SUPPLEMENTARY INFORMATION

Recoding RNA editing of *antizyme inhibitor 1* predisposes to hepatocellular carcinoma

Leilei Chen[1, 2, 3*], Yan Li[1, 3*], Chi Ho Lin[4], Tim Hon Man Chan[1, 2, 3], Raymond Kwok Kei Chow[1], Yangyang Song[1, 3], Ming Liu[1, 3], Yun-Fei Yuan[5], Li Fu[1], Kar Lok Kong[1], Lihua Qi[2], Yan Li[5], Na Zhang[4], Amy Hin Yan Tong [4], Dora Lai-Wan Kwong[1], Kwan Man[6], Chung Mau Lo[6], Si Lok[4], Daniel G. Tenen[2], & Xin-Yuan Guan[1, 3, 5]

[1]*Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China;*

[2]*Cancer Science Institute of Singapore, National University of Singapore, Singapore;*

[3]*State Key Laboratory of Liver Research, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China;*

[4]*Genome Research Centre, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China;*

[5]*State Key Laboratory of Oncology in Southern China, Sun Yat-sen University Cancer Center, Guangzhou, China;*

[6]*Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China;*

*These authors contributed equally to this work.

This PDF file includes:

- Supplementary Data
- Supplementary Reference
- Supplementary Figures 1-7 (legends included)
- Supplementary Tables 1-6

SUPPLEMENTARY DATA

Identification of A-to-I (G) RNA editing in *AZIN1* by RNA-Seq

In order to detect abnormal RNA editing events during neoplastic transformation, we applied high-throughput RNA-Seq to investigate three pairs of HCC tumor tissues and their adjacent non-tumor (NT) tissues (**Supplementary Table 1**). The high-quality variants were identified using VarScan[42] (**Supplementary Fig. 1a**). A total of 568, 545 and 494 potential somatic single nucleotide variants (SNVs), including 94, 89 and 101 coding somatic SNVs (cSNVs), were identified in the three tumor samples (**Supplementary Tables 2** and **3**). The intersecting SNVs (**Supplementary Fig. 1b**) are distributed in several functional categories (non-synonymous, synonymous, UTR, intron, splicing sites, intergenic region and pseudo or non-coding RNA [ncRNA]) (**Supplementary Fig. 1c**). Approximately 60~80% of the intersecting SNVs can be found in the NCBI dbSNP (Single Nucleotide Polymorphism database) (Build 131) (**Supplementary Fig. 1b**). To facilitate the search for functional tumor-associated SNVs, we

focused on the intersecting cSNVs that were identified by all three of the different software packages utilized in the analysis (**Supplementary Tables 3** and **4**).

Validation analysis was performed for 10 of the intersecting cSNVs (all are –non-synonymous substitutions) within selected genes of interest, and the majority were confirmed. Two genes, *translationally-controlled tumor protein* (*TCTP*) and *albumin*, demonstrated confirmed somatic mutations, resulting in variant protein sequences (**Supplementary Fig. 1d**). Loss of heterozygosity (LOH) of two known tumor suppressor genes, *KN motif and ankyrin repeat domains 1* (*KANK1*) and *oxidative stress induced growth inhibitor 1* (*OSGIN1*), were also validated in tumor samples (**Supplementary Fig. 1e**). By genotyping validation of the DNA samples, a*ntizyme inhibitor 1* (*AZIN1*), demonstrated a high frequency of non-synonymous A-to-I transcript editing, leading to a serine (Ser) to glycine (Gly) amino acid substitution (**Supplementary Fig. 1f–h**).

Analysis of conservation of *AZIN1* editing and RNA secondary structure

If RNA editing of *AZIN1* is likely to be of biological significance, it may not be unique to *Homo sapiens* but conserved over a long evolutionary period. For this purpose, we checked if this editing event can be observed in *Mus musculus*. As a result, A-to-I (G) editing of *azin1* was detected in mouse (*Mus musculus*) liver tissue (**Supplementary Fig. 2a**), suggesting that the *AZIN1* editing site is evolutionarily conserved.

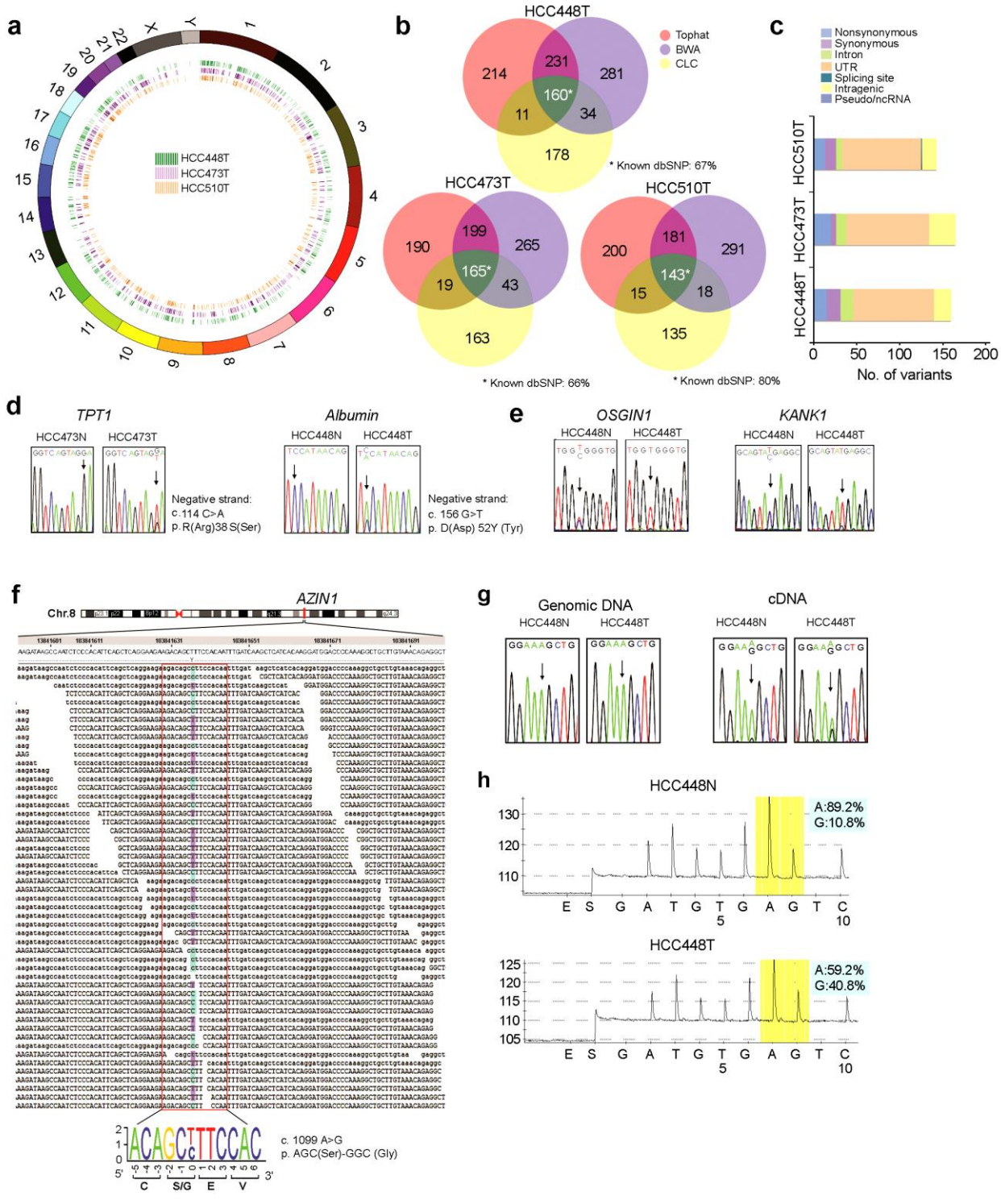
We hypothesized that the *AZIN1* transcript would undergo A-to-I editing by a mechanism similar to that occurring in *GluR-B* transcripts[8,9]. This mechanism involves an extensive RNA secondary structure configured from complementary edited exonic and downstream intronic

sequences. Because *AZIN1* editing is conserved between *Homo sapiens* and *Mus musculus*, the sequence encompassing RNA secondary structure should also be conserved. We compared *Homo sapiens* and *Mus musculus* genomic DNA sequences including the edited exon (exon 11) and downstream intron (intron 11). As shown, *AZIN1* coding sequences are highly conserved between the two species, approaching 100% identity at the nucleotide level (**Supplementary Fig. 2b**). In contrast, intron sequences are not highly conserved with only 48.3% identity. However, we observed an exception extending approximately 100bp in intron 11 immediately downstream of exon 11, particularly the first 16bp region of 100% identity (indicated by a box in **Supplementary Fig. 2b**). Intriguingly, a highly base-paired RNA secondary structure was predicted by the Centriod fold[44] program using the entire exon 11 and the first 16bp of intron 11 as input sequence (**Supplementary Fig. 2c**). This RNA secondary structure may juxtapose the *AZIN1* editing site and the potential exon complementary sequence (ECS) for A-to-I conversion.

SUPPLEMENTARY REFERENCES

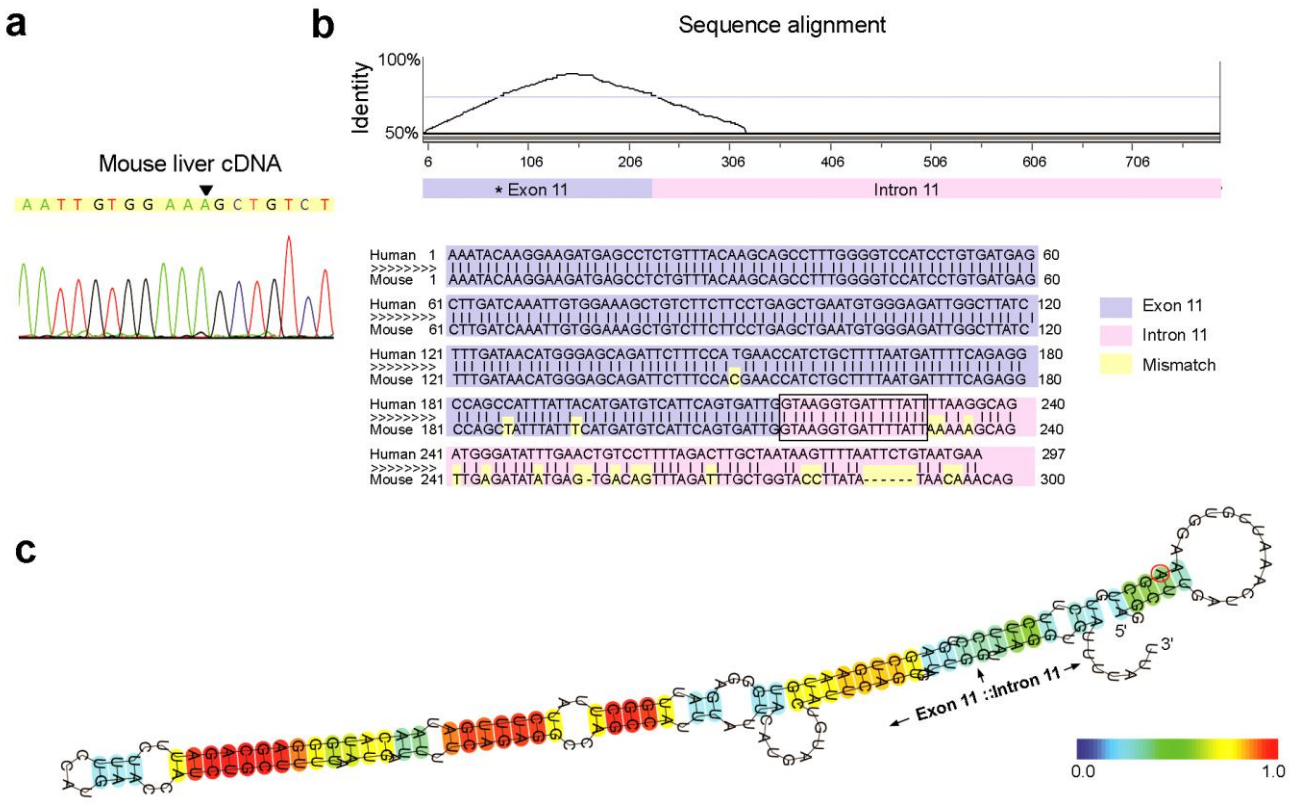
44. Hamada, M., Sato, K., Kiryu, H., Mituyama, T. & Asai, K. CentroidAlign: fast and accurate aligner for structured RNAs by maximizing expected sum-of-pairs score. *Bioinformatics* **25**, 3236-43 (2009).

SUPPLEMENTARY FIGURES AND LEGENDS



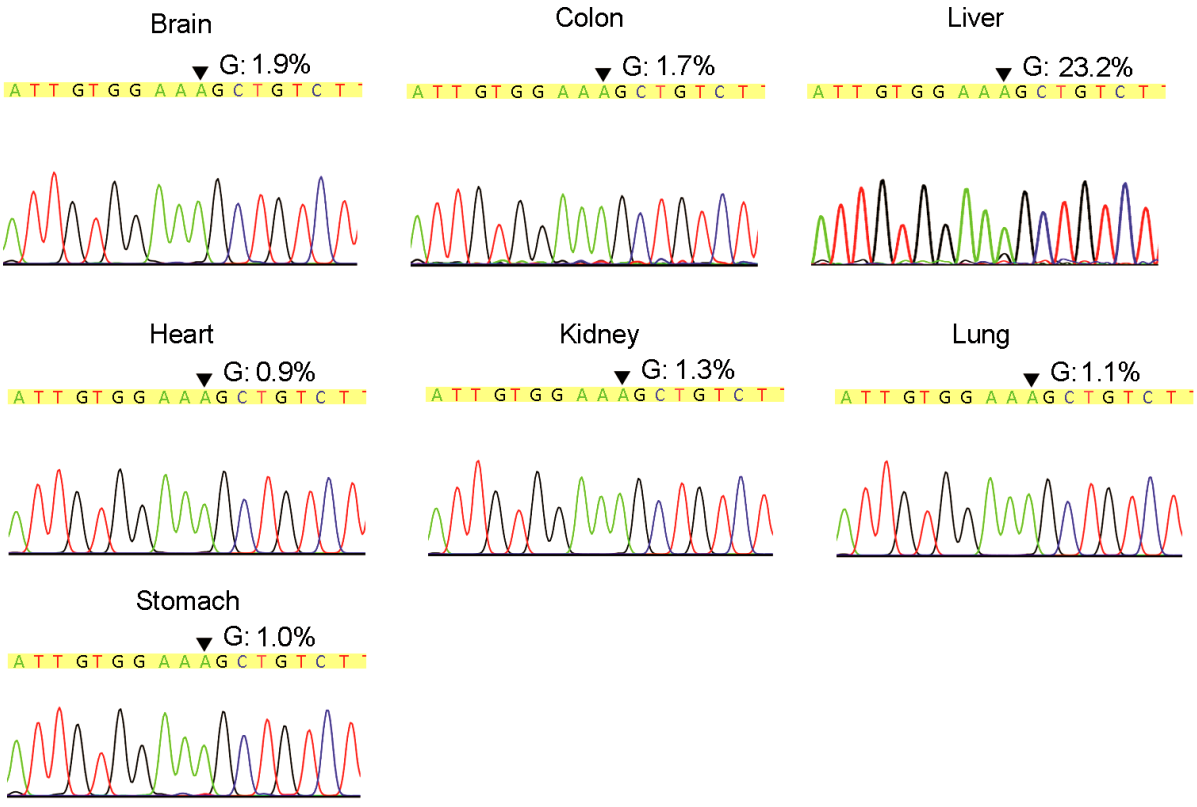
Supplementary Figure 1 Identification of *AZIN1* RNA editing by RNA sequencing (RNA-Seq).

(**a**) Distribution of potential somatic single nucleotide variations (SNVs) across all of the chromosomes in three tumor samples (HCC448T, HCC473T and HCC510T). In outer circle, green, purple and orange blocks correspond to HCC448T, HCC473T and HCC510T, respectively. (**b**) Venn diagram of potential SNVs in HCC tumor specimens based on TopHat, the Burrows-Wheeler Aligner (BWA) and the CLC Genomics Workbench (CLC). (**c**) Number of intersecting SNVs distributed in each functional category. (**d**) Chromatograms show somatic mutations in two genes, *TPT1* and *albumin*. (**e**) Chromatograms show the loss of heterozygosity (LOH) of two genes, *KANK1* and *OSGIN1*. (**f**) Sequences of individual reads were aligned to the published human genomic sequence of the *AZIN1* gene. Editing or reference events are highlighted by blue or purple shading, respectively. A sequence logo representation of the editing event in the tumor sample is shown below. The height of each letter is proportional to its frequency. (**g**) Matching *AZIN1* DNA and cDNA sequences in the HCC448T sample. The arrows indicate the editing site. (**h**) Quantitative pyrosequencing of A-to-I (G) editing frequencies for the HCC448N and HCC448T specimens.

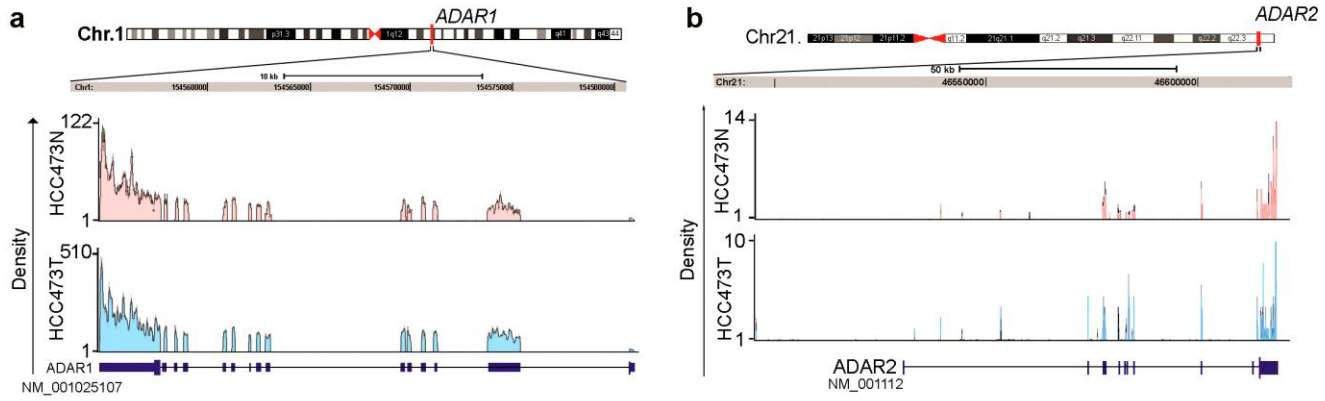


Supplementary Figure 2 Identification of a hypothetical RNA secondary structure within *AZIN1* pre-mRNA. (**a**) Sequence chromatogram of *AZIN1* transcripts in mouse liver tissue. The arrows indicate the editing site. (**b**) Sequence analysis of *AZIN1* genomic DNA region encompassing the edited exon (exon 11) and downstream intron (intron 11) of *Homo sapiens* and *Mus musculus*. The sequences were aligned and plotted as percent identity by VISTA (VISualization Tools for Alignments; Ernest Orlando Lawrence Berkeley National Laboratory, U.S.A.). The boxed region indicates 16-bp intronic sequences of 100% identity between the two species. (**c**) Secondary structure of the *AZIN1* transcript in the region of the RNA editing site predicted by Centriod Fold[44]. Exonic and intronic sequences were used to predict an RNA secondary structure for the edited site based on conserved sequences as described in (**b**). Each predicted base pair is indicated by a gradation of color (red to blue) corresponding to the base-pairing probability from 1 (red) to 0 (blue) according to Centroid Fold.

The 5’ and 3’ ends of the transcript are indicated. An arrow indicates the position of 5’ exonic donor site. The editing adenosine is indicated by red circle near the loop at the extreme right.

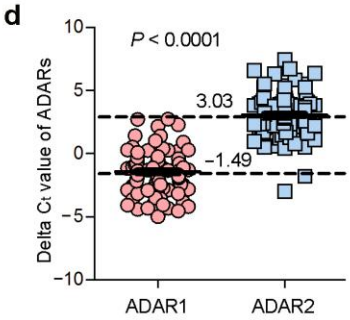


Supplementary Figure 3 *AZIN1* editing levels in different human organs. Direct sequencing of the *AZIN1* transcripts from the indicated human tissues. Percentage of edited (G) transcripts was detected by pyrosequencing. An arrow indicates the editing position.

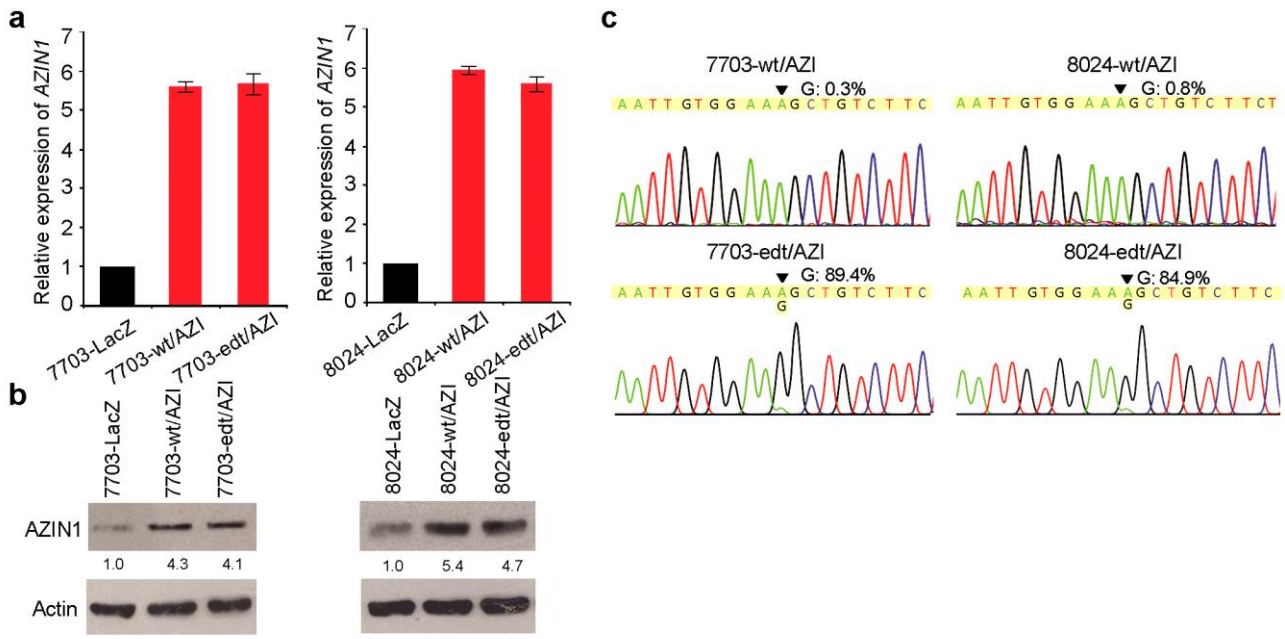


c RNA-Seq profiling of *ADARs*

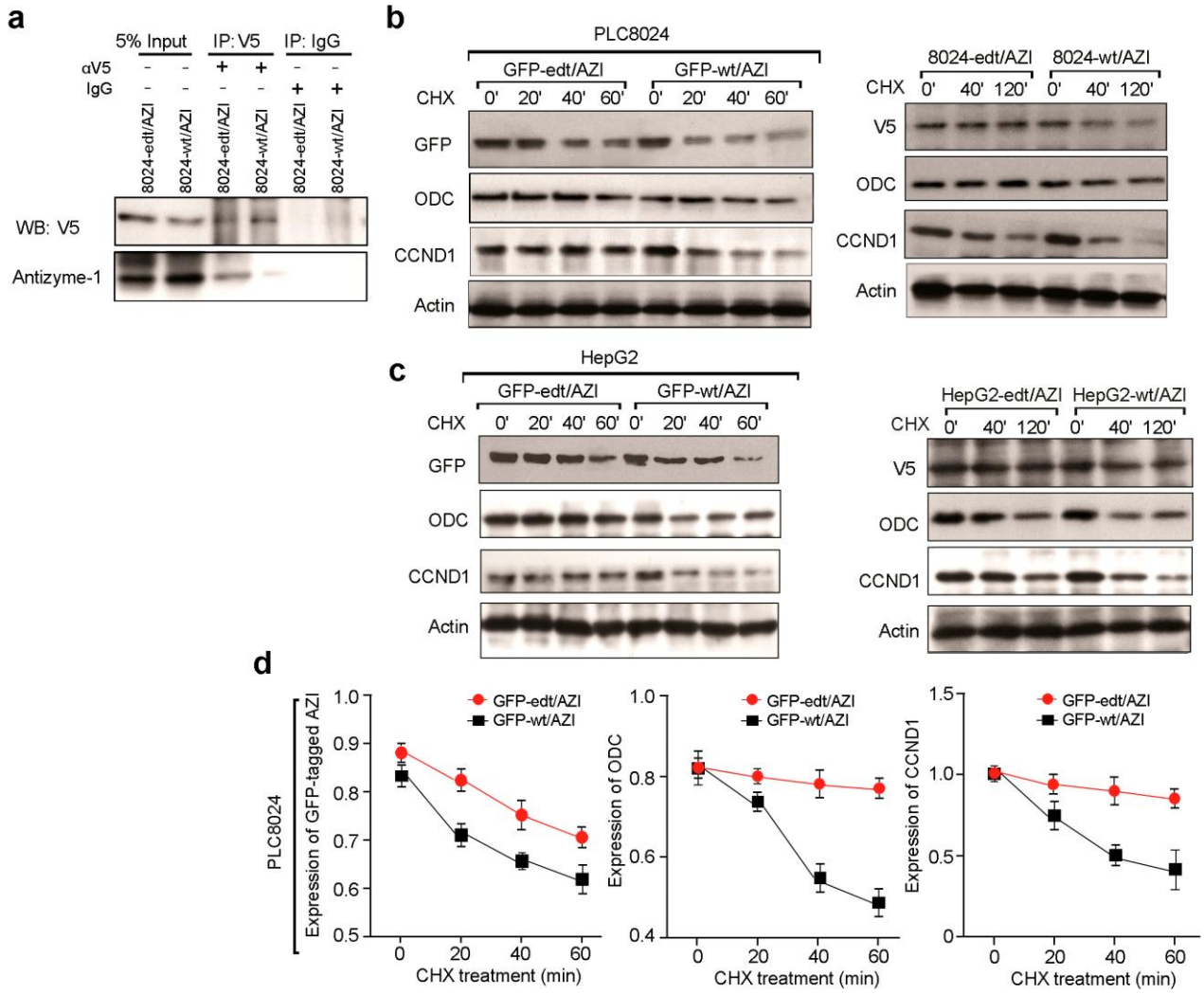
Case	Accession_No	Gene symbol	Locus	Protein size	Non-tumor (FPKM)	Tumor (FPKM)	Fold_change
HCC448	NM_001025107	ADAR1	chr1:154554535-154600437	932aa (p110)	8.68001	27.9852	3.224097668
	NM_015840	ADAR1	chr1:154554535-154600437	1201aa (p150)	3.01754	16.7396	5.547432677
	NM_001112	ADAR2	chr21:46494492-46646478	702aa	0.379977	0.304712	1.247003728
	NM_001160230	ADAR2	chr21:46494492-46646478	675aa	0.283501	0.294844	1.040010441
	NM_015833	ADAR2	chr21:46494492-46646478	742aa	0.354111	0.343997	1.029401419
	NM_015834	ADAR2	chr21:46494492-46646478	715aa	0.282164	0.301061	1.06697169
HCC473	NM_001025107	ADAR1	chr1:154554535-154600437	932aa (p110)	8.25445	46.0126	5.574278117
	NM_015840	ADAR1	chr1:154554535-154600437	1201aa (p150)	5.61722	19.3039	3.436557585
	NM_001112	ADAR2	chr21:46494492-46646478	702aa	0.356338	0.693934	1.947403869
	NM_001160230	ADAR2	chr21:46494492-46646478	675aa	0.231175	0.394604	1.706949281
	NM_015833	ADAR2	chr21:46494492-46646478	742aa	0.292772	0.59534	2.033459484
	NM_015834	ADAR2	chr21:46494492-46646478	715aa	0.171572	0.494779	2.883798056
HCC510	NM_001025107	ADAR1	chr1:154554535-154600437	932aa (p110)	19.4624	27.8095	1.428883385
	NM_015840	ADAR1	chr1:154554535-154600437	1201aa (p150)	10.3819	16.3241	1.572361514
	NM_001112	ADAR2	chr21:46494492-46646478	702aa	0.28323	0.400637	1.414528828
	NM_001160230	ADAR2	chr21:46494492-46646478	675aa	0.231851	0.32162	1.38718401
	NM_015833	ADAR2	chr21:46494492-46646478	742aa	0.288264	0.3744	1.298809425
	NM_015834	ADAR2	chr21:46494492-46646478	715aa	0.232148	0.349815	1.506862002



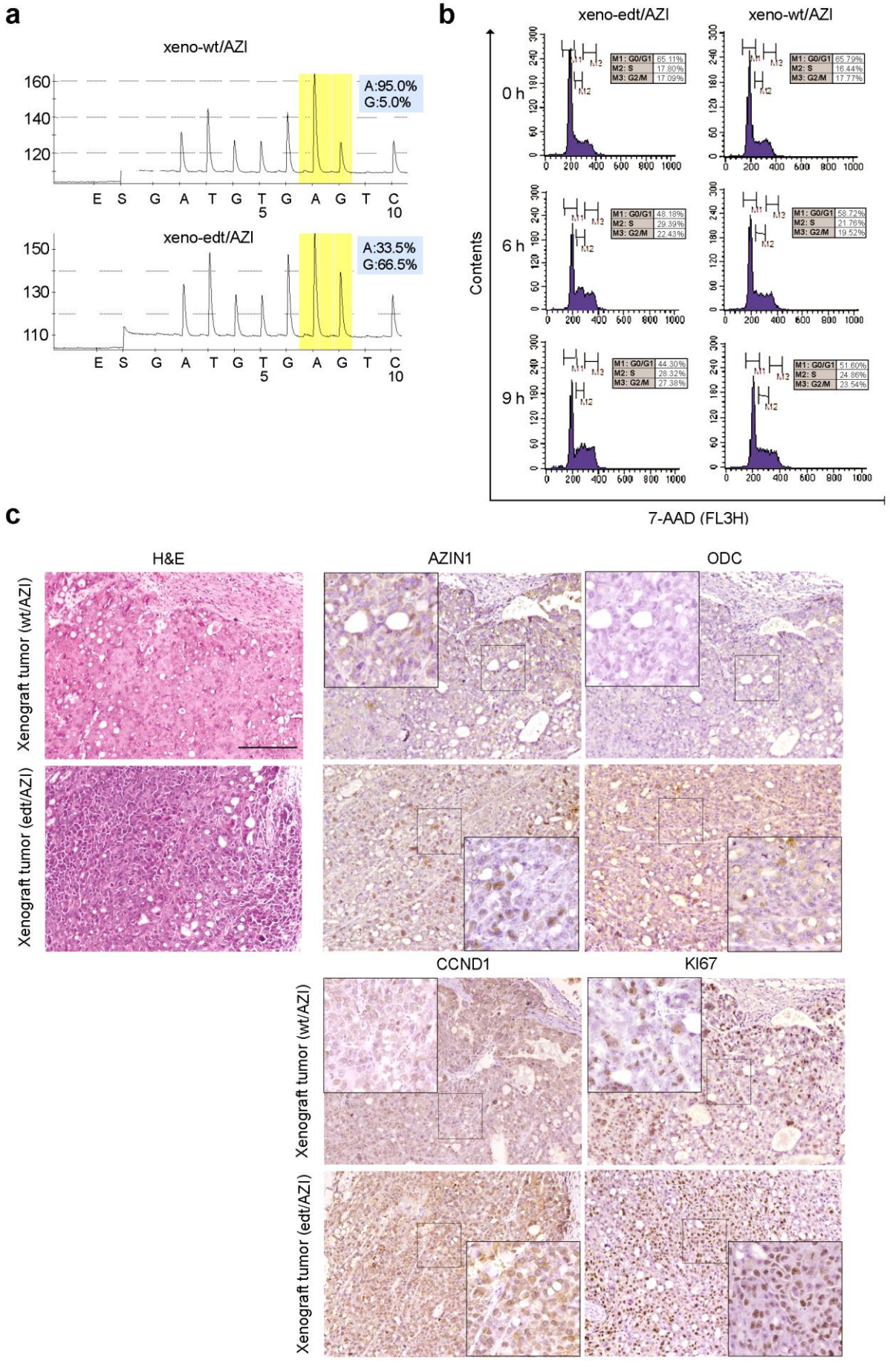
Supplementary Figure 4 Transcript abundances of *ADAR1* and *ADAR2* in human liver specimens. (**a, b**) RNA-Seq read mapping to the University of California, Santa Cruz (UCSC) reference genome (hg19) of the *ADAR1* (NM_001025107) (**a**) and *ADAR2* (NM_001112) (**b**) genes. The absolute read counts for each sample are indicated on the Y-axis. (**c**) RNA-Seq profiling of *ADARs* in clinical HCC samples. aa, amino acid. FPKM, fragments per kilobase of exon per million fragments mapped. (**d**) Dot plots represent the *ADAR1* and *ADAR2* delta Ct (ΔCt) values that were detected in tumors and the matched non-tumor (NT) specimens of 94 HCC patients from GZ cohort (Mann-Whitney U test). The dotted lines indicate the average ΔCt values of *ADAR1* and *ADAR2*.



Supplementary Figure 5 AZIN1expression and editing levels after the introduction of wild-type or edited *AZIN1* constructs into liver cell lines. (**a, b**) AZIN1 mRNA and protein levels were detected in the indicated cells by QPCR (**a**) and Western blot (**b**), respectively. The AZIN1 and β-actin protein levels were quantified using ImageJ. Data are presented as mean ± s.e.m of three independent experiments. (**c**) Direct sequencing of *AZIN1* transcripts from the indicated cell lines. Percentage of edited (G) transcripts was detected by pyrosequencing. An arrow indicates the editing position.



Supplementary Figure 6 Edited AZIN1 neutralizes the antizyme-mediated degradation of target oncoproteins. (**a**) Results of Co-IP assays of antizyme-1, which was immunoprecipitated by a V5-specific antibody (IP: V5) or mouse IgG (IP: IgG) in the 8024-edt/AZI and 8024-wt/AZI stable cell lines. (**b**–**d**) Western blot demonstrating expression of tagged AZIN1, ODC and CCND1 proteins in PLC8024 (**b**) and HepG2 (**c**) cells that were transfected with the GFP-edt/AZI or GFP-wt/AZI or in stable cell lines after cycloheximide (CHX) treatment (50 μg ml^{-1}) for the indicated times in minutes. Data are presented as mean ± s.e.m of three independent experiments (**d**).


a
xeno-wt/AZI
A:95.0%
G:5.0%
xeno-edt/AZI
A:33.5%
G:66.5%
E S G A T G T G A G T C
b
xeno-edt/AZI
xeno-wt/AZI
0 h
6 h
9 h
Contents
7-AAD (FL3H)
c
H&E
AZIN1
ODC
CCND1
KI67
Xenograft tumor (wt/AZI)
Xenograft tumor (edt/AZI)

Supplementary Figure 7 A functional e*dt*/AZI-ODC/CCND1-cell proliferation axis in tumor progression. (**a**) Quantitative pyrosequencing of A-to-I (G) editing frequencies of xeno-wt/AZI and xeno-edt/AZI cells isolated from xenograft tumors. (**b**) Flow cytometry histogram of DNA content (7-ADD staining) for each cell line following mimosine treatment (400 μmol l^{-1}) for 24 hours (0 h) and then released into complete medium containing BrdU (1μmol l^{-1}) for 6 (6 h) and 9 (9 h) hours. (**c**) Immunohistochemical (IHC) staining of AZIN1, ODC, CCND1, and KI67 on serial sections of xenograft tumors derived from 8024-wt/AZI and 8024-edt/AZI cells. The tissue sections were also stained with Hematoxylin and Eosin (H&E) as a reference. Insets show the boxed region at higher magnification. Scale bar, 200 μm.

Supplementary Table 1 Summary of sequencing statistics and mapping of reads for three paired HCC and their adjacent non-tumor samples

Sample ID	Raw Reads	Filter (rRNA+Adapters +Pipeline)[1]	% reads passing filters	Total mappable reads[2]	Uniquely Mapped Reads[3]	% of unique reads
HCC 473N	39,667,680	29,734,216	77.60%	26,781,730	22,207,580	82.92%
HCC 473T	37,944,360	29,189,562	80.10%	26,008,467	21,495,347	82.65%
HCC 510N	42,808,920	30,816,256	74.10%	27,534,858	22,452,003	81.54%
HCC 510T	37,717,200	29,299,026	81.40%	26,515,440	22,055,080	83.18%
HCC 448N	38,347,320	29,523,749	80.20%	26,546,578	22,321,772	84.09%
HCC 448T	35,981,160	28,824,374	82.00%	26,032,546	21,817,765	83.81%
Average	38,744,440	29,564,531	79.23%	26,569,937	22,058,258	83.03%

[1]Filtered reads are the reads from Illumina GAIIx pipeline with polymer track, adaptor, primer and ribosomal DNA complete repeating unit sequences removed.

[2]Sequencing reads were mapped to NCBI Build 37.1 (GRCh37) with at most 2 mismatches.

[3] % of unique reads are the percentages of unique reads over total mappable reads.

Supplementary Table 2 Summary of detected somatic single nucleotide variations (SNVs) in three HCC samples

Variant Alignment software*	# of variants called			# of variants found in dbSNP			% of variants found in dbSNP		
	448T	**473T**	**510T**	**448T**	**473T**	**510T**	**448T**	**473T**	**510T**
TopHat	616	573	539	261	241	291	42%	42%	54%
BWA	706	672	633	256	243	256	36%	36%	40%
CLC	381	390	311	212	203	196	56%	52%	63%
Average	568	545	494	243	229	248	45%	43%	52%

*High-quality reads were mapped to the human reference sequence (GRCh37/hg19) using three different software systems: TopHat, Burrows-Wheeler Aligner (BWA) and CLC Genomics Workbench (CLC). High quality variants were identified using VarScan[10] with the following parameters: a threshold sequence coverage depth of 10, 30% variation frequency and base quality of more than 15.

Supplementary Table 3 Summary of detected somatic coding SNVs in three HCC samples

Variant Alignment software*	# of variants called			# of variants found in dbSNP			% of variants found in dbSNP		
	448T	**473T**	**510T**	**448T**	**473T**	**510T**	**448T**	**473T**	**510T**
TopHat	130	104	119	82	72	92	63%	69%	77%
BWA	87	78	104	57	53	75	66%	68%	72%
CLC	66	86	80	35	46	55	53%	53%	69%
Average	94	89	101	58	57	74	61%	63%	72%

* High-quality reads were mapped to the human reference sequence and high quality variants identified as described in Supplementary Table 2.