Material and Methods

Integration, curation and filtering of publically available sequencing data of esophageal cancer samples.

We curated somatic mutations identified in 1,048 tumor-germline matched genome pairs from 602 and 446 ESCC and EAC individuals, respectively. Out of these samples, 818 were subjected to whole-exome sequencing (WES), and the rest 230 were examined by whole-genome sequencing (WGS). We retrieved the mutation results from The Cancer Genome Atlas (TCGA) [1], International Cancer Genome Consortium (ICGC, part of the results have been published [2 3]) and recent large-scale genomic reports [4-7] (**Supplementary Table 1**). Extensive curation and filtering steps were performed. Specifically, mutations from studies where variants were reported using hg18 reference assembly were converted to GRCHh37 (hg19) using Liftover [8]. To avoid bias due to differences in annotation methods among studies, all variants were re-annotated using Variant Effect Predictor (VEP) version 85 and assigned to a single prioritized canonical transcript using vcf2maf version 1.6.8 [9]. Intergenic and noncoding variants were discarded when analyzing WGS resources. We further filtered potential false positive variants (might be resulted from technology- or laboratory- specific sequencing artefacts) or germline variants based on their minor allele frequencies in the general healthy population. Briefly, i) variants with allele frequencies >1% in dbSNP database were removed; ii) variants with minor allele frequency >0.04% in any of the ExAC subpopulation were removed while retaining those classified as "pathogenic" or "risk factor" in ClinVar [10]. Final data contain a total of 169,162 variants comprising of 1,045 samples (91,858 missense; 63,395 silent; 6,677 nonsense; 2,868 frameshift deletions; 2,237 splice site; 1,502 frameshift insertions; 713 inframe insertions; 130 inframe deletions; 232 affecting translational start or end sites). Processed Mutation Annotation Format (MAF) files can be downloaded at https://cedars.app.box.com/.

Identification of significantly mutated genes (SMGs)

Disease associated driver genes were detected using MutSigCV (v1.4) [11], which was performed separately on ESCC and EAC cohorts, resulting in a total of 55 SMGs ($q<0.1$; 25 and 34 in

ESCC and EAC, respectively). We further required that: i) SMGs occurred in at least 4 samples in corresponding cohorts; and ii) the median level of SMGs > 1 FPKM (Fragments Per Kilobase of transcript per Million mapped reads) in corresponding TCGA RNA-seq data. Finally, a total of 19 and 17 genes were significantly mutated in ESCC and EAC, respectively, with four SMGs being commonly shared (*TP53, CDKN2A, FBXW7* and *PIK3CA*). Importantly, all but one gene (*ATF5*) were mutated in at least 2 different individual cohorts, and the majority SMGs (28/32) had mutations in 3 or more individual cohorts. Moreover, we compared the mutational frequency of every SMG in each individual cohort to the combined cohorts, and high concordance was observed in all comparisons (all Pearson correlation coefficient > 0.97, all P < 0.0001, **Supplementary Fig. 3**). Together, these data strongly suggest that cohort- or batch-bias had a minimal impact in this analysis. The results also indicate that the discovery of a number of novel SMGs in the present study was empowered by the increased sample size, in line with a recent saturation analysis estimating that the number of SMGs with low frequency (<5%) rose steadily with increased sample size [12].

Mutational signature analysis

We applied the signature deconvolution method described by Alexandrov et al [13] using Bioconductor package SomaticSignatures [14]. First, the mutational profiles from both ESCC and EAC cohorts were used to infer de novo consensus mutational signatures based on Negative Matrix Factorization (NMF) method. A total of 6 mutational signatures appearing in at least 20 samples with the contribution of at least 30% in each individual were retained and shown in **Fig. 2**. For survival and enrichment analyses, as well as Pearson correlation between individual cohorts and combined cohort, samples with at least 10% contribution of a specific signature were selected. The enrichment analysis for a specific signature was performed using hypergeometric test with the background as all other samples not having that signature. To validate the results determined by NMF method, we performed pmsignature [15], a different statistical framework, to extract de novo mutational signatures in the present cohorts.

WGBS and data analysis

Primary tumor tissues from 6 ESCC patients, and 2 matched nonmalignant esophageal mucosa, were collected at Linzhou Esophagus Cancer Hospital, China. All patients signed separate informed consent forms for molecular analysis. This study has been approved by the ethics committee of the Cancer Hospital/Institute, Chinese Academy of Medical Sciences. WGBS library preparation and high-throughput sequencing were performed by Novogene, Inc. Briefly, 3-5 microgram gDNA spiked with 26 ng lambda DNA were fragmented. Cytosine-methylated barcodes were ligated to the sonicated DNA, which were treated twice with bisulfite using EZ DNA Methylation-Gold^{TM} Kit (Zymo Research). The resulting DNA fragments were PCR amplified using HiFi HotStart Uracil + ReadyMix (Kapa Biosystems). The clustering of the index-coded samples was performed on Illumina cBot Cluster Generation System according to the manufacturer's instructions. Finally, DNA libraries were sequenced on an Illumina Hiseq 2000/2500 platform and 100bp paired-end reads were generated.

WGBS FASTQ files, after preprocessing using FastQC (https://www. bioinformatics.babraham.ac.uk/projects/fastqc/), Trimgalore (https://www. bioinformatics. babraham.ac.uk/projects/trim_galore/) and Picardtools (https://broadinstitute.github.io/picard/), were mapped using BWA aligner (https://github.com/zwdzwd/biscuit) to human reference genome GRCHh37 (hg19) using default parameters. WGBS data of selected normal tissues and tumors were downloaded from TCGA data portal (level 3 data) and Roadmap Epigenomics Consortium (http :// www. roadmapepigenomics.org/). These data were visualized in Integrative Genomics Viewer [16] with bigWig files.

Chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq)

ChIP-seq was performed as described previously [17]. Briefly, cells were first crosslinked with 1% formaldehyde solution and followed by nuclear extraction. Chromatin was sheared in a Bioruptor Sonicator (Diagenode). Sonicated lysates were cleared and incubated overnight at 4℃ with magnetic beads bound with H3K27ac antibody (Abcam, ab4729). Precipitated immunocomplexes were washed, and DNA was eluted and sequenced in Illumina Hiseq 4000 at Beijing Genomic Institute (BGI). ChIP-seq reads were analyzed as described previously [17]. Briefly, they were aligned to human reference genome (build GRCh37/hg19) with Bowtie Aligner, and peaks were identified using MACS (Model-based Analysis of ChIP-seq). Wiggle files were generated and normalized in terms of reads per million reads (rpm). These files were then converted into bigwig files using wigToBigWig tool (http://hgdownload.cse.ucsc.edu/admin/exe/) and visualized in IGV. ChIP-seq data from KYSE510 and TE7 cell lines were generated previous by our group [17]. ChIP-seq data from nonmalignant esophagus epithelium and primary keratinocytes were retrieved from Roadmap Epigenomics Consortium.

Assessment of variant clonality

Cancer cell fraction (CCF) of variant was determined through an integrative analysis of tumor purity, variant allele frequency (VAF) and variant copy number as we and others described previously [18-20]. Because of the availability of genome-wide copy number data, we restricted our analysis to TCGA cohort which contains matched copy number values measured by SNP-6.0 array. We obtained level-3 segment profiles using TCGA Biolinks Bioconductor package [21]. Segmented copy number data along with mutational profiles were fed into ABSOLUTE program to assess CCF of each variant taking account tumor purity, read depth coverage and variant allele frequencies [22]. All of the solutions were the "top" solutions generated by the ABSOLUTE program (**Supplementary Table 2**). A variant was classified as clonal if the upper 95% confidence interval of its CCF was equal to or above 1, as previously described [18-20].

Expression vectors and siRNAs

Expression vectors including pcDNA3-myc-CUL3 (Addgene, Plasmid #19893) and pANT7-cGST-ZFP36L2 (DNASU, HsCD00079067) were obtained commercially. ZFP36L2 was further subcloned into pcDNA3.1 vector. Pooled siRNA (which have been confirmed to reduce off-target effect [23 24]) against ATF5, PTCH1, CUL3, ZFP36L2 were obtained from GE Dharmacon. To further excluding off-target effects, individual siRNAs against CUL3 and ZFP36L2 were used (GenePharma). DNA vectors and siRNAs were transfected into cells using either Lipofectamine 2000 or Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions. Sequences of PCR primers and siRNAs were listed in **Supplementary Table 4**.

Cell culture and phenotypical assays

ESCC and EAC cell lines were cultured in Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific). All media were supplemented with 10% fetal bovine serum (Omega Scientific, Inc), penicillin (100 U/mL) and streptomycin (100 mg/mL). All cell lines were authenticated by short tandem repeat analysis recently. In short-term proliferation assay, 2,000-4,000 cells were seeded onto 96-well plates, and cultured for 3-5 days. Cell viability was assessed using MTT (3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide) staining method. For the colony formation assay, 500-1,000 cells were seeded onto 6-well plates, and cultured for 2 weeks. Cells were fixed with methanol and stained with crystal violet. Cell migration assay used Borden Chamber, and $1x10^5$ cells in serum free medium were seeded onto the membrane with 8-μm pores of the top chamber (Thermo Fisher Scientific, Inc), and the bottom chamber was filled with medium containing 10% fetal bovine serum. After 24-48 hours, the membranes were washed, fixed and stained with crystal violet, and migrated cells were quantified.

Xenograft assay in nude mice

Xenograft assay was performed as described previously [25]. In brief, ten 6-week-old nude mice (Taconic Biosciences) were subcutaneously injected with $2x10^6$ TE7 cells on their dorsal flanks, with each mouse carrying four explants. Tumor growth was monitored and tumor size was measured every three days. After 20 days, mice were sacrificed and the tumors were dissected, weighed and analyzed. This animal study was approved by the Institutional Animal Care and Use Committee (IACUC) at Medical College of Shantou University.

Antibodies and Chemicals

Santa Cruz Biotechnology: CUL3 (C-3) sc-166054, β-catenin (H-102) sc-7199, c-Myc (9E10) sc-40, cyclin D1 (72-13G) sc-450, TCF4 (367.2) sc-101095, p27 (C-19) sc-528. Cell signaling technology: NRF2 (D1Z9C) #12721**,** GAPDH (14C10) #2118

Chemicals: ICG001 was kindly provided by Dr. Michael Kahn. Trigonelline was obtained from Sigma-Aldrich, Inc.

Supplementary figure legend

Supplementary Figure 1. Comparison of clinical-pathological parameters among different individual cohorts

Dot plot and column plots comparing important clinical-pathological parameters retrieved from individual cohorts.

Supplementary Figure 2. Comparison of mutational patterns among different individual cohorts

Bar plots displaying the fraction of all possible 96 trinucleotide substitutions extracted in different individual cohorts.

Supplementary Figure 3. Highly consistent SMG frequency across different individual cohorts

Scatter plot showing the Pearson correlation of SMG frequency between each individual cohort (X axis) and the combined cohort (Y axis). Red dots denote novel SMGs.

Supplementary Figure 4. Pearson correlation of overlapping SMGs of esophageal cancer and 21 pan cancer types

Abbreviations: LUSC, lung squamous cell carcinoma; HNSC, head and neck squamous cell carcinoma; BLCA, bladder urothelial Carcinoma; LUAD, lung adenocarcinoma; MEL, melanoma; GBM, glioblastoma multiforme; BRCA, breast invasive carcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; KIRC, kidney renal clear cell carcinoma; CRC, colon and rectal; STAD, stomach adenocarcinoma; DLBCL, diffuse large b-cell lymphoma; PRAD, prostate adenocarcinoma; CLL, chronic lymphocytic leukemia; LAML, acute myeloid leukemia; NB, neuroblastoma; RHAB, rhabdomyosarcoma; CARC, carcinoid tumor; MED, medulloblastoma.

Supplementary Figure 5. Pearson correlation test of mutational signature analysis

(A) Scatter plot showing the Pearson correlation coefficient of the contribution from each mutational signature determined by NMF (X axis) and pmsignature (Y axis). (B) Scatter plot showing the Pearson correlation coefficient of the contribution from each mutational signature in each individual cohort (X axis) and the combined cohort (Y axis).

Supplementary Figure 6. Survival-associated clinical parameters in esophageal cancer

Supplementary Figure 7. Functional screen of novel ESCC SMGs

Four novel ESCC SMGs (ATF5, PTCH1, CUL3 and ZFP36L2) were individually silenced via siRNAs in 4 different ESCC cell lines and subjected to (A) qRT-PCR analysis and (B) short-term cell proliferation assay. Data represent mean + SD (n = 3). *, P < 0.05.

Supplementary Figure 8. Functional test of CUL3 and ZFP36L2 in EAC cell lines

CUL3 and ZFP36L2 were silenced via siRNAs individually in 2 different EAC cell lines and subjected to (A) qRT-PCR analysis and (B) short-term cell proliferation assay. Data represent mean + SD (n = 3).

Supplementary Figure 9. Validation of the on-target effects of pooled siRNA.

Independent individual siRNAs against CUL3 and ZFP36L2 were designed and utilized to verify the effects produced by pooled siRNAs. (A) qRT-PCR, (B) short-term cell proliferation and (C) cell migration assays were performed in both TE7 and KYSE510 cells.

Supplementary Figure 10. CUL3 inhibits both NRF2 and WNT/β-catenin pathways

(A) Mutually exclusive mutational patterns of *CUL3/KEAP1/NFE2L2* in ESCC. (B) CUL3 was either silenced by siRNA (left panel) or ectopically expressed (right panel) in ESCC cells, followed by qRT-PCR assay. (C) Short-term proliferation assay testing the combinational use of ICG001 and Trig.

Supplementary Figure 11. Lineage-specific ZFP36L2 hypermethylation in SCC samples

WGBS IGV snapshot displaying DNA methylation level in *ZFP36L2* gene loci in nonmalignant tissues (esophagus, lung, stomach), as well as cancer samples from ESCC, LUSC and STAD. REMC, Roadmap Epigenomics Consortium.

Reference

1. Cancer Genome Atlas Research N, Analysis Working Group: Asan U, Agency BCC, Brigham, Women's H, Broad I, et al. Integrated genomic characterization of oesophageal carcinoma. Nature 2017;**541**(7636):169-75
2. Secrier M, Li X, de Silva N, Eldridge MD, Contino G, Bornschein J, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. Nature genetics 2016;**48**(10):1131-41
3. 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**(7498):91-5
4. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nature genetics 2013;**45**(5):478-86
5. Gao YB, Chen ZL, Li JG, Hu XD, Shi XJ, Sun ZM, et al. Genetic landscape of esophageal squamous cell carcinoma. Nature genetics 2014;**46**(10):1097-102
6. Lin DC, Hao JJ, Nagata Y, Xu L, Shang L, Meng X, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. Nature genetics 2014;**46**(5):467-73
7. 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**(5):1171-82
8. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, et al. The UCSC Genome Browser database: 2014 update. Nucleic acids research 2014;**42**(Database issue):D764-70
9. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010;**26**(16):2069-70
10. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;**536**(7616):285-91
11. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;**499**(7457):214-8
12. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;**505**(7484):495-501
13. Alexandrov LB, Nik-Zainal S, Siu HC, Leung SY, Stratton MR. A mutational signature in gastric cancer suggests therapeutic strategies. Nature communications 2015;**6**:8683
14. Gehring JS, Fischer B, Lawrence M, Huber W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics 2015;**31**(22):3673-5
15. Shiraishi Y, Tremmel G, Miyano S, Stephens M. A Simple Model-Based Approach to Inferring and Visualizing Cancer Mutation Signatures. PLoS genetics 2015;**11**(12):e1005657
16. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature biotechnology 2011;**29**(1):24-6
17. Jiang YY, Lin DC, Mayakonda A, Hazawa M, Ding LW, Chien WW, et al. Targeting super-enhancer-associated oncogenes in oesophageal squamous cell carcinoma. Gut 2016
18. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Science translational medicine 2015;**7**(283):283ra54
19. 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. Nature genetics 2016;**48**(12):1500-07
20. Lin DC, A. M, Dinh HQ, Huang P, Lin L, Liu X, et al. Genomic and epigenomic heterogeneity of hepatocellular carcinoma. Cancer research 2017;**10.1158/0008-5472**
21. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic acids research 2016;**44**(8):e71
22. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nature biotechnology 2012;**30**(5):413-21

23. Kittler R, Surendranath V, Heninger AK, Slabicki M, Theis M, Putz G, et al. Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. Nature methods 2007;**4**(4):337-44

24. Jackson AL, Linsley PS. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. Nature reviews Drug discovery 2010;**9**(1):57-67

25. 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 2016