Supplemental materials

Methods

Patients with esophageal benign disease

The detail of 44 patients with esophageal benign disease in test cohort: 17 cases of esophageal leiomyoma, 11 cases of upper gastrointestinal polyps, 4 cases of esophageal cyst, 4 cases of esophageal papilloma, 3 cases of low grade atypical hyperplasia, 2 cases of esophageal varices, 1 cases of reflux esophagitis, 1 cases of esophageal gastric mucosal endometriosis, and 1 cases of mycoderma thickening. 74 patients with esophageal benign disease in the validation cohort containing 27 cases of reflux esophagitis, 10 cases of Barrett's esophagus, 5 cases of esophageal cyst, 2 cases of low grade atypical hyperplasia, 5 cases of esophageal gastric mucosal endometriosis, 2 cases of esophageal varices, 9 cases of esophageal papilloma and 14 cases of esophageal leiomyomas.

RNA isolation, cDNA library preparation, RNA-seq and analysis

Total RNA was extracted from frozen tissues using the Trizol reagent (Invitrogen, USA) according to the manufacture’s instruction. Beads with oligo(dT) were used to isolate poly(A) mRNA. First-strand cDNA was synthesized using random hexamer-primer and reverse transcriptase (Invitrogen). The second-strand cDNA was synthesized using RNase H (Invitrogen) and DNA polymerase I (New England BioLabs). Then the cDNA libraries were prepared according to Illumina's protocols
and sequenced by Illumina HiSeq™ 2000. Sequence data from genomic DNA and complementary DNA were mapped to the reference human genome (hg19) using the Burrows-Wheeler Aligner and were processed using the publicly available SAMtools, Picard, and Genome Analysis Toolkit. The quantity of gene expression was calculated by the RPKM method (Reads Per Kb per Million reads)\(^1\). The genes with FDR less than 0.001 and change fold more than 2 fold were considered as the DEG (Differentially expressed gene).

**Immunohistochemistry**

Formalin-fixed, paraffin-embedded ESCC sections were incubated with antibodies to CHI3L1 (Abcam, UK), MMP13 (R&D systems, USA), SPP1 (Abcam, UK) overnight at 4°C. After washing in PBST, the tissue sections were treated with a horseradish peroxidase-conjugated anti-rabbit secondary antibody (1:1000, Zymed). The tissue sections were then developed with 3-diaminobenzidine tetrahydrochloride for 10 seconds, followed by counterstaining with 10% Mayer’s hematoxylin. The degree of immunostaining was reviewed by two independent observers.

**Results**

**Cut-off values**

As shown in Table S5, for each stage, to reach 90% sensitivity, the cut-off values for CHI3L1, MMP13, and SPP1 in the test cohort were 25.19, 0.62 and 19.08 in stage I; 27.15, 0.43 and 19.87 in stage II; 42.78, 0.61 and 32.74 in stage III; and 42.29, 0.30
and 31.78 in stage IV, respectively. In the validation cohort, meanwhile, the breakpoints for CHI3L1, MMP13, and SPP1 were 31.37, 0.58 and 12.48 in stage I; 33.49, 0.70 and 13.08 in stage II; 34.62, 0.41 and 12.38 in stage III; and 34.01, 0.66 and 15.55 in stage IV, respectively.

The comparison between CHI3L1, MMP13, or SPP1 combination and CEA

In comparison to the traditional ESCC marker CEA, as shown in Figure S6A and S6B, the AUC of CEA was 0.646 (95% CI: 0.583–0.710) in the test cohort and 0.633 (95% CI: 0.573–0.694) in the validation cohort. Moreover, when we applied 5.0 ng/ml, according to the manufacturer’s protocol, for CEA as the cut-off values, as shown in Figure S6B, the sensitivity of CEA was 6.7% and 13.0% in the test and validation cohort, respectively, which is significantly lower than that for the combination (90.00%), which was rather low to apply for the early detection of ESCC. However, the specificity was slightly higher. Moreover, the combination exhibited a higher NPV compared with CEA (87.07% vs. 48.91% in the test cohort, and 93.52% vs. 49.83% in the validation cohort) without an obvious change in the PPV (77.59% vs. 62.5% in the test cohort, and 62.14% vs. 73.33% in the validation cohort).

The association between serum CHI3L1, MMP13 and SPP1 and clinicopathological characteristics

The associations between the median serum CHI3L1, MMP13, SPP1 levels and the clinicopathological parameters are presented in Table S6. Serum CHI3L1 and MMP13 was not significantly correlated with gender, T classification, N classification, metastasis, or clinical stage, but it was significantly associated with age ($P = 0.0002$
and $P = 0.0267$, respectively). The serum level of CHI3L1 was higher in elderly patients ($\geq 60$ years) than in patients aged less than 60 years, whereas the serum level of MMP13 was lower in elderly patients ($\geq 60$ years) than in patients aged less than 60 years. Serum SPP1 was not significantly correlated with age, N classification, metastasis, or clinical stage, but it was correlated with gender and T classification ($P = 0.0037$ and $P = 0.0165$, respectively). The level of serum SPP1 was elevated in male compared with female patients, and T4 patients exhibited the highest level of SPP1.
References:

[1] Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq[J]. Nat Methods. 2008,5(7): 621-628.

[2] Bilgin DE, Dizdar Y, Aksit E, et al. EMMPRIN and ADAM12 in prostate cancer: Preliminary results of a prospective study[J]. Tumour Biol. 2014,35(11): 11647-11653.

[3] Sole X, Crous-Bou M, Cordero D, et al. Discovery and validation of new potential biomarkers for early detection of colon cancer[J]. PLoS One. 2014,9(9): e106748.

[4] Tanwar MK, Gilbert MR, Holland EC. Gene expression microarray analysis reveals YKL-40 to be a potential serum marker for malignant character in human glioma[J]. Cancer Res. 2002,62(15): 4364-4368.

[5] Johansen JS, Christensen IJ, Riisbro R, et al. High serum YKL-40 levels in patients with primary breast cancer is related to short recurrence free survival[J]. Breast Cancer Res Treat. 2003,80(1): 15-21.

[6] Zheng X, Xing S, Liu XM, et al. Establishment of using serum YKL-40 and SCCA in combination for the diagnosis of patients with esophageal squamous cell carcinoma[J]. Bmc Cancer. 2014,14: 490.

[7] Yoneda K, Iida H, Endo H, et al. Identification of Cystatin SN as a novel tumor marker for colorectal cancer[J]. Int J Oncol. 2009,35(1): 33-40.

[8] Kosanam H, Prassas I, Chryostoa CC, et al. Laminin, gamma 2 (LAMC2): A promising new putative pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues[J]. Mol Cell Proteomics. 2013,12(10): 2820-2832.

[9] Jiao XL, Chen D, Wang JG, et al. Clinical significance of serum matrix metalloproteinase-13 levels in patients with esophageal squamous cell carcinoma (ESCC)[J]. Eur Rev Med Pharmacol Sci. 2014,18(4): 509-515.

[10] Etoh T, Inoue H, Yoshikawa Y, et al. Increased expression of collagenase-3 (MMP-13) and MT1-MMP in esophageal cancer is related to cancer aggressiveness[J]. Gut. 2000,47(1): 50-56.

[11] Sasaki H, Dai M, Auclair D, et al. Serum level of the peristin, a homologue of an insect cell adhesion molecule, in thymoma patients[J]. Cancer Lett. 2001,172(1): 37-42.

[12] Fersching DM, Nagel D, Siegele B, et al. Apoptosis-related biomarkers sFAS, MIF, ICAM-1 and PAI-1 in serum of breast cancer patients undergoing neoadjuvant chemotherapy[J]. Anticancer Res. 2012,32(5): 2047-2058.

[13] Shih YL, Hsieh CB, Lai HC, et al. SFRP1 suppressed hepatoma cells growth through Wnt canonical signaling pathway[J]. Int J Cancer. 2007,121(5): 1028-1035.

[14] Mor G, Visintin I, Lai Y, et al. Serum protein markers for early detection of ovarian cancer[J]. Proc Natl Acad Sci U S A. 2005,102(21): 7677-7682.

Figure S1. 175 genes differently expressed in at least 5 out of 6 tissue pairs form RNA sequencing. Red:upregulated genes; Green:downregulated genes. The X axis is measured by log2(T/N). T:ESCC tissue, N: normal tissue.
**Figure S2.** 39 differently expressed genes that were upregulated by at least 5-fold in 4 out of 6 tumor tissues compared to the non-cancerous tissues. The X axis is measured by log2(T/N). T: ESCC tissue, N: normal tissue.

**Figure S3.** 32 secretory proteins identified through SignalP4.1 and SecretomeP 2.0.

**Figure S4.** The expression level of 32 candidates in three PubMed GEO database. A. GSE23400(53N,53T); B. GSE20347(17N,17T); C. GSE33810(GSM836193: mixture of 10 normal tissues; GSM836194: mixture of 10 ESCC tissues). N: normal tissue, T: ESCC tissue. The X axis is measured by log2(T/N).

**Figure S5.** Serum levels of candidate biomarkers in the preliminary screening phase.

Levels of serum CA9, CST1, LAMC2, POSTN, SERPINE1, and SFRP4, were compared between 40 ESCC patients (ESCC) and 40 healthy controls (HC). The Mann-Whitney U test was performed for comparisons between groups. P < 0.05 was considered statistically significant.

**Figure S6.** Diagnostic effect of serum CEA in the test cohort and the validation cohort.

A: The comparison of ROC curves for CEA and the combination (Logit(p=ESCC)=-4.583+0.017×CHI3L1+0.018×SPP1+0.821×MMP13) in discriminating patients with 150 ESCC from 140 controls. B: The comparison of ROC curves for CEA and the combination in discriminating patients with 169 ESCC from 154 controls. C: The diagnostic performance of CEA in discriminating ESCC and controls (healthy controls and patients with esophageal benign disease) in the test
cohort and validation cohort. ESCC: esophageal squamous cell carcinoma; Sen, sensitivity; Spe, specificity; PPV, positive predictive value; NPV, negative predictive value; HC, healthy control.

Table S1. Primers for real-time RT-PCR

| Gene          | primer                          | length (bp) |
|---------------|---------------------------------|-------------|
| GAPDH pF      | 5′-GACTCATGACCACAGTCCATGC-3′     | 113         |
| GAPDH pR      | 5′-AGAGGCAAGGGATGATGTCTG-3′      |             |
| CHI3L1 pF     | 5′-GAGGATGGAACCTTTGGGTCTC-3′    | 182         |
| CHI3L1 pR     | 5′-TCATTTCCCTGATAGGCTGCT-3′     |             |
| MMP13 pF      | 5′-CCAAGGACCTGGGACACTC-3′       | 173         |
| MMP13 pR      | 5′-CAAGGGATAAGGAAGGGTCAC-3′     |             |
| SPP1 pF       | 5′-GACCTGCCAGCAACCGAAG-3′       | 155         |
| SPP1 pR       | 5′-GGTGATGTCTCTGCTCTTAGC-3′     |             |

Table S2. Pathway analysis of significantly differentially expressed genes.

| Markers     | Pathway                          |
|-------------|----------------------------------|
| ADAM12      | EGF, NOTCH                       |
| CA9         | Angiogenesis; Hypoxia            |
| CHI3L1      | NF-KB                            |
| COL11A1     | PI3K-Akt                         |
| CST1        | Wnt                              |
| CTHRC1      | Wnt                              |
Table S3. GO Biological process analysis of significantly differentially expressed genes.

| Markers   | GO Biological process                                           |
|-----------|-----------------------------------------------------------------|
| INHBA     | TGF-β                                                           |
| LAMC2     | PI3K-Akt                                                        |
| MFAP2     | NOTCH                                                           |
| MMP13     | Degradation of the extracellular matrix                         |
| POSTN     | NOTCH                                                           |
| SERPINE1  | P53                                                             |
| SFRP4     | Wnt                                                             |
| SPP1      | PI3K-Akt                                                        |
| WISP1     | NOTCH Wnt                                                       |
| HMGA2     | Transcriptional misregulation in cancer                         |
| HOXD10    | Stem Cell Transcription Factors in human, Proteoglycans in cancer, MicroRNAs in cancer |
| SIX1      | Wnt; Transcriptional misregulation in cancer                    |

Stem Cell Transcription Factors in human, Proteoglycans in cancer, MicroRNAs in cancer

| Markers   | GO Biological process                                           |
|-----------|-----------------------------------------------------------------|
| ADAM12    | cell adhesion                                                   |
| CA9       | response to hypoxia                                             |
| CHI3L1    | activation of NF-kappaB-inducing kinase activity                 |
| COL11A1   | extracellular matrix organization                                |
| CST1      | negative regulation of endopeptidase activity                   |
| CTHRC1    | cell migration                                                   |
Table S4. Evidence sourced from literature about markers existence in blood.

| Markers | Evidence in literature |
|---------|------------------------|
| INHBA   | negative regulation of cell cycle, extrinsic apoptotic signaling, cell differentiation, growth |
| LAMC2   | cell junction assembly, cell adhesion |
| MFAP29  | extracellular matrix organization |
| MMP13   | extracellular matrix organization |
| POSTN   | cell adhesion |
| SERPINE1| angiogenesis |
| SFRP4   | negative regulation of cell proliferation |
| SPP1    | cell adhesion |
| WISP1   | regulation of cell growth |
| HMGA2   | cell proliferation, epithelial to mesenchymal transition |
| HOXD10  | epithelial to mesenchymal transition |
| SIX1    | apoptotic process |

ADAM12    | Prostate cancer\(^2\) |
CA9       | Breast cancer\(^3\) |
CHI3L1    | Melanoma\(^4\), breast cancer\(^5\), ESCC\(^6\) |
CST11     | Colon cancer\(^7\) |
LAMC2     | Pancreatic cancer\(^8\) |
MMP13     | ESCC\(^9, 10\) |
POSTN     | Thymoma\(^11\) |
| Case number of ESCC | CHI3L1  (ng/ml) | MMP13  (ng/ml) | SPP1   (ng/ml) |
|--------------------|-----------------|----------------|----------------|
| **Test cohort**    |                 |                |                |
| Stage I            | 20              | 25.19          | 0.62           | 19.08          |
| Stage II           | 51              | 27.15          | 0.43           | 19.87          |
| Stage III          | 71              | 42.78          | 0.61           | 30.74          |
| Stage IV           | 8               | 42.29          | 0.39           | 29.78          |
| **Validation cohort** |             |                |                |
| Stage I            | 13              | 31.37          | 0.58           | 17.48          |
| Stage II           | 71              | 33.49          | 0.70           | 17.08          |
| Stage III          | 69              | 34.62          | 0.41           | 19.38          |
| Stage IV           | 16              | 34.01          | 0.66           | 25.55          |
| Characteristics | case numbers | CHI3L1 (ng/ml) | MMP13 (ng/ml) | SPP1 (ng/ml) | P   |
|-----------------|-------------|----------------|--------------|-------------|-----|
| Age, years      |             |                |              |              |     |
| <60             | 160         | 73.03 (11.56-360.50) | 4.83 (0.03-29.50) | 78.36 (7.04-203.40) | 0.224 |
| ≥60             | 159         | 100.8 (8.05-430.80)  | 4.08 (0.01-26.00) | 84.25 (5.12-233.40)  |     |
| Gender          |             |                |              |              |     |
| Male            | 237         | 78.70 (11.56-430.80) | 4.55 (0.01-29.50) | 86.68 (5.12-233.40)  | 0.0037 |
| Female          | 82          | 84.46 (8.05-419.20)  | 4.55 (0.15-26.00) | 67.78 (7.65-225.10)  |     |
| pT status       |             |                |              |              |     |
| pT1             | 39          | 78.18 (15.40-264.70) | 3.97 (0.01-14.92) | 72.81 (8.07-187.50)  | 0.0165 |
| pT2             | 50          | 92.96 (19.20-430.80) | 4.55 (0.11-17.13) | 90.71 (8.99-174.60)  |     |
| pT3             | 197         | 79.23 (8.05-421.30)  | 4.55 (0.02-27.41) | 77.91 (5.12-233.40)  |     |
| pT4             | 33          | 95.80 (27.32-361.00) | 5.11 (0.16-29.50) | 111.10 (14.92-200.50) |     |
| pN status       |             |                |              |              |     |
| pN0             | 143         | 78.70 (11.56-430.80) | 5.16 (0.01-29.50) | 83.05 (8.90-225.10)  | 0.4014 |
| pN1             | 125         | 81.64 (8.05-421.30)  | 4.04 (0.16-27.41) | 86.40 (7.04-228.40)  |     |
| pN2             | 38          | 85.48 (30.08-361.00) | 4.65 (0.17-16.99) | 66.35 (11.91-211.40) |     |
| pN3             | 12          | 61.62 (24.27-185.60) | 3.14 (0.42-11.97) | 68.83 (5.12-157.00)  |     |
| pM status       |             |                |              |              |     |
| pM0             | 295         | 81.44 (8.05-430.80)  | 4.55 (0.01-29.50) | 80.01 (5.12-228.40)  | 0.6224 |
| pM1             | 24          | 67.76 (21.64-419.20) | 4.30 (0.02-13.45) | 88.26 (7.04-233.40)  |     |
| pTNM status     |             |                |              |              |     |
| Stage I         | 33          | 80.48 (15.40-264.70) | 4.36 (0.01-14.92) | 75.98 (9.50-187.50)  | 0.8656 |
| Stage II        | 122         | 78.52 (11.56-430.80) | 4.57 (0.03-26.00) | 78.32 (7.65-225.10)  |     |
| Stage III       | 140         | 85.00 (8.05-421.30)  | 4.54 (0.16-29.50) | 83.29 (5.12-228.40)  |     |
| Stage IV        | 24          | 67.76 (21.64-419.20) | 4.30 (0.02-13.45) | 88.26 (7.04-233.40)  |     |
| Marker (ng/ml)   | AUC  | Cut-off | Sen(%) | Spe(%) | PPV(%) | NPV(%) |
|-----------------|------|---------|--------|--------|--------|--------|
| Test-CEA        | 0.646| 5.0     | 6.7    | 95.7   | 62.5   | 48.91  |
| Validation-CEA  | 0.633| 5.0     | 13.0   | 94.8   | 73.33  | 49.83  |