Supplementary Materials and Methods

Cell Lines

NCI-N87, KATOIII, and HEK293T cell lines were purchased from the American Type Culture Collection (Manassas, VA). MKN28 and NUGC4 were obtained from Japanese Collection of Research Bioresources Cell Bank. All GC cell lines were cultured in RPMI medium (Gibco BRL, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS) (Gibco BRL). HEK293T cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% FBS. All cell lines used in this study were regularly authenticated by morphologic observation and tested for mycoplasma contamination (MycoAlert; Lonza Rockland, Rockland, ME). Cells were incubated at 37°C in a humidified incubator containing 5% CO_2.

Clinical Tissue Samples

Primary GCs and matched normal gastric tissues were obtained from the SingHealth Tissue Repository (Singapore) or National University Hospital (Singapore), with approvals from Institutional Review Board, National University of Singapore, and signed patient informed consent. In this study, "normal" (ie, non-tumor) samples refer to samples harvested from the stomach, from sites distant from the tumor, and exhibiting no visible evidence of tumor or intestinal metaplasia/dysplasia upon surgical assessment. Tumor samples were confirmed by cryosectioning to contain >40% tumor cells. Clinicopathologic data of the GC patients are shown in Supplementary Table 5.

Transcriptome Sequencing (RNA-Seq)

All RNA-Seq libraries were prepared using the Illumina Tru-Seq RNA Sample Preparation v2 protocol (according to manufacturer's instructions). All libraries were validated with an Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA), diluted to 11 pM and were then applied to an Illumina flow cell using the Illumina Cluster Station. Sequencing was performed on an Illumina HiSeq2000 sequencer at the Duke-NUS Genome Biology Facility using paired-end 76-bp read option. RNA-Seq data is available at the European Genome-phenome Archive under Accession No: EGAS00001001128.

Global Identification of RNA Editing Sites

A 2-stage process was used for the discovery and quantification of RNA editing levels in the entire transcriptome. In the first stage, a previously published pipeline was used for the discovery of high-confidence RNA editing sites.[1] For each RNA-Seq data set, the RNA-Seq data was mapped onto a combined fasta file based on the hg19 reference genome and a junction database generated from transcript annotations derived from UCSC, Refseq, Ensembl and GencodeV19 using the mapper BWA (v0.7.5a-r405).[2] After which, PCR duplicates were removed using SAMtools (v0.1.19) and junction-mapped reads were converted to genomic-based coordinates. Reads with a mapping quality <20 were then discarded. Variant calling was then performed and variant sites supported by at least 2 variant reads and with a variant allelic frequency >0.1 were then retained. Sites of variants that were present in databases of DNA variants (1000 genome, whole genome sequencing project and dbSNP138), within 4 bp of the intronic side of the intron-exon boundary or in homopolymeric regions were all removed. Candidate variants that were found to be in non-uniquely mapped reads as determined by BLAT were also excluded. As the whole genome sequencing data for most of the samples used in this study were not available, we utilized the methodology of using multiple samples for the identification of higher-confidence RNA editing sites based on principles described previously.[3]

In the second phase of the analysis, SAMtools mpileup and custom written Perl and Python scripts were then used to extract RNA editing levels of all candidate sites identified in the first phase of the analysis across all samples used in the study. Subsequently, to identify the strand on which the variant occurred, annotations based on UCSC known genes were utilized. Variants with overlapping annotations on both strands were discarded, as it was not possible to uniquely identify the strand from which the variant was derived. The strand information of the variant was subsequently used to determine the nucleotide change from the reference nucleotide at the RNA level.

Assessment of Effectiveness of a Multi-Sample Method in the Removal of Rare Germline Variants

As described here, all candidate editing sites found in the primary list were prefiltered based on a number of DNA variant databases. However, this might be inadequate for removing rare and unannotated germline variants. To assess the effectiveness of the multi-sample methodology in removing rare germline variants, the whole genome sequencing of 4 matched pairs of GC tumors and NT samples. The sequence data from whole-genome profiles were aligned to human reference genome (hs37d5) using BWA-MEM, version 0.7.9a (http://bio-bwa.sourceforge.net/). PCR duplicates were removed using SAMTools (http://samtools.sourceforge.net/). The mapped reads were subsequently realigned around insertion/deletion using Genome Analysis Toolkit (pmid: 20644199) (GATK) version 1.0. The realigned data were used as inputs for GATK Unified Genotyper for calling single-nucleotide variants. Candidate RNA editing variants were subsequently overlapped with this list of single nucleotide DNA variants to assess the number of contaminating DNA variants in our candidate list of RNA editing variants.

Gastric Cancer Microarray Datasets

Details of GC microarray datasets used (SG and Samsung) are listed in Supplementary Table 6. ADAR1/2 levels were compared between normal and GC tissues using the "limma" Bioconductor R package.[4] For survival analysis, GCs were classified into 3 groups based on their ADAR1/2 expression levels: "normal ADAR1 and ADAR2 expression" group, "ADAR1 overexpression (OE) or ADAR2

down-regulation (DR)" group, and "ADAR1 OE and ADAR2 DR" group. For the SG cohort, fold changes (FC) of ADAR1/2 were calculated using either gene expression of their matched NT gastric sample or the median expression of all NT samples if the tumor sample did not have a matched normal tissue. Matched gastric samples with ADAR1 $\log_2$(FC) $>75^{th}$ percentile were classified as ADAR OE and those with ADAR2 $\log_2$(FC) $<25^{th}$ percentile were classified as ADAR2 DR. Unmatched GC samples with ADAR1 $\log_2$(FC) $>70^{th}$ percentile were classified as ADAR1 OE and those with ADAR2 $\log_2$(FC) $<20^{th}$ percentile were classified as ADAR2 DR. Similarly, for the Samsung cohort, ADAR1 OE was defined as expression $\geq 70^{th}$ percentile of ADAR1 expression, while ADAR2 DR was defined as expression $\leq 20^{th}$ percentile of ADAR2 expression. Kaplan-Meier plots and log-rank tests were used for OS analysis between the 3 ADAR1 or ADAR2 groups. OS time was calculated from the date of surgery to the last follow-up data. For each dataset, univariate and multivariate survival analyses were performed using the Cox proportional hazards regression model with OS defined as the outcome measure. A *P* value of $<.05$ was considered to be statistically significant.

Copy Number Variation Data Analysis

We analyzed the copy number variation dataset for TCGA stomach adenocarcinoma cohort of 413 patients. Copy number change was analyzed by GISTIC based on the single nucleotide polymorphism array data and retrieved from the Broad Firehose platform, while mRNA expression data (RNAseqV2) was extracted from TCGA data portal. The copy number status of each patient's tumor sample at the *ADAR1* and *ADAR2* loci was correlated with the mRNA expression of ADAR1 and ADAR2, respectively. A copy number increase ranging from 1 to 4 is defined as a low-level copy number gain, while anything >4 is defined as a high-level copy number gain. The normalized mRNA expression of ADAR1 and ADAR2 was analyzed based on the normalized gene expression value RPKM (reads per million per kilobase).[5]

Plasmid Construction

Wild-type ADAR1 p110 and ADAR2 expression constructs were generated as described previously.[6] ADAR1 and ADAR2 catalytic mutants carry point mutations introduced into the deaminase domain (H910Y and E912A for ADAR1, and E396A for ADAR2). PCR-directed mutagenesis was performed using an inner forward or reverse primer containing nucleotide alterations (AR1-CM-forward: 5′- GA GAAACTGTCAATGACTGCTATGCAGCAATAATCTCCCGGAGAG GCTT -3′; AR1-CM-reverse: 5′- AAGCCTCTCCGGGAGATTAT TGCTGCATAGCAGTCATTGACAGTTTCTC -3′; AR2-CM-forward: 5′- CATTAAATGACTGCCATGCAGCAATAATATCTCGGA GATCCTT -3′; AR2-CM-reverse: 5′- AAGGATCTCCGAGATATT ATTGCTGCATGGCAGTCATTTAATG -3′) with the corresponding external primers used for amplifying the wt form.

The wt (unedited) and edited *PODXL* expression constructs were generated using the pLenti6/V5-TOPO vector (Invitrogen, Carlsbad, CA) according to manufacturer's instructions.

Establishment of Stable Cell Lines Using a Lentiviral Expression System

As described previously,[6] expression constructs were transfected into the 293FT cell line for virus generation. Virus-containing supernatants were collected for subsequent transduction into GC cells. At 48 hours after virus transduction, half of the cells were collected for sequencing and expression analysis. The remaining cells were cultured in complete medium containing blasticidin (2 μg/mL; Sigma-Aldrich, St Louis, MO) for establishing stably transduced cell lines.

ADAR1 and ADAR2 Knockdown Experiments

The pLKO.1-puro vector was purchased from Sigma-Aldrich. shRNAs against *ADAR1* were designed following RNA consortium guidelines (http://www.broadinstitute.org/rnai/public/) (#1: CCGGAGTTTCCTGCTTAAGCAAATA CTGCAGTATTTGCTTAAGCAGGAAACTTTTTTG; #5:CCTGGTG GAATCCAGTGACATTGTGCCTACTCAAGAGGTAGGCACAATGT CACTGGATTCCACCAGGTTTTTG). Due to alternative exons at the 5′ end of the mRNA, the p150 and p110 isoforms initiate at a methionine (Met) at position 1 and 296, respectively, such that the *ADAR1* p150 and p110 isoforms share 100% identity of the overlapped mRNA sequences. These 2 specific shRNAs against *ADAR1* gene could target both p150 and p110 isoforms. shRNAs were cloned into the PLKO.1-puro vector following Addgene's PLKO.1 protocol (http://www.addgene.org/tools/protocols/plko/). For ADAR2, Validated shRNA vectors were directly purchased from Sigma-Aldrich (TRCN0000050939: CCGGCCCGTGATG ATCTTGAACGAACTCGAGTTCGTTCAAGATCATCACGGGTTTT TG and TRCN0000050942: CCGGCCCAGGACTCAAGTAT-GACTTCTCGAGAAGTCATACTTGAGTCCTGGGTTTTTG). At 48 hours after virus transduction, half of control and knockdown cells were collected for subsequent analysis. The remaining cells were cultured in complete medium containing puromycin (0.2 μg/mL; Sigma-Aldrich) for establishing stably knockdown cell lines.

Complementary DNA Synthesis and Quantitative Real-Time Polymerase Chain Reaction

To quantify ADAR1 and ADAR2 expression levels in GC cells, equal amounts of cDNA were synthesized using the Advantage RT-for-PCR kit (Clontech, Mountain View, CA) and used for quantitative real-time PCR analysis. Quantitative real-time PCR was performed with GoTaq qPCR Master Mix (Promega) using a Rotor-Gene 6000 (Qiagen, Germany) and the following primers:

qADAR1-F(5′-CCCTTCAGCCACATCCTTC-3′),qADAR1-R(5′ GCCATCTGCTTTGCCACTT-3′), qADAR2-F(5′- CTGACACGCT CTTCAATGGTT-3′) and qADAR2-R(5′- GGCGCAGTTCGTTCA AGAT-3′). 18S was amplified as an internal control using the following primers: q18S-F: 5′-CTCTTAGCTGAGTGTCCCGC-3′; q18S-R: 5′-CTGATCGTCTTCGAACCTCC-3′. Relative expression levels of ADAR1 and ADAR2 were compared as described previously.[6,7]

Foci Formation Assay

Briefly, 1×10^3 cells were seeded in a 6-well plate. After culture for 10 days, surviving colonies were counted and stained with crystal violet (Sigma-Aldrich). Triplicate independent experiments were performed and the data were expressed as the mean ± SD of triplicate wells within the same experiment.

Anchorage-Independent Assay (Colony Formation in Soft Agar)

Briefly, 2×10^3 of cells were suspended in 2 mL top medium containing 0.4% (w/v) agarose, on 2 mL bottom medium with 0.6% agarose in 6-well plate. The number of colonies was determined at 14 days after plating. Triplicate independent experiments were performed and the data were expressed as the mean ± SD of triplicate wells within the same experiment.

Xenograft Tumor Formation Assay

In vivo tumorigenesis was investigated by tumor xenograft experiments. Approximately 6×10^6 of MKN28 cells transduced with EV, wt ADAR1 or ADAR2, catalytically inactive ADAR1 or ADAR2, and wt or edited PODXL lentivirus were injected subcutaneously into the left and right rear flank of 4-week-old NSG (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ) mice. Tumor formation in NSG mice was monitored during a 12-week period. The tumor volume was calculated by the formula V (volume) = 0.5 × L (length) × W (width)^2 at the end point.

In Vitro Cell Proliferation Assay

Briefly, 1×10^3 cells were seeded per well in a 96-well plate. A 50-μL volume of the XTT working solution provided by the kit (Roche Diagnostics, Indianapolis, IN) was added to each well. The cells were incubated at 37°C in the dark for 4 hours, followed by detection of the absorbency using a scanning multi-well spectrophotometer (Tecan Sunrise, Tecan Trading AG, Switzerland) at a test wavelength of 492 nm and a reference wavelength of 630 nm. The data are expressed as the mean ± SD of triplicate wells within the same experiment. Three independent experiments were conducted.

Matrigel Invasion Assay

We performed invasion assays using 24-well BioCoat Matrigel Invasion Chambers (BD Biosciences, San Jose, CA) according to the manufacturer's instructions. Briefly, 0.5×10^5 cells in FBS-free RPMI were added to the top chamber, and 10% FBS in RPMI was added to the bottom chamber as a chemoattractant. After 22 hours of incubation, cells that invaded the Matrigel were fixed and stained with crystal violet (Sigma-Aldrich). The number of cells was counted and imaged using SPOT imaging software (Nikon).

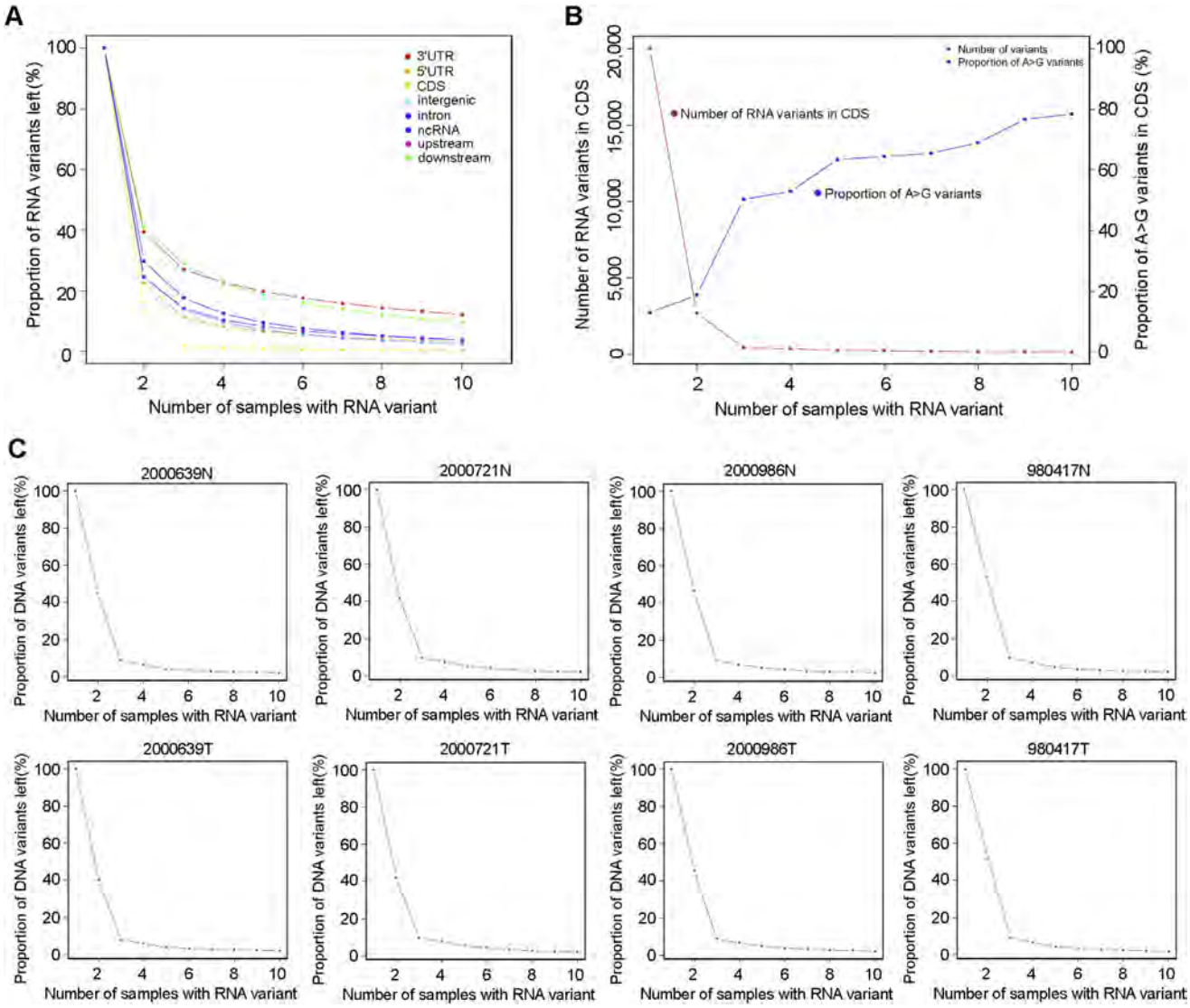
Antibodies and Western Blot Analysis

Mouse anti-ADAR1, anti–β-actin and rabbit anti-PODXL antibodies were purchased from Abcam (Cambridge, MA). The mouse anti-ADAR2 antibody was purchased from Sigma-Aldrich. Protein lysates were quantified and resolved on a sodium dodecyl sulfate polyacrylamide gel electrophoresis gel, transferred onto a polyvinylidene difluoride membrane (Millipore, Billerica, MA, USA) and immunoblotted with a primary antibody, followed by incubation with a secondary antibody. The blots were visualized by enhanced chemiluminescence (GE Healthcare, Buckinghamshire, UK).

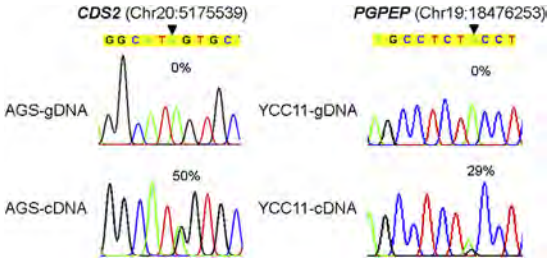
References

1. **Ramaswami G, Lin W, Piskol R**, et al. Accurate identification of human Alu and non-Alu RNA editing sites. Nat Methods 2012;9:579–581.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754–1760.
3. **Ramaswami G, Zhang R**, Piskol R, et al. Identifying RNA editing sites using RNA sequencing data alone. Nat Methods 2013;10:128–132.
4. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
5. **Mortazavi A, Williams BA**, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621–628.
6. **Chan TH, Lin CH**, Qi L, et al. A disrupted RNA editing balance mediated by ADARs (Adenosine DeAminases that act on RNA) in human hepatocellular carcinoma. Gut 2014;63:832–843.
7. **Chen L, Li Y**, Lin CH, et al. Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. Nat Med 2013;19:209–216.

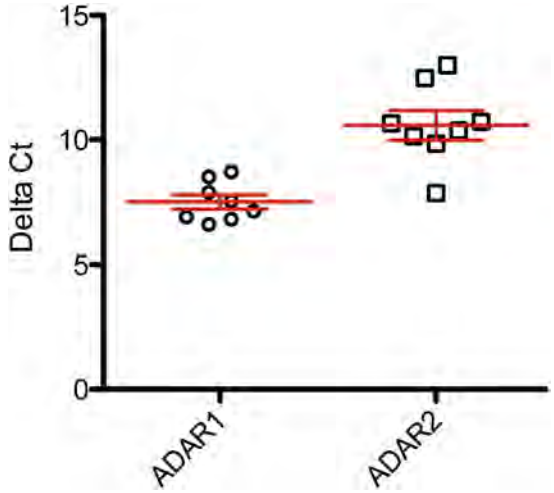
Author names in bold designate shared co-first authorship.



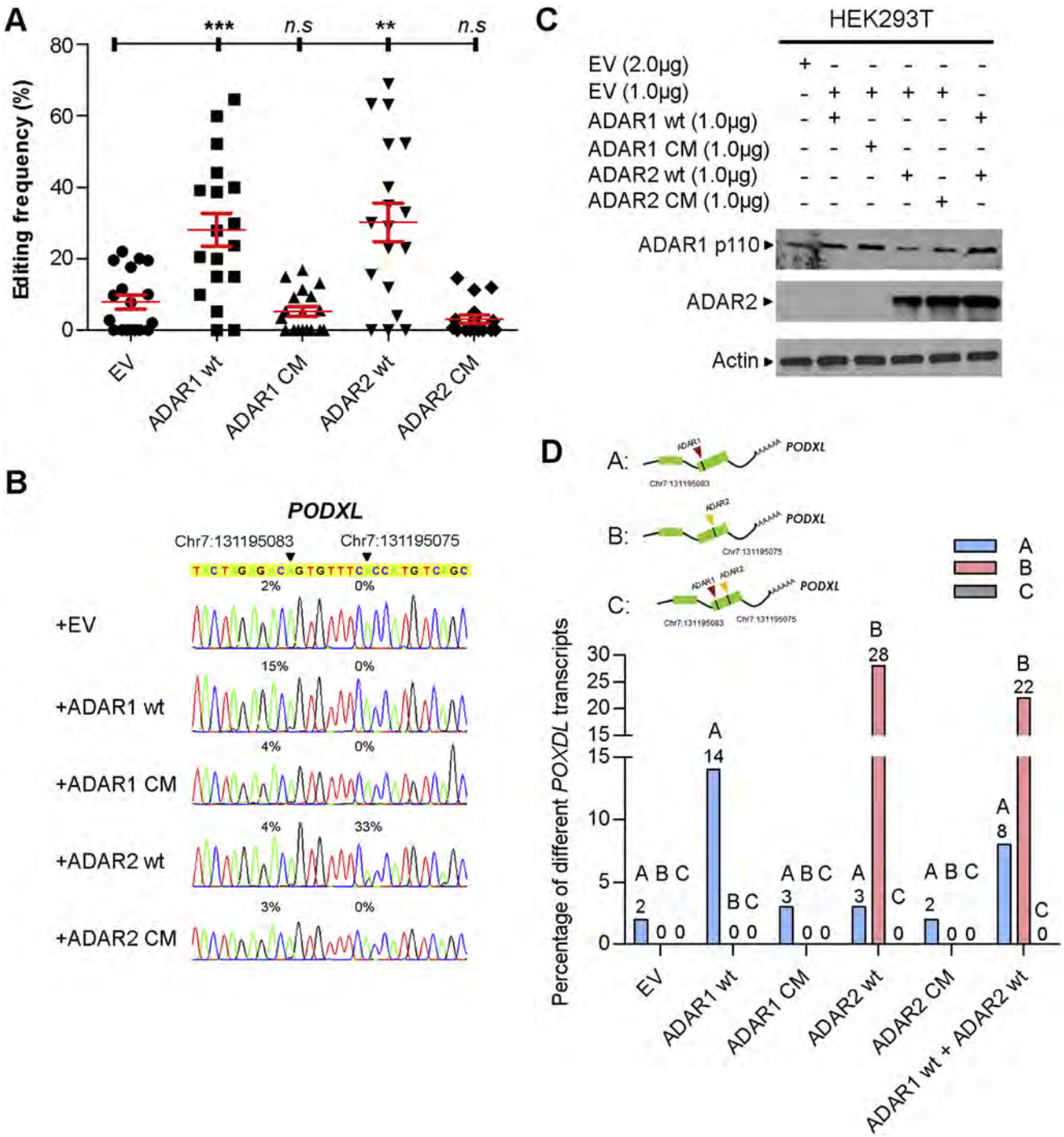
Supplementary Figure 1. Global identification of A-to-I (G) RNA editing sites in GC by RNA-Seq. (*A*) Proportion of the remaining RNA variants in different genomic regions in multiple samples ranging from 2 to 10, relative to those in one single samples. (*B*) Proportion of A-to-G RNA variants among the total RNA variants in CDS, in multiple samples ranging from 2 to 10, relative to those in one single sample. (*C*) Proportion of the contaminating DNA variants shown in the list of RNA variants detected in multiple samples, in 4 matched pairs of GC tumors and NT specimens (2000639N/T, 2000721N/T, 2000986N/T, and 980417N/T).



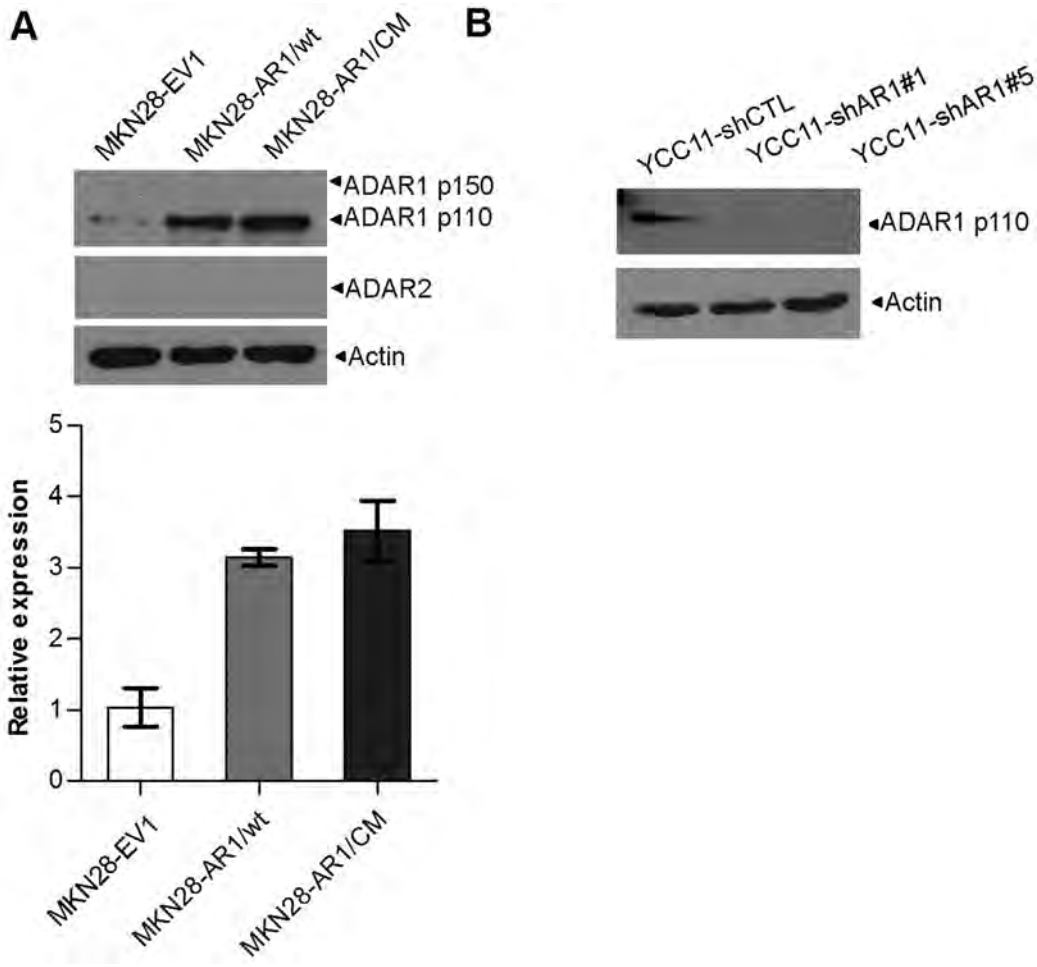
Supplementary Figure 2. *CDS2* and *PGPEP* are bona fide RNA editing targets. The sequence chromatograms of the corresponding genomic DNA (gDNA) (*upper*) and complementary DNA (cDNA) (*bottom*) sequences in GC cells are shown. An *arrow* indicates the editing position and value indicates the editing frequency of the corresponding editing site.



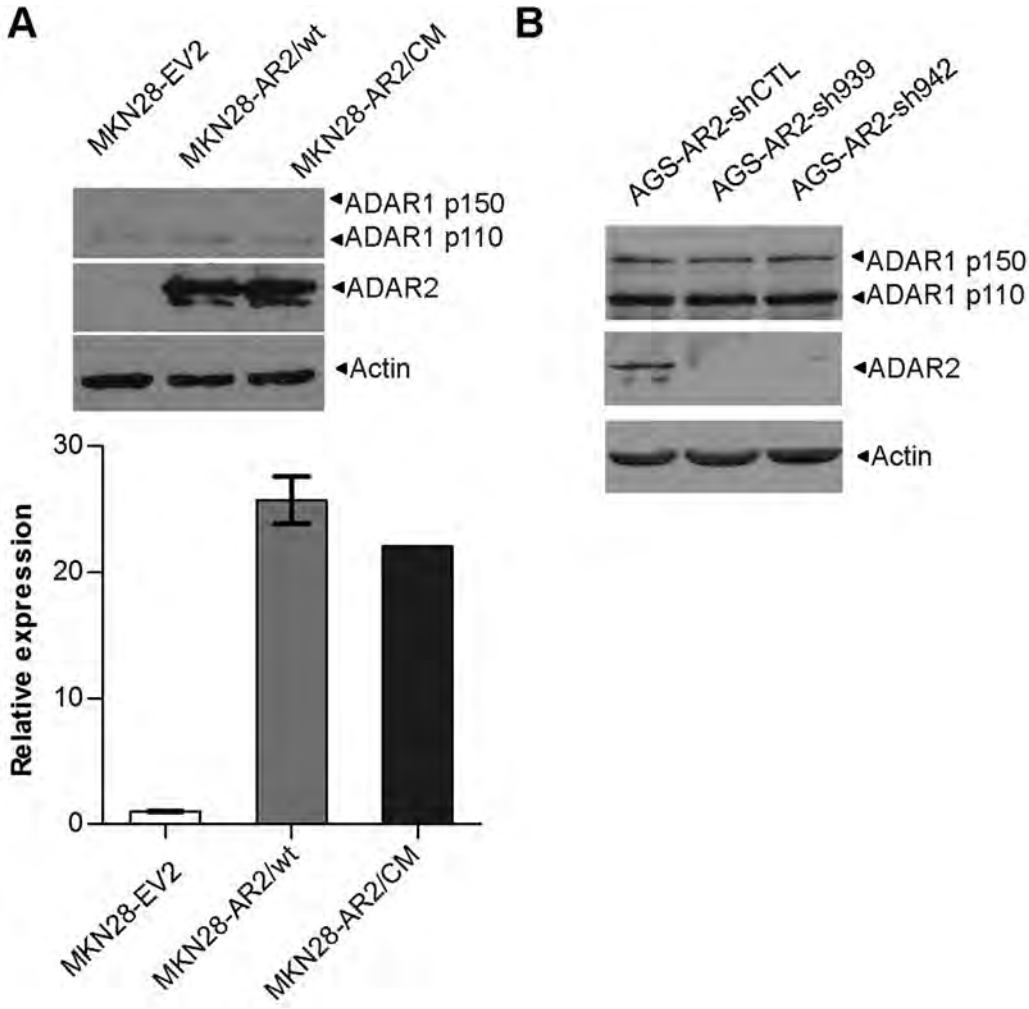
Supplementary Figure 3. ADAR2 is less abundantly expressed in GC cell lines. *Scatter plots* represent the ΔCt value of ADAR1 and ADAR2 in 8 GC cell lines, as detected by quantitative real-time PCR.



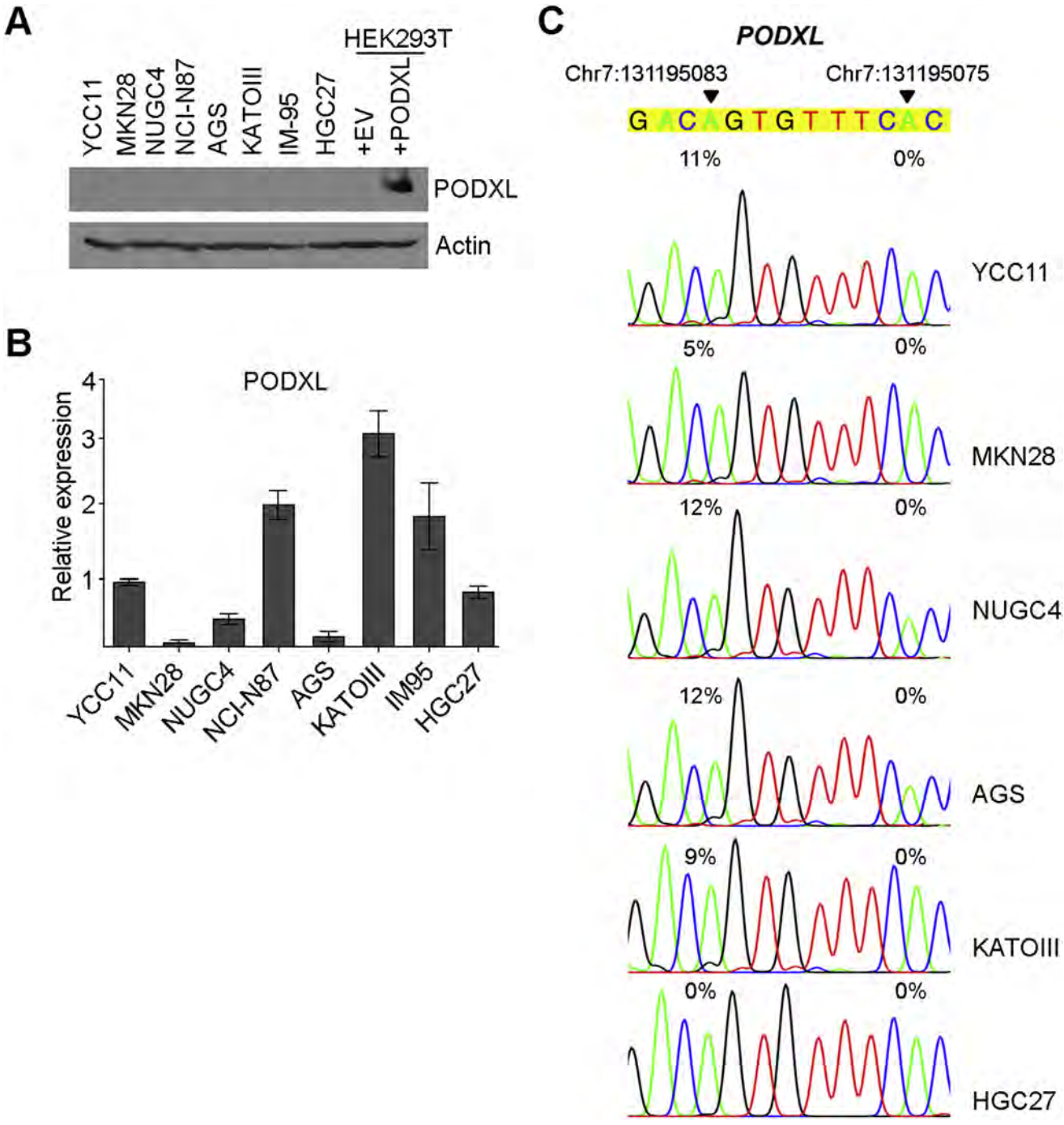
Supplementary Figure 4. Catalytic mutants of ADAR1 and ADAR2 are devoid of deaminase function. (*A*) *Dot plots* represent editing frequencies of 20 verified A-to-I RNA editing sites in HEK293T cells transiently transfected with EV, ADAR1 wt, ADAR1 catalytic mutant (ADAR1 CM), ADAR2 wt, and ADAR2 catalytic mutant (ADAR2 CM). (*ns*, no significance; $^{**}P < .01$; $^{***}P < .001$; Mann-Whitney U test) (*B*) Sequence chromatograms of the *PODXL* transcript in 5 groups of cells (EV, ADAR1 wt, ADAR1 CM, ADAR2 wt, and ADAR2 CM). An *arrow* indicates the editing position and value indicates the editing frequency of the corresponding editing site. (*C*) Western blot analysis showing expression of ADAR1 and ADAR2 proteins in HEK293T cells transfected with the indicated expression construct(s). β-Actin was the loading control. (*D*) *Bar chart* showing the percentage of *PODXL* transcripts harboring the edited ADAR1 ("A") *or* ADAR2 ("B")-regulated site alone, or both edited sites ("C") in each group of cells described. RNA editing analysis was conducted using the Sanger sequencing of 50 individual bacterial colonies for each sample as described in Materials and Methods.



Supplementary Figure 5. Establishment of ADAR1 overexpression and knockdown cell models. (*A*) Western blot analysis showing expression of ADAR1 and ADAR2 proteins in the indicated stable cell lines. β-Actin was the loading control. (*B*) Western blot analysis showing ADAR1 protein in YCC11 cells stably transfected with the scrambled shRNA and 2 specific shRNAs against *ADAR1* (YCC11-shCTL, YCC11-shAR1 #1 and #5).



Supplementary Figure 6. Establishment of ADAR2 overexpression and knockdown cell models. (*A*) Western blot analysis showing expression of ADAR1 and ADAR2 proteins in the indicated stable cell lines. β-Actin was the loading control. (*B*) Western blot analysis showing ADAR1 and ADAR2 proteins in AGS-AR2 cells stably transfected with the scrambled shRNA, 2 specific shRNAs against *ADAR2* (AGS-AR2-shCTL, AGS-AR2-sh939, and sh942).



Supplementary Figure 7. Expression level and editing frequency of *PODXL* in 8 GC cell lines. (*A*) Western blot analysis of PODXL protein in 8 GC cell lines. Antibody specificity was verified by overexpressing *PODXL* expression construct in HEK293T cells. β-Actin was the loading control. (*B*) Quantitative real-time PCR (qPCR) measurement of the *PODXL* transcripts in the indicated cell lines. The relative expression level of *PODXL* in each cell line as indicated is shown in a *bar chart* (mean $\pm$ SD of 3 independent experiments). (*C*) Sequence chromatograms of the *PODXL* transcript in the indicated cell lines. *Arrows* indicate the editing positions and value indicates the editing frequency of the corresponding editing site.