recent study, Xiao and co-workers found that CSNK2B attenuates the inhibition of NF-κB in hepatocellular carcinoma [82].

By gene set enrichment analysis, we examined the data to identify pathways potentially underlying the set of differentially expressed gene products. On the basis of the significant up-regulation of 19 proteins, the most significantly enriched pathways identified were the *Processing of Capped Intron-Containing Pre-mRNA* pathway and the sub-pathway mRNA Major Splicing, which both are part of the RNA metabolism super-pathway. These pathways stood out for their highly significant enrichment (FDR = 9e−9) and differed distinctly in this respect from other pathways (e.g., 4th pathway, *T cell receptor signaling*: FDR = 4e−6). Strong enrichment for spliceosome components was earlier found in a study by Francavilla and co-workers in epithelial ovarian cancer by using a mass-spectrometry-based proteomics approach [83]. A meta-analysis comparing four publicly available BE- and EAC-associated micro-array datasets, published by Nangraj et al. revealed that RNA metabolism and spliceosome are critical in the formation and development of EAC [84]. The

manifold and complex associations of spliceosome and cancer have been reviewed in detail by Srebrow et al. and El Marabti et al. [85, 86] What it basically comes down to is that due to mutations and alterations of expression levels of the splicing factors, the cancer is able to affect splicing and, thus, is potentially able to promote the selection of certain splicing variants. This observation is relevant for the pathology of cancer because, on the one hand, the functions of a protein are often related to the splicing form and, on the other hand, alternative splicing affects more than 90% of human genes [87]. Thus, changes of splicing factors can affect the splicing isoform selection—and hence processes related to cancer [85]. Dysregulation of splicing has in multiple studies been linked to cancer development, involving both oncogene and tumor suppressor activities [88]. Jiménez-Vacas et al. showed that the up-regulation of splicing factor SF3B1 is associated with the expressions of oncogenic splicing variants and the progression of prostate cancer [89]. Highly relevant in the development of EAC is the expression of different p53 protein isoforms as a result of different TP53 splicing forms [90]. Equally relevant is that MYC, an EAC driver gene [26], regulates the splicing of selected genes via the activation of alternative splicing factors or components of the core spliceosome [91, 92]. Spliceosome core components have been suggested as potential therapeutic target in various types of cancer, such as lung, breast, ovarian and prostate cancer [89, 93]. In the present study, we found the highest up-regulation within the spliceosome for the *cleavage and polyadenylation complex*, due to elevated expression of CPSF1, CSTF1 and SYMPK, potentially leading to deregulation of alternative polyadenylation (APA), which further yields mRNA 3′ untranslated region (UTRs) isoforms with modified characteristics including oncogenic activities [94]. For example, an SNP in the 3′ UTR of TP53 that is transcribed as a consequence of APA (lengthening) forms a risk factor for different types of cancer, such as prostate cancer, glioma, and colorectal adenoma [94, 95]. The regulation of proliferation marker Ki-67, which we found up-regulated, is as well mediated by APA in breast cancer [96].

IHC staining of nuclear MSH6 and nuclear and cytoplasmic XPO5 was observed in all samples. This observation partly reflects the results of the MS-based discovery study, in which XPO5 and MSH6 were found almost exclusively in dysplastic samples—and not in non-dysplastic samples. The discrepancy between absence in the discovery experiment and presence in the validation experiment might be primarily due to limited sensitivity of mass spectrometric detection, but could also be related to limited selectivity of the antibodies used. Nevertheless, the latter consideration seems unlikely, especially in the case of MSH6, because we used clinically validated antibodies. Hence, the relatively high fold-change ratios of the discovery study relative to those found during IHC validation are likely a result of zero imputation during statistical analysis. It must be noted, however, that the significance determined was not affected by zero imputation because a non-parametric test was applied. Furthermore, a linear quantitative relationship need not necessarily be assumed between the IHC score and the actual protein concentration of the tissue or the protein concentration determined in the discovery phase by mass spectrometry. In summary, both proteins had significantly higher IHC scores in dysplastic/EAC tissue, and thus positively validated the result of the discovery experiment.

In summary, this study provides insights in the alteration of epithelial proteomes during progression from NDBE into EAC. We determined a set of differentially expressed proteins that overall are up-regulated in dysplasia/EAC and point to increased activity of DNA mismatch repair, microRNA transport and RNA splicing. We showed increased immunostaining of MSH6 and XPO5 to confirm these findings. The proteomic finding of associations of spliceosome and polyadenylation activity with dysplastic progression of BE confirms recent novel findings, and extends the current knowledge of Barrett's carcinogenesis.

Acknowledgements We thank Ko Hagoort for critical reading of the manuscript and the Erasmus MC Cancer Computational Biology Center (CCBC) for providing their IT resources and sharing their software.

Author contributions JB and TL designed research; ABG, NP, and JB performed the clinical research; CS and CG acquired proteomic data; MD, GB, and SM acquired and interpreted pathological data; CS analyzed the data and drafted the manuscript; JB and TL supervised the study. All authors revised and finally approved the manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Sharma P. Clinical practice. Barrett's esophagus. N Engl J Med. 2009;361:2548–56.
2. Zhang HY, Spechler SJ, Souza RF. Esophageal adenocarcinoma arising in Barrett esophagus. Cancer Lett. 2009;275:170–7.
3. Kuipers EJ, Spaander MC. Natural history of Barrett's esophagus. Dig Dis Sci. 2018;63:1997–2004.
4. Shiota S, Singh S, Anshasi A, et al. Prevalence of Barrett's esophagus in Asian Countries: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2015;13:1907–18.
5. Ronkainen J, Aro P, Storskrubb T, et al. Prevalence of Barrett's esophagus in the general population: an endoscopic study. Gastroenterology. 2005;129:1825–31.
6. Kambhampati S, Tieu AH, Luber B, et al. Risk factors for progression of Barrett's esophagus to high grade dysplasia and esophageal adenocarcinoma. Sci Rep. 2020;10:4899.
7. Thrift AP. The epidemic of oesophageal carcinoma: where are we now? Cancer Epidemiol. 2016;41:88–95.
8. Zhang Y. Epidemiology of esophageal cancer. World J Gastroenterol. 2013;19:5598–606.
9. Boyle P, Levin B. World cancer report 2008. International Agency for Research on Cancer; 2008.
10. Stewart BW, Wild CP. World cancer report 2014, non-series publication. IARC; 2014. p. 630.
11. Thrift AP. Barrett's esophagus and esophageal adenocarcinoma: how common are they really? Dig Dis Sci. 2018;63:1988–96.
12. Lim YC, Fitzgerald RC. Diagnosis and treatment of Barrett's oesophagus. Br Med Bull. 2013;107:117–32.
13. Wani S, Rubenstein JH, Vieth M, et al. Diagnosis and management of low-grade dysplasia in Barrett's esophagus: expert review from the clinical practice updates committee of the American Gastroenterological Association. Gastroenterology. 2016;151:822–35.
14. Kountourakis P, Ajani JA, Davila M, et al. Barrett's esophagus: a review of biology and therapeutic approaches. Gastrointest Cancer Res. 2012;5:49–57.
15. Vennalaganti P, Kanakadandi V, Goldblum JR, et al. Discordance among pathologists in the United States and Europe in diagnosis of low-grade dysplasia for patients with Barrett's esophagus. Gastroenterology. 2017;152:564-570.e4.
16. Shaheen NJ, Fennerty MB, Bergman JJ. Less is more: a minimalist approach to endoscopy. Gastroenterology. 2018;154:1993–2003.
17. Kerkhof M, Kusters JG, van Dekken H, et al. Biomarkers for risk stratification of neoplastic progression in Barrett esophagus. Cell Oncol. 2007;29:507–17.
18. Huang Q, Hardie LJ. Biomarkers in Barrett's oesophagus. Biochem Soc Trans. 2010;38:343–7.
19. Qureshi AP, Stachler MD, Haque O, et al. Biomarkers for Barrett's esophagus—a contemporary review. Expert Rev Mol Diagn. 2018;18:939–46.
20. Li X, Galipeau PC, Sanchez CA, et al. Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression. Cancer Prev Res (Phila). 2008;1:413–23.
21. Galipeau PC, Prevo LJ, Sanchez CA, et al. Clonal expansion and loss of heterozygosity at chromosomes 9p and 17p in premalignant esophageal (Barrett's) tissue. J Natl Cancer Inst. 1999;91:2087–95.
22. Rygiel AM, Milano F, Ten Kate FJ, et al. Assessment of chromosomal gains as compared to DNA content changes is more useful to detect dysplasia in Barrett's esophagus brush cytology specimens. Genes Chromosomes Cancer. 2008;47:396–404.
23. Gregson EM, Bornschein J, Fitzgerald RC. Genetic progression of Barrett's oesophagus to oesophageal adenocarcinoma. Br J Cancer. 2016;115:403–10.
24. Khara HS, Jackson SA, Nair S, et al. Assessment of mutational load in biopsy tissue provides additional information about genomic instability to histological classifications of Barrett's esophagus. J Gastrointest Cancer. 2014;45:137–45.
25. 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.
26. Frankell AM, Jammula SG, Li X, et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. Nat Genet. 2019;51:506–16.
27. Snyder P, Dunbar K, Cipher DJ, et al. Aberrant p53 immunostaining in Barrett's esophagus predicts neoplastic progression: systematic review and meta-analyses. Dig Dis Sci. 2019;64:1089–97.
28. Reid BJ, Prevo LJ, Galipeau PC, et al. Predictors of progression in Barrett's esophagus II: baseline 17p (p53) loss of heterozygosity identifies a patient subset at increased risk for neoplastic progression. Am J Gastroenterol. 2001;96:2839–48.
29. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. Gut. 2014;63:7–42.
30. Zhao J, Chang AC, Li C, et al. Comparative proteomics analysis of Barrett metaplasia and esophageal adenocarcinoma using two-dimensional liquid mass mapping. Mol Cell Proteomics. 2007;6:987–99.
31. Elsner M, Rauser S, Maier S, et al. MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100–A10 as novel prognostic markers in Barrett's adenocarcinoma. J Proteomics. 2012;75:4693–704.
32. Zaidi AH, Gopalakrishnan V, Kasi PM, et al. Evaluation of a 4-protein serum biomarker panel-biglycan, annexin-A6, myeloperoxidase, and protein S100–A9 (B-AMP)-for the detection of esophageal adenocarcinoma. Cancer. 2014;120:3902–13.
33. O'Neill JR, Pak H-S, Pairo-Castineira E, et al. Quantitative shotgun proteomics unveils candidate novel esophageal adenocarcinoma (EAC)-specific proteins. Mol Cell Proteomics. 2017;16:1138–50.
34. Stingl C, Van VFGI, Guzel C, et al. Reproducibility of protein identification of selected cell types in Barrett's esophagus analyzed by combining laser-capture microdissection and mass spectrometry. J Proteome Res. 2011;10:288–98.
35. Sharma P, Dent J, Armstrong D, et al. The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. Gastroenterology. 2006;131:1392–9.
36. Pouw RE, van Vilsteren FGI, Peters FP, et al. Randomized trial on endoscopic resection-cap versus multiband mucosectomy for piecemeal endoscopic resection of early Barrett's neoplasia. Gastrointest Endosc. 2011;74:35–43.
37. Vizcaíno JA, Côté RG, Csordas A, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2013;41:D1063–9.
38. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367–72.
39. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11:2301–19.
40. Cox J, Neuhauser N, Michalski A, et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10:1794–805.

41. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013.
42. Braakman RBH, Tilanus-Linthorst MMA, Liu NQ, et al. Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. J Proteomics. 2012;75:2844–2854.
43. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
44. Doncheva NT, Morris JH, Gorodkin J, et al. Cytoscape StringApp: network analysis and visualization of proteomics data. J Proteome Res. 2019;18:623–32.
45. Morris JH, Apeltsin L, Newman AM, et al. ClusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinform. 2011;12:1–14.
46. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38:W214–20.
47. Hinkelbein J, Jansen S, Iovino I, et al. Thirty minutes of hypobaric hypoxia provokes alterations of immune response, haemostasis, and metabolism proteins in human serum. Int J Mol Sci. 2017;18: 1882.
48. Montojo J, Zuberi K, Rodriguez H, et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics. 2010;26:2927–8.
49. Salomao MA, Lam-Himlin D, Pai RK. Substantial interobserver agreement in the diagnosis of dysplasia in Barrett esophagus upon review of a patient's entire set of biopsies. Am J Surg Pathol. 2018;42:376–81.
50. Duits LC, Lao-Sirieix P, Wolf WA, et al. A biomarker panel predicts progression of Barrett's esophagus to esophageal adenocarcinoma. Dis Esophagus. 2019;32:doy102.
51. Tschanz ER. Do 40% of patients resected for Barrett esophagus with high-grade dysplasia have unsuspected adenocarcinoma? Arch Pathol Lab Med. 2005;129:177–80.
52. Yantiss RK. Diagnostic challenges in the pathologic evaluation of Barrett esophagus. Arch Pathol Lab Med. 2010;134:1589–600.
53. Braakman RBH, Stingl C, Tilanus-Linthorst MMA, et al. Proteomic characterization of microdissected breast tissue environment provides a protein-level overview of malignant transformation. Proteomics. 2017;17: 1600213.
54. Güzel C, Govorukhina NI, Wisman GBA, et al. Proteomic alterations in early stage cervical cancer. Oncotarget. 2018;9:18128–47.
55. Arul GS, Moorghen M, Myerscough N, et al. Mucin gene expression in Barrett's oesophagus: an in situ hybridisation and immunohistochemical study. Gut. 2000;47:753–61.
56. Warson C, Van de Bovenkamp JHB, Korteland-Van Male AM, et al. Barrett's esophagus is characterized by expression of gastric-type mucins (MUC5AC, MUC6) and TFF peptides (TFF1 and TFF2), but the risk of carcinoma development may be indicated by the intestinal-type mucin, MUC2. Hum Pathol. 2002;33:660–8.
57. Ahrens TD, Lutz L, Lassmann S, et al. Turning skyscrapers into town houses: insights into Barrett's esophagus. Pathobiology. 2017;84:87–98.
58. Katoh M. Trefoil factors and human gastric cancer (review). Int J Mol Med. 2003;12:3–9.
59. Xiao P, Ling H, Lan G, et al. Trefoil factors: gastrointestinal-specific proteins associated with gastric cancer. Clin Chim Acta. 2015;450:127–34.
60. Schellnegger R, Quante A, Rospleszcz S, et al. Goblet cell ratio in combination with differentiation and stem cell markers in Barrett esophagus allow distinction of patients with and without esophageal adenocarcinoma. Cancer Prev Res. 2017;10:55–66.
61. Lao-Sirieix P, Boussioutas A, Kadri SR, et al. Non-endoscopic screening biomarkers for Barrett's oesophagus: from microarray analysis to the clinic. Gut. 2009;58:1451–9.
62. Ross-Innes CS, Debiram-Beecham I, O'Donovan M, et al. Evaluation of a minimally invasive cell sampling device coupled with assessment of Trefoil factor 3 expression for diagnosing Barrett's esophagus: a multi-center case-control study. PLOS Med. 2015;12: e1001780.
63. Kadri SR, Lao-Sirieix P, O'Donovan M, et al. Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care: cohort study. BMJ. 2010;341: c4372.
64. Ständker L, Schrader M, Kanse SM, et al. Isolation and characterization of the circulating form of human endostatin. FEBS Lett. 1997;420:129–33.
65. Xing M, Oksenych V. Genetic interaction between DNA repair factors PAXX, XLF, XRCC4 and DNA-PKcs in human cells. FEBS Open Bio. 2019;9:1315–26.
66. Haffner MC, Aryee MJ, Toubaji A, et al. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. Nat Genet. 2010;42:668–75.
67. Wilczak W, Rashed S, Hube-Magg C, et al. Up-regulation of mismatch repair genes MSH6, PMS2 and MLH1 parallels development of genetic instability and is linked to tumor aggressiveness and early PSA recurrence in prostate cancer. Carcinogenesis. 2017;38:19–27.
68. Shahi A, Lee J-H, Kang Y, et al. Mismatch-repair protein MSH6 is associated with Ku70 and regulates DNA double-strand break repair. Nucleic Acids Res. 2011;39:2130–43.
69. Jentzsch T, Robl B, Husmann M, et al. Expression of MSH2 and MSH6 on a tissue microarray in patients with osteosarcoma. Anticancer Res. 2014;34:6961–72.
70. Farris AB 3rd, Demicco EG, Le LP, et al. Clinicopathologic and molecular profiles of microsatellite unstable Barrett esophagus-associated adenocarcinoma. Am J Surg Pathol. 2011;35:647–55.
71. Dudley JC, Lin M-T, Le DT, et al. Microsatellite instability as a biomarker for PD-1 blockade. Clin Cancer Res. 2016;22:813–20.
72. Wu K, He J, Pu W, et al. The role of Exportin-5 in MicroRNA biogenesis and cancer. Genomics Proteomics Bioinform. 2018;16:120–6.
73. Clark RJ, Craig MP, Agrawal S, et al. microRNA involvement in the onset and progression of Barrett's esophagus: a systematic review. Oncotarget. 2018;9:8179–96.
74. Craig MP, Rajakaruna S, Paliy O, et al. Differential MicroRNA signatures in the pathogenesis of Barrett's esophagus. Clin Transl Gastroenterol. 2020;11: e00125.
75. Shigeyasu K, Okugawa Y, Toden S, et al. Exportin-5 functions as an oncogene and a potential therapeutic target in colorectal cancer. Clin Cancer Res. 2017;23:1312–22.
76. Chiosea S, Jelezcova E, Chandran U, et al. Up-regulation of dicer, a component of the MicroRNA machinery, in prostate adenocarcinoma. Am J Pathol. 2006;169:1812–20.
77. Duijvesz D, Rodriguez-Blanco G, Hoogland AM, et al. Differential tissue expression of extracellular vesicle-derived proteins in prostate cancer. Prostate. 2019;79:1032–42.
78. Contino G, Vaughan TL, Whiteman D, et al. The evolving genomic landscape of Barrett's esophagus and esophageal adenocarcinoma. Gastroenterology. 2017;153:657-673.e1.
79. Dulak AM, Schumacher SE, van Lieshout J, et al. Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. Cancer Res. 2012;72:4383–93.
80. Rabalski AJ, Gyenis L, Litchfield DW. Molecular pathways: emergence of protein kinase CK2 (CSNK2) as a potential target to inhibit survival and DNA damage response and repair pathways in cancer cells. Clin Cancer Res. 2016;22:2840–7.

81. Ortega CE, Seidner Y, Dominguez I. Mining CK2 in cancer. PLoS ONE. 2014;9: e115609.
82. Xiao Y, Huang S, Qiu F, et al. Tumor necrosis factor α-induced protein 1 as a novel tumor suppressor through selective down-regulation of CSNK2B blocks nuclear factor-κB activation in hepatocellular carcinoma. EBioMedicine. 2020;51: 102603.
83. Francavilla C, Lupia M, Tsafou K, et al. Phosphoproteomics of primary cells reveals druggable kinase signatures in ovarian cancer. Cell Rep. 2017;18:3242–56.
84. Nangraj AS, Selvaraj G, Kaliamurthi S, et al. Integrated PPI- and WGCNA-retrieval of hub gene signatures shared between Barrett's esophagus and esophageal adenocarcinoma. Front Pharmacol. 2020;11:881.
85. Srebrow A, Kornblihtt AR. The connection between splicing and cancer. J Cell Sci. 2006;119:2635–41.
86. El Marabti E, Younis I. The cancer spliceome: reprograming of alternative splicing in cancer. Front Mol Biosci. 2018;5:80.
87. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456:470–6.
88. Dvinge H, Kim E, Abdel-Wahab O, et al. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer. 2016;16:413–30.
89. Jiménez-Vacas JM, Herrero-Aguayo V, Gómez-Gómez E, et al. Spliceosome component SF3B1 as novel prognostic biomarker and therapeutic target for prostate cancer. Transl Res. 2019;212:89–103.
90. Bourdon J-C, Surget S, Khoury MP. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. Onco Targets Ther. 2013;7:57.
91. Hsu TY-T, Simon LM, Neill NJ, et al. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. Nature. 2015;525:384–8.
92. Koh CM, Bezzi M, Low DHP, et al. MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. Nature. 2015;523:96–100.
93. Quidville V, Alsafadi S, Goubar A, et al. Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. Cancer Res. 2013;73:2247–58.
94. Erson-Bensan AE, Can T. Alternative polyadenylation: another foe in cancer. Mol Cancer Res. 2016;14:507–17.
95. Stacey SN, Sulem P, Jonasdottir A, et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. Nat Genet. 2011;43:1098–103.
96. Yan H, Tian R, Wang W, et al. Aberrant Ki-67 expression through 3'UTR alternative polyadenylation in breast cancers. FEBS Open Bio. 2018;8:332–8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.