BASIC AND TRANSLATIONAL—ALIMENTARY TRACT

Hypomethylation of Noncoding DNA Regions and Overexpression of the Long Noncoding RNA, *AFAP1-AS1*, in Barrett's Esophagus and Esophageal Adenocarcinoma

WENJING WU,[1,2,*] TUSHAR D. BHAGAT,[3,*] XUE YANG,[2] JEE HOON SONG,[2] YULAN CHENG,[2] RACHANA AGARWAL,[2] JOHN M. ABRAHAM,[2] SARIAT IBRAHIM,[2] MATTHIAS BARTENSTEIN,[3] ZULFIQAR HUSSAIN,[3] MASAKO SUZUKI,[3] YITING YU,[3] WEI CHEN,[1] CHARIS ENG,[4] JOHN GREALLY,[3] AMIT VERMA,[3] and STEPHEN J. MELTZER[2]

[1]Center for Laboratory Medicine, The First Affiliated Hospital, School of Medicine, Xi'an Jiaotong University, Xi'an, China; [2]Division of Gastroenterology, Departments of Medicine and Oncology and Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland; [3]Albert Einstein College of Medicine, Bronx, New York; and [4]Cleveland Clinic, Cleveland, Ohio

BACKGROUND & AIMS: Alterations in methylation of protein-coding genes are associated with Barrett's esophagus (BE) and esophageal adenocarcinoma (EAC). Dysregulation of noncoding RNAs occurs during carcinogenesis but has never been studied in BE or EAC. We applied high-resolution methylome analysis to identify changes at genomic regions that encode noncoding RNAs in BE and EAC. **METHODS:** We analyzed methylation of 1.8 million CpG sites using massively parallel sequencing-based HELP tagging in matched EAC, BE, and normal esophageal tissues. We also analyzed human EAC (OE33, SKGT4, and FLO-1) and normal (HEEpic) esophageal cells. **RESULTS:** BE and EAC exhibited genome-wide hypomethylation, significantly affecting intragenic and repetitive genomic elements as well as noncoding regions. These methylation changes targeted small and long noncoding regions, discriminating normal from matched BE or EAC tissues. One long noncoding RNA, *AFAP1-AS1*, was extremely hypomethylated and overexpressed in BE and EAC tissues and EAC cells. Its silencing by small interfering RNA inhibited proliferation and colony-forming ability, induced apoptosis, and reduced EAC cell migration and invasion without altering the expression of its protein-coding counterpart, *AFAP1*. **CONCLUSIONS: BE and EAC exhibit reduced methylation that includes noncoding regions. Methylation of the long noncoding RNA *AFAP1-AS1* is reduced in BE and EAC, and its expression inhibits cancer-related biologic functions of EAC cells.**

Keywords: Esophageal Cancer Progression; Tumor Development; Gene Regulation; Noncoding RNA.

Esophageal adenocarcinoma (EAC) is one of the fastest-growing cancers in the Western world. Ninety-five percent of EACs occur without any antecedent diagnosis of Barrett's esophagus (BE), and the 5-year survival rate is only 15% in this group of patients.[1] The rising incidence and poor prognosis of EAC have intensified research efforts into earlier methods to detect this disease.

Recently, increasing evidence has shown that eukaryotic transcriptomes and genomes are not the simple, well-ordered substrates of gene transcription that they were once believed to be. It is now known that genomes are transcribed into a broad spectrum of RNA molecules, ranging from long protein-encoding messenger RNAs (mRNAs) to short noncoding transcripts, which frequently overlap or are interleaved on either strand.[2] "Noncoding regions" refer to RNAs that are transcribed into RNA but not translated to protein. These noncoding regions are interspersed throughout genomic DNA. One subcategory of these transcripts, called long noncoding RNAs (lncRNAs), comprise noncoding RNA more than 200 nucleotides in length. lncRNAs are pervasively transcribed in the genome, but our understanding of the functions of these lncRNAs is limited. lncRNA transcription was previously believed to represent random transcriptional noise. However, expression levels of lncRNA transcripts have been observed to vary spatially, temporally, and in response to various stimuli.[2,3] Moreover, several lncRNAs exhibit very precise expression patterns in different tissues. For example, Mercer et al observed exquisite patterning of lncRNA expression in the mouse brain, both in the tissue as a whole and in subcellular locations.[4] Similarly, the expression of some lncRNAs has also been shown to be developmentally regulated.[5]

[*]Authors share co-first authorship.

Abbreviations used in this paper: bp, base pairs; BE, Barrett's esophagus; EAC, esophageal adenocarcinoma; FDR, false discovery rate; lncRNA, long noncoding RNA; mRNA, messenger RNA; NE, normal esophagus; PCR, polymerase chain reaction; siRNA, small interfering RNA.

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2013.01.019

Despite this remarkable diversity in RNA species, only a few dysregulated lncRNAs have been implicated in cancer in humans.[6,7] Examples include MALAT-1 in lung cancer,[8] HULC in hepatocellular carcinoma,[9] and PCGEM1 in prostate cancer,[10] suggesting that lncRNAs may be involved in tumorigenesis or tumor progression. However, to our knowledge, studies of lncRNAs in EAC have not yet been reported.

In addition, emerging research has suggested mechanisms underlying the regulation of coding gene expression by lncRNAs. For example, lncRNAs can regulate chromosome structure in cis (XIST)[11] or in trans (HOTAIR).[12] Other lncRNAs modulate the activity of protein-binding partners.[13,14] Many lncRNAs are antisense to protein-coding genes and may function by regulating splicing, editing, transport, translation, or degradation of their corresponding coding mRNA transcripts.[15] In addition, lncRNAs may be posttranscriptionally processed into short non–protein-coding RNAs, which in turn regulate gene expression.[16]

In our previous study,[17] unsupervised hierarchical clustering analyses showed that, at the level of the transcriptome, squamous mucosa clustered discretely from "glandular" epithelium (including gastric cardia as well as all stages of progression of BE); in contrast, at the level of the epigenome, "normal" mucosa (including both squamous and gastric cardia) clustered discretely from all "abnormal" (ie, BE) epithelia. These results showed similarity of epigenetic profiles between otherwise normal gastrointestinal tissues, despite obvious morphological differences. Having established this finding previously, our focus in the present study was to study epigenetic differences between normal esophagus (NE) and BE at a much higher resolution on the whole-genome level. Following this initial step, we sought to characterize lncRNAs that were both differentially methylated and differentially expressed in EAC versus NE. We found that one such differentially regulated and methylated lncRNA, *AFAP1-AS1*, was derived from the antisense strand of DNA at the *AFAP1* coding gene locus and was hypomethylated and up-regulated in EAC tissues and cell lines. Inhibition of its expression in EAC cells resulted in diminished cell growth, migration, and invasion, as well as in increased apoptosis, thereby establishing, to our knowledge for the first time, a functional cancer-related consequence of epigenetic alteration at a lncRNA genomic locus. A schematic summary of experiments and a diagram of proposed *AFAP1-AS1* mechanisms of action are shown in Supplementary Figure 1*A–B*, respectively.

Materials and Methods

Cell Culture

This study used 3 established human EAC cell lines (OE-33, SK-GT-4, and FLO-1) as well as human primary normal nonimmortalized esophageal epithelial cells (HEEpic; ScienCell Research Laboratories, Carlsbad, CA).

Tissue Specimens

Primary tissue samples were obtained at endoscopy performed for clinical diagnostic indications. All patients provided written informed consent under protocols approved by institutional review boards at The Johns Hopkins University School of Medicine, University of Maryland School of Medicine, or Baltimore Veterans Affairs Medical Center. All tissue samples were pathologically confirmed as NE, BE, or EAC. Specimens were stored in liquid nitrogen before RNA extraction. Three sets of NE/BE samples were studied by HELP-tagging analysis. Twelve pairs of NE/BE samples and 20 pairs of NE/EAC samples were also studied for differential expression of both *AFAP1* and *AFAP1-AS1*.

HELP Tagging for Genome-Wide Methylation Analysis

The HELP-tagging assay applies massively parallel sequencing to analyze the status of 1.8 million CpGs distributed across the entire genome.[18] To perform HELP-tagging assays,[18] DNA samples were digested with *Hpa*II and ligated to customized Illumina (San Diego, CA) adapters with a complementary cohesive end. These adapters also contain an *EcoP15*I site that cuts into the adjacent sequence 27 base pairs (bp) away, allowing us to polish that end and ligate the other Illumina adapter for library generation by polymerase chain reaction (PCR). The presence of the CCGG and *EcoP15*I sequences at the ends of the reads allowed us to remove spurious sequences. We normalized the *Hpa*II signal with that of the deeply sequenced *Msp*I profiles, as performed previously.[18] Results were generated using the WASP system and linked to a local mirror of the UCSC Genome Browser for visualization.

Methylation Analysis

HELP-tagging data were analyzed using an automated pipeline as described previously.[18] Loci were defined in a continuous variable model, given the quantitative nature of this and comparable published assays.[19] Methylation values were depicted from a range of 0 to 100, with 0 representing fully methylated to 100 representing fully hypomethylated loci. Mean methylation values for noncoding regions were obtained by averaging values over the whole transcript region.

Quantitative DNA Methylation Analysis by MassArray Epityping

Validation of HELP microarray findings was performed by matrix-assisted laser desorption/ionization time of flight mass spectrometry using EpiTyper by MassArray (Sequenom, San Diego, CA) on bisulfite-converted DNA as previously described.[17,20,21] MassArray primers were designed to cover the flanking *Hpa*II sites for a given locus, as well as any other *Hpa*II sites found up to 2000 bp upstream of the downstream site and up to 2000 bp downstream of the upstream site, to cover all possible alternative sites of digestion.

Genomic Annotations

Genomic coordinates were obtained from HG18 build of the human genome from the UCSC browser using RefSeq annotations. Genomic regions 2 kilobases upstream and downstream of the transcription start sites were annotated as promoters. Two-kilobase flanking regions around the edges of CpG islands were annotated as CpG shores. RefSeq annotations with an NR prefix were categorized as noncoding transcripts. A size cutoff of 200 bp was used to distinguish between small and large noncoding transcripts.[22]