Original Article

Qingjie Min, Min Zhang, Dongmei Lin, Weimin Zhang, Xianfeng Li, Lianmei Zhao, Huajing Teng, Tao He, Wei Sun, Jiawen Fan, Xiying Yu, Jie Chen, Jinting Li, Xiaohan Gao, Bin Dong, Rui Liu, Xuefeng Liu, Yongmei Song, Yongping Cui, Shih-Hsin Lu, Ruiqiang Li, Mingzhou Guo*, Yan Wang* and Qimin Zhan*

Genomic characterization and risk stratification of esophageal squamous dysplasia

https://doi.org/10.1515/mr-2024-0008
Received January 29, 2024; accepted April 15, 2024;
published online May 15, 2024

Abstract

Objectives: The majority of esophageal squamous dysplasia (ESD) patients progress slowly, while a subset of patients can undergo recurrence rapidly or progress to invasive cancer even after proper treatment. However, the molecular mechanisms underlying these clinical observations are still largely unknown.

Methods: By sequencing the genomic data of 160 clinical samples from 49 tumor-free ESD patients and 88 esophageal squamous cell carcinoma (ESCC) patients, we demonstrated lower somatic mutation and copy number alteration (CNA) burden in ESD compared with ESCC.

Results: Cross-species screening and functional assays identified *ACSM5* as a novel driver gene for ESD progression. Furthermore, we revealed that miR-4292 promoted ESD progression and could serve as a non-invasive diagnostic marker for ESD.

Qingjie Min, Min Zhang, Dongmei Lin, Weimin Zhang, and Xianfeng Li contributed equally to this work.

Qimin Zhan is the lead contact.

***Corresponding authors: Mingzhou Guo**, Department of Gastroenterology & Hepatology, Chinese PLA General Hospital, #28 Fuxing Road, Beijing 100853, China, E-mail: mzguo@hotmail.com; **Yan Wang**, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China, E-mail: wy305@126.com; and **Qimin Zhan**, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China; Peking University International Cancer Institute, Peking University, Beijing 100191, China; Soochow University Cancer Institute, Suzhou 215000, China; State Key Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China; and Research Unit of Molecular Cancer Research, Chinese Academy of Medical Sciences, Beijing 100021, China, E-mail: zhanqimin@bjmu.edu.cn. https://orcid.org/0000-0002-1731-938X

Qingjie Min, Weimin Zhang, Jiawen Fan, Jie Chen, Jinting Li and Rui Liu, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing, China

Min Zhang and Ruiqiang Li, Novogene Co. Ltd, Beijing, China

Dongmei Lin and Wei Sun, Department of Pathology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Peking University Cancer Hospital & Institute, Beijing, China

Xianfeng Li, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing, China; and Department of Gastroenterology, Daping Hospital, Army Medical University (Third Military Medical University), Chongqing, China

Lianmei Zhao, Research Center, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

Huajing Teng, Department of Radiation Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Peking University Cancer Hospital & Institute, Beijing, China

Tao He, Department of Gastroenterology & Hepatology, Chinese PLA General Hospital, Beijing, China; and Department of Pathology, Characteristic Medical Center of Chinese People's Armed Police Force, Tianjin, China

Xiying Yu and Shih-Hsin Lu, Department of Etiology and Carcinogenesis and State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Xiaohan Gao and Yongmei Song, State Key Laboratory of Molecular Oncology, Cancer Institute and Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Bin Dong, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Central Laboratory, Peking University Cancer Hospital & Institute, Beijing, China

Xuefeng Liu, Institute of Cancer Stem Cell, Dalian Medical University, Dalian, Liaoning, China

Yongping Cui, Shenzhen Peking University-The Hong Kong University of Science and Technology (PKU-HKUST) Medical Center, Peking University Shenzhen Hospital, Shenzhen, Guangdong, China

Conclusions: These findings largely expanded our understanding of ESD genetics and tumorigenesis, which possessed promising significance for improving early diagnosis, reducing overtreatment, and identifying high-risk ESD patients.

Keywords: esophageal squamous dysplasia; genomic alteration; *ACSM5*; early diagnosis; miR-4292

Introduction

Esophageal cancer is the sixth leading cause of cancer-related death worldwide, with over 500,000 deaths annually [1–3]. More than 70 % of the global esophageal cancer cases occur in China, with esophageal squamous cell carcinoma (ESCC) being the most predominant histological subtype. The five-year survival rate for ESCC patients remains less than 25 % [4], however, if patients are diagnosed at an early stage, the five-year survival rate may reach 80 % or higher [5, 6]. Thus, to improve the clinical outcomes of ESCC patients, an effective early detection is urgently needed.

Molecular alterations identified in precancerous lesions will be beneficial to early diagnosis and stratification of high-risk populations [7]. Esophageal squamous dysplasia (ESD) is the precursor lesion of ESCC, and multiple steps ranging from low-grade dysplasia (LD) to high-grade dysplasia (HD), are involved in its pathogenesis [5]. An intriguing and distinctive feature of ESD is that it can either progress to malignancies or regress to normal epithelia. In clinical practice, ESD patients have to undergo surgical resection or endoscopy every 3–6 months, which may pose a risk of overtreatment. But without appropriate treatment, ESD patients may take the risk of recurrence and ultimately progress to invasive cancer. However, accurately predicting the fate of ESD is difficult, due to the shortage of ideal biopsy materials for identifying biomarkers which can be used for risk assessment. In recent decades, extensive application of endoscopy incorporated with Lugol's iodine staining or narrow-band imaging (NBI) provides an opportunity to identify early lesions, thus deciphering genetic alterations associated with risk evaluation.

However, to date, genetic alterations of ESD were mainly derived from different regions of surgical patients with invasive cancer (Ib or higher stage) [8, 9]. Given that tumor cells can remodel the neighboring microenvironment through "field cancerization" [10], it is difficult to distinguish whether the identified genetic alterations of ESDs are the initiation or the consequence of malignant transformation.

To uncover the genomic alterations underlying these clinical phenomena, we sequenced and analyzed 72 lesions from 49 tumor-free ESD patients (Table S1), and collected the long-term prognosis data with a 34.77 months' median follow-up (range 3.33–106.27 months), and finally explored the feasibility of clinical application of non-invasive early detection.

Materials and methods

Samples characteristics and cohort design

In total, 66 healthy donors, 141 ESD patients, 88 ESCC patients and 13 NMBzA rats were used in this study. The flow chart of the present study and the number of samples and/or cases for each analysis are shown in Figure S1. Briefly, we collected 72 ESD samples and corresponding 49 normal samples from 49 esophageal squamous cell carcinoma-free patients (ESD cohort) for whole-exome sequencing (WES) and genomic analyses, then compared with 88 ESCC dataset (ESCC cohort) and 13 NMBzA-induced rat models. Samples were collected from 2007 to 2016 and no patients had received chemotherapy or radiotherapy (Table S1). 1) For somatic single nucleotide variants (SNVs) and insertions and deletions (InDels) analysis, 72 ESD samples with matched normal samples, including 3 SH, 27 LD and 42 HD samples (Figure S1), were used, and then extended analyses with 88 ESCC samples from our published dataset [11]; 2) For CNA analysis, 72 ESD samples and 88 ESCC samples were used; 3) The major findings were confirmed sequentially in our independent ESD cohort (66 patients) by Sanger sequencing, and a cohort of NMBzA model of 13 rats by WGS; 4) For the clinical application studies, we used sera of 26 HD patients as well as 66 healthy controls to explore the potential of genomic alterations in ESD non-invasive diagnosis. All samples were obtained with the approval of the ethics committee of Chinese People's Liberation Army General Hospital, Peking University Cancer Hospital and The Fourth Hospital of Hebei Medical University. All patients provided written informed consent before enrollment in this study.

Establishment of esophageal cancer model in rat using NMBzA

NMBzA (purity>98 %) was generously provided by Professor Shih-Hsin Lu. Dimethyl sulfoxide (DMSO) was purchased from Sigma-Aldrich Inc (St. Louis, MO, USA). All procedures involving rats were carried out in accordance with the standards approved by the Ethical Committee of National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences. Male SD rats (3–4 weeks old) were purchased from Vital River Laboratory Animal Technology Co., Ltd (Beijing, China). Rats were housed two animals per cage under standard conditions (24 ± 2 °C, 40 %–70 % relative humidity, and 12 h light/dark cycles) and with unlimited access to standard rodent maintenance feed (KEAO XIELI FEED Co., Ltd, Beijing, China) and water. Male SD rats were administrated with NMBzA (0.30 mg/kg) three times/week by subcutaneous injections, and lasted for 12 weeks to induce the development of ESD. From week 16 to 44, rats were euthanized at a

four-week interval. Each esophagus was opened longitudinally, epithelium lesion and tumors were collected.

Cell lines and cell culture

Immortalized esophageal epithelium cell line HET-1A was cultured in high glucose DMEM (Gibco) with 10 % fetal bovine serum (FBS). All of these cells were maintained at 37 °C with 5 % CO_2.

Whole-exome and whole-genome sequencing

For human whole exome sequencing, capture libraries were prepared from 200 ng genomic DNA (gDNA) using Agilent SureSelect Human All Exon V6 kit (Agilent Technologies, CA, USA) following manufacturer's recommendations. Fragmentation was carried out by hydrodynamic shearing system (Covaris, Massachusetts, USA) to generate 280 bp fragments. DNA fragments with ligated adapters on both ends were selectively enriched in a PCR reaction which hybridized with liquid phase with biotin-labeled probes. All of the constructed libraries were sequenced on HiSeqX platform with 2×150 bp paired-end reads.

For whole genome sequencing (WGS) of rat, paired-end DNA library were prepared according to manufacturer's instructions (Illumina Truseq Library Construction). Two microgram genomic DNA was sheared into 350 bp fragments, followed by end-polished, A-tailed, and ligation with the adaptors. The concentration and the size distribution of the libraries were determined on an Agilent Bioanalyzer DNA 1000 chip. DNA libraries were sequenced on Illumina Hiseq X according to manufacturer's instructions and paired-end reads were generated.

Somatic mutations detection

All data were rigorously filtered to remove any reads containing adapter sequence and low-quality reads. High quality paired-end reads were aligned to GRCh38 (for human samples) or rn6 (for rat samples) reference genome using BWA (v0.7.17) [12]. Picard (v1.119) was used to mark duplicate reads caused by PCR. Somatic mutations were identified using consensus calling pipeline from Pan-Cancer Analysis of Whole Genomes (PCAWG) [13], which consisted of multiple calling approaches, including GATK4 (v4.1.0.0, MuTect2) [14, 15], CaVEMan (v1.15.3) [16], Bcftools (v1.16) [17], MuSE (v2.0) [18] for calling somatic SNVs, and MuTect2, cgpPindel (v3.10.0) [19], Platypus (v0.8.1) [20] for calling somatic InDels. All mutations were annotated using ANNOVAR (v2020Jun7) [21]. Somatic mutations were subjected to a rigorous filtering, including coverage of mutated sites ≥10 reads, at least four mutated reads, and variant allele frequency of somatic mutations ≥0.02 were retained. To filter possible polymorphic sites of rat, we removed any variants that present in rat dbSNP (build 149) or control samples and those with at least two occupies in different rat individuals. Only variants on Cancer Gene Census (CGC) genes or rat orthologues of human genes were selected for further analysis. Rat:human ortholog pair assignment was retrieved from Ensembl compara [22] through BioMart, and rat orthologues with high confidence were used for screening. Somatic CNAs were identified by CNVkit (v0.9.6) while adjusting for the tumor cell fraction or purity (ABSOLUTE v1.0.6) [23, 24]. All control samples were used to construct the copy number reference.

We used weighted-genomic integrity index (wGII) method to assess CIN dynamic during ESD progression [25]. Briefly, we calculated integer ploidy number of each ESD and ESCC sample, then evaluated the percentage of genomic alteration (gain or loss) for every autosomal chromosome. The mean percentage of genomic alteration was then calculated across all 22 autosomal chromosomes, and determined as wGII score.

Identification of significantly mutated genes (SMGs)

Significantly mutated genes were identified with SMG module of the MuSiC2 (v0.2) tools suite [26]. Firstly, MuSiC2 calculated the background mutation rate (BMR) of seven mutational categories, including AT transitions, AT transversions, CpG transitions, CpG transversions, CG (non-CpG) transitions and transversions, and indel category. After comparison between those respective mutation rate and BMR, seven ps were generated for each gene. Three SMG test, convolution test (CT), Fisher's combined p test (FCPT) and likelihood ratio test (LRT), were applied to summarize all ps. We also compared identified SMGs with previously published driver genes in ESCC studies.

Mutational signature analysis

SigProfilerExtractor [27] was applied to extract *de novo* mutational signatures, to decompose *de novo* mutational signatures to known COSMIC signatures (v3.3) [28], then assigned *de novo* mutational signatures to each sample by YAPSA [29]. For ESCC cohort, only somatic mutations overlapped with exonic regions were used for signature analysis. For ESD cases with multiple samples, only higher grade sample was selected for signature analysis. Possible sequencing artefacts were removed according to COSMIC signatures. Clustering of mutational signatures and clinical data was performed by R package pheatmap (v1.0.12) [30].

Validation of somatic mutations

Somatic mutations were validated by PCR amplification and Sanger sequencing. The PCR primers for putative somatic mutations were designed by primer 5 in silicon. PCR was performed on a Dual 96-well GeneAmp PCR System 9700 (Applied Biosystems), 20 ng template DNA from each sample was used per reaction. The products were sequenced by a 3730xl DNA Analyzer (Applied Biosystems). All sequences were analyzed by the Sequencing Analysis Software Version 5.2 (Applied Biosystems). If the mutations were successfully confirmed in the tumors but not identified in the matched normal DNA, the somatic statuses would be determined as successful validation.

Identifying significant frequent copy number alterations

GISTIC (v2.0.23) [31] algorithm was used to infer recurrently amplified or deleted genomic regions. G-scores were calculated for genomic and gene-coding regions on the basis of the frequency and amplitude of amplification or deletion affecting each gene. A significant CNA region was defined as having amplification or deletion with a G-score>0.1, corresponding to a p threshold of 0.05 from the permutation-derived null distribution.

Colony formation assay and MTS assay

Cells transfected with *ACSM5* siRNAs or stable expressed wild-type or mutated *ACSM5* were seeded into 6-well culture plates and incubated at 37 °C with 5 % CO_2 for 10 days. Culture plates were performed in triplicate. After washing with pre-cooled PBS, cultures were fixed with pre-cooled methanol for 20 min and stained with crystal violet for 15 min. Colonies were examined and automatically calculated by G:box (Syngene). MTS assay was performed according to the manufacturer's instructions.

RNA isolation and quantitative real-time PCR

RNA isolation was performed as described in our previous study [32]. Briefly, total RNA of cultured cells was extracted with TRIzol reagent (Invitrogen, Carlsbad, California, USA) and 100 μL blood plasma was applied to RNA isolation by acid phenol-chloroform. The cDNA was synthesized with the PrimeScript RT reagent Kit (Promega, Madison, WI, USA). Real-time PCR was carried out using an ABI 7500 real time PCR system (Applied Biosystems, Foster City, California, USA). Bulge-Loop hsa-miR-4292 qRT–PCR Primer Set (Ribobio, miRQ0016919-1-2) was used for the measurement of the relative quantity of hsa-miR-4292.

Western blot

Western blot was performed as described in our previous study [32]. The antibodies included antibodies to ACSM5 (Sigma, St Louis, MO, USA), Flag tag (Abcam, USA), GFP (Cell Signaling, Boston, MA, USA), β-actin (Sigma, St Louis, MO, USA). Secondary antibodies such as goat-anti-mouse IgG (1:2,000) and goat-anti-rabbit IgG (1:3,000) conjugated with horseradish peroxidase (HRP) were used to probe membrane for 1 h. The membrane was rinsed in 1 × PBS with 0.1 % Tween. After incubation with the Chemiluminescence substrate, photographs were taken by Image Reader LAS-4000 (Fujifilm) and analyzed by the Multi Gauge V3.2 software.

Transwell migration assays

Migration assay was performed as described in our previous study [32]. In brief, migration of cells was assayed in Transwell cell culture chambers with 6.5 mm diameter polycarbonate membrane filters containing 8 μm pore size (Neuro Probe, Gaithersburg, MD, United States). After transfected with has-miR-4292 mimic or control, 1×10^5 HET-1A cells in 100 μL of serum-free medium were added to the upper chamber of the device, and the lower chamber was filled with 600 μL fresh media (added 20 % FBS). After 12 h of incubation at 37 °C, the non-migration cells were removed from the upper surface of the membrane with a cotton swab. The filters were then fixed in methanol for 10 min, stained with crystal violet solution for 1 h, and counted. Five random microscopic fields (×100) were counted per well and the mean was determined.

Immunohistochemistry

Immunohistochemistry (IHC) analysis was performed and diagnosed by two pathologists blindly on the paraffin-embedded ESCC tissue sections (Tissue Microarray) and the xenograft tumor tissues. In brief, the sections were deparaffinized with xylenes and rehydrated in graded ethanol. Sections were submerged into EDTA antigenic retrieval buffer (pH=8.0) and microwaved for antigenic retrieval. The sections were then treated with 3 % hydrogen peroxide in methanol to quench the endogenous peroxidase activity, followed by incubation with 1 % goat serum albumin to block nonspecific binding. The tissue sections were incubated with rabbit anti-ACSM5 (1:200; Sigma) overnight at 4 °C. After washing, the tissue sections were treated with goat anti-mouse/rabbit IgG HRP-polymer (ZSGB-BIO, Beijing, China) for 20 min. 3,3′-Diaminobenzidine was used as the chromogen. The scores were determined by combining the intensity of staining and the proportion of positively stained tumor cells as described previously [33].

Xenograft tumor formation in nude mice model

For xenograft tumor formation study, 5×10^6 HET-1A cells with stable expressed wild-type *ACSM5*, mutated *ACSM5* (p.C135delinsCMR) and control cells were injected subcutaneously into the right dorsal flank of BALB/c nude mice (10 mice per group). The tumor volume and weight of mice were measured twice per week. Tumors were then excised and embedded in paraffin for hematoxylin and eosin staining (H&E) and immunohistochemistry (IHC) analysis.

Statistical analysis

Two tailed Student's t-test was used to analyze the results expressed as Mean ± S.E.M. The survival curves were plotted by using Kaplan-Meier analysis and compared by log-rank test. Survival data were evaluated by multivariate Cox regression analysis. Differences were considered significant when the p-value was less than 0.05. Fisher's exact test was used to analyze the significance of the correlation between clinical data and mutated events. For mutational burden and signature, Wilcoxon rank sum test was applied. All statistical analyses were performed using the R (version 3.6.1) environment.

Results

Comparison of exonic mutational burden and signatures between ESD and ESCC

As part of the International Cancer Genome Consortium-Accelerating Research in Genomic Oncology (ICGC-ARGO) project, we recruited 141 ESD patients without invasive cancer by endoscopic submucosal dissection (Figure 1A, Figure S1). Slides stained with H&E were reviewed independently by three experienced pathologists to identify the consensus areas of morphological classification according to WHO fifth edition, and grouped into one of the following histological categories: simple hyperplasia (SH), low-grade dysplasia (LD) and high-grade dysplasia (HD) (Figure 1B). Seventy two ESD samples of different histological grades from 49 patients were subjected to whole-exome sequencing (WES) (Figure S1), with

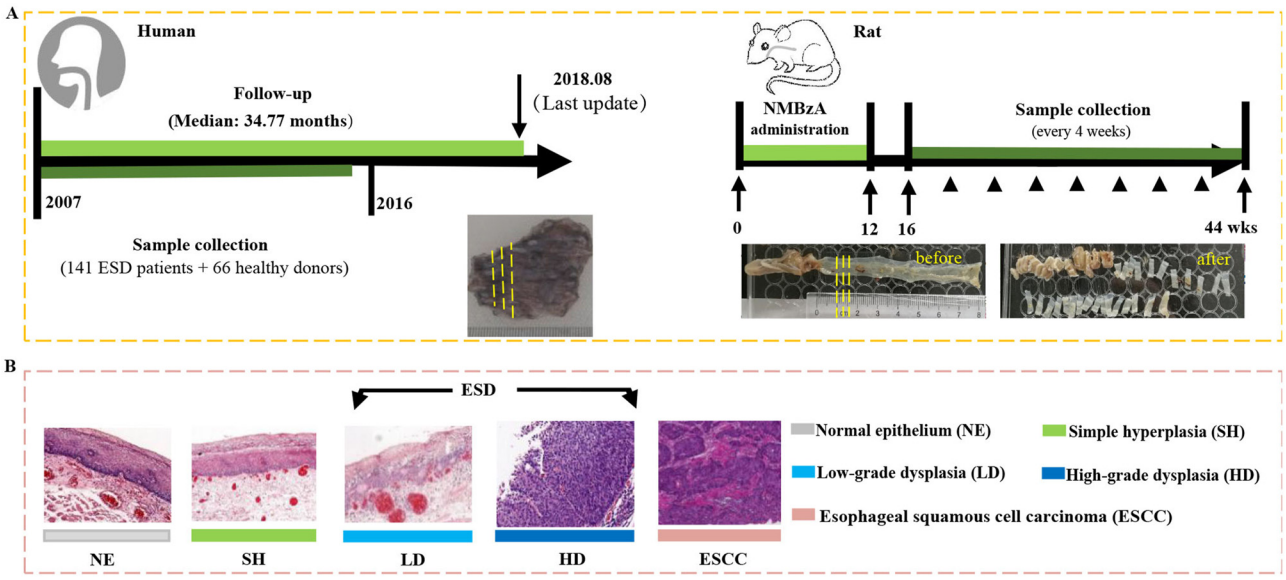

Figure 1: Overview of the study and the representative hematoxylin and eosin staining of esophageal epithelium and lesions. **(A)** Schematic representation of sample collection of patients and NMBzA rat model. **(B)** The representative hematoxylin and eosin staining of esophageal epithelium and lesions.

a median sequencing depth of 182× (range 106×–317×) (Table S2). The somatic mutations per megabase (Mb) ranged from 0.03 to 11.26, with a median of 0.70 mutations/Mb (Table S3). The mutational burden of LDs (0.34 mutations/Mb) was lower than that of HDs (1.21 mutations/Mb; Wilcoxon rank sum test, p=0.01), all of which were lower than previously reported LDs (3.3 mutations/Mb) and HDs (4.1 mutations/Mb) from invasive tumor-bearing patients [8]. Although SHs represented a lower mutational burden (0.18 mutations/Mb), we excluded SHs for mutation burden comparisons due to small sample size (n=3). Interestingly, when compared with our published dataset of 88 ESCC patients (2.84 mutation/Mb) [11], we observed both LDs and HDs harbored lower mutational burden than ESCC (Figure 2A; Wilcoxon rank sum test, $p=1.1 \times 10^{-8}$ and 2.1×10^{-9}, respectively), and this difference was not driven by the purity levels of ESD and tumor samples, since there was no significant correlation between the number of single nucleotide variants (SNVs) and the purity of those samples (Spearman's rho, $r=-0.159$, p=0.18).

We then used SigProfilerExtractor to extract *de novo* mutational signatures [27], to decompose *de novo* mutational signatures to known COSMIC signatures (v3.3) [28], and to assign *de novo* mutational signatures to each sample by YAPSA [29]. After excluding possible sequencing artefacts, we identified three *de novo* mutational signatures (Figure 2B). Each *de novo* signatures was matched to a combination of COSMIC signatures (Figure S2). Signature A was characterized by clock-like signatures (SBS5 and SBS1) and SBS18 signature (Damage by reactive oxygen species). Activity of APOBEC family of cytidine deaminases (SBS13 and SBS2) was found as the predominant signature in Signature B. Signature C matched a combination of COSMIC signatures SBS5, SBS37 (unknown aetiology), SBS22 (Aristolochic acid exposure), and SBS1. In addition, significantly different contributions of Signatures B and C were observed between ESD and ESCC, suggesting that various mutational processes contribute to tumor progression at different stages (Figure 2C). Clustering analysis based on mutational signatures displayed four distinct clusters across ESD and ESCC samples (Figure 2D). Interestingly, we observed that patients with a history of drinking or combined with smoking, were significantly enriched in cluster 4 (Fisher's exact test, p=0.002) which was dominated by Signature C. Collectively, these results revealed that ESD had distinct mutational burden and signatures compared with ESCC, and exogenous and endogenous factors contributed to mutagenesis during tumor progression.

Mutational landscape associated with ESD progression across species

We next investigated significantly mutated genes (SMGs) that facilitated ESD progression. Several well-known ESCC SMGs were identified in both ESD and ESCC patients

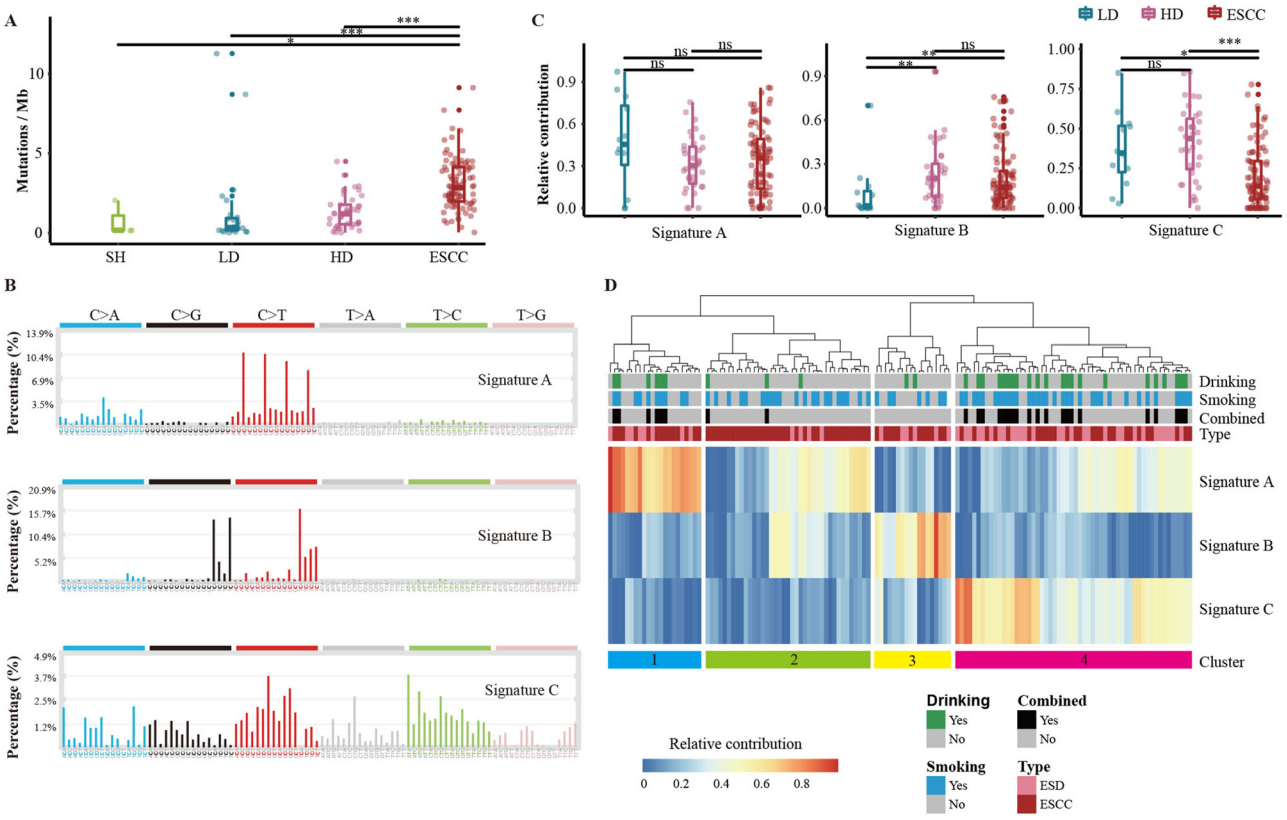

Figure 2: Comparison of mutational burdens and signatures of ESD and ESCC. **(A)** Mutational burden of simple hyperplasia (SH, n=3), low-grade dysplasia (LD, n=27), high-grade dysplasia (HD, n=42) and ESCC (n=88). The data were analyzed by the Wilcoxon rank-sum test, two-sided. *p<0.05, **p<0.01, ***p<0.001. **(B)** Base substitutions of three *de novo* mutational signatures. Base substitutions of each signature are shown as color bars. The height of each bar represents the proportion of substitution pattern. **(C)** Comparison of relative contributions of three *de novo* mutational signatures between dysplasia and ESCC samples. Wilcoxon rank-sum test. **(D)** Clustering of three *de novo* mutational signatures. Top, hierarchical clustering analysis of mutational signatures of ESD and ESCC patients; middle, clinical features indicated by color blocks, combined means the combination of drinking and smoking; bottom, relative contribution of three *de novo* mutational signatures in each individual.

(Figure 3A), including *TP53*, *NOTCH1*, *ZNF750*, *CDKN2A*, *FAT1*, and *PIK3CA* [11, 34–38], implying that these genes may play an important role in tumor development from initiation to invasive tumor. Subsequently, we analyzed ESD samples of two patients, each had multiple histological grades, including one SH, two LDs and one HD from Patient P040, one SH and one HD from Patient P023. Multiple mutations were found at SH or LD stage, and persisted until HD stage (Figure S3), for example, *TP53* non-sense mutation (p.R64X, Patient P023) and *NOTCH1* splice-site mutation (c.2969+2T>C, Patient P040) (Table S4). Furthermore, three shared *TP53* mutations were maintained and present in all stages during tumor progression (Figure 3B, Table S5). These results suggested that *TP53* and *NOTCH1* mutations may occur at early stages of tumorigenesis even at SH stage, and may confer a selective advantage on the cells with those mutations. Notably, *TP53* showed lower mutation frequencies in ESD compared with ESCC (odd ratio=0.1, p-value<0.0001, fisher's exact test), while *NOTCH1* showed a reverse pattern (odd ratio=2.83, p=0.011, fisher's exact test), indicating their differential contributions to tumor progression.

About 75 % of HDs progress to ESCC without proper treatment [5], which inspired us to figure out the mutations that occurred in HDs and ESCCs concurrently. We identified *ZNF750*, *EYS*, *MUC17* and *SMPD1* which emerged in HDs and persisted in ESCCs, but were absent in LDs. A previous study reported that *ZNF750* was a tumor suppressor, and the mutant *ZNF750* promoted cell proliferation, foci formation, migration, and xenograft tumor growth in multiple squamous cell carcinomas [39]. *MUC17* mutations was associated with poor prognosis in glioma cohorts [40]. It is possible that those mutations occurring at key transitional points will promote extensive mutagenesis and the development of invasive carcinoma [41].

In addition, new SMGs were identified across ESD samples, including *EYS*, *RERE*, *CDH9*, *CACNA2D1*, *PITPNB*, *RBP3*, *ACSM5*. We then employed a well-established rat ESCC model [42, 43] to investigate whether these driver events

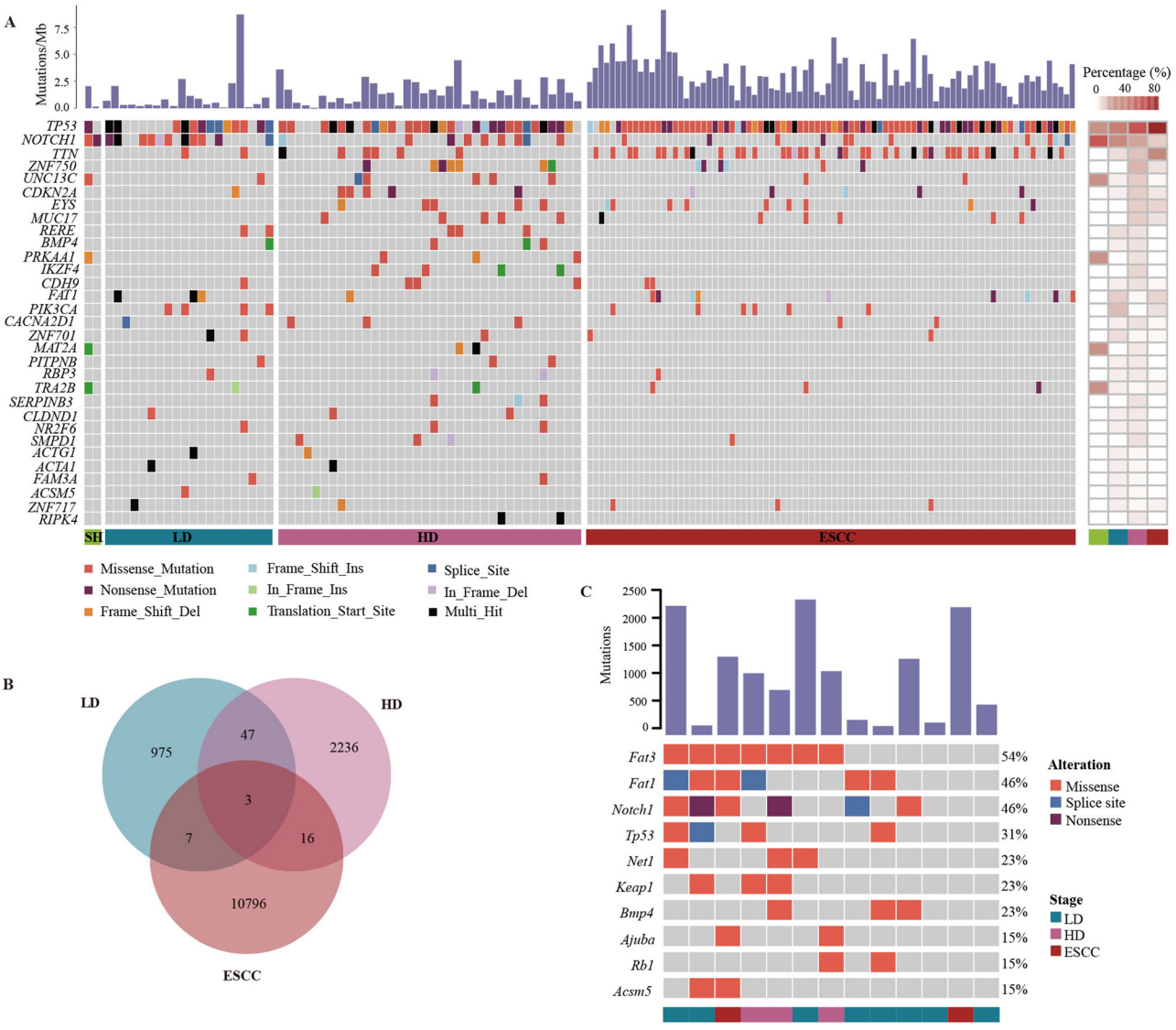

Figure 3: Significantly mutated genes in ESD and ESCC samples of patients and NMBzA rat model. **(A)** Mutational landscape of SH, LD, HD and ESCC patients. Top, mutations per megabase (Mb) of each sample; middle, significantly mutated genes colored by mutational types; right, the percentage of samples with specific mutated genes according to different stages. **(B)** Venn plot shows the shared and unique somatic mutations among LDs, HDs and ESCCs. **(C)** Mutational landscape of NMBzA rat model. Top, the number of non-synonymous mutations in each sample; bottom, significantly mutated genes colored by mutational types.

were conserved across species in the progression of ESD to ESCC. In brief, male SD rats were treated with NMBzA to induce the development of esophageal lesions (Figure 1A). At week 16–44, rats were euthanized at a four-week interval, and epithelium lesion and tumors were collected. In total, 13 samples of NMBzA induced rat esophagus lesions (n=8 for LD; n=3 for HD; n=2 for ESCC) and four normal samples were conducted with whole-genome sequencing (Tables S6 and S7). The mutation spectrum was dominated by C>T/G>A transitions (Figure S4 and Table S8). As expected, well-known ESCC SMGs, such as *Tp53*, *Notch1*, *Fat1*, were present in NMBzA model, and new SMGs were also identified, including *Net1*, *Bmp4* and *Acsm5* (Figure 3C). None of the two ESCC samples of NMBzA model borne *Tp53* mutations, the possible reason may be the small sample size (n=2) of ESCC samples of NMBzA model. Strikingly, about 70 % (9/13) of samples harbored mutations in Wnt signaling pathway, including *Fat1* and *Fat3*. This suggested the essential role of Wnt signaling pathway in the initiation of dysplasia. Of the new SMGs, *Acsm5* was detected in both ESD and ESCC samples of the NMBzA model. Driver genes shared between ESD patients and NMBzA model highlighted the importance of

these mutated genes in the malignant transformation of esophageal epithelial cells.

ACSM5 is a novel driver gene for ESD progression

Our cross-species genomic profiles identified a novel SMG, *ACSM5*, which was not been linked to tumorigenesis previously. *ACSM5* (acyl-CoA synthetase medium chain family member 5), catalyzes the first step of fatty acid oxidation with CoA to produce acyl-CoA, the activation status of fatty acid [44]. We observed that *ACSM5* was mutated in three of 72 ESD samples (2/49 patients, 4.08 %). Among the three *ACSM5* somatic mutations, p.C135delinsCMR reoccurred in our independent ESD cohort (66 patients) (Table S9), which prompted us to decipher the function of this mutation. Interestingly, knockdown of *ACSM5* in immortalized normal esophageal cell lines, HET-1A, promoted cell proliferation and colony formation (Figure 4A, C and D). Additionally, mutated *ACSM5* (p.C135delinsCMR) remarkably promoted the malignant phenotypes of immortalized normal esophageal cells *in vitro* and *in vivo* (Figure 4B–G).

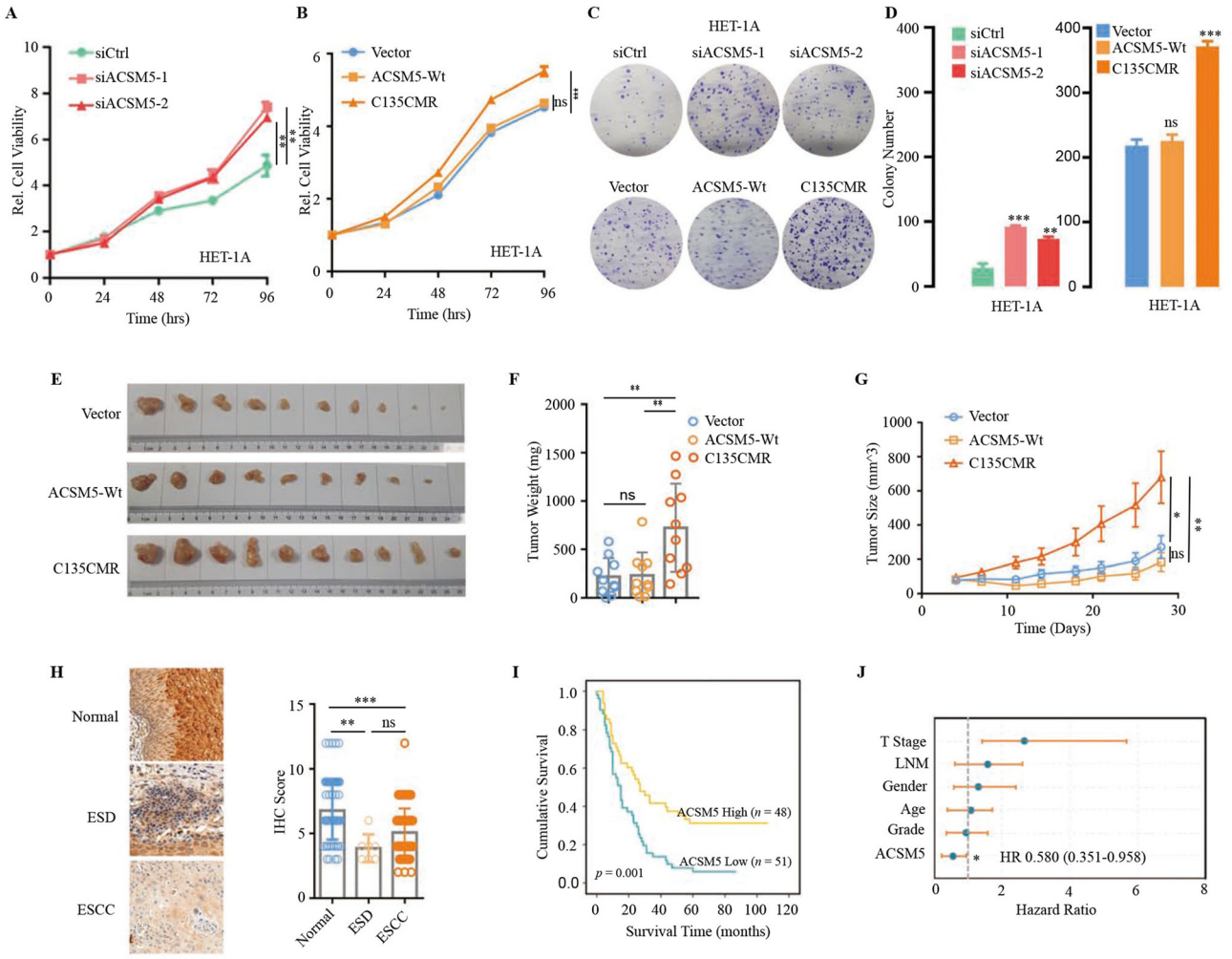

Figure 4: Altered expression of *ACSM5* or its mutant contributes to the malignant transformation of immortalized normal esophageal epithelial cell. **(A)** Knockdown of *ACSM5* by siRNAs promoted HET-1A cell proliferation (MTS assay). Three independent experiments of each siRNAs (siACSM5-1 and siACSM5-2) were performed. **(B)** *ACSM5* mutation (p.c135delinsCMR) increased HET-1A cell proliferation (MTS assay). (**C** and **D**) Knockdown of *ACSM5* by siRNAs enhanced the colony formation ability of HET-1A (up), *ACSM5* mutation (p.c135delinsCMR) enhanced the colony formation ability of HET-1A (bottom). **(C)** Representative results of colony formation. **(D)** Quantitative data of colony formation. (**E-G**) Mutated *ACSM5* (p.c135delinsCMR) significantly promoted tumor growth *in vivo*. **(E)** Representative tumor xenografts. **(F)** Tumor weight. **(G)** Tumor volume. **(H)** Expression level of *ACSM5* in normal, ESD and ESCC tissues (IHC assay). **(I)** Kaplan–Meier survival analysis of ACSM5 expression in ESCCs (Log-rank test). **(J)** Multivariate Cox regression survival analysis. *p<0.05, **p<0.01, ***p<0.001, ns, not significant. ESD, Esophageal squamous dysplasia; ESCC, esophageal squamous cell carcinoma.

Subsequently, when interrogating the whole exonic region of *ACSM5* in our independent ESD cohort (66 patients), we identified another four coding mutations (Table S9). Additionally, we found six coding mutations of *ACSM5* in our previously published WGS cohort of ESCC patients (Table S10) [38]. We next examined protein expression of *ACSM5* in normal, ESD and tumor tissues. *ACSM5* was highly expressed in normal tissues compared to ESDs and ESCCs (Figure 4H). Statistical analysis revealed that the expression level of *ACSM5* was negatively correlated with American Joint Committee on Cancer (AJCC) stage (p=0.041), lymph node metastasis (LNM; p=0.027) and gender (p=0.044) (Tables S11 and S12). Notably, low expression of *ACSM5* was strongly associated with poor survival of ESCC patients (p=0.001, Kaplan–Meier survival analysis and log-rank test, Figure 4I), and the five-year survival rate in the low expression group of *ACSM5* (7.69 %) was substantially lower than that in the high expression group (31.25 %). The median survival in the low expression group of *ACSM5* was 15 months (95 % CI 10.048–19.952), while 27 months (95 % CI 17.309–36.691) in the high expression group. Multivariate Cox regression survival analysis adjusting for age, tumor stage, LNM, pathological grade and gender also reported the strong correlation between low expression of *ACSM5* and shorter survival (p=0.033, HR=0.580, 95 % CI 0.351–0.958; Figure 4J), indicating that the expression of *ACSM5* was an independent prognostic factor for clinical outcome in ESCC. Collectively, our findings suggested that *ACSM5* may act as a tumor suppressor, and mutated *ACSM5* or altered expression of *ACSM5* probably exerted tumor-promoting potential during tumor progression.

Profiling of copy number alterations and clinical application of miR-4292 as a diagnostic marker

Somatic CNA analysis revealed that the CNA burden of LDs and HDs were significantly lower than that of ESCCs (Wilcoxon rank sum test, p=0.025 and 2.4e-06, respectively) (Figure 5A). And this trend was also observed in chromosome instability (CIN) (Figure S5). LDs and HDs accumulated more CNAs during tumor development (Figure 5A and B). For well-defined CNAs in ESCC, we noticed that 29.63 % of LDs and 47.62 % of HDs, while 65.91 % of ESCCs had the most prevalent amplification region 11q13.3, which harbored *CCND1*, *FGF4*, *FGF19*, *MYEOV*, *TPCN2*, and our previously identified microRNA (*MIR548K*) (Figure 5B) [11, 32]. Other CNAs identified in our study included amplifications in 3q, 5p, 9q, and 8q, and deletions in 3p and 9p (Figure 5B).

We next analyzed the focal amplified region of 9q34.3, which harbored *NOTCH1*, and several miRNAs including miR-4292. Fluorescence *in situ* hybridization (FISH) assay validated that miR-4292 and *NOTCH1* were amplified in HD regions of patient (P025) and patient P030 (Figure 5C). Furthermore, overexpression of miR-4292 mimic significantly promoted cell proliferation, colony formation, cell mobility and invasiveness (Figure 5D–F). Cell-free circulating miRNAs are promising diagnostic biomarkers for early detection of cancer [45]. Thus, we sought to evaluate the diagnostic value of miR-4292 in serum samples of 26 ESD patients and 66 health donors. Interestingly, the serum abundance of miR-4292 was significantly higher in ESD patients than that in health controls (Figure 5G, Student's t-test, p<0.001), and also higher in publicly independent ESCC serum cohort and ESCC tissue cohort (Figure S6, Wilcoxon rank sum test, p<0.001) [46, 47]. Receiver operator characteristic (ROC) curve was used to determine the diagnostic accuracy of miR-4292 in this cohort, and the AUC was 0.774 with 95 % CI (0.678–0.870) (Figure 5H). Collectively, these findings illustrate the potential significance of miR-4292 in the progression of ESD and highlight its utility as a non-invasive diagnostic marker for early detection of ESD.

Discussion

This study delineated the landscape of genomic alterations in ESD, and revealed molecular dynamics during ESD progression. We optimized the tissues selection by using samples of the endoscopic submucosal dissection of ESD. Taking micro-environment into consideration, we recruited ESD patients without invasive cancer when tissue samples were taken.

Our results highlighted significant differences of mutation frequency between ESD and invasive carcinoma. Intriguingly, we demonstrated that *TP53* and *NOTCH1* mutations emerged at the early stage of dysplasia, even in simple hyperplasia stage. Based on analyses of the multi-region ESD patients, we noticed that these somatic mutations persisted during tumor evolution from ESD to ESCC, indicating their important roles during tumor initiation and progression. In addition, integrative analyses of NMBzA-induced rat progression model from ESD to ESCC and clinical cohorts of ESD patients revealed that major driver genes were shared across species. We also identified several novel SMGs of ESD, including *ACSM5*. Knockdown of *ACSM5* or mutated *ACSM5* remarkably promoted the malignant phenotypes of immortalized normal esophageal cells. Moreover, *ACSM5* was highly expressed in normal tissues compared to ESDs and ESCCs, and low expression of *ACSM5* was an

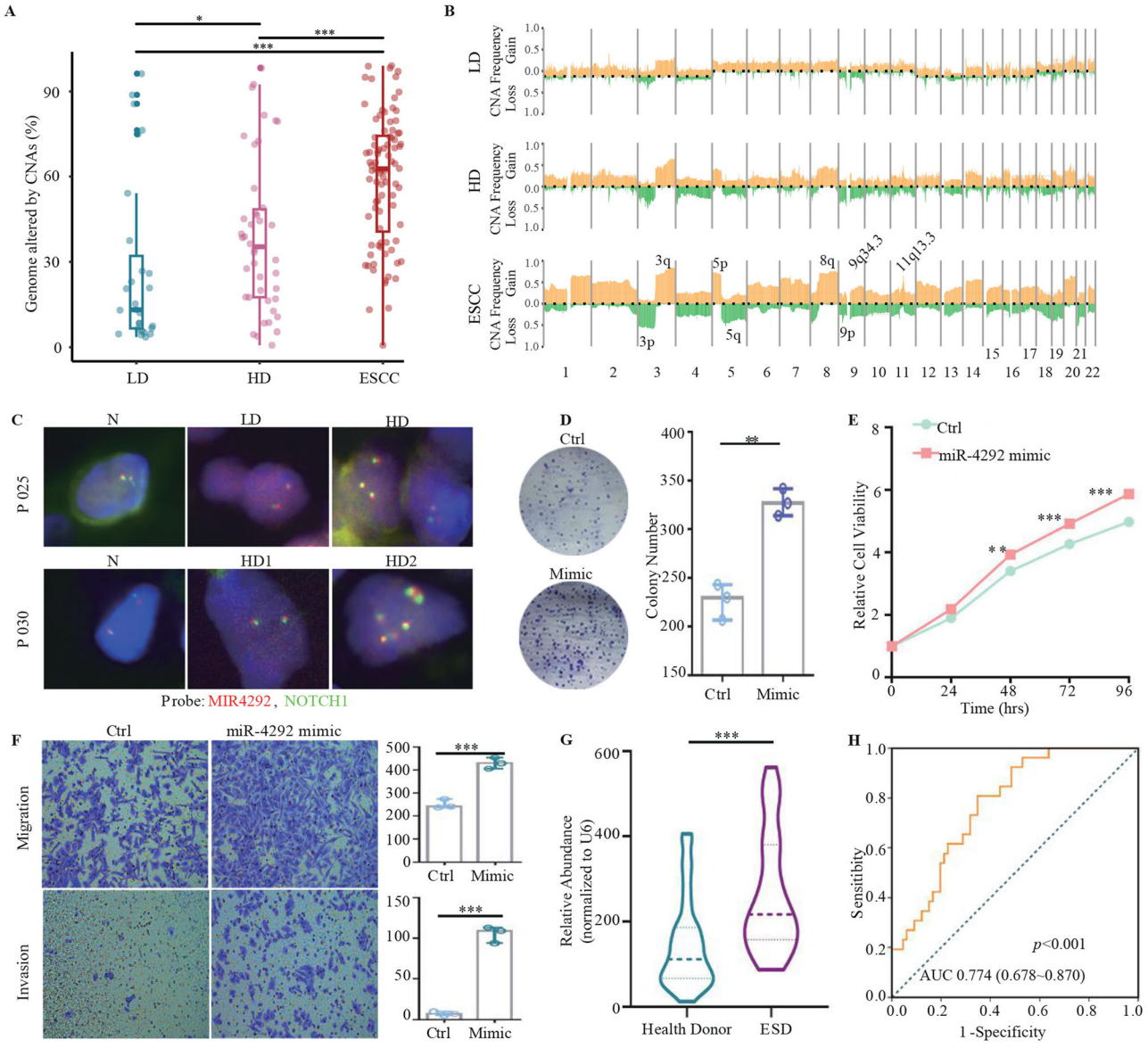

Figure 5: The landscape of CNAs. **(A)** Comparison of CNA burden between ESCC (n=88) and LD (n=27), HD (n=42), respectively. **(B)** CNAs profiling in SH, LD, HD and ESCC samples. Amplifications are denoted by yellow, and deletions are denoted by green. Y axis indicates CNA frequency. **(C)** Representative fluorescence *in situ* hybridization (FISH) of miR-4292 and NOTCH1 status. **(D)** miR-4292 mimic overexpression increased colony formation of HET-1A cells. **(E)** miR-4292 mimic overexpression promoted cell proliferation of HET-1A cells (MTS assay). **(F)** miR-4292 mimic overexpression enhanced cell mobility and invasiveness of HET-1A cells (Transwell assay). **(G)** The abundance of miR-4292 in serum across ESD patients and health donors were detected by real time PCR. Relative expression level was normalized to U6. **(H)** ROC curve evaluated the diagnostic accuracy of miR-4292 abundance. ROC, receiver operator characteristic. AUC, area under the curve. Error bars indicate SEM of three independent experiments. *p<0.05, **p<0.01, ***p<0.001.

independent prognostic factor for poor clinical outcome of ESCC patients. Taken together, these functional data indicate that *ACSM5* may act as a tumor suppressor, and altered expression of *ACSM5* or its mutant contributes to the malignant transformation of esophageal epithelial cells.

The genomic features of ESD obtained in this study were quite different from esophageal adenocarcinoma (EAC), another major pathological type of esophageal cancer in western countries. Matthew Stachler [48] and colleagues have reported that Barrett's esophagus, the precancerous lesion of EAC, even without dysplasia, still harbored substantial mutations. The discrepancy between our findings and the precancerous lesions of EAC may be mainly due to the differences in disease locations and pathogenic factors. Previous genomic studies reported that the tumor-bearing ESDs had comparable mutation/CNA burden and similar mutation spectrum when compared to ESCCs from the same individual [8, 9]. Significant correlations between atypical

and inflammatory levels of paired ESD and ESCC were also reported [9]. Since the majority of previous studies of ESDs were obtained from the patients who already had invasive cancer, it was difficult to rule out the effects of long-term inflammation and other factors released by invasive carcinoma. Chen et al. compared mutational burden of non-invasive ESD cohort to invasive ESD cohort, and found that non-invasive ESD cohort had significantly lower mutational burden, CNA level and genome doubling events than invasive ESD cohort [8]. In general, CNA level was positively correlated with gene expression [49], and we observed well-known ESCC amplification regions, such as 11q13.3 (*CCND1*, *MIR548K*) and 9q34.3 (*NOTCH1*, *MIR4292*) in our ESD cohort. We also validated that the copy numbers of these genes were positively correlated with corresponding gene expressions. A previous study reported that *NOTCH1* amplification and increased expression promoted the expansion of cancer associated fibroblast [50]. The underlying functions of *NOTCH1* amplification and other amplifications in ESD progression remain further investigation. As CNA level and genome doubling are indicator of chromosome instability (CIN), and ongoing CIN is pervasive during tumor initiation and development, which may modulate tumor heterogeneity and karyotypic complexity, CIN may be a valuable prognostic marker for tumorigenesis, and potential target for clinical therapy [51–53]. Thus, this study, using dysplasia without invasive cancer, provided genomic characterization of ESDs before progression to ESCC, and expanded our understanding of tumorigenesis and development.

Another highlight of this study was that we integrated the long-term follow-up data (median 34.77 months) and the clinical prognosis information of the ESD patients to identify high risk ESD. Although endoscopic submucosal dissection is the primary treatment for ESD to interdict its progression to ESCC [54], it is expensive and invasive, other effective and non-invasive method must be discovered to evaluate and monitor disease development. In the present study, we demonstrated that miR-4292 promoted malignant phenotypes of immortalized esophageal epithelium cells, and serum miR-4292 may be a non-invasive diagnostic marker for ESD.

Despite using the largest genomic dataset, we realized that our study remained underpowered to detect rare genetic alterations. In subsequent studies, particular attention should be paid to multi-omic analyses, such as RNA, protein and microenvironment level, to achieve a better understanding of the pathogenesis and development of ESD.

Acknowledgments: We appreciated Yiping Li (The University of Alabama at Birmingham), Bo Chen (Conmed Biosciences Co. Ltd.), and Yan Wang (Beijing Institutes of Life Science, Chinese Academy of Sciences) for revising this manuscript and offering constructive advices.

Informed consent: All patients provided written informed consent before enrollment in this study.

Ethical approval: All samples were obtained with approval of the ethics committee of Chinese People's Liberation Army General Hospital, Peking University Cancer Hospital and The Fourth Hospital of Hebei Medical University.

Author contributions: Conceptualization and funding acquisition, Q. Z. and Y. W.; investigation, Y. W., W. S., J. F., J. L., J. C., X. G., R. L. and B. D.; data curation and validation, M. G., D. L., L. Z., T. H., X. L., Y. S., Y. C., S. L. and R. L.; methodology, Y. W., Q. M., M. Z., X. L., W. Z., and X. Y.; writing, review, and/or editing of the manuscript, Q. Z., Y. W., Q. M., H. T., W. Z., and X. L.; project administration and supervision, M. G. and Q. Z.

Competing interests: Authors state no conflict of interest. Professor Qimin Zhan is a member of Medical Review editorial board and is not involved in the peer review and decision process of this article.

Research funding: This work was supported by the National Natural Science Foundation of China (81988101 and 81830086 to Q. Zhan, 82173152 to Y. Wang), 'Beijing Municipal Administration of Hospitals' Mission Plan (SML20181101), Guangdong Basic and Applied Basic Research Foundation (No. 2019B030302012), Science Foundation of Peking University Cancer Hospital (17-01, 2022-28), CAMS Innovation Fund for Medical Sciences (2019-I2M-5-081), PKU-Baidu Fund (Project 2019BD012), Suzhou Top-Notch Talent Groups (ZXD2022003), and also supported by National Institute of Health Data Science of Peking University.

Data availability: The raw ESD sequencing datasets for all human and rat samples of this study have been deposited in the BIG Data Center (https://ngdc.cncb.ac.cn/gsa-human/) at Beijing Institute of Genomics (BIG) of the Chinese Academy of Sciences [55, 56] with accession code HRA000144. Other datasets that support the findings of this study are available from the corresponding author (Qimin Zhan) upon reasonable request. Additionally, 71 WES and 17 WGS sequencing datasets of ESCC were obtained from our previous study with European Genome-phenome Archive accession number EGAS00001000709 [11]. The source code for this study is available in GitHub repository and can be accessed via this link https://github.com/minqing1/ESD.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
2. Qiu H, Cao S, Xu R. Cancer incidence, mortality, and burden in China: a time-trend analysis and comparison with the United States and United

Kingdom based on the global epidemiological data released in 2020. Cancer Commun 2021;41:1037–48.
3. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. Lancet Glob Health 2018;6:e555–67.
4. Min Q, Wang Y, Wu Q, Li X, Teng H, Fan J, et al. Genomic and epigenomic evolution of acquired resistance to combination therapy in esophageal squamous cell carcinoma. JCI Insight 2021;6:e150203.
5. Wang GQ, Jiao GG, Chang FB, Fang WH, Song JX, Lu N, et al. Long-term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening. Ann Thorac Surg 2004; 77:1740–4.
6. Yang S, Wu S, Huang Y, Shao Y, Chen XY, Xian L, et al. Screening for oesophageal cancer. Cochrane Database Syst Rev 2012;12: CD007883.
7. Campbell JD, Mazzilli SA, Reid ME, Dhillon SS, Platero S, Beane J, et al. The case for a pre-cancer genome atlas (PCGA). Cancer Prev Res (Phila) 2016;9:119–24.
8. Chen XX, Zhong Q, Liu Y, Yan SM, Chen ZH, Jin SZ, et al. Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by multi-region whole-exome sequencing. Nat Commun 2017; 8:524.
9. Liu X, Zhang M, Ying S, Zhang C, Lin R, Zheng J, et al. Genetic alterations in esophageal tissues from squamous dysplasia to carcinoma. Gastroenterology 2017;153:166–77.
10. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. Nat Rev Cancer 2018;18:19–32.
11. Song Y, Li L, Ou Y, Gao Z, Li E, Li X, et al. Identification of genomic alterations in oesophageal squamous cell cancer. Nature 2014;509:91–5.
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.
13. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature 2020;578:82–93.
14. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31: 213–9.
15. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20:1297–303.
16. Jones D, Raine KM, Davies H, Tarpey PS, Butler AP, Teague JW, et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. Curr Protoc Bioinf 2016; 56:15.0.1–.0.8.
17. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience 2021;10:giab008.
18. Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal PA, et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. Genome Biol 2016;17:178.
19. Raine KM, Hinton J, Butler AP, Teague JW, Davies H, Tarpey P, et al. cgpPindel: identifying somatically acquired insertion and deletion events from paired end sequencing. Curr Protoc Bioinf 2015;52: 15.7.1–.7.2.
20. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, et al. Integrating mapping-assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 2014;46: 912–8.
21. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
22. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 2009;19:327–35.
23. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016;12:e1004873.
24. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012;30:413–21.
25. Burrell RA, McClelland SE, Endesfelder D, Groth P, Weller M-C, Shaikh N, et al. Replication stress links structural and numerical cancer chromosomal instability. Nature 2013;494:492–6.
26. Dees ND, Zhang QY, Kandoth C, Wendl MC, Schierding W, Koboldt DC, et al. MuSiC: identifying mutational significance in cancer genomes. Genome Res 2012;22:1589–98.
27. Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genom 2022;2:100179.
28. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101.
29. Huebschmann D, Jopp-Saile L, Andresen C, Gu Z, Schlesner M. YAPSA: yet another package for signature analysis. R package; 2020.
30. Kolde R. Pretty heatmaps. R package; 2019.
31. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12:R41.
32. Zhang W, Hong R, Li L, Wang Y, Du P, Ou Y, et al. The chromosome 11q13.3 amplification associated lymph node metastasis is driven by miR-548k through modulating tumor microenvironment. Mol Cancer 2018;17:125.
33. Xue L, Ren L, Zou S, Shan L, Liu X, Xie Y, et al. Parameters predicting lymph node metastasis in patients with superficial esophageal squamous cell carcinoma. Mod Pathol 2012;25:1364–77.
34. Lin DC, Hao JJ, Nagata Y, Xu L, Shang L, Meng X, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. Nat Genet 2014;46:467–73.
35. Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, et al. Genomic landscape of esophageal squamous cell carcinoma in a Japanese population. Gastroenterology 2016;150:1171–82.
36. Hao JJ, Lin DC, Dinh HQ, Mayakonda A, Jiang YY, Chang C, et al. Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. Nat Genet 2016;48: 1500–7.
37. Yan T, Cui H, Zhou Y, Yang B, Kong P, Zhang Y, et al. Multi-region sequencing unveils novel actionable targets and spatial heterogeneity in esophageal squamous cell carcinoma. Nat Commun 2019;10:1670.
38. Cui Y, Chen H, Xi R, Cui H, Zhao Y, Xu E, et al. Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. Cell Res 2020;30: 902–13.
39. Hazawa M, Lin DC, Handral H, Xu L, Chen Y, Jiang YY, et al. ZNF750 is a lineage-specific tumour suppressor in squamous cell carcinoma. Oncogene 2017;36:2243–54.

40. Machado GC, Ferrer VP. MUC17 mutations and methylation are associated with poor prognosis in adult-type diffuse glioma patients. J Neurol Sci 2023;452:120762.
41. Weaver JMJ, Ross-Innes CS, Shannon N, Lynch AG, Forshew T, Barbera M, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. Nat Genet 2014;46:837–43.
42. Labuc GE, Archer MC. Esophageal and hepatic microsomal metabolism of N-nitrosomethylbenzylamine and N-nitrosodimethylamine in the rat. Cancer Res 1982;42:3181–6.
43. Fan H, Yu X, Zou Z, Zheng W, Deng X, Guo L, et al. Metformin suppresses the esophageal carcinogenesis in rats treated with NMBzA through inhibiting AMPK/mTOR signaling pathway. Carcinogenesis 2019;40: 669–79.
44. Watkins PA, Maiguel D, Jia Z, Pevsner J. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. J Lipid Res 2007; 48:2736–50.
45. Zhou Z, Wu Q, Yan Z, Zheng H, Chen CJ, Liu Y, et al. Extracellular RNA in a single droplet of human serum reflects physiologic and disease states. Proc Natl Acad Sci USA 2019;116:19200–8.
46. Zheng D, Ding Y, Ma Q, Zhao L, Guo X, Shen Y, et al. Identification of serum MicroRNAs as novel biomarkers in esophageal squamous cell carcinoma using feature selection algorithms. Front Oncol 2019;8:674.
47. Hee-Jin J, Hyun-Sung L, Bryan MB, Geon Kook L, Kyong-Ah Y, Yun-Yong P, et al. Integrated genomic analysis of recurrence-associated small non-coding RNAs in oesophageal cancer. Gut 2017;66:215.
48. Stachler MD, Taylor-Weiner A, Peng S, McKenna A, Agoston AT, Odze RD, et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. Nat Genet 2015;47:1047–55.
49. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. BMC Med Genet 2019;20:175.
50. Katarkar A, Bottoni G, Clocchiatti A, Goruppi S, Bordignon P, Lazzaroni F, et al. NOTCH1 gene amplification promotes expansion of Cancer Associated Fibroblast populations in human skin. Nat Commun 2020;11:5126.
51. Watkins TBK, Lim EL, Petkovic M, Elizalde S, Birkbak NJ, Wilson GA, et al. Pervasive chromosomal instability and karyotype order in tumour evolution. Nature 2020;587:126–32.
52. Drews RM, Hernando B, Tarabichi M, Haase K, Lesluyes T, Smith PS, et al. A pan-cancer compendium of chromosomal instability. Nature 2022;606:976–83.
53. Bach D-H, Zhang W, Sood AK. Chromosomal instability in tumor initiation and development. Cancer Res 2019;79:3995–4002.
54. Bourke MJ, Neuhaus H, Bergman JJ. Endoscopic submucosal dissection: indications and application in western endoscopy practice. Gastroenterology 2018;154:1887–900.e5.
55. Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The genome sequence archive family: toward explosive data growth and diverse data types. Genomics, Proteomics Bioinf 2021;19:578–83.
56. Members C-N, Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2022. Nucleic Acids Res 2022;50:D27–38.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/mr-2024-0008).