Genomic Analyses Reveal Mutational Signatures and Frequently Altered Genes in Esophageal Squamous Cell Carcinoma

Ling Zhang,[1,2,18] Yong Zhou,[3,18] Caixia Cheng,[1,2,4,18] Heyang Cui,[1,2,3,18] Le Cheng,[5,18] Pengzhou Kong,[1,2,18] Jiaqian Wang,[3,18] Yin Li,[6,18] Wenliang Chen,[3] Bin Song,[1,2] Fang Wang,[1,2] Zhiwu Jia,[1,2] Lin Li,[3] Yaoping Li,[1,7] Bin Yang,[1,7] Jing Liu,[1,8] Ruyi Shi,[1,2] Yanghui Bi,[1,2] Yanyan Zhang,[1,8] Juan Wang,[1,2] Zhenxiang Zhao,[1,2] Xiaoling Hu,[1,2] Jie Yang,[1,2] Hongyi Li,[1,2] Zhibo Gao,[3] Gang Chen,[3] Xuanlin Huang,[3] Xukui Yang,[3] Shengqing Wan,[3] Chao Chen,[3] Bin Li,[3] Yongkai Tan,[3] Longyun Chen,[3] Minghui He,[3] Sha Xie,[3] Xiangchun Li,[3] Xuehan Zhuang,[3] Mengyao Wang,[3] Zhi Xia,[3] Longhai Luo,[3] Jie Ma,[9] Bing Dong,[9] Jiuzhou Zhao,[9] Yongmei Song,[10] Yunwei Ou,[10] Enming Li,[11] Liyan Xu,[11] Jinfen Wang,[12] Yanfeng Xi,[12] Guodong Li,[12] Enwei Xu,[12] Jianfang Liang,[4] Xiaofeng Yang,[13] Jiansheng Guo,[8] Xing Chen,[14] Yanbo Zhang,[15] Qingshan Li,[16] Lixin Liu,[17] Yingrui Li,[3] Xiuqing Zhang,[3] Huanming Yang,[3] Dongxin Lin,[10] Xiaolong Cheng,[1,2] Yongjun Guo,[9] Jun Wang,[3] Qimin Zhan,[10,*] and Yongping Cui[1,2,*]

Esophageal squamous cell carcinoma (ESCC) is one of the most common cancers worldwide and the fourth most lethal cancer in China. However, although genomic studies have identified some mutations associated with ESCC, we know little of the mutational processes responsible. To identify genome-wide mutational signatures, we performed either whole-genome sequencing (WGS) or whole-exome sequencing (WES) on 104 ESCC individuals and combined our data with those of 88 previously reported samples. An *APOBEC*-mediated mutational signature in 47% of 192 tumors suggests that APOBEC-catalyzed deamination provides a source of DNA damage in ESCC. Moreover, *PIK3CA* hotspot mutations (c.1624G>A [p.Glu542Lys] and c.1633G>A [p.Glu545Lys]) were enriched in *APOBEC*-signature tumors, and no smoking-associated signature was observed in ESCC. In the samples analyzed by WGS, we identified focal (<100 kb) amplifications of *CBX4* and *CBX8*. In our combined cohort, we identified frequent inactivating mutations in *AJUBA*, *ZNF750*, and *PTCH1* and the chromatin-remodeling genes *CREBBP* and *BAP1*, in addition to known mutations. Functional analyses suggest roles for several genes (*CBX4*, *CBX8*, *AJUBA*, and *ZNF750*) in ESCC. Notably, high activity of hedgehog signaling and the PI3K pathway in approximately 60% of 104 ESCC tumors indicates that therapies targeting these pathways might be particularly promising strategies for ESCC. Collectively, our data provide comprehensive insights into the mutational signatures of ESCC and identify markers for early diagnosis and potential therapeutic targets.

Introduction

Esophageal carcinoma (EC) is the eighth most common and the sixth most lethal cancer worldwide. It tends to have a very poor prognosis as a result of the limited clinical approaches for early diagnosis[1]—the 5-year survival rate ranges from 10% to 25%.[1] Clinically, this heterogeneous disease is categorized into two subtypes: esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC).[1] Approximately 70% of the worldwide cases of ESCC occur in China, where the highest incidences are in the Taihang Mountains of north-central China.[1] The risk of developing ESCC in China has been linked to factors such as dietary habits (e.g., hot food and betelnut chewing) and family history.[2] Alcohol abuse and tobacco consumption explain nearly 90% of ESCC cases in Western countries but represent minor factors in China.[2]

The processes of DNA damage and repair, which produce the somatic mutations in a cancer genome, leave a signature of mutation.[3] Recent analyses of comprehensive mutational catalogs from a malignant melanoma and a lung cancer revealed the characteristic mutational patterns of UV light and tobacco carcinogens, respectively.[4,5] Exome-wide investigations have characterized somatic point

[1]Translational Medicine Research Center, Shanxi Medical University, Taiyuan, Shanxi 030001, China; [2]Key Laboratory of Cellular Physiology, Ministry of Education, Shanxi Medical University, Taiyuan, Shanxi 030001, China; [3]BGI-Shenzhen, Shenzhen, Guangdong 518083, China; [4]Department of Pathology, First Hospital, Shanxi Medical University, Taiyuan, Shanxi 030001, China; [5]BGI-Yunnan, Kunming, Yunnan 650000, China; [6]Department of Thoracic Surgery, Henan Cancer Hospital, Zhengzhou, Henan 450008, China; [7]Department of Tumor Surgery, Shanxi Cancer Hospital, Taiyuan, Shanxi 030001, China; [8]Department of General Surgery, First Hospital, Shanxi Medical University, Taiyuan, Shanxi 030001, China; [9]Department of Molecular Pathology, Henan Cancer Hospital, Zhengzhou, Henan 450008, China; [10]State Key Laboratory of Molecular Oncology, Cancer Institute and Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; [11]Institute of Oncologic Pathology, Shantou University Medical College, Shantou 515041, China; [12]Department of Pathology, Shanxi Cancer Hospital, Taiyuan, Shanxi 030001, China; [13]Department of Urology, First Hospital, Shanxi Medical University, Taiyuan, Shanxi 030001, China; [14]Department of Endoscopy, Shanxi Provincial People's Hospital, Taiyuan, Shanxi 030001, China; [15]Department of Statistics, Shanxi Medical University, Taiyuan, Shanxi 030001, China; [16]School of Pharmaceutical Sciences, Shanxi Medical University, Taiyuan, Shanxi 030001, China; [17]Experimental Center of Science and Research, First Hospital, Shanxi Medical University, Taiyuan, Shanxi 030001, China
[18]These authors contributed equally to this work
*Correspondence: zhanqimin@pumc.edu.cn (Q.Z.), cuiy0922@yahoo.com (Y.C.)
http://dx.doi.org/10.1016/j.ajhg.2015.02.017.

mutations in ESCC and identified the ESCC-associated genes *ZNF750* (MIM 610226), *FAT1* (MIM 600976), *FAT2* (MIM 604269), and *FAM135B*,[6,7] and point mutations in genes such as *PLCE1* (MIM 608414), *XPF* (MIM 133520), *ALDH2* (MIM 100650), and *MTHFR* (MIM 607093) show different signatures between smokers and non-smokers with ESCC.[8–10] However, such studies have been limited to a few genes, and it is unclear how representative these findings are of mutational processes operative in ESCC across the whole genome. In this study, we performed whole-genome sequencing (WGS) of 14 and whole-exome sequencing (WES) of 90 ESCC tumors and adjacent normal tissue from individuals recruited from the Taihang Mountains of north-central China and combined our data with those of 88 previously reported samples to extract the mutational signatures that cause somatic mutations in ESCC and identify driver genes or pathways contributing to this highly fetal disease in Chinese individuals.

Material and Methods

Samples and Clinical Data

Tumor samples and adjacent, histologically normal tissues were obtained from 104 ESCC-affected individuals recruited from the Taihang Mountains of north-central China. All individuals gave their informed consent, and all samples were obtained before treatment according to the guidelines of the institutional review board of Shanxi Medical University (approval no. 2009029) and the ethics committee of Henan Cancer Hospital (approval no. 2009xjs12). The tumor samples of all affected individuals had at least 40%–50% of tumor cell content. This study was approved by the ethics committees of the Shanxi Medical University and Henan Cancer Hospital. The ESCC individuals collected for this study were staged according to the Cancer Staging Standards of the American Joint Committee on Cancer (seventh edition, 2010). The cohort of 104 ESCC individuals included 96 smoking and 8 non-smoking individuals. Different subsets of individuals were assayed on each platform: 14 tumors and 14 matched normal samples had data available on WGS (65× coverage), and 90 samples had data available on WES (132× coverage); in addition, 96 of the 104 samples had target capture deep resequencing (365× coverage). A detailed description of the clinical characteristics of the analyzed samples is presented in Table S1. We also analyzed our previously published ESCC mutation dataset[6] of 17 WGS and 71 WES samples recruited from the Chaoshan District of Gongdong Province, another area of high ESCC prevalence in China. This cohort included 57 smoking and 31 non-smoking individuals.[6] The summary of next-generation sequencing analyses in this study is shown in Figure S1.

Sequencing

For WGS, genomic DNAs extracted from 14 tumors and matched normal tissues were randomly fragmented and purified. Standard paired-end adaptors were ligated according to the manufacturer's (Illumina) protocol. Adaptor-ligated fragments were purified with preparatory gel electrophoresis, and identical bands were excised, resulting in two libraries per sample with inserts averaging 500 bp. Four lanes of each of the resulting WGS libraries were subjected to WGS on an Illumina HiSeq 2000. Target depth (65× for tumors and normal samples) and at least 30× haploid coverage were achieved in all samples.

For WES, the qualified genomic DNAs from 90 tumors and matched normal tissues were randomly fragmented by Covaris, ligated to Illumina sequenced adapters, and selected for lengths from 150 to 200 bp. Extracted DNAs (150–200 bp) were then amplified by ligation-mediated PCR (LM-PCR), purified, and hybridized to the NimbleGen SeqCap EZ Exome (44M) array for enrichment. Hybridized fragments were bound to the streptavidin beads, whereas non-hybridized fragments were washed out after 24 hr. We then subjected captured LM-PCR products to the Agilent 2100 Bioanalyzer to estimate the magnitude of enrichment. We independently loaded each captured library from the process described above on three lanes of an Illumina HiSeq 2000 platform with 90-bp paired-end reads for high-throughput sequencing to ensure that each sample met the desired average coverage. Raw image files were processed by Illumina base-calling software (v.1.7) with default parameters. The mean coverage achieved was 130× in tumors and 133× in normal tissue. A detailed description is presented in Table S2.

Mutation Detection

For detection of somatic point mutations, sequencing reads from a Illumina HiSeq 2000 were aligned to the human reference genome (UCSC Genome Browser hg19) with the Burrows-Wheeler Aligner. After duplicate reads (redundant information produced by PCR) were removed with SAMtools, an in-house pipeline was used to call somatic mutations. In brief, we implemented SAMtools (v.0.1.18) and VarScan (v.2.2.5) to call somatic variants. We required a minimum depth of 10× and variant frequency of 10% for both normal and tumor samples in order to call a specific variant at that locus. A single-nucleotide variant (SNV) was labeled highly confident if it met the following requirements: (1) the locus was not enriched with reads of low mapping quality, (2) reads that supported the SNV were not significantly overrepresented with bases of low quality, (3) reads that supported the SNV showed no bias toward the read end, (4) no gaps were found near the SNV locus, and (5) the SNV was not encompassed in short repeat regions.

The indel-calling step was performed by the Genome Analysis Toolkit SomaticIndel Detector with default parameters. The highly confident indels were identified by an in-house pipeline and further annotated as germline or somatic on the basis of whether any evidence of the event at the same locus was observed in the normal data. Finally, highly confident SNVs were annotated with ANNOVAR and used in follow-up analysis. A full list of mutation events is presented in Table S3.

Analysis of DNA Copy Number

To detect DNA copy-number alterations (CNAs), we performed SegSeq[11] to infer somatic CNA in ESCC genomes on the basis of WGS reads. Copy numbers ≤ 1.5 were considered to indicate deletions, and those ≥ 2.5 were considered amplifications. To infer recurrently amplified or deleted genomic regions, we re-implemented the GISTIC algorithm[12] by using copy numbers in 1-kb windows as markers instead of SNP array probes. 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 value threshold of 0.05 from the permutation-derived null distribution. A full list of CNAs is presented in Table S4.

Identification of Significantly Mutated Genes

To analyze mutation data and identify significantly mutated genes (SMGs), we applied the analytical methodology MutSigCV (mutation significance with covariates) to facilitate the significance analysis with default parameters.

Pathway-Enrichment Analysis

We extended our significance analysis beyond single genes by looking at gene sets. Pathway-enrichment analyses of genes with non-synonymous mutations were performed with KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment. In brief, we performed the pathway-enrichment analysis with the Database for Annotation, Visualization, and Integrated Discovery (v.6.7) by examining the distribution of the non-synonymously mutated genes identified within KEGG. Significantly altered pathways were determined by p values calculated on the basis of hypergenometric distribution with Benjamini correction.

Cell Lines

Normal esophageal epithelium cell lines SHEE and HEEPIC and the following ESCC lines were used in this study: EC9706, Eca109, TE1, TE12, TE13, Caes17, KYSE2, KYSE30, KYSE140, KYSE150, KYSE410, KYSE510, KYSE450, and KYSE680. All were tested and found to be free of mycoplasma contamination. HEK293T cells were used as a packaging cell line to produce virus. All cells were grown in DMEM with nutrient mixture F-12 (DMEM/F-12) supplemented with 10% fetal bovine serum at 37°C in 5% CO_2. Endogenous products encoded by the genes of interest were detected through real-time PCR and/or immunoblotting analyses. For functional analysis of special genes, the ESCC lines with high endogenous gene expression were used for knockdown experiments, and ESCC lines with low gene expression were used for forced-expression experiments.

Knockdown and Overexpression of Genes of Interest in ESCC Lines

Lentiviral vector pLKO.1-puro and its packaging plasmids pMD2.G and psPAX2 were obtained from Addgene. Knockdown experiments of the special genes were performed in at least two ESCC lines with high endogenous protein levels. Two independent short hairpin RNAs (shRNAs) were cloned into the pLKO.1-puro vector as described previously.[13] Specifically, to produce virus production, titration, and infection lentiviruses, we transfected HEK293T cells with the packaging plasmids along with the lentiviral shRNA vector by using Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions, and we changed the medium after 6 hr. Virus was harvested 24 hr after transfection, passed through 0.22-μm filters, and used fresh for shRNA infection. To perform lentiviral infections, we plated the target ESCC cells at 40%–50% confluence and incubated them overnight (16 hr). On the day of infections, the culture medium was replaced with the appropriately titered viral supernatant (1.5 ml/well) and incubated at 37°C for 24 hr; afterward, the viral supernatant was replaced with fresh media. Forty-eight hours later, infected cell populations were selected in puromycin (2 μg/ml). After 5 days of selection, shRNA-knockdown efficiency was determined by immunoblot analysis for the respective proteins with the use of specific antibodies. For knockdown of special genes in ESCC cells, two independent shRNA constructs that had been cloned into the pLKO.1-puro vector were used (Table S5). A non-specific targeting shRNA was also cloned into the pLKO.1-puro vector with the use of a scrambled control (SCR). Relative amounts of special gene product were normalized to β-actin levels.

pMSCV-puro empty vector and wild-type pMSCV-puro-*ZNF750* were generous gifts from Paul A. Khavari (Stanford University). pcDNA3-RFP empty vector and wild-type pcDNA3-RFP-*AJUBA* (MIM 609066) were generous gifts from Alejandra Garcia-Cattaneo (National Heart and Lung Institute, Imperial College London). The wild-type versions of these genes of interest were cloned into the pLV-EGFP(2A)-puro-GFP vector and validated by sequencing. For overexpression experiments, we used the pLV-EGFP(2A)-puro-GFP vector as a control. Viruses were produced as previously described. ESCC cells with low endogenous protein levels were infected with viruses as previously described.[14] Twenty-four hours after infection, cells were subjected to subsequent experiments. The mutants of genes of interest were generated with the QuikChange II Site-Directed Mutagenesis Kit (Agilent).

Fluorescence In Situ Hybridization Analysis

Tumor and matched normal tissues of ESCC individuals were cut into pieces in PBS, swollen in 65 mmol/l KCl for 5 min at 37°C, fixed in cold acetic acid and methanol for 5 min at 4°C, dropped onto slides, and dried at room temperature. For interphase fluorescence in situ hybridization (FISH) analysis, slides were stained with Cytocell enumeration probes against chromosomal region 5q, *CBX8* (chr17: 77,768,175–77,770,915), or *CBX4* (MIM 603079; chr17: 77,806,954–77,813,213). These probes were conjugated with fluorescein isothiocyanate (FITC) or Cy3.5 (Rainbow Scientific). Probes against chromosomal region 5q or *TMC8* (MIM 605829; located near the *CBX4* and *CBX8* regions) were used as controls for verification of focal CNAs of *CBX4* or *CBX8*. Staining was carried out according to the manufacturer's protocol. FISH samples were viewed with a fully automated, upright Zeiss Axio-ImagerZ.1 microscope with a 20× objective and DAPI, FITC, and Rhodamine filter cubes. Images were produced with the AxioCam MRm CCD camera and the Axiovision v.4.5 software suite. p values were calculated with a two-sample test for equality of proportions with continuity correction.

qPCR Copy-Number Analysis

CBX4 and *CBX8* copy number was assessed in seven frozen tumor samples and matched normal tissues. Copy numbers were determined by real-time PCR with DNA binding dye SYBR Green I with the use of specific primer pairs that flanked coding exons of each gene. In a final volume of 25 μl, 20 ng DNA was amplified with SYBR Green PCR Master Mix (QIAGEN) in triplicate. *RPPH1* (ribonuclease P RNA component H1 [MIM 608513]; Life Technologies, 4403328) was used as a diploid control, and *TMC8* (chr17: 76,126,858–76,139,049) was used as a control located in the region near genes *CBX4* and *CBX8*. Data were analyzed via the comparative (delta-Ct) Ct method.

Immunoblotting

Cells were lysed for 30 min in Triton buffer (1% Triton X-100, 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% sodium deoxycholate, and 0.1% SDS) supplemented with protease and phosphatase inhibitors (1 mM PMSF, 2 mM sodium pyrophosphate, 2 mM sodium betaglycerophosphate, 1 mM sodium fluoride, 1 mM sodium orthovanadate, 10 μg/ml leupeptin, and 10 μg/ml aprotinin). Lysates were cleared by centrifugation at 15,000 × *g* at 4°C for 15 min, and protein concentrations were determined by the

Bradford method. Fifty micrograms of protein was separated by SDS-PAGE and transferred onto Immobilon-P membranes. Proteins were detected with special antibodies. Antibody binding was detected with horseradish-peroxidase-labeled anti-mouse (Sigma) or anti-rabbit (Cell Signaling) antibodies, and chemiluminescence was detected with a LAS4000 device (Fuji). Equal protein loading was confirmed with antibodies against β-actin (Transgen). Detailed information on antibodies is shown in Table S5.

MTT Assay

A total of 4×10^3 cells were seeded in 96-well plates and incubated in normal conditions for 24 hr. Cells were treated with 100 μl of 5 mg/ml of MTT (Invitrogen) solution for 4 hr at 37°C until crystals were formed. MTT solution was removed from each well, and 100 μl of DMSO was added to each well to dissolve the crystals. Color intensity was measured by Microplate Reader (Bio-Rad) at 490 nm. Each experiment consisted of four replications, and at least three independent experiments were carried out. For cell-death analysis, cells were treated, in duplicate, with BKM120 (10 μM), GANT61 (20 μM) or both for up to 72 hr prior to flow cytometric analysis for determining the extent of cell death by Annexin V/PI staining.

Migration and Invasion Assays

Migration and invasion assays were performed in 16-well CIM plates in an xCELLigence RTCA DP System (ACEA Biosciences) with Matrigel Basement Membrane Matrix (BD) for real-time cell-migration analysis as described previously.[15] In brief, 30,000 cells per well were seeded as five duplicates in serum-free medium at the upper compartment of the CIM plates coated with or without Matrigel. Serum-complemented medium was added to the lower compartment of the chamber, and then measurement began in the xCELLigence RTCA DP system. We analyzed the cell-index curves to determine cell-invasion activity. For negative controls, we added serum-free medium at both upper and bottom chambers. The cell index representing the amount of migrated cells was calculated with RTCA Software from ACEA Biosciences. At least three independent experiments were carried out; for each independent experiment, five duplicates were performed for each group.

Colony-Formation Assay

The assay was performed as described previously.[16] In brief, cells were seeded at 300–500 cells per well in 6-well plates containing complete DMEM/F-12 on day 0 and incubated at 37°C and 5% CO_2 for 10 days. On day 10, cells were fixed with 4% polyformaldehyde for 15 min and stained with 1% crystal violet before quantification. The experiments were triplicated, and the numbers of colonies containing more than 50 cells were microscopically counted.

Real-Time qPCR

Real-time qPCR was used for measuring expression levels of genes of interest in a subset of tumor samples or ESCC lines. The probes or kits used in this study are shown in Table S5. All qPCR reactions were performed in triplicate with an Applied Biosystems StepOnePlus. The relative expression of genes of interest was determined by normalization to *GAPDH* expression via a standard-curve method with ten serial dilutions according to the manufacturer's instructions. All real-time PCR experiments included a no-template control and were done in triplicate.

Immunohistochemistry Analysis

Immunohistochemistry was performed as previously described.[17] In brief, sections were incubated with special antibody at an ideal dilution for 14 hr at 4°C and then detected with PV8000 (Zhongshan) and the DAB detection kit (Maixin), producing a dark-brown precipitate. Slides were counterstained with hematoxylin. All images were captured at 100×. The nuclear amount of the protein of interest was analyzed with Aperio Nuclear v.9 software, and the cytoplasmic protein amount was quantified with Aperio Cytoplasma 2.0 software. Statistical analyses were performed with GraphPad Prism 5.0. The staining intensity was scored as 0 (negative), 1 (weak), 2 (moderate), or 3 (strong). The immunoreactive score (IRS) was determined by the product of the extent score and the intensity score. IRS values ranged from 0 to 9, which were graded as follows: 0 (negative), 1–3 (weak), 4–6 (moderate), and 7–9 (strong). The median IRS was chosen to define the individuals, and the ratio of tumor to matched normal tissue (T_{IRS}/N_{IRS}) was used to compare the protein of interest with significantly high amounts ($T_{IRS}/N_{IRS} > 2$), high amounts ($1 < T_{IRS}/N_{IRS} < 2$), low amounts ($T_{IRS}/N_{IRS} < 1$), or no change ($T_{IRS}/N_{IRS} = 1$) in tumor tissue to that in matched normal tissue. All antibodies used in this study are shown in Table S5.

Mouse Xenograft Assay and Immunohistochemistry

To determine the effects of ZNF750 on tumorigenesis in vivo, we used a mouse xenograft assay with 12- to 14-week-old BALB/c nude female mice. We injected 2×10^6 KYSE150 cells stably depleted of ZNF750, scrambled control vector, wild-type ZNF750, or p.Ser70* ZNF750 into nude mice ($n = 6$ mice/group). The growth rates of xenograft tumors were measured for 4 weeks, after which mice were sacrificed. After 28 days, tumors were removed, snap frozen in liquid nitrogen, and stored at –80°C. Tumor size was measured with calipers. Additionally, formalin-fixed paraffin-embedded xenograft tumors were immunohistochemically stained with a monoclonal mouse Ki-67 antibody (Zhongshan). In brief, sections were incubated with the Ki67 antibody working dilution for 14 hr at 4°C and then detected with PV8000 (Zhongshan) and the DAB detection kit (Maixin), producing a dark-brown precipitate. Slides were counterstained with hematoxylin. All images were captured at 20×. The nuclear amount of Ki-67 was analyzed with Aperio Nuclear v.9 software. Statistical analyses were performed with GraphPad Prism 5.0.

Statistical Analysis

The SPSS Statistics 17.0 package was employed to correlate clinical and biological variables by means of Fisher's test or a non-parametric test when necessary. Experiments were performed in triplicate, and data were presented as mean ± SD. Student's t test was used for statistical analysis, and data from more than two groups were analyzed by one-way ANOVA in SPSS Statistics 19.0 and a subsequent Fisher's least significant difference t test. Results were considered significant when $p < 0.05$. Association tests on ZNF750 genotype and levels were performed on log-transformed expression values by linear regression or t test.

Results

Mutational Signatures of ESCC

To extract the mutational signatures that cause somatic mutations in ESCC and identify driver genes or pathways

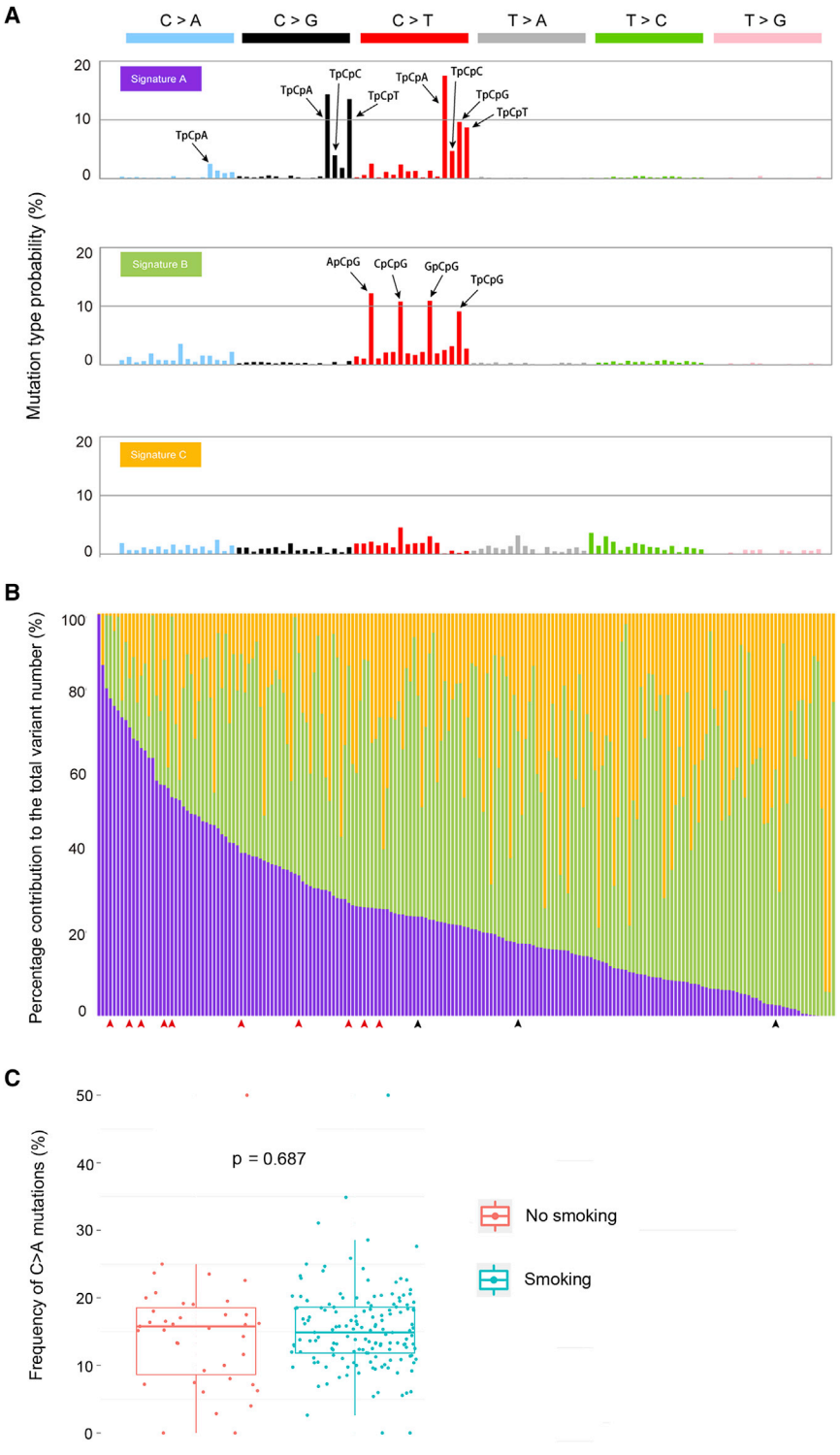

Figure 1. Three Mutational Signatures Identified in ESCC and Their Incidence across the 192 Tumor Samples

(A) Ninety-six substitution classifications from WGS or WES data derived from the 192 pairs of ESCC samples, including 88 samples from Song et al.[6] Mutation types are displayed in different colors on the horizontal axis. The vertical axis depicts the percentage of mutations attributed to a specific mutation type.

(B) The contributions of mutational signatures to individual tumors. Each bar represents a sample, and samples are ordered by the proportion of signature A found in them. The three mutational signatures are represented by the different colors shown in (A). For signature A, a significant number of mutations in a sample is defined as more than 100 substitutions in total or more than 25% of all mutations in that sample. Red or black triangles indicate samples harboring the *PIK3CA* c.1624G>A (p.Glu542Lys) or c.1633G>A (p.Glu545Lys) mutation, respectively.

(C) Boxplot of the proportion of C>A transversions in individuals with or without a smoking history. Each dot represents the proportion of C>A mutations in one individual; Student's t test was used to compare the difference between the two groups. Data represent the mean ± SD.

contributing to ESCC in Chinese individuals, we sequenced the genome of 104 ESCC tumors and matched adjacent normal tissues from individuals recruited from the Taihang Mountains in north-central China (Table S1). WGS (median coverage of 65×) of 14 tumors and WES (median coverage of 132×) of 90 tumors were performed (Figure S1). The average mutation rate was 3.9 coding mutations/Mb in WGS samples and 2.4 non-silent mutations/Mb in WES samples (Table S2). This rate is consistent with recently published mutation rates in ESCC.[6] A high frequency of C>T transitions was identified in the overall dataset[18] (Figure S2A), and C>G transversions occurred more frequently in ESCC than in EAC[19] (Figure S2B). We selected candidate non-silent mutations identified in 96 tumors for validation by using the deep target capture system (at least 365×). Validation rates were 97.8% for identified SNVs and 58% for indels. We also analyzed our previously published ESCC mutation dataset[6] of 17 WGS and 71 WES samples recruited from the Chaoshan District of Gongdong Province, another area of high ESCC prevalence in China (Figure S1E).

To identify the mutational signatures within ESCC genomes, we applied the non-negative matrix-factorization method[20] to a combined mutation set of 192 ESCC tumors (14 WGS and 90 WES samples from this study and 88 from Song et al.[6]) and uncovered three mutational signatures (Figure 1 and Figure S3). Signature A was characterized by C>G, C>T, and C>A mutations at TpCpX trinucleotides (suggesting collateral damage following DNA-element

retrotransposition or exogenous viruses) and was associated with mutations in the *APOBEC* family of cytidine deaminases.[21–25] Moreover, hotspot mutations (c.1624G>A [p.Glu542Lys] and c.1633G>A [p.Glu545Lys]) on the SMG *PIK3CA* (MIM 171834) were significantly enriched in ESCC tumors that had an *APOBEC* signature ($p = 0.0028$, Fisher's exact test, one-sided), implicating APOBEC activity as a key driver of *PIK3CA* mutagenesis in ESCC. Signature B was characterized by an enrichment of C>T mutations at XpCpG trinucleotides as a result of an elevated rate of spontaneous 5-methyl-cytosine deamination.[26] The elevated C>T mutation rate at XpCpG trinucleotides is a well-recognized mutational mechanism probably due to deamination to thymine of methylated cytosines, which are usually at XpCpGs.[26] Signature C was represented by types that, to our knowledge, are not yet known.

Tobacco smoking is consistently reported as an important risk factor for esophageal cancer, especially for squamous cell carcinoma[27] and gastric cancer.[28] The combined cohort of 192 ESCC individuals (104 from this study and 88 from Song et al.[6]) included 153 smoking and 39 non-smoking individuals (Table S1). Notably, we observed no smoking-associated signature characterized by C>A mutations, which was defined by Alexandrov et al.,[29] within ESCC genomes. Furthermore, we compared the proportion of C>A transversions between smoking and non-smoking individuals and observed no statistically significant difference ($p = 0.687$, Figure 1C). Additionally, we found no smoking-associated signature in 149 EACs (at least 49 of 149 individuals with a history of smoking) from a pan-cancer study.[19,29] These results indicate that the smoking-associated signature of C>A mutations is limited in EC. Considering that epidemiological studies suggest that tobacco consumption might be associated with EC,[30] we speculate that other smoking-associated signatures that have not been recognized might contribute to malignancy of ESCC.

CNAs of ESCC

To investigate somatic CNAs in ESCC, we applied a modified GISTIC method to profile genome segments with CNAs in the 14 WGS tumors (Figure S4A). This analysis yielded 126 significantly altered regions (Table S4). One of the most significantly altered regions, 3q26.1–q29, contains *SOX2* (MIM 184429), encoding a member of the SRY-related HMG-box family of transcription factors involved in embryonic development regulation and in cell-fate determination.[31] This gene is overexpressed in ESCC and associated with chemoresistance. Downregulation of *SOX2* might inhibit ESCC tumorigenesis and increase sensitivity to chemotherapy.[32] Notably, of the nine tumors harboring *SOX2* amplifications, we observed one focal (defined as <100 kb) CNA containing only *SOX2* (Figure S5), indicating that *SOX2* is the most likely target of 3q26 amplification in ESCC.

Focal CNAs are more likely to contain driver genes that confer clonal advantage, are causally implicated in oncogenesis, and have been positively selected during the evolution of the cancer.[33] Therefore, to identify genes affected by recurrent CNAs, we used the Integrative Genomics Viewer to manually inspect the 126 significantly altered regions. This approach identified recurrent focal CNAs, including amplified chromosomal segments containing *CBX4* and *CBX8*[34,35] (Figure S4B). Amplification of these genes was validated by FISH and qPCR copy-number assay (Figure 2A). Of the 104 ESCC specimens, 39% and 51% of tumors had an immunoreactivity score for CBX4 and CBX8 levels, respectively, which is at least double that of matched normal tissue ($T_{IRS}/N_{IRS} > 2$; Figure 2B). Moreover, silencing *CBX4* and *CBX8* in KYSE2 and KYSE510 cells (which harbor high endogenous expression) significantly inhibited cell proliferation, colony formation, and cell invasion (Figures 2C–2E; Figure S6). These results suggest that *CBX4* and *CBX8* amplification and the resultant protein upregulation contribute to the development of ESCC and that they might thus serve as potential drug targets for ESCC treatment.

SMGs

We next applied the MutSigCV method[36] to identify SMGs in the combined 192 ESCC tumors (Figure S1E) and discovered nine such genes driven by point mutations (false discovery rate, $q < 0.1$) and six further genes with $p < 0.01$ (Figure 3A). Eleven of these 15 genes—including *AJUBA*, *ZNF750*, *FAT1*, *FBXW7* (MIM 606278), and *PTCH1* (MIM 601309) and the chromatin-remodeling genes *CREBBP* (MIM 600140) and *BAP1* (MIM 603089)—harbored frequent inactivating mutations. Although *AJUBA*, *ZNF750*, *FAT1*, and *FBXW7* were recently implicated as tumor suppressors in ESCC,[7,37] their roles in mice models and the mechanisms by which they function as tumor suppressors are limited. As in other cancers,[38,39] particularly EAC,[19] this analysis also identified well-known cancer-associated genes, such as *TP53* (MIM 191170), *PIK3CA*, and *CDKN2A* (MIM 600160), as SMGs in ESCC, thus providing evidence of common dysfunctions in cell-cycle control and apoptotic signaling.

AJUBA, which encodes a LIM domain protein, inhibits the ATR-dependent DNA-damage response and is involved in several cellular processes, such as cell-fate determination, cytoskeletal organization, transcriptional repression, mitotic commitment, cell-cell adhesion, and migration.[40] Along with *LIMD1* (MIM 604543), *AJUBA* has been proposed to be a major component of the miRNA-mediated gene-silencing machinery and might have a tumor-suppressive function.[41] However, the biological functions of *AJUBA* in ESCC tumorigenesis have never been reported. The following *AJUBA* mutations were identified and verified in four tumor samples: two stop-gain mutations (c.985G>T [p.Glu329*] and c.1057C>T [p.Gln353*]) and two frameshift indels

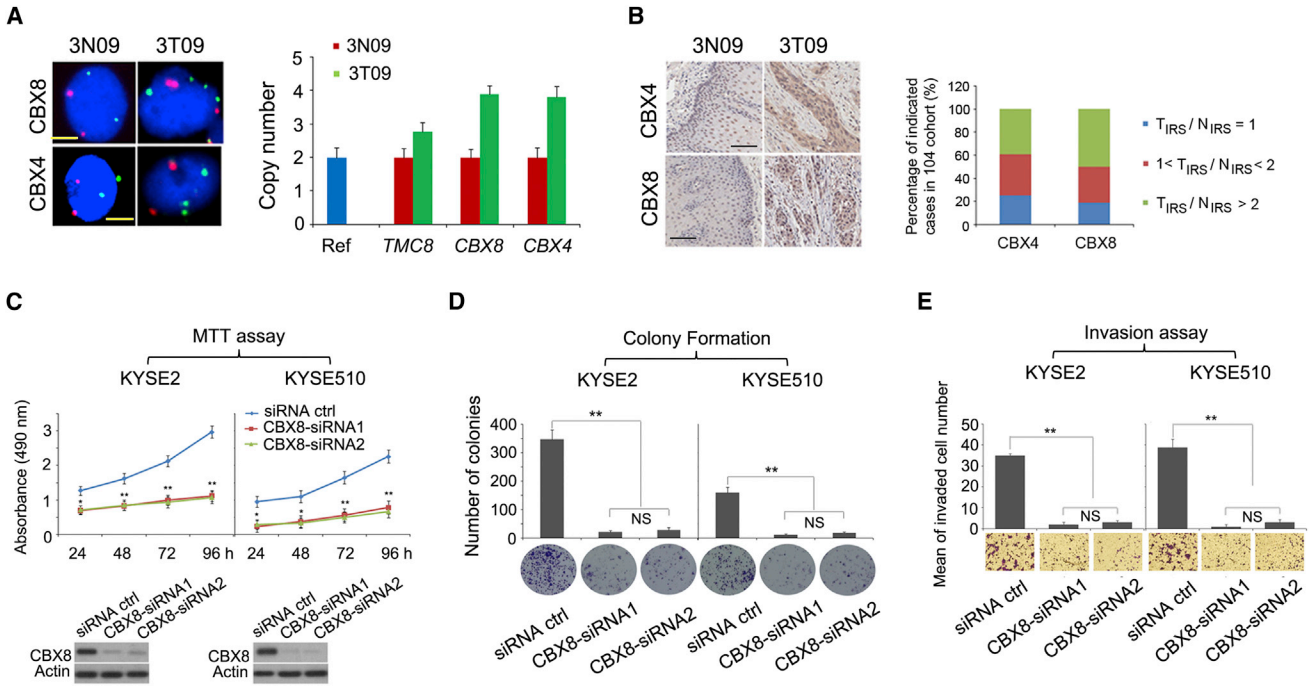

Figure 2. Amplified Genes Identified in ESCC by GISTIC

(A) *CBX4* and *CBX8* were amplified in ESCC tumors. Left panel: immunofluorescence images show signals produced from FISH analyses using probes specific to chromosomal region 5q (red, control) and *CBX4* or *CBX8* (green) in ESCC samples 3N09 and 3T09. Scale bars represent 5 μm. Right panel: *CBX4* and *CBX8* copy number was assayed by qPCR in 3N09 and 3T09. *RPPH1* was used as a normal reference, and *TMC8* (located near this region) was used as a focal CNA control. Data represent the mean ± SD. All assays were performed in triplicate.

(B) The amounts of CBX4 and CBX8 were increased in ESCC tumors. Representative immunohistochemistry images show CBX4 or CBX8 levels in the same tumor. The bar graph shows the percentage of indicated individuals with different CBX4 or CBX8 levels in the 104-individual ESCC cohort. CBX4 or CBX8 levels were based on subjective assessment of immunohistochemical staining intensity (see Material and Methods). Scale bars represent 400 μm.

(C–E) Depletion of *CBX8* in KYSE2 and KYSE510 cells significantly inhibited cell proliferation, as measured by MTT assay (C), colony-formation assays (D), and cell invasion (E). Knockdown of *CBX8* is demonstrated by immunoblotting (bottom of C); β-actin was used as a loading control. Data represent the mean ± SD. At least three independent experiments were performed in triplicate. Statistical analysis was performed with a two-sided t test. $^{**}p < 0.01$, $^{*}p < 0.05$.

(c.790_791insT [p.Val264fs*] and c.152delG [p.Gly51fs*]) in the 104-individual cohort and one frameshift insertion (c.1249_1250insA [p.Ala417fs*]) and one splice-site mutation in two individuals from our previous cohort[6] (Figure 3B; Figure S7A; Table S3). These mutational events result in truncated or disrupted protein products that lack proper LIM domains, indicating that they are loss-of-function mutations, and these mutations in *AJUBA*, encoding the truncated or disrupted protein, are expected to promote ESCC oncogenesis. In support of this possibility, immunoblot analysis showed the presence of truncated AJUBA in ESCC tumors (Figure 4A). *AJUBA* knockdown in KYSE140 and KYSE510 cells led to increased cell growth, colony formation, cell migration, and cell invasion (Figure S8). Moreover, exogenous levels of wild-type AJUBA in KYSE30 and KYSE150 cells with low endogenous AJUBA levels suppressed cell growth, cell migration, and cell invasion; these effects were abrogated by AJUBA alterations (p.Gln353* and p.Val264fs*; Figures 4B–4F). Together with our genetic observations, these functional data indicate that *AJUBA* might act as a tumor suppressor in ESCC and that the *AJUBA* mutations observed in ESCC abrogate its tumor-suppressive function.

Mutations in *ZNF750*, a tumor-associated gene located at 17q25,[7] were identified in 6% of ESCC tumors. We identified 11 somatic mutations, of which we verified (via Sanger sequencing) six, including two nonsense mutations (c.620G>A [p.Trp207*] and c.209C>A [p.Ser70*]), one indel (c.108_111del [p.Glu37Lysfs*]), and three missense mutations (c.96T>A [p.Phe32Leu], c.1520A>C [p.Asp507Ala], and c.1508C>G [p.Ser503Cys]) in *ZNF750* in the 104-individual cohort. In addition, five (c.414C>A [p.Cys138*], c.770C>A [p.Ser257*], c.85C>T [p.Gln29*], c.625_626insAA [p.Ala209fs*], and c.1621G>A [p.Ala541Thr]) out of 11 somatic mutations were verified by Sanger sequencing in our previous cohort[6] (n = 88; Figure 3B). Sixty-four percent of the mutations identified in *ZNF750* are inactivating. Additionally, *ZNF750* deletions, but no somatic mutations, were observed in 3 out of 14 tumor samples in the WGS set (21%, G-score > 0.23) and validated with a qPCR copy-number assay. The qPCR copy-number assay also determined that four out of six tumor samples harboring

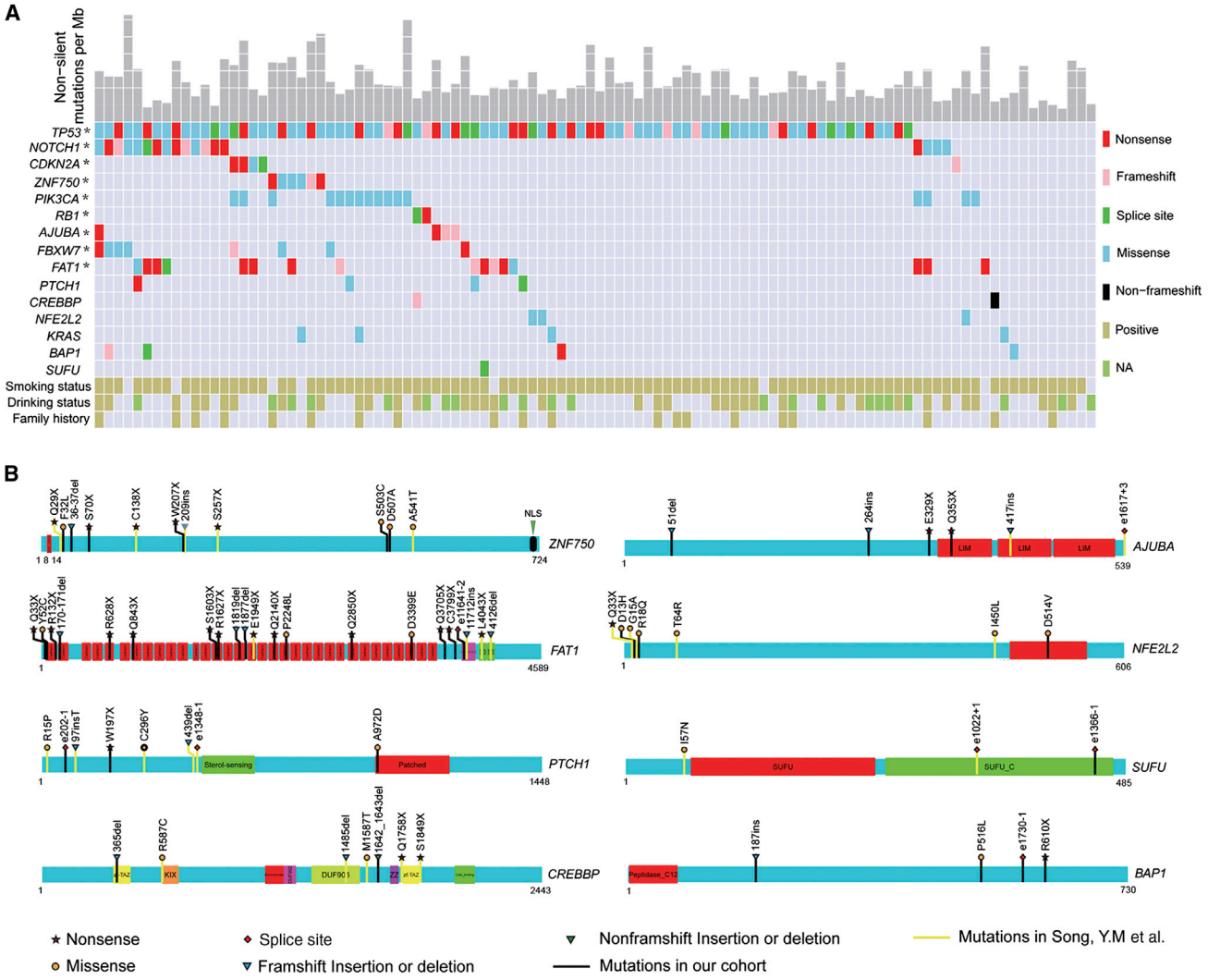

Figure 3. Somatic Mutations of Candidate Cancer-Associated Genes, Ranked by Significance across 192 Tumors

(A) Candidate driver genes were identified via MutSigCV significance analysis. The gray histogram at the top shows the number of mutations per megabase in each sample, and each rectangle represents 1 Mb. The main section shows each mutation type for every sample, including the total number of mutated samples per gene; mutation subtypes are denoted by color. If multiple mutations were observed within a gene in a single sample, only one is shown. The candidate cancer-associated genes are shown on the left: genes marked by an asterisk have a q value of <0.1. The bottom three rows indicate smoking status, drinking status, and family history. NA means that the condition remains unknown, and "positive" means that the individual has the condition shown on the left. The full list of mutated genes is provided in Table S3.

(B) A schematic representation of the domain structure of proteins (ZNF750, AJUBA, FAT1, NFE2L2, PTCH1, SUFU, CREBBP, and BAP1) encoded by SMGs shows the location of somatic variants identified in ESCC tumors. DNA mutations and amino acid changes are shown. The C-terminal nuclear localization sequence (NLS) of ZNF750 is shown in black.

ZNF750 mutations or indels in the 90-sample WES set were affected by deletions, indicating that inactivation of both alleles had occurred in these individuals (Figures S9A–S9C). Notably, frequent loss of heterozygosity at 17q25.3 has been previously reported in ESCC,[42] but no gene at or near the 17q25.3 region was identified. Our genetic data strongly indicate that *ZNF750* is the missing piece of this puzzle.

We next used the immunohistochemical method to determine whether a correlation exists between these genetic changes and the amount of ZNF750. As expected, ZNF750, a nuclear factor, was strongly stained in the nuclei of normal esophageal epithelial cells. Surprisingly, though, it was dramatically upregulated in the cytoplasm of ESCC tumor cells in comparison to that of normal tissue cells. Moreover, individuals with tumors harboring *ZNF750* mutations had higher cytoplasmic expression levels than did those lacking *ZNF750* mutations (Figure 5A). Mislocalization of a truncated (p.Ser70*) ZNF750 (Figure 5B) indicated that cytoplasm translocation was partly caused by loss of the C-terminal nuclear localization sequence. ZNF750 regulates the gene program controlling terminal epidermal differentiation.[43] For investigating a function for ZNF750 in tumorigenesis, shRNA-mediated stable *ZNF750* depletion was followed

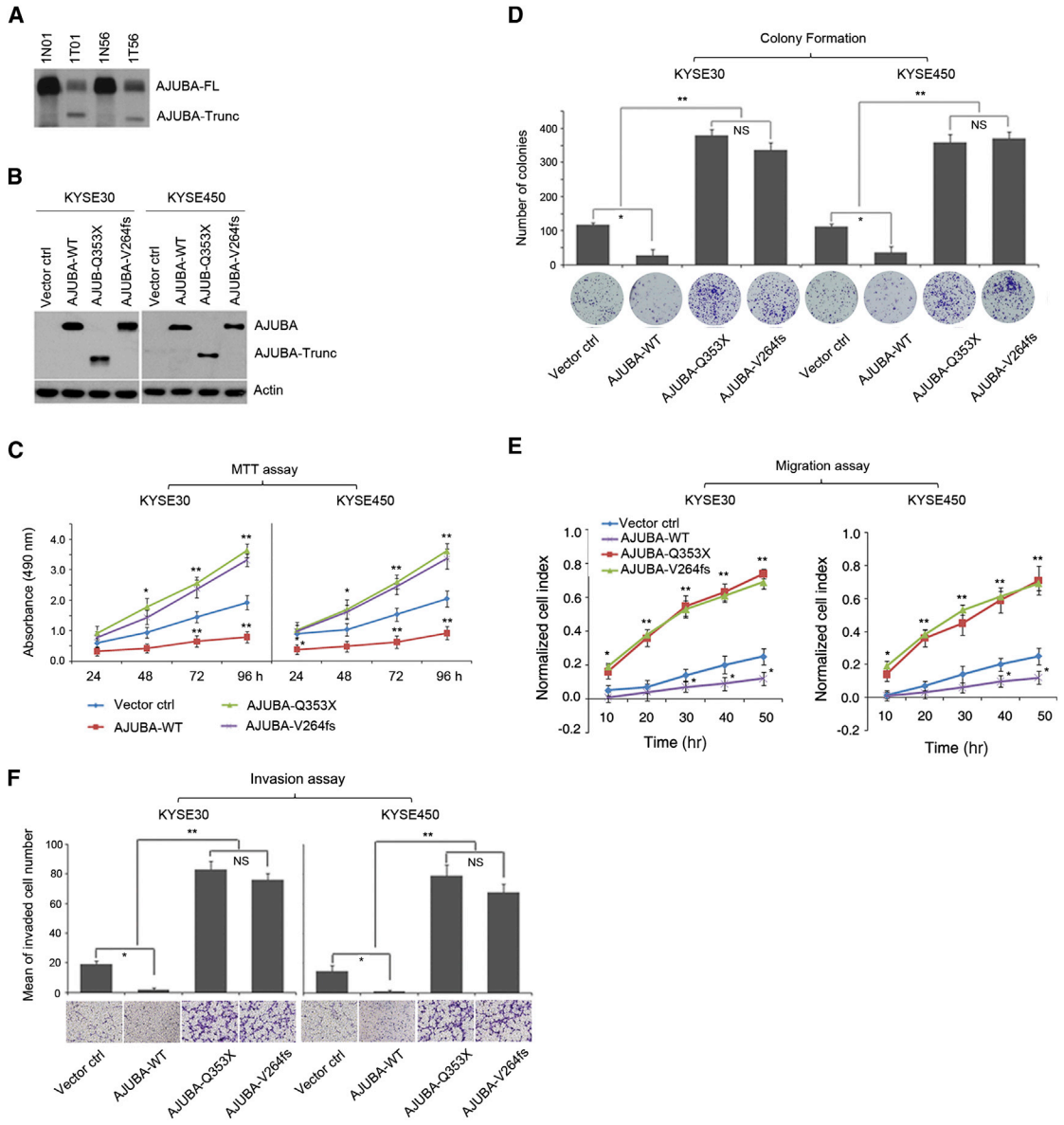

Figure 4. ***AJUBA*** **Acts as a Tumor Suppressor in ESCC**

(A) Immunoblotting validation of truncated AJUBA proteins in two identified ESCC tumors.

(B–F) The effect of wild-type and altered AJUBA on cell growth, cell migration, and cell invasion, as monitored by MTT assay (C), colony-formation assay (D), and cell-migration and cell-invasion assays in an xCELLigence RTCA DP system (E and F). The presence of wild-type AJUBA or AJUBA altered by p.Gln353* (p.Q353X) or p.Val264fs* (p.V264fs) in KYSE30 and KYSE150 cells was confirmed by immunoblotting with anti-AJUBA antibody, which recognizes the N terminus of AJUBA (B). These cell lines all have a low level of endogenous AJUBA.

Data represent the mean ± SD; three independent experiments were performed in triplicate. Statistical analysis was performed with a two-sided t test. **$p < 0.01$, *$p < 0.05$.

by transfection with wild-type or altered (p.Ser70* or p.Trp207*) ZNF750 in KYSE150 and KYSE140 cells. *ZNF750* knockdown strongly promoted KYSE150 and KYSE140 cell proliferation, migration, and invasion. Moreover, functional studies demonstrated that wild-type ZNF750 inhibited cell growth, migration, and invasion, and this effect was abrogated by altered (p.Ser70* or p.Trp207*) ZNF750 (Figure 5C; Figures S9D and S9E). Finally, *ZNF750* depletion and p.Ser70* ZNF750 markedly increased tumor size in the xenograft system in mice, whereas wild-type ZNF750 significantly decreased it (Figure 5D). Our genetic observations and functional data therefore suggest that *ZNF750* acts as a tumor suppressor in ESCC.

FAT1, which encodes a cadherin-like protein commonly expressed in epithelial tissues,[44] was mutated in 15% of ESCC tumors. Notably, 13 mutations (77%) were truncating (stop-gain and frameshift) (Table S3). *FAT1* is reported to regulate cell-cell adhesion and other cell behavior by controlling actin polymerization.[44] Recently,

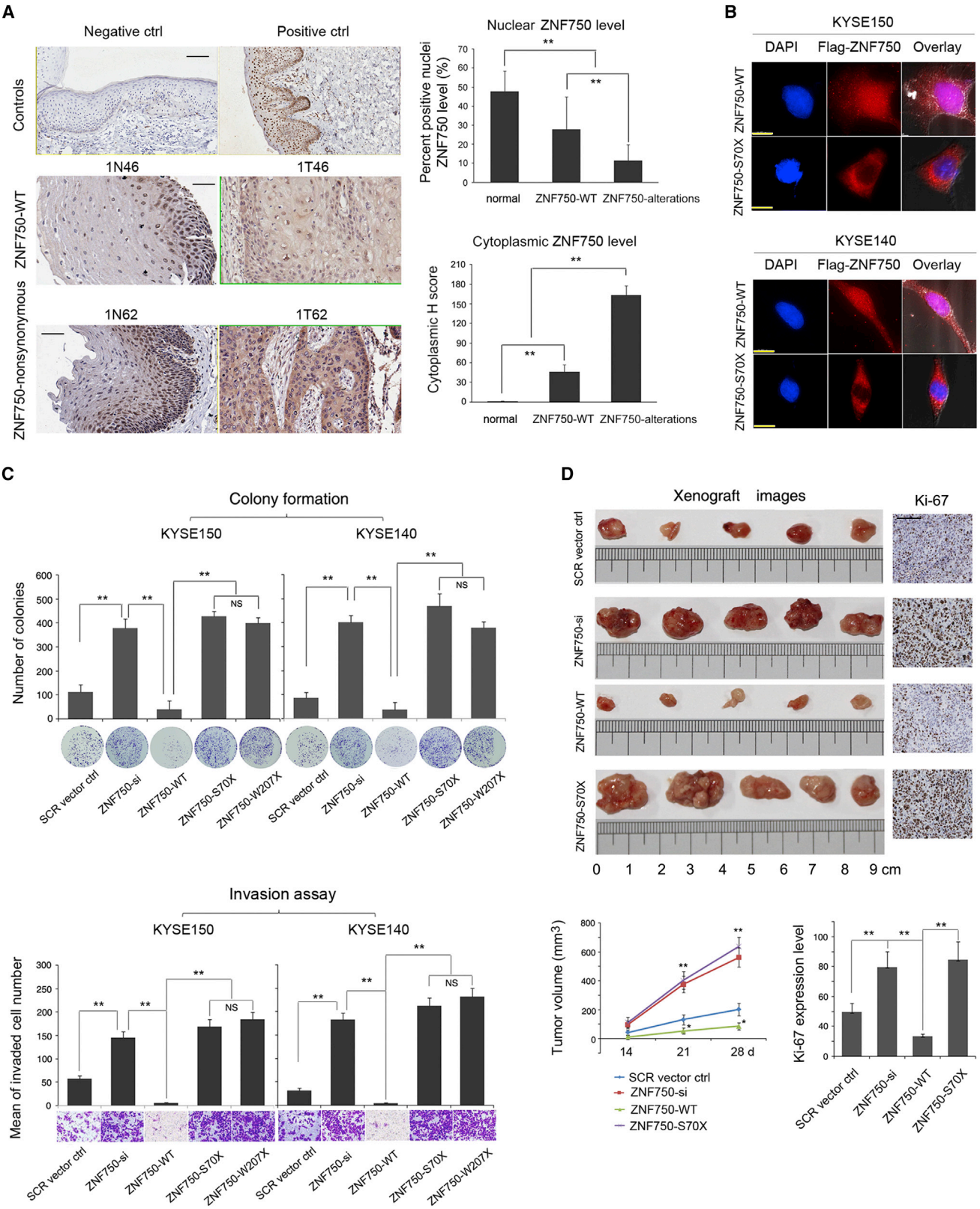

Figure 5. *ZNF750* Acts as a Tumor Suppressor in ESCC

(A) Immunohistochemical analysis shows ZNF750 staining in esophageal tissues and matched normal tissue of ESCC-affected individuals harboring wild-type ZNF750 (ZNF750-WT, second row) or nonsynonymous alterations (ZNF750-nonsynonmyous, third row). Negative and positive controls are normal skin samples (first row). Graphs (right) show nuclear (top) and cytoplasmic (bottom) amounts of ZNF750 in ESCC and normal esophageal tissue on the basis of a subjective assessment of immunohistochemical staining intensity (see Material and Methods, χ^2 test, $^{**}p < 0.01$, $^{*}p < 0.05$). Scale bars represent 400 μm.

(legend continued on next page)

Morris et al. reported that *FAT1* suppresses cancer cell growth by binding β-catenin and preventing nuclear localization.[44] *FAT1* inactivation, via mutations that affect the cytoplasmic domain, leads to aberrant Wnt/β-catenin signaling in multiple cancer types.[45] In our cohort, we discovered mutations affecting cadherin repeats and laminin G domains but no mutations affecting the cytoplasmic domain. This observation indicates that mutations affecting FAT1 extracellular domains might disrupt cell-cell associations and increase invasiveness and thus potentially contribute to ESCC tumorigenesis. *FBXW7* mutations were observed in eight ESCC tumors in our cohort: these included two nonsense mutations (c.1005T>A [p.Cys335*] and c.409G>T [p.Glu137*]), one frameshift deletion (c.736_739del [p.Gly247Profs*] and inactivating mutations), and five missense mutations predicted to be deleterious by SIFT and PolyPhen-2 analyses (Table S3).[46] *FBXW7* mutations and copy-number loss and a subsequent decreased FBXW7 level have been observed in various cancer types.[46,47] Decreased amounts of FBXW7 are reported to correlate with poor prognosis;[47] however, we observed no correlation between *FBXW7* mutations and prognosis in our cohort.

Inactivating mutations in several chromatin-remodeling genes, including *CREBBP* and *BAP1*, frequently occurred in our 104 ESCC samples. *CREBBP* and *EP300* (MIM 602700) mutations occurred in 11/104 ESCC tumors (Table S3). Inactivating *CREBBP* and *EP300* mutations have been reported in various human cancer types.[6,48] We also identified mutations in *BAP1*, which encodes a nuclear deubiquitinase involved in chromatin remodeling.[49] *BAP1* mutations have been reported in renal carcinoma and uveal melanoma,[49] but not in ESCC to date. Moreover, frequent truncating mutations were observed in the chromatin-remodeling genes *KMT2D* (*MLL2* [MIM 602113]) and *KMT2C* (*MLL3* [MIM 606833]) (14% together) and *KDM6A* (MIM 300128) (3%) (Table S3). At least one chromatin-remodeling gene was altered in 33/104 ESCC tumors.

Altered Pathways in ESCC

In our screen, SMGs were enriched in four pathways known to be important in cancer, including the cell cycle, NOTCH signaling, PI3K signaling, and cell adhesion. In addition, analyses of the KEGG pathway revealed overrepresented mutations of hedgehog (Hh) signaling and MAPK signaling, and a high percentage of truncating mutations were observed in chromatin-remodeling genes (Figure 6A). Eleven recurrent mutated genes involving the cell-cycle regulatory pathway were identified, and *TP53*, *CDKN2A*, and *RB1* (MIM 614041) accounted for 88%, 8%, and 2%, respectively (Table S3). Moreover, frequent truncating mutations were observed in *KMT2D* and *KMT2C* (14%), *CREBBP* and *EP300* (13%), and *KDM6A* (3%). Genes involved in the PI3K-AKT-mTOR pathway were mutated in 29% of 104 tumors, and *PIK3CA* was the most significantly altered gene (17%). Immunohistochemical analysis confirmed the presence of PIK3CA, AKT1, and GLI1 in tumors (compared with matched normal tissue) in the 104-individual WGS and WES cohorts (Figure 6B). Hence, these data shed light on the essential role of dysregulation of these critical pathways in tumorigenesis of ESCC.

Discussion

In this study, we report the detailed, next-generation analysis of mutational processes underlying ESCC. By combining our findings with a previous dataset, we found three genome-wide mutational signatures for ESCC. Notably, an APOBEC-mediated mutational signature indicates that the APOBEC family might get access to the nucleus and cause ESCC-associated genomic damage or mutation. The APOBEC family can deaminate cytosine to uracil within DNA, leading to mutation clusters in various types of cancer.[22] Several APOBEC family members have access to the nuclear compartment and can impede the cell cycle, most likely through DNA deamination and the ensuing DNA-damage response. Such genomic damage might contribute to carcinogenesis.[23,24] APOBEC3B, acting on lentiviral replication intermediates constituting an innate pathway of antiretroviral defense,[24,25] was reported to be present in high amounts in ESCC,[7] suggesting that it might be an attractive candidate for the mechanisms underlying mutation signature A in ESCC. In our cohort and that from Song et al.,[6] at least 47% of ESCC tumors had an *APOBEC* signature, suggesting that *APOBEC*-catalyzed genomic uracil lesions are responsible for a large proportion of mutations in ESCC. Thus, the *APOBEC* signature might be a potential oncogenic pathway underlying mutational mechanisms in ESCC development.

Moreover, we identified several significantly amplified focal CNAs containing two ESCC-associated genes (*CBX4*

(B) Immunofluorescence of FLAG-tagged wild-type ZNF750 (ZNF750-WT, first and third rows) and p.Ser70* ZNF750 (ZNF750-S70X, second and fourth rows) in KYSE150 (top) and KYSE140 (bottom) cells. DAPI labels the nucleus. Scale bars represent 10 μm.
(C) Endogenous *ZNF750* was stably knocked down in KYSE150 and KYSE140 cells and then forced to encode FLAG-tagged wild-type ZNF750 (ZNF750-WT) or the p.Ser70* (ZNF750-S70X) or p.Trp207* (ZNF750-W207X) variant. Cell proliferation was monitored by colony-formation assay. SCR indicates scramble control. The data represent the mean ± SD; three independent experiments were performed in triplicate. Data were statistically analyzed by a two-sided t test. $^{**}p < 0.01$, $^{*}p < 0.05$.
(D) Tumor xenografts show significant growth inhibition of cells overexpressing WT ZNF750. Five representative tumors are shown: (from top to bottom) the SCR vector control, *ZNF750* knockdown (ZNF750-si), wild-type ZNF750 (ZNF750-WT), and the p.Ser70* variant (ZNF750-S70X) (n = 6 per group). The graph at the bottom left shows tumor volumes. Immunohistochemical analyses of Ki-67 are also shown (in the images on the right [the scale bar represents 200 μm] and the graph at the bottom right). Data represent the mean ± SD. $^{**}p < 0.01$, $^{*}p < 0.05$.

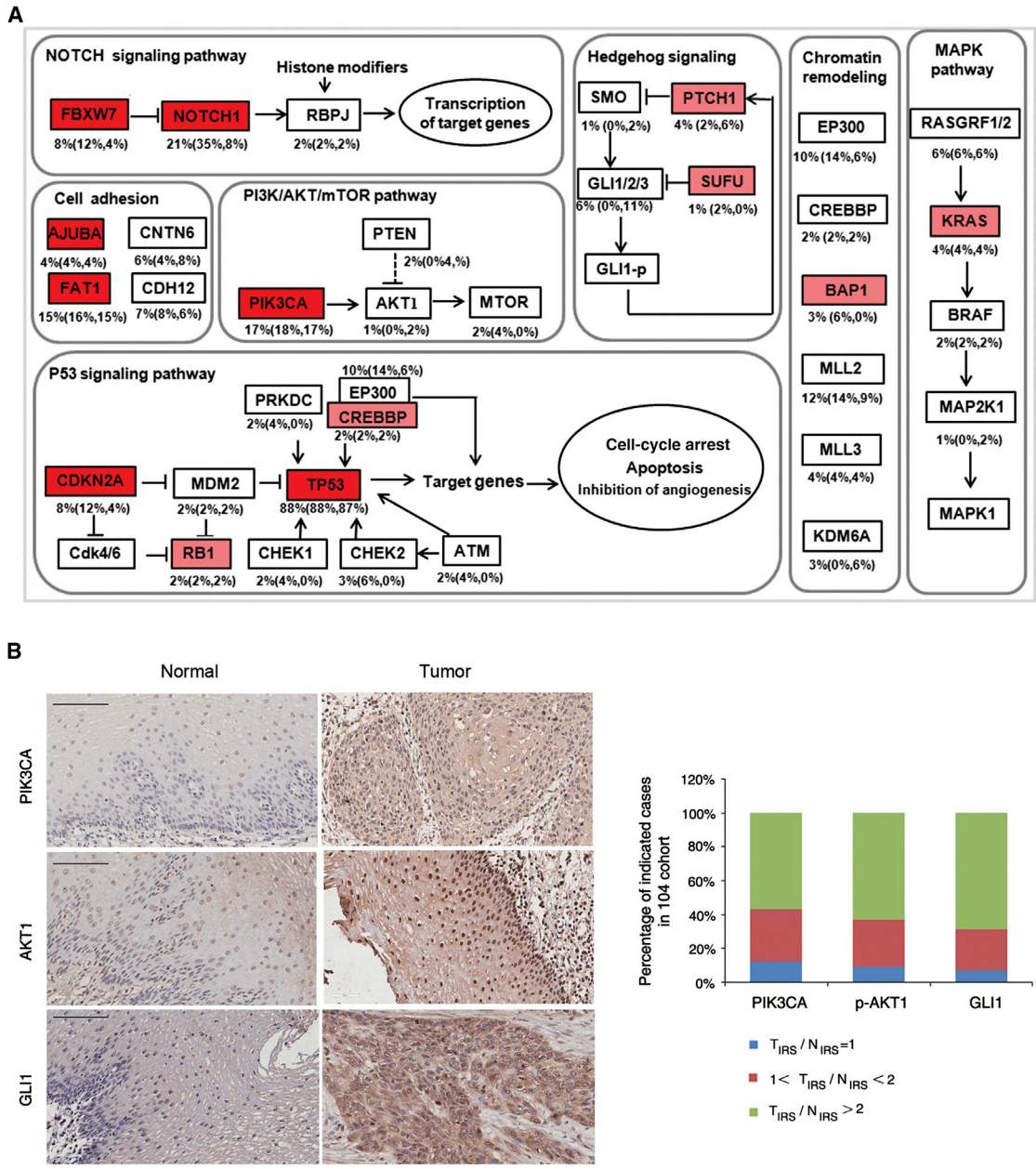

Figure 6. Altered Pathways in ESCC

(A) Key cancer pathway components altered in ESCC. ESCC potential driver genes and their mutation frequencies were mapped to the following major groups: NOTCH signaling, PI3K pathway, p53 pathway, Hh pathway, MAPK pathway, chromatin remodeling, and regulation of cell adhesion. Alteration frequencies are shown as a percentage of those in all samples; genes identified as SMGs by the MutSigCV analytical method are colored in red. Excitatory (arrows) and inhibitory (black lines) interactions were taken from the KEGG pathway database.

(B) Representative immunohistochemistry images show PIK3CA, AKT1, or GLI1 levels in tumors and matched normal tissues. The bar graph shows PIK3CA, AKT1, or GLI1 levels on the basis of a subjective assessment of immunohistochemical staining intensity. Scale bars represent 400 μm.

and *CBX8*) not previously observed in ESCC. *CBX4* and *CBX8* are major transcriptional repressors that epigenetically modify chromatin and exert important functions in cell-cycle regulation, DNA repair, cell differentiation, cell senescence, and cell death.[34] *CBX4* and *CBX8* expression has been demonstrated to be upregulated in certain types of human cancer.[35] However, their amplification and biological roles have not previously been reported in ESCC. Our functional and clinical analyses indicate that these genes might act as oncogenes in ESCC and thus represent potential therapeutic targets.

In addition to identifying the well-defined cancer-associated genes *TP53*, *NOTCH1* (MIM 190198), *PIK3CA*, *RB1*, *CDKN2A*, and *FBXW7*, we identified recently implicated ESCC-associated genes *AJUBA*, *ZNF750*, and *FAT1* in our combined cohort. Recently, pan-cancer studies across 21

cancer types predicted that *AJUBA* and *ZNF750*, two SMGs in our cohort, are associated with some types of cancer.[50] Using functional studies, we found that *AJUBA* and *ZNF750* (previously known but with unknown function in ESCC) act as tumor suppressors and that mutations in these genes abrogated their tumor-suppressive effects. Therefore, *AJUBA* and *ZNF750* mutations encoding truncated or disrupted proteins might contribute to ESCC tumorigenesis.

Our previously published paper identified *TP53*, *RB1*, *CDKN2A*, *PIK3CA*, *NOTCH1*, *NFE2L2* (MIM 600492), *ADAM29* (MIM 604778), and *FAM135B* as SMGs on the basis of the dataset of WGS in 17 ESCC individuals and WES in 71 individuals recruited from the Chaoshan District of Guangdong Province, an area of high ESCC prevalence in southern China.[6] In this study, in addition to identifying *TP53*, *PIK3CA*, *NOTCH1*, and *CDKN2A*, which have been implicated previously, we further identified *FBXW7*, *FAT1*, *AJUBA*, and *ZNF750* as SMGs from the dataset of WGS in 14 ESCC individuals and WES in 90 individuals recruited from Shanxi and Henan provinces in the Taihang Mountains, an area of high ESCC prevalence in northern China. When we combined these two cohorts, we identified more SMGs, including *PTCH1* and the chromatin-remodeling genes *CREBBP* and *BAP1*. Of these SMGs, *AJUBA*, *ZNF750*, *FAT1*, and *FBXW7* were identified in ESCC individuals recruited from northern China, but not in those from southern China, which was confirmed in the cohorts from Lin et al.[7] and Gao et al.[37] (individuals in their cohorts were from the Cancer Institute and Hospital of the Chinese Academy of Medical Sciences, where individuals are mostly from northern China). Epidemiologic studies have shown that in addition to smoking, alcohol consumption, and family history, the known risk factors for ESCC were thought to be hot food and N-nitroso compounds in northern China, whereas chewing of fermented areca nut has been shown to be independently associated with ESCC in southern China.[1,2] We speculate that the different environmental carcinogens (e.g., dietary and lifestyle patterns) between northern and southern China might cause genetic diversity in ESCC individuals.

In summary, we used genomics data to pinpoint the predominant underlying mutational processes linked to ESCC. A signature attributed to the *APOBEC* family of cytidine deaminases was predominant in ESCC individuals, reflecting that this gene family contributes to cytosine mutation clusters in ESCC and cancer susceptibility through *APOBEC*-dependent mutational processes. Unlike the situation with lung cancer, smoking tobacco was not strongly associated with ESCC in the Chinese population, giving insights into the mutational burden not associated with this lifestyle choice. Furthermore, we have provided functional evidence of inactivating mutations in *AJUBA* and *ZNF750*, amplification of *CBX4* and *CBX8*, and genomic aberrations targeting the cell cycle and NOTCH, PI3K, and Hh signaling pathways. These results highlight the substantial genetic alterations contributing to ESCC tumorigenesis and provide a set of potential therapeutic targets toward its prevention and chemotherapies. Further studies are required to explore how *APOBEC* family members and oncogenic *ZNF750*, *AJUBA*, *CBX4*, and *CBX8* contribute to mutagenesis in ESCC, whether these genes are promising for drug discovery, and whether the therapies targeting PI3K and/or Hh signaling pathways are particularly promising strategies for ESCC.

Accession Numbers

Data, including sequence data and analyses, are available for download from the NCBI Sequence Read Archive under accession number SRA112617.

Supplemental Data

Supplemental Data include nine figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.02.017.

Acknowledgments

This work was supported by funding from the 973 National Fundamental Research Program of China (2015CB553904 to Q.Z.), the National Natural Science Foundation of China (81021061 and 81230047 to Q.Z., 81330063 and 81272189 to Y.C., 81272694 to X. Cheng, 81201956 to J.L., and 81402342 to L.Z.), the Key Project of Chinese Ministry of Education (NO213005A to Y.C.), the Specialized Research Fund for the Doctoral Program of Higher Education (20121417110001 to Y.C.), a research project supported by the Shanxi Scholarship Council of China (2013-053 to Y.C.), the Innovative Team in Science & Technology of Shanxi (2013131023 to Y.C.), and the Program for the Outstanding Innovative Teams of Higher Learning Institutions of Shanxi (to Y.C.).

Received: December 30, 2014
Accepted: February 26, 2015
Published: April 2, 2015

Web Resources

The URLs for data presented herein are as follows:

Agilent SureSelect Solution, http://www.genomics.agilent.com/en/home.jsp
Database for Annotation, Visualization, and Integrated Discovery (DAVID), http://david.abcc.ncifcrf.gov/
Kyoto Encyclopedia of Genes and Genomes (KEGG), http://www.genome.jp/kegg
MutSigCV, http://www.broadinstitute.org/cancer/cga/mutsig
OMIM, http://omim.org/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
Sequence Read Archive (SRA), http://www.ncbi.nlm.nih.gov/sra
SIFT, http://sift.jcvi.org/
UCSC Genome Browser, http://genome.ucsc.edu/

References

1. Pennathur, A., Gibson, M.K., Jobe, B.A., and Luketich, J.D. (2013). Oesophageal carcinoma. Lancet *381*, 400–412.

2. Engel, L.S., Chow, W.H., Vaughan, T.L., Gammon, M.D., Risch, H.A., Stanford, J.L., Schoenberg, J.B., Mayne, S.T., Dubrow, R., Rotterdam, H., et al. (2003). Population attributable risks of esophageal and gastric cancers. J. Natl. Cancer Inst. *95*, 1404–1413.
3. Jia, P., Pao, W., and Zhao, Z. (2014). Patterns and processes of somatic mutations in nine major cancers. BMC Med. Genomics *7*, 11.
4. Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordóñez, G.R., Bignell, G.R., et al. (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. Nature *463*, 191–196.
5. Pleasance, E.D., Stephens, P.J., O'Meara, S., McBride, D.J., Meynert, A., Jones, D., Lin, M.L., Beare, D., Lau, K.W., Greenman, C., et al. (2010). A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature *463*, 184–190.
6. Song, Y., Li, L., Ou, Y., Gao, Z., Li, E., Li, X., Zhang, W., Wang, J., Xu, L., Zhou, Y., et al. (2014). Identification of genomic alterations in oesophageal squamous cell cancer. Nature *509*, 91–95.
7. Lin, D.C., Hao, J.J., Nagata, Y., Xu, L., Shang, L., Meng, X., Sato, Y., Okuno, Y., Varela, A.M., Ding, L.W., et al. (2014). Genomic and molecular characterization of esophageal squamous cell carcinoma. Nat. Genet. *46*, 467–473.
8. Duan, F., Xie, W., Cui, L., Wang, P., Song, C., Qu, H., Wang, K., Zhang, J., and Dai, L. (2013). Novel functional variants locus in PLCE1 and susceptibility to esophageal squamous cell carcinoma: based on published genome-wide association studies in a central Chinese population. Cancer Epidemiol. *37*, 647–652.
9. Liu, Y., Cao, L., Chang, J., Lin, J., He, B., Rao, J., Zhang, Z., and Zhang, X. (2014). XPF-673C>T polymorphism effect on the susceptibility to esophageal cancer in Chinese population. PLoS ONE *9*, e94136.
10. Fang, Y., Xiao, F., An, Z., and Hao, L. (2011). Systematic review on the relationship between genetic polymorphisms of methylenetetrahydrofolate reductase and esophageal squamous cell carcinoma. Asian Pac. J. Cancer Prev. *12*, 1861–1866.
11. Chiang, D.Y., Getz, G., Jaffe, D.B., O'Kelly, M.J., Zhao, X., Carter, S.L., Russ, C., Nusbaum, C., Meyerson, M., and Lander, E.S. (2009). High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat. Methods *6*, 99–103.
12. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010). Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. USA *107*, 16910–16915.
13. Takenaka, C., Nishishita, N., Takada, N., Jakt, L.M., and Kawamata, S. (2010). Effective generation of iPS cells from CD34+ cord blood cells by inhibition of p53. Exp. Hematol. *38*, 154–162.
14. Liu, J., Cheng, X., Zhang, Y., Li, S., Cui, H., Zhang, L., Shi, R., Zhao, Z., He, C., Wang, C., et al. (2013). Phosphorylation of Mps1 by BRAFV600E prevents Mps1 degradation and contributes to chromosome instability in melanoma. Oncogene *32*, 713–723.
15. Lal, S., La Du, J., Tanguay, R.L., and Greenwood, J.A. (2012). Calpain 2 is required for the invasion of glioblastoma cells in the zebrafish brain microenvironment. J. Neurosci. Res. *90*, 769–781.
16. Vagnarelli, P., Hudson, D.F., Ribeiro, S.A., Trinkle-Mulcahy, L., Spence, J.M., Lai, F., Farr, C.J., Lamond, A.I., and Earnshaw, W.C. (2006). Condensin and Repo-Man-PP1 co-operate in the regulation of chromosome architecture during mitosis. Nat. Cell Biol. *8*, 1133–1142.
17. Zhang, L., Shi, R., He, C., Cheng, C., Song, B., Cui, H., Zhang, Y., Zhao, Z., Bi, Y., Yang, X., et al. (2013). Oncogenic B-Raf(V600E) abrogates the AKT/B-Raf/Mps1 interaction in melanoma cells. Cancer Lett. *337*, 125–132.
18. Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjöblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., et al. (2007). The genomic landscapes of human breast and colorectal cancers. Science *318*, 1108–1113.
19. Dulak, A.M., Stojanov, P., Peng, S., Lawrence, M.S., Fox, C., Stewart, C., Bandla, S., Imamura, Y., Schumacher, S.E., Shefler, E., et al. (2013). Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nat. Genet. *45*, 478–486.
20. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013). Deciphering signatures of mutational processes operative in human cancer. Cell Rep. *3*, 246–259.
21. Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.
22. Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol. Cell *10*, 1247–1253.
23. Caval, V., Suspène, R., Vartanian, J.P., and Wain-Hobson, S. (2014). Orthologous mammalian APOBEC3A cytidine deaminases hypermutate nuclear DNA. Mol. Biol. Evol. *31*, 330–340.
24. Yu, Q., Chen, D., König, R., Mariani, R., Unutmaz, D., and Landau, N.R. (2004). APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. J. Biol. Chem. *279*, 53379–53386.
25. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C., and Fenton, T.R. (2014). APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. Cell Rep. *7*, 1833–1841.
26. Waters, T.R., and Swann, P.F. (2000). Thymine-DNA glycosylase and G to A transition mutations at CpG sites. Mutat. Res. *462*, 137–147.
27. Lin, Y., Totsuka, Y., He, Y., Kikuchi, S., Qiao, Y., Ueda, J., Wei, W., Inoue, M., and Tanaka, H. (2013). Epidemiology of esophageal cancer in Japan and China. J. Epidemiol. *23*, 233–242.
28. Pandey, A., Tripathi, S.C., Mahata, S., Vishnoi, K., Shukla, S., Misra, S.P., Misra, V., Hedau, S., Mehrotra, R., Dwivedi, M., and Bharti, A.C. (2014). Carcinogenic Helicobacter pylori in gastric pre-cancer and cancer lesions: association with tobacco-chewing. World J. Gastroenterol. *20*, 6860–6868.
29. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.
30. Wu, C., Hu, Z., He, Z., Jia, W., Wang, F., Zhou, Y., Liu, Z., Zhan, Q., Liu, Y., Yu, D., et al. (2011). Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nat. Genet. *43*, 679–684.

31. Hoogstraat, M., de Pagter, M.S., Cirkel, G.A., van Roosmalen, M.J., Harkins, T.T., Duran, K., Kreeftmeijer, J., Renkens, I., Witteveen, P.O., Lee, C.C., et al. (2014). Genomic and transcriptomic plasticity in treatment-naive ovarian cancer. Genome Res. *24*, 200–211.
32. Gen, Y., Yasui, K., Nishikawa, T., and Yoshikawa, T. (2013). SOX2 promotes tumor growth of esophageal squamous cell carcinoma through the AKT/mammalian target of rapamycin complex 1 signaling pathway. Cancer Sci. *104*, 810–816.
33. Stratton, M.R. (2011). Exploring the genomes of cancer cells: progress and promise. Science *331*, 1553–1558.
34. Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., et al. (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nat. Cell Biol. *15*, 353–362.
35. Li, G., Warden, C., Zou, Z., Neman, J., Krueger, J.S., Jain, A., Jandial, R., and Chen, M. (2013). Altered expression of polycomb group genes in glioblastoma multiforme. PLoS ONE *8*, e80970.
36. Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.
37. Gao, Y.B., Chen, Z.L., Li, J.G., Hu, X.D., Shi, X.J., Sun, Z.M., Zhang, F., Zhao, Z.R., Li, Z.T., Liu, Z.Y., et al. (2014). Genetic landscape of esophageal squamous cell carcinoma. Nat. Genet. *46*, 1097–1102.
38. Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., McKenna, A., et al. (2011). The mutational landscape of head and neck squamous cell carcinoma. Science *333*, 1157–1160.
39. Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., Lawrence, M.S., Sivachenko, A.Y., Sougnez, C., Zou, L., et al. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. Nature *486*, 405–409.
40. Jiang, S., Katayama, H., Wang, J., Li, S.A., Hong, Y., Radvanyi, L., Li, J.J., and Sen, S. (2010). Estrogen-induced aurora kinase-A (AURKA) gene expression is activated by GATA-3 in estrogen receptor-positive breast cancer cells. Horm. Cancer *1*, 11–20.
41. James, V., Zhang, Y., Foxler, D.E., de Moor, C.H., Kong, Y.W., Webb, T.M., Self, T.J., Feng, Y., Lagos, D., Chu, C.Y., et al. (2010). LIM-domain proteins, LIMD1, Ajuba, and WTIP are required for microRNA-mediated gene silencing. Proc. Natl. Acad. Sci. USA *107*, 12499–12504.
42. Iwaya, T., Maesawa, C., Ogasawara, S., and Tamura, G. (1998). Tylosis esophageal cancer locus on chromosome 17q25.1 is commonly deleted in sporadic human esophageal cancer. Gastroenterology *114*, 1206–1210.
43. Sen, G.L., Boxer, L.D., Webster, D.E., Bussat, R.T., Qu, K., Zarnegar, B.J., Johnston, D., Siprashvili, Z., and Khavari, P.A. (2012). ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. Dev. Cell *22*, 669–677.
44. Morris, L.G., Kaufman, A.M., Gong, Y., Ramaswami, D., Walsh, L.A., Turcan, Ş., Eng, S., Kannan, K., Zou, Y., Peng, L., et al. (2013). Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. Nat. Genet. *45*, 253–261.
45. Tukachinsky, H., Lopez, L.V., and Salic, A. (2010). A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu-Gli protein complexes. J. Cell Biol. *191*, 415–428.
46. Kuhn, E., Wu, R.C., Guan, B., Wu, G., Zhang, J., Wang, Y., Song, L., Yuan, X., Wei, L., Roden, R.B., et al. (2012). Identification of molecular pathway aberrations in uterine serous carcinoma by genome-wide analyses. J. Natl. Cancer Inst. *104*, 1503–1513.
47. Iwatsuki, M., Mimori, K., Ishii, H., Yokobori, T., Takatsuno, Y., Sato, T., Toh, H., Onoyama, I., Nakayama, K.I., Baba, H., and Mori, M. (2010). Loss of FBXW7, a cell cycle regulating gene, in colorectal cancer: clinical significance. Int. J. Cancer *126*, 1828–1837.
48. Le Gallo, M., O'Hara, A.J., Rudd, M.L., Urick, M.E., Hansen, N.F., O'Neil, N.J., Price, J.C., Zhang, S., England, B.M., Godwin, A.K., et al.; NIH Intramural Sequencing Center (NISC) Comparative Sequencing Program (2012). Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. Nat. Genet. *44*, 1310–1315.
49. Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat. Genet. *46*, 225–233.
50. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.