Supplementary Methods

Our multi-step approach to ESCC diagnostic biomarker panel development is schematically illustrated in Supplementary Figure 1*A*.

First, after searching for "methylation" and "esophageal squamous cell carcinoma" in PubMed >500 abstracts were perused. Next, DNA methylation data on 93 ESCCs were compared with 14 normal adjacent esophagi and 120 normal tissues from 12 different tissue types in The Cancer Genome Atlas. To optimize candidate biomarker discovery, we sought only CpG sites with methylation >30% in ≥50% of cancer tissues but <5% in all normal esophageal and other normal tissues. This yielded 56 CpG sites in ESCC vs normal tissues, esophagus and 37 in ESCC vs nonesophageal normal tissues, with 16 sites shared. From data from the publications and The Cancer Genome Atlas, 32 candidate genes were selected for further testing, with all genes containing ≥2 CpG sites. After primer design based on post-bisulfite–modified methylated sequences and non–real-time polymerase chain reaction testing on universal methylated and unmethylated DNAs (Chemicon), 15 loci were chosen for probe design and quantitative methylation-specific polymerase chain reaction (qMSP) testing; the remaining 17 sites failed polymerase chain reaction amplification despite multiple attempts and primer redesigns. These 15 genes underwent qMSP testing on 15 sets of ESCC and normal esophageal tissue DNAs for primer/probe performance and cancer specificity. From these tissue DNA methylation data, the top-ranked 6 genes were selected for evaluation in 48 pairs of ESCC vs adjacent normal tissues. Five genes performed well in these 48 tissue pairs and were tested on sponge-capsule cytology samples from 35 patients with ESCC and 56 control subjects.

Fresh-frozen clinically indicated endoscopic biopsies were obtained with consent under approved Institutional Review Board protocols at Johns Hopkins University, University of Maryland, or Baltimore Veterans Affairs hospitals. All ESCC and adjacent normal tissues were evaluated by expert pathologists to confirm histologic diagnoses.

DNA was extracted from frozen tissue (Qiagen DNeasy Blood & Tissue Kit) before bisulfite conversion using methylation on beads.[e1] Before testing on tissue samples, real-time qMSP assays with all candidate genes were quality-control–tested on fully methylated and unmethylated control DNAs. Methylation levels in bisulfite-converted tissue DNAs were then measured via real-time qMSP (7900HT Fast Real-Time PCR System). Unmethylated β-actin served as an internal control for normalization. Standard curves were generated via serial dilutions of universal methylated DNA for absolute quantification. Methylation indices comprised fractional methylation of each sample's β-actin reference. Wilcoxon matched-pairs signed-rank testing was used to analyze methylation differences between paired ESCCs and matched normal tissues for each candidate gene.

This was a multi-site observational case–control study conducted at the Johns Hopkins University (Baltimore, MD) and Mbarara Regional Referral Hospital (Mbarara, Uganda) with approval by the Institutional Review Boards at these 2 entities. All molecular assays were performed at Johns Hopkins University School of Medicine. Patients 18 years or older and undergoing EGD for clinical indications were recruited between January 2018 and July 2021. Patients with extra-esophageal malignancies and/or those who had undergone esophagectomy were excluded. Patients with esophageal strictures preventing capsule swallowing and those with severe dysphagia/odynophagia were also excluded. Written informed consent was obtained face-to-face by clinic staff. Patient samples were taken before undergoing EGD or in clinic follow-up after initial diagnostic EGD. Study participants were also assessed for nausea, vomiting, pain, or hematochezia after sample retrieval. All subjects had a pathology-confirmed diagnosis of either ESCC or a noncancer condition by means of esophageal biopsy, without any Barrett's esophagus.

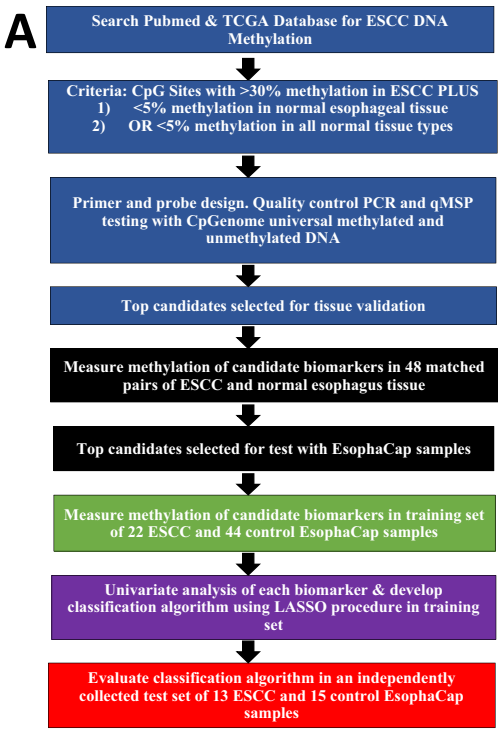
The tethered, gelatin-encapsulated sponge device was used to collect esophageal cells. Patients were given the option of local pharyngeal anesthesia with lidocaine spray before swallowing the encapsulated sponge with a few sips of water. The string tethered to the capsule was held outside the patient's mouth while the sponge was swallowed. A minimum of 5 minutes was required to elapse before sponge retrieval via string, to allow adequate time for the capsule to dissolve and the sponge to expand. The esophageal cytological material collected on the sponge was stored in ThinPrep PreservCyt Solution for preservation and transport. For DNA extraction, the container with preserving solution and the collected sponge was agitated to dislodge remaining cells and then centrifuged. The resulting pellet was lysed with proteinase K (NEB P8107S), and DNA was extracted using DNeasy Kit (Qiagen). Each sample was bisulfite-treated using the methylation-on-beads method[e2] and assayed by real-time qMSP, as described previously for esophageal cancer tissue samples.[e3]

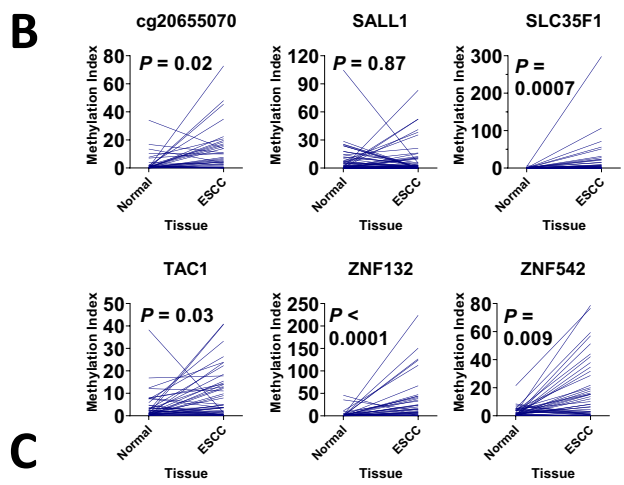
In the tissue sample validation stage, the distribution differences in methylated genes between matched-normal and tumor samples were tested via Wilcoxon signed-rank test. Differences in patient characteristics between ESCC and controls were tested via χ^2 test or Fisher exact test for categorical variables, or by Wilcoxon rank-sum test for continuous variables in the training and test sets. Distribution of methylation levels and baseline patient characteristics were examined separately in the training and test sets. The Wilcoxon rank-sum test was used to evaluate methylation differences between patients with ESCC and control subjects for each biomarker gene. In the training set, univariate analysis was conducted via logistic regression and the classification performance was evaluated via area under the receiver operating characteristic curve. An optimum multivariable model was generated in the training set using the LASSO (least absolute shrinkage and selection operator[e4]) method in a logistic regression setting, where the tuning parameter was chosen via 5-fold cross-validation to minimize the mean squared error. Candidate risk factors for the multivariable model included our candidate methylation markers, and the individual's age and sex. The classification algorithm was built by a linear sum of penalized weights, which were obtained from

the LASSO procedure to reduce overfitting, and with a cut point chosen on the basis of maximization of Youden's index.[e5] An independent test set was used to validate our classification performance with sensitivity, specificity, positive predictive value, and negative predictive value reported. Analyses were conducted in R version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria), with a 2-sided P value $< .05$ as the significance level.

Supplementary References

e1. Keeley B, et al. Clin Chim Acta 2013;425:169–175.
e2. Bailey VJ, et al. Clin Chem 2010;56:1022–1025.
e3. Schulmann K, et al. Oncogene 2005;24:4138–4148.
e4. Tibshirani R. https://pubag.nal.usda.gov/catalog/2337405.
e5. Youden WJ. Cancer 1950;3:32–35.





C

Univariate ROC Analysis			
Marker	AUC	*P* value	95% CI
cg20655070	0.733	0.0022	0.610 to 0.856
SLC35F1	0.839	< 0.0001	0.722 to 0.955
TAC1	0.780	0.0002	0.655 to 0.905
ZNF132	0.857	< 0.0001	0.751 to 0.964
ZNF542	0.846	< 0.0001	0.742 to 0.950

Supplementary Figure 1. *(A)* Summary of study design. There were five principal steps: 1) candidate biomarker search; 2) tissue validation; 3) biomarker testing with samples obtained using EsophaCap device; 4) multivariate model development based on methylation data from this training set; and 5) multivariate model applied to independent test set. *(B)* Methylation index levels of the six strongest candidate biomarkers in 48 esophageal squamous cell carcinoma (ESCC) and matched normal biopsy tissue pairs. Paired methylation indices of each subject's normal esophageal and tumor tissues are shown as *blue lines*. Methylation indices were calculated based on each sample's corresponding β-actin reference level. *cg20655070* ($P = 0.02$), *SLC35F1* ($P = 0.0007$), *TAC1* ($P = 0.03$), *ZNF132* ($P < 0.0001$), and *ZNF542* ($P = 0.009$) exhibited significantly higher methylation levels in tumor vs normal esophageal biopsy tissues. *(C)* Biomarker performance for each gene including the AUC, *P* value, and 95% CI.