ORIGINAL ARTICLE

A disrupted RNA editing balance mediated by ADARs (Adenosine DeAminases that act on RNA) in human hepatocellular carcinoma

Tim Hon Man Chan,[1,2,3] Chi Ho Lin,[4] Lihua Qi,[1] Jing Fei,[1] Yan Li,[2,3,5] Kol Jia Yong,[1] Ming Liu,[2,3] Yangyang Song,[2,3] Raymond Kwok Kei Chow,[2] Vanessa Hui En Ng,[1] Yun-Fei Yuan,[6] Daniel G Tenen,[1,7] Xin-Yuan Guan,[2,3,5] Leilei Chen[1]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2012-304037).

[1]Cancer Science Institute of Singapore, National University of Singapore, Singapore
[2]Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China
[3]State Key Laboratory of Liver Research, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China
[4]Genome Research Centre, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China
[5]State Key Laboratory of Oncology in Southern China, Cancer Centre, Sun Yat-sen University Cancer Centre, Guangzhou, China
[6]Department of Hepatobiliary Oncology, Sun Yat-sen University Cancer Centre, Guangzhou, China
[7]Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts, USA

Correspondence to
Dr L Chen, Cancer Science Institute of Singapore, National University of Singapore, 14 Medical Drive, Singapore 117599, Singapore; csicl@nus.edu.sg

THMC and CHL contributed equally to this work.

Received 30 October 2012
Revised 23 May 2013
Accepted 23 May 2013
Published Online First 13 June 2013

Linked

▶ http://dx.doi.org/10.1136/gutjnl-2013-305334

To cite: Chan THM, Lin CH, Qi L, *et al*. *Gut* 2014;**63**:832–843.

ABSTRACT

Objective Hepatocellular carcinoma (HCC) is a heterogeneous tumour displaying a complex variety of genetic and epigenetic changes. In human cancers, aberrant post-transcriptional modifications, such as alternative splicing and RNA editing, may lead to tumour specific transcriptome diversity.

Design By utilising large scale transcriptome sequencing of three paired HCC clinical specimens and their adjacent non-tumour (NT) tissue counterparts at depth, we discovered an average of 20 007 inferred A to I (adenosine to inosine) RNA editing events in transcripts. The roles of the double stranded RNA specific ADAR (Adenosine DeAminase that act on RNA) family members (ADARs) and the altered gene specific editing patterns were investigated in clinical specimens, cell models and mice.

Results HCC displays a severely disrupted A to I RNA editing balance. *ADAR1* and *ADAR2* manipulate the A to I imbalance of HCC via their differential expression in HCC compared with NT liver tissues. Patients with *ADAR1* overexpression and *ADAR2* downregulation in tumours demonstrated an increased risk of liver cirrhosis and postoperative recurrence and had poor prognoses. Due to the differentially expressed *ADAR1* and *ADAR2* in tumours, the altered gene specific editing activities, which was reflected by the hyper-editing of *FLNB* (filamin B, β) and the hypo-editing of *COPA* (coatomer protein complex, subunit α), are closely associated with HCC pathogenesis. In vitro and in vivo functional assays prove that *ADAR1* functions as an oncogene while *ADAR2* has tumour suppressive ability in HCC.

Conclusions These findings highlight the fact that the differentially expressed ADARs in tumours, which are responsible for an A to I editing imbalance, has great prognostic value and diagnostic potential for HCC.

Significance of this study

What is already known about this subject?

- RNA editing is a widespread post-transcriptional process contributing to greater cellular transcriptome diversity in eukaryotes.
- In humans, the most frequent type of editing is the conversion of A to I, which is catalysed by the dsRNA specific ADAR family of RNA editing enzymes.
- Until now, only a few recoding RNA editing events (eg, Q/R site editing in the glutamate receptor) have been verified, and no apparent causal relationship between altered RNA editing levels and cancer progression exists.
- Accumulating evidence has indicated that a hypo-editing (editing deficiency) phenotype is found in brain tumours and tumour tissues, such as prostate, lung, kidney and testis, and the hypo-editing phenotype is linked to several cancer phenotypes in paediatric astrocytomas and malignant gliomas. It has also been reported that a gene specific hyper-editing phenotype is found in metastatic lobular breast cancer and acute myeloid leukaemia.

What are the new findings?

- We provide the first extensive analysis of RNA editing in the human liver cancer transcriptome.
- HCC, distinct from most cancer types, is neither a hypo- nor a hyper-editing cancer; instead, HCC displays a severely disrupted A to I RNA editing balance induced by the differentially expressed ADARs (ADAR1 and ADAR2).
- Clinically, the differentially expressed ADARs, which are characterised by ADAR1 overexpression and ADAR2 downregulation in tumours, have great prognostic value and diagnostic potential for HCC.
- ADAR1 has oncogenic ability while ADAR2 functions as a tumour suppressor in HCC.

INTRODUCTION

RNA editing is an integral step in generating the diversity and plasticity of cellular RNA signatures. The best characterised type of RNA editing found in mammals converts C to U (cytosine to uracil) and A to I (adenosine to inosine). In humans, the most frequent type of editing is the conversion of A to I, which is catalysed by the double stranded RNA (dsRNA) specific ADAR (Adenosine DeAminase that act on RNA) family of proteins. Because the translation machinery reads inosine as guanosine (G), the ADARs may recode transcripts, which results in a proteome that is divergent from the genome[1–4] and thus modulates the protein sequence and function of several gene products.

Significance of this study

How might it impact on clinical practice in the foreseeable future?

- These findings suggest a widespread occurrence of transcript variation at the single nucleotide level in the human liver transcriptome and highlight a link between the RNA editome imbalance in the forms of defective and excess gene specific RNA editing activity to HCC pathogenesis. Monitoring expression levels of ADARs or the global activity of RNA editing represents a useful early biomarker than can be utilised to detect disorders in HCC even before clinical symptoms become apparent.

Editing of a specific adenosine is rarely 100% efficient; consequently, the ADARs can generate different protein isoforms in the same cell. Due to the diverse impact of RNA editing on gene expression and function, it is possible that the misregulation of RNA editing may play a role in tumorigenesis by either inactivating tumour suppressor or activating genes that promote tumour progression.

Hepatocellular carcinoma (HCC) is one of the most common types of cancer; it is the third most common cause of cancer related deaths worldwide and it is associated with a poor clinical outcome.[5] Furthermore, the incidence of HCC is continually increasing in the USA and Western Europe.[6] As with many other solid tumours, HCC development is believed to be a multistep process involving the accumulation of genetic and epigenetic alterations.[7 8] The recent advent of RNA sequencing (RNA-Seq) has provided a powerful tool that can be used to study changes in transcriptomes and genomes.[9–11] In this study, we used RNA-Seq to identify post-transcriptional editing events in three matched pairs of patient derived HCC clinical specimens and their adjacent non-tumour (NT) liver tissue counterparts. We identified an average of 20 007 inferred A to I RNA editing events in non-coding genes and introns, untranslated regions (UTRs) and coding regions of protein coding genes. Because the editing sites occur in coding regions and may result in amino acid substitutions affecting protein properties and interactions, we were particularly interested in the identification and characterisation of RNA editing events within coding regions; these events may be involved or responsible for HCC initiation and progression. Our recent study has reported that the hyper-editing of a gene called antizyme inhibitor 1 (*AZIN1*) predisposes to human HCC.[12] In this study, we performed an extensive analysis of RNA editing in the human liver cancer transcriptome and demonstrated that RNA editing differs between HCC and matched NT liver tissues as follows: (1) unlike most cancer types, HCC displays a disrupted A to I RNA editing balance that is characterised by gene specific hypo-editing and hyper-editing; (2) *ADAR1* and *ADAR2* but not *ADAR3* are responsible for the disrupted editing balance in HCC through their differential expression in HCC compared with NT liver tissues; (3) patients with *ADAR1* overexpression and *ADAR2* downregulation in tumours demonstrated an increased risk of liver cirrhosis and postoperative recurrence and had poor prognoses; (4) specific recoding events in two genes, *FLNB* (filamin B, β) and *COPA* (coatomer protein complex, subunit α) which display the altered editing patterns in tumours compared with normal tissues; and (5) ADAR1 has an oncogenic ability while ADAR2 functions as a tumour suppressor in HCC.

MATERIALS AND METHODS

Detailed methods are included in the online supplementary materials and methods.

Clinical samples

Guangzhou cohort

A total of 125 paired human HCC and adjacent NT tissues that were surgically removed, snap frozen in liquid nitrogen (for protein, RNA and DNA extraction) and embedded in a paraffin block (for tissue microarray construction) were obtained from the Sun Yat-Sen University Cancer Centre (Guangzhou, China), along with their associated clinicopathological summaries, between 2002 and 2007.

Shanghai cohort

A total of 46 paired human HCC and matched NT specimens were obtained from the hepatectomy specimen archives at the Eastern Hepatobiliary Surgery Hospital (Shanghai, China). None of these patients received preoperative chemotherapy or radiotherapy.

Healthy human liver tissues were obtained from donor livers that had not been used for transplantation; the tissues were provided by Dr Man K and Dr Lo CM (Department of Surgery, University of Hong Kong). All of the samples were immediately frozen in liquid nitrogen or fixed in 10% formalin for paraffin embedding. All of the patients provided written informed consents for the use of their clinical specimens for medical research. All of the samples used in this study were approved by the committees for ethics review for research involving human subjects at Sun Yat-Sen University, University of Hong Kong and National University of Singapore.

Cell lines

The SNU-423, 449, 475, 182, 387 and 398 cell lines were obtained from the American Type Culture Collection. All of these cells were maintained in RPMI medium (Gibco BRL, Grand Island, New York, USA) supplemented with 10% fetal bovine serum (Gibco BRL). All of the cell lines used in this study were regularly authenticated by morphological observation and tested for mycoplasma contamination (MycoAlert, Lonza Rockland, Rockland, Maine, USA). The cells were incubated at 37°C in a humidified incubator containing 5% CO_2.

Discovery of RNA editing sites

Briefly, aligned files were processed with SAMtools and subsequently VarScan (V.2.2)[13] for the detection of A to G (positive strand) or T to C (negative strand) substitutions. Variance calling was constrained to locations within gene regions containing at least 10× coverage, a variation frequency of greater than 10% and a base quality of more than 15. We first filtered the editing sites against the known sites in the NCBI dbSNP database (Build 135) to eliminate germline variants. Second, we eliminated single nucleotide variations (SNVs) for which more than two types of nucleotide sequences were found because these likely represent false positives. Finally, we excluded polymorphic sites with a variation frequency of 100% because they may have resulted from intrinsic mapping errors. After validation, the false positive rate of RNA editing site detection was approximately 40%.

Analysis of RNA editing

Direct sequencing was performed on PCR products, and editing was calculated with the Discovery Studio Gene (DSGene) 1.5 programme (Accelrys Inc, San Diego, California, USA). The reliability of this method was further verified by cloning of individual sequences. PCR products were subcloned into the T-easy vector (Promega), and approximately 50 individual plasmids were sequenced for each sample. The percentage of edited clones was determined and compared with the DSGene quantification. For each sample, 2–3 independent RT-PCR reactions were performed.

Statistical analyses

Unless otherwise indicated, the data are presented as the mean ±SD of three independent experiments. The SPSS statistical package for Windows (V.16; SPSS) was used to perform the data analyses. The clinicopathological features of patients with a different status for the differential expression of ADAR1 and ADAR2 were compared using a non-parametric crosstabs analysis (χ^2 test) for categorical variables. *ADAR1* or *ADAR2* expression levels in any two groups of clinical samples (eg, tumours and matched NT liver tissues) were compared using the Mann–Whitney U test. Kaplan–Meier plots and log rank tests were used for disease free survival (DFS) analysis. DFS times were calculated from the data of curative surgery to HCC recurrence, death or the last follow-up data. For the tissue microarray (TMA) analysis, which was based on immunohistochemical (IHC) scores, ADAR1 expression levels in the primary HCC tissues and their matched NT liver tissues were compared using a Wilcoxon signed rank test. A paired Student's t test was used to compare editing levels of *FLNB* or *COPA* in HCC and matched NT liver specimens of patients from the Guangzhou (GZ) and Shanghai (SH) cohorts. Editing levels of *FLNB* and *COPA* between two preselected groups were compared using the Mann–Whitney U test. An unpaired two tailed Student's t test was used to compare the number of foci, number of migrative and invasive cells, tumour volume and the relative expressions of target genes between any two preselected groups. A p value of <0.05 was considered to be statistically significant.

RESULTS

Global identification of potential A to I editing sites by RNA-Seq

The high throughput transcriptome sequencing (RNA-Seq) of three pairs of primary HCC and matched NT liver tissues (HCC448N/T, HCC473N/T and HCC510N/T) from HCC patients of Chinese origin (GZ cohort) generated 132.3 million reads that could be uniquely aligned to the human genome (hg19). The aligned reads provided substantial coverage (an average of 83.03%) for the vast majority of the identified mRNA transcripts.[12] The global identification of potential A to I (G) editing sites was called using VarScan,[13] with the following parameters: a minimum coverage depth of 10, a variation frequency of more than 10% and a base quality of more than 15 (figure 1A). To facilitate the detection of actual A to I (G) editing events, we first filtered the editing sites against known events in the NCBI dbSNP database (Build 135) to eliminate germline variants. Next, we eliminated SNVs for which more than two types of nucleotide sequence were found because they likely represented false positives. Finally, we excluded polymorphic sites with a variation frequency of 100% because they may have resulted from an intrinsic mapping error. The distribution of the remaining editing sites in each functional category (coding sequence, UTR, intron, splicing sites, intergenic region and pseudo/ncRNA) is shown in figure 1B and in online supplementary table S1. The majority (approximately 80%) of the inferred editing sites in each of the samples had a low editing level, which ranged from 0% to 20% (see online supplementary figure S1A). Functional enrichment analysis was performed on the genes edited in all three tumours or three NT liver specimens. In both NT liver and tumour samples, the edited genes were found to be most significantly enriched with acute inflammatory response, metabolic process, protein processing and response to wound (see online supplementary tables S2 and S3).

To experimentally validate our calls, we verified a subset of potential sites by performing Sanger sequencing of both DNA and RNA from the same samples that were utilised for RNA-Seq. We validated 30 inferred editing sites, and the majority (21/30, 70%) were verified (see online supplementary figure S1B). Notably, *UTP14C* mRNA contains previously undescribed A to G substitution at multiple sites (figure 1C). We also compared sites identified in this study with editing sites in the DARNED database[14] and a human B cell dataset.[15] An example of an editing target, *FLNB*, exhibited confirmed, extensive, non-synonymous editing at the same site as in the human B cell dataset. However, no transcript alteration was identified at the position described in the DARNED database (figure 1D). These results support the current notion that RNA editing frequency can be regulated in a tissue or cell type specific manner.[16]

HCC displays a disrupted editing balance

To better understand the link between the RNA editing process and hepatocarcinogenesis, we applied several adjustments and additional filters to facilitate the identification of editing sites within coding regions that may have highly tumour suppressive or oncogenic potentials. When the coverage depth was not less than 15, the editing sites in the coding regions detected in more than one NT liver but not in HCC tissues were assigned to the 'NT specific editing' category (figure 1E and see online supplementary table S4); conversely, those sites found in more than one tumour tissue but not in NT liver specimens were placed in the 'tumour specific editing' category (figure 1E and see online supplementary table S5). The editing sites that were found in more than one pair of HCC and matched NT liver tissues were placed in the 'common editing events' category (figure 1E and see online supplementary table S6). Interestingly, two recoding editing sites in *AZIN1*[12] and *FLNB*, which are in the 'common editing events' category, demonstrated higher A to I (G) editing frequencies in HCC448T and HCC473T than in HCC448N and HCC473N, respectively (figure 1F and online supplementary figure S2). As a member of the 'NT specific editing' category, an A to I (G) editing site at codon 164 (Ile→Val) of the *COPA* gene was completely absent in all three tumour samples (figure 1F and see online supplementary figure S2). Conversely, a 'tumour specific' editing site within the *UTP14C* coding region only exhibited an A to G substitution in the tumour samples (figure 1F and see online supplementary figure S2). These results suggest that HCC is neither a hypo-editing nor a hyper-editing cancer; instead, HCC displayed a severely disrupted A to I RNA editing balance.

In the past few years, bioinformatic and experimental studies have revealed that the A to I editing events occur in non-coding repetitive sequences, mostly *Alu* elements, and tend to undergo multi-editing in tight clusters.[17] To date, it is commonly accepted that a reduced A to I editing in general may be involved in the pathogenesis of cancer, and a significant global hypo-editing of *Alu* repetitive elements was observed in brain,

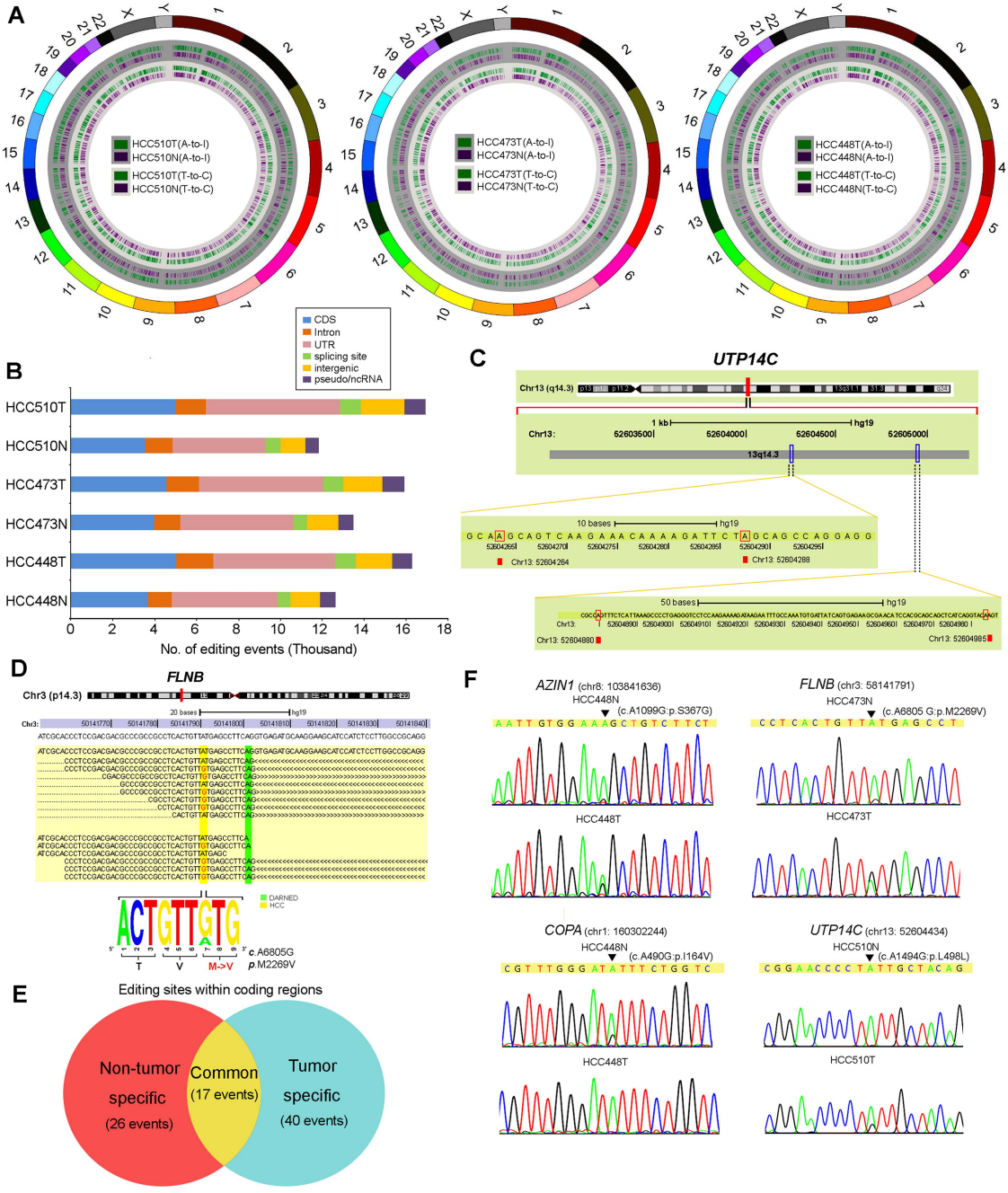


Figure 1 Global identification of potential A to I (G) editing sites by RNA sequencing (RNA-Seq). (A) Distribution of potential editing sites across all of the chromosomes (shown by exterior circle) in three paired hepatocellular carcinoma (HCC) and matched non-tumour (NT) liver samples (HCC448N/T, HCC473N/T and HCC510N/T). The green and purple blocks in the inner circles (deep gray) indicate the A to G substitutions in the tumour and matched NT liver samples, respectively. The green and purple blocks in the inner circle (light gray) indicate the T to C substitutions in the tumour and matched NT liver samples, respectively. UTR, untranslated region. (B) Number of editing sites distributed in each functional category. (C) One example of the *UTP14C* gene with multiple edits. The RNA editing sites identified from the RNA-Seq data are highlighted by the red boxes. (D) Sequences of individual reads were aligned to the published human genomic sequence of the *FLNB* (filamin B, β) gene. An A to G conversion was found in the HCC473T sample. The editing and reference events are highlighted by yellow shading. The green boxes denote the editing site reported by DARNED.[14] A sequence logo representation of the editing event in the tumour sample is shown below. The height of each letter is proportional to its frequency. (E) Venn diagram illustrating the numbers of exonic editing events which were classified into the three indicated categories. (F) Sequence chromatograms of the *AZIN1*, *FLNB*, *COPA* and *UTP14C* transcripts in the indicated tumour and matched NT liver samples. An arrow indicates the editing position. The sequence chromatograms of the matching genomic DNA (gDNA) sequences of each gene are shown in online supplementary figure S2.

prostate, lung, kidney and testis tumours.[17] In our study, the numbers of potential A to I (G) editing sites within the *Alu* sequences in three tumours were higher than their corresponding NT liver samples (see online supplementary figure S3A). We validated 30 editing sites within *Alu* sequences and all were verified. The editing levels of two editing sites within the *Alu* repetitive element of the gene *TTPA* (tocopherol (α) transfer protein) were dramatically decreased in tumour samples (see

online supplementary figure S3B). In contrast, five editing sites in *Alu* sequences of two genes called *MAGT1* (magnesium transporter 1) and *PAICS* (phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase) were highly edited in tumour samples (see online supplementary figure S3B). All of these data suggest that there is a disrupted A to I editing balance in coding regions and non-coding *Alu* repetitive elements in human HCC.

Differentially expressed RNA editing enzyme ADAR1 and ADAR2 in HCC

A to I RNA editing is a post-transcriptional modification in stem loop structures within precursor mRNA, which is catalysed by dsRNA specific ADAR enzymes.[18] In humans, the ADAR family is composed of three independent genes, ADAR1–3. ADAR1 and ADAR2 are expressed in many tissues whereas ADAR3 is specifically expressed in the brain.[3] As described previously, RNA-Seq profiling of the *ADARs* indicated that two *ADAR1* transcript variants (NM_001025107 and NM_015840) encoding 110 kDa (p110) and 150 kDa (p150) isoforms, respectively, demonstrated relatively high abundances in liver tissue.[12] However, *ADAR2* and *ADAR3* were expressed at extremely low levels and were undetectable in all samples.[12] In this study, we constructed a panel of TMAs consisting of 92 surgically resected primary HCCs and their matched NT liver tissues from the GZ cohort. By performing IHC staining, we observed the differential nuclear expression of ADAR1 between the primary HCC and matched NT liver tissues. A detailed analysis of the IHC data revealed the ADAR1 was overexpressed in 71.7% (66/92) of the analysed HCC tissues ($p<0.001$, Wilcoxon signed rank test) (figure 2A and see online supplementary table S7). However, it was found that the sensitivity of IHC was too low to detect ADAR2 expression in both the primary HCC and matched NT liver tissues. Therefore, we determined expression of ADAR2 and ADAR1

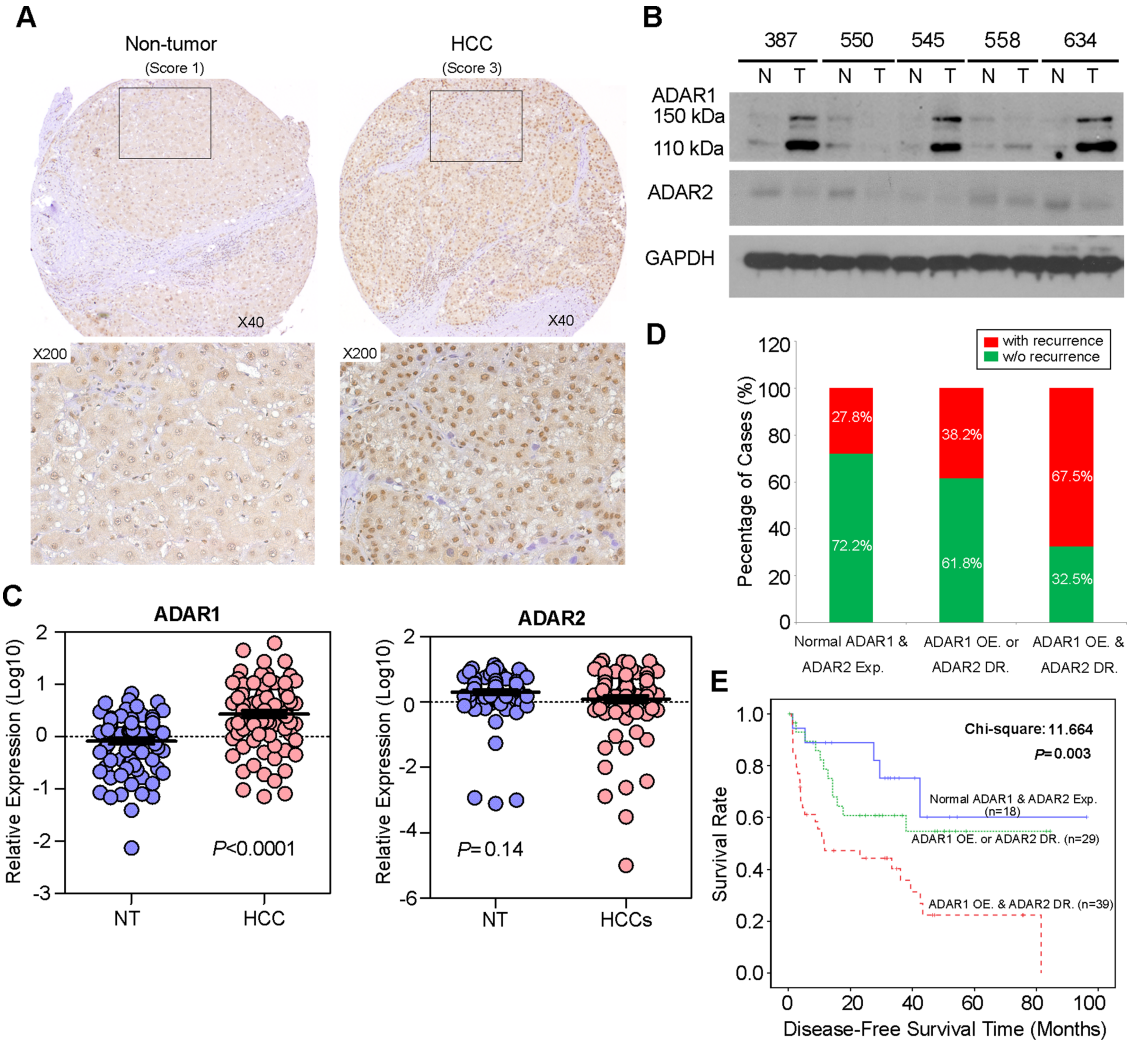


Figure 2 Differential expression of ADAR (Adenosine DeAminase that acts on RNA) 1 and ADAR2 in hepatocellular carcinoma (HCC) and its clinical implication. (A) Example of ADAR1 expression level detected in a primary HCC tumour and its matched non-tumour (NT) liver specimen. Based on staining intensities, ADAR1 immunoreactivities were scored as strong expression (3) and weak (1) expression in the primary HCC and matched NT liver specimens, respectively (see the online supplementary materials and methods section for details). The boxed regions are magnified and displayed in the lower panels. (B) Western blot analyses of ADAR1 and ADAR2 expression levels in five paired HCC and matched NT liver specimens. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a loading control. (C) Dot plots represent relative *ADAR1* (left) and *ADAR2* (right) expression levels in HCC and corresponding NT liver tissue samples, as detected by quantitative real time PCR (mean±SD, n=92; Mann–Whitney U test). (D) Association between the postoperative recurrence rate and status of expression of ADARs (p=0.004, χ^2 test). (E) Kaplan–Meier plots for disease free survival rate of patients demonstrating ADAR1 overexpression (OE) and ADAR2 downregulation (DR) (red line; n=39), ADAR1 OE or ADAR2 DR (green line; n=44) and normal ADAR1 and ADAR2 expression in tumours (blue line; n=23) (log rank test).

using a western blot analysis and found that ADAR2 expression in approximately 50% (15/30) of the HCC samples was lower than in the matched NT liver specimens (figure 2B). Consistently, ADAR1 protein expression (both p110 and p150 isoforms) in approximately 73% (22/30) of the HCC specimens was higher than in the matched NT liver specimens (figure 2B). In order to obtain the expression levels of both ADAR1 and ADAR2 in all 92 paired HCCs and their matched NT liver tissues, we subsequently used quantitative real time PCR to examine *ADAR1* and *ADAR2* expression in 92 paired specimens that were utilised for TMA construction. As a result, *ADAR1* was significantly overexpressed in the HCC specimens compared with the NT liver specimens ($p<0.0001$, Mann-Whitney U test) (figure 2C). However, *ADAR2* expression was obviously decreased in approximately 47% (43/92) of the tumour samples compared with the NT samples ($p=0.14$, Mann-Whitney U test) (figure 2C).

Based on the quantitative real time PCR analysis of ADAR1and ADAR2 expression, patients with ADAR1 overexpression (defined as a twofold increase in ADAR1 expression in tumours) and ADAR2 downregulation (defined as a twofold decrease in ADAR2 expression in tumours), demonstrated higher incidences of tumour recurrence ($p=0.004$) and liver cirrhosis ($p=0.016$) and shorter DFS times ($p = 0.003$) than patients who had ADAR1 overexpression or ADAR2 downregulation and patients who had neither ADAR1 overexpression nor ADAR2 downregulation (figure 2D, E and table 1). In the univariate Cox analyses, the statistically significant predictors for a patient's DFS were liver cirrhosis, American Joint Committee on Cancer tumour staging and the differentially expressed ADARs in tumours (see online supplementary table S8). In the multivariate Cox analyses, differentially expressed ADARs in tumours were shown to be an independent prognostic factor for DFS ($p=0.025$, HR 1.725; 95% CI 1.071 to 2.777; see online supplementary table S8). We conclude that the differentially expressed ADAR1 and ADAR2 in HCC, as shown by ADAR1 overexpression and ADAR2 downregulation in tumours, predicts a poor prognosis for HCC patients.

Differentially expressed ADARs contribute to the altered gene specific editing patterns in HCC

Due to the differentially expressed ADARs, the A to I editing balance could be disrupted in HCC. In this study, we were

Table 1 Clinicopathological analyses of the differentially expressed ADARs in the Guangzhou cohort of 92 primary hepatocellular carcinoma patients

Clinical feature	No	Normal ADAR1 and ADAR2 expression (n=18)†	ADAR1 OE or ADAR2 DR (n=34)‡	ADAR1 OE and ADAR2 DR (n=40)§	p Value
Gender (n (%))					
Female	19	5 (27.8)	3 (8.9)	11 (27.5)	
Male	73	13 (72.2)	31 (91.2)	29 (72.5)	0.073
Age (years) (n (%))					
≦60	74	12 (66.7)	29 (85.3)	33 (82.5)	
>60	18	6 (33.3)	5 (14.7)	7 (17.5)	0.305
HbsAg¶ (n (%))					
Negative	18	6 (33.3)	6 (19.4)	6 (15.4)	
Positive	70	12 (66.7)	25 (80.6)	33 (84.6)	0.290
Serum AFP (ng/mL)¶ (n (%))					
≤400	52	11 (64.7)	22 (71.0)	19 (48.7)	
>400	35	6 (35.3)	9 (29.0)	20 (51.3)	0.152
Tumour size (cm)¶ (n (%))					
≦5	23	4 (22.2)	6 (19.4)	13 (32.5)	
>5	66	14 (77.8)	25 (80.6)	27 (67.5)	0.421
Cirrhosis¶ (n (%))					
Absent	26	10 (55.6)	9 (30.0)	7 (17.9)	
Present	61	8 (44.4)	21 (70.0)	32 (82.1)	0.016*
Cell differentiation¶ (n (%))					
Well differentiated (I–II)	17	4 (25.0)	4 (13.8)	9 (24.3)	
Moderately differentiated (III)	48	7 (43.8)	19 (65.5)	22 (59.5)	
Poorly differentiated (IV)	17	5 (31.3)	6 (20.7)	6 (16.2)	0.540
Recurrence or metastasis (n (%))					
Absent	47	13 (72.2)	21 (61.8)	13 (32.5)	
Present	45	5 (27.8)	13 (38.2)	27 (67.5)	0.004*
Tumour stage (AJCC)¶ (n (%))					
Stage I	60	14 (82.4)	21 (70.0)	25 (62.5)	
Stage II	20	3 (17.6)	8 (26.7)	9 (22.5)	
Stage III	7	0 (0)	1 (3.3)	6 (15.0)	0.224
Mean DFS (months)		68.6 (48.3–89.0)	52.6 (39.0–66.3)	30.1 (19.3–40.9)	0.003*

*Statistical significance ($p<0.05$).
†Cases fitting neither of these two criteria are regarded as 'Normal ADAR1 and ADAR2 expression': (1) a twofold increase in ADAR1 expression level in tumours; (2) a twofold decrease in ADAR2 expression level in tumours compared with their matched non-tumour liver samples.
‡Cases fitting either one of the two criteria listed above are defined as 'ADAR1 OE or ADAR2 DR' (OE, overexpression; DR, downregulation).
§Cases fitting both criteria, as listed above, are classified as 'ADAR1 OE and ADAR2 DR' (OE, overexpression; DR, downregulation).
¶Partial data are not available and statistics were based on available data.
ADARs, Adenosine DeAminases that act on RNA; AFP, α fetoprotein; AJCC, American Joint Committee on Cancer; DFS, disease free survival time; HbsAg, hepatitis B surface antigen.

particularly interested in two exemplary editing targets, *FLNB* and *COPA*. To clarify which *ADAR*s are responsible for *FLNB* and *COPA* editing, the tumour samples were divided into 'high level' and 'low level' groups based on the average relative quantification (RQ) values of *ADAR1* or *ADAR2* in all 92 tumour specimens (Avg[$_{ADAR1}$]: 5.84; Avg[$_{ADAR2}$]: 4.98). Tumours with 'high level' *ADAR1* expression (RQ≥5.84) demonstrated a significantly higher editing level of *FLNB* (p=0.003) but not *COPA* (p=0.81; Mann–Whitney U test) (figure 3A). Tumours with 'high level' *ADAR2* expression (RQ≥4.98) were found to have higher editing degrees of both *COPA* (p=0.013) and *FLNB* (p=0.091; Mann–Whitney U test) (figure 3B). The

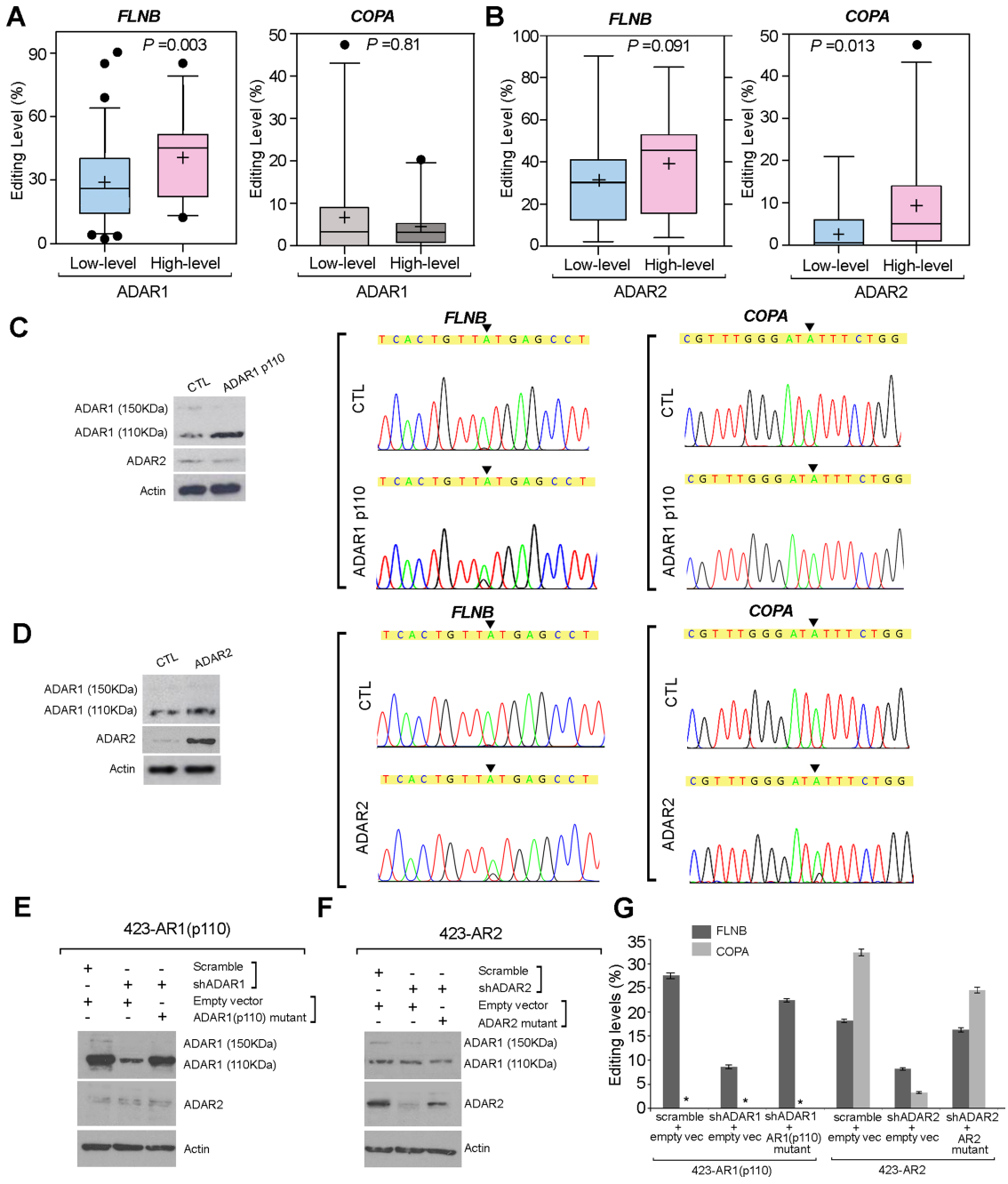


Figure 3 *FLNB* (filamin B, β) editing is catalysed by both ADAR (Adenosine DeAminase that acts on RNA) 1 and ADAR2, while *COPA* (coatomer protein complex, subunit α) editing is specifically catalysed by ADAR2. (A, B) *FLNB* and *COPA* editing level in tumours with 'high level' and 'low level' expression of *ADAR1* (A) or *ADAR2* (B). The data are presented as box plots with median (horizontal line), 25–75% (box) and 5–95% (error bar) percentiles for each group. The mean is indicated as '+' (Mann–Whitney U test) and the dots indicate the outliers. (C, D) Left: Western blot analyses of *ADAR1* and *ADAR2* expression in SNU-423 cells that were transiently transfected with an *ADAR1* p110 variant expression construct (ADAR1 p110) or empty vector (CTL) (C) or SNU-423 cells that were transiently transfected with an *ADAR2* expression construct (ADAR2) or empty vector (CTL) (D). (C, D) Right: sequence chromatograms of the *FLNB* and *COPA* transcripts in the indicated cell lines. The percentages of edited *FLNB* or *COPA* transcripts were detected as described in the Materials and Methods section. An arrow indicates the editing position. (E, F) Following transfection with the indicated constructs into SNU-423 cells that stably expressed the ADAR1 p110 isoform (E) or ADAR2 (F) (see the online supplementary materials and methods section for details), ADAR1 and ADAR2 expression levels were detected by western blot analysis. (G) Editing levels of *FLNB* and *COPA* in each cell line as indicated are shown in a bar chart (mean±SD of three independent experiments; *undetectable).

ADAR1 p150 isoform is presumably responsible for the A to I editing of viral RNAs produced by viruses[18 19] but not of the nuclear pre-mRNAs.[20] To directly determine whether *ADARs* regulate *FLNB* and *COPA* editing, either the *ADAR1* p110 isoform or *ADAR2* was overexpressed in the HCC cell line SNU-423. SNU-423 cells overexpressing ADAR1 p110 displayed enhanced *FLNB* editing whereas ADAR1 overexpression did not affect *COPA* editing (figure 3C). The introduction of *ADAR2* into SNU-423 cells resulted in increased *COPA* and *FLNB* editing levels (figure 3D). To further confirm our findings, we conducted knockdown/rescue experiments with SNU-423 cells that stably expressed ADAR1 p110 (423-AR1) or ADAR2 (423-AR2). Silencing of *ADAR1* expression using a small hairpin RNA (shRNA) against *ADAR1* (shADAR1) in 423-AR1 cells dramatically decreased the *FLNB* editing level from 27.4% to 8.2%; this editing was effectively rescued by overexpressing an *ADAR1* p110 mutant that preserves the native amino acid sequence but contains six point mutations within the *ADAR1* shRNA targeting site (figure 3E,G). Using the same strategy, *ADAR2* knockdown in 423-AR2 cell lines dramatically decreased the editing levels of both *FLNB* and *COPA*, and the editing was effectively abolished by reintroducing the *ADAR2* mutant into 423-AR2 cells (figure 3F, G). Together, these results likely suggest that the differential expression of *ADAR1* and *ADAR2* in tumours, which is tightly associated with the altered gene specific editing pattern, may explain the A to I editing imbalance in HCC.

To further investigate the relationship between RNA editing and HCC progression, we also examined the editing frequencies of two editing targets, *FLNB* and *COPA*, in healthy human liver tissues (n=8) and two different cohorts of primary HCC and NT liver samples from GZ (n=125) and SH (n=47) cohorts. Two individual cohorts of HCC and matched NT liver samples demonstrated dramatically higher *FLNB* editing degrees compared with the healthy liver specimens (figure 4A). *COPA* editing levels were remarkably lower in both the HCC and matched NT liver specimens from two cohorts than those in the healthy liver specimens (figure 4B). Moreover, two individual cohorts of HCC samples displayed significantly higher and lower editing levels of *FLNB* and *COPA*, respectively, compared with the matched NT liver tissues (*FLNB*: $p_{GZ}<0.0001$, $p_{SH}<0.0001$; *COPA*: $p_{GZ}<0.0001$, $p_{SH}<0.0001$; Mann–Whitney U test) (figure 4A,B). All of these data suggest that the altered gene specific editing activities are closely associated with HCC pathogenesis from normal to adjacent non-tumour to clinically verified HCC.

Because the average *FLNB* or *COPA* editing level was approximately 10% different between the NT liver and tumour specimens in the GZ cohort (*FLNB*: 18.9 vs 31.6%; *COPA*: 20.7 vs 7.1%), we set an increase of not less than 10% of the

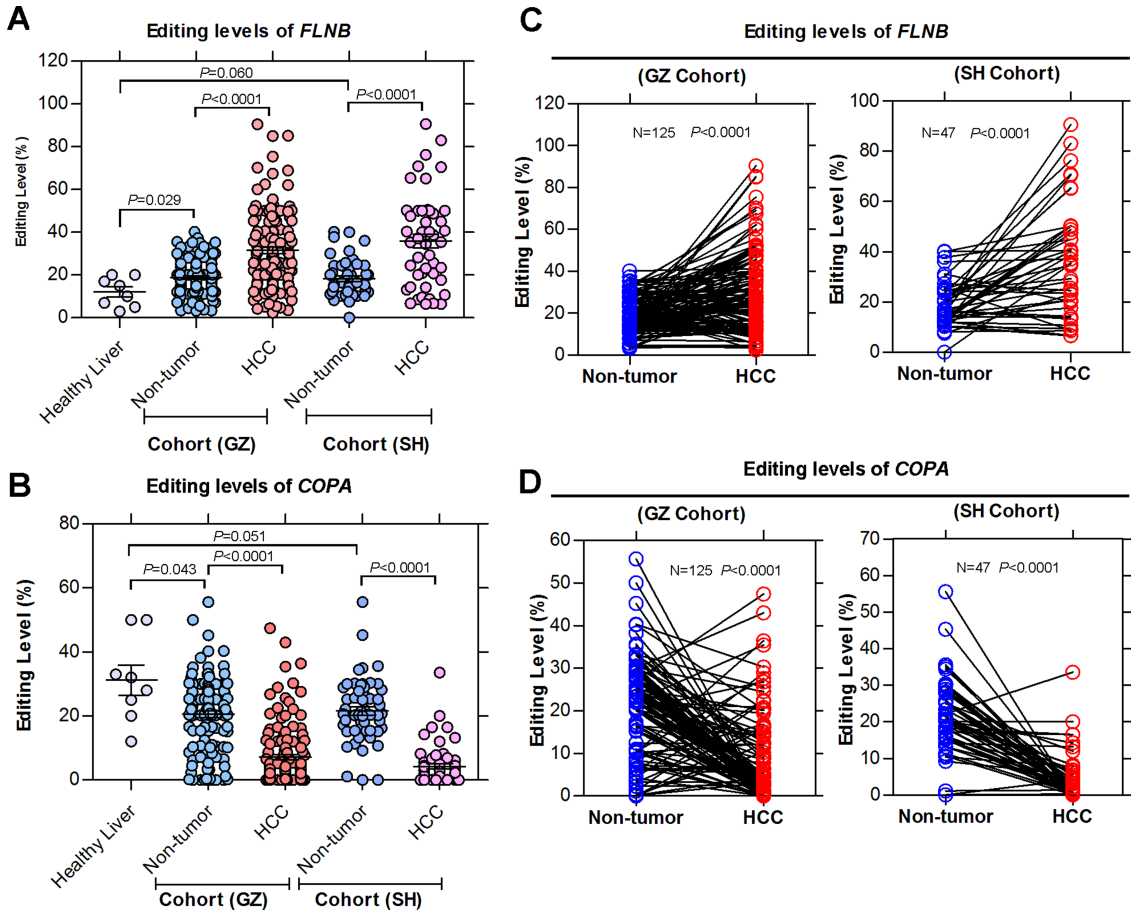


Figure 4 Altered gene specific editing patterns induced by the differentially expressed ADARs (Adenosine DeAminases that act on RNA) is closely associated with hepatocellular carcinoma (HCC) pathogenesis. (A, B) Dot plots represent *FLNB* (filamin B, β) (A) and *COPA* (coatomer protein complex, subunit α) (B) editing levels in healthy human liver tissues (mean±SD, n=8) and in 125 matched primary HCC and non-tumour (NT) liver samples in the Guangzhou (GZ) cohort and in 47 matched primary HCC and NT liver samples in the Shanghai (SH) cohort (Mann–Whitney U test). (C, D) *FLNB* (C) and *COPA* (D) editing levels in HCC and matched NT liver specimens from 125 and 47 patients in the GZ and SH cohorts (paired Student's t test).

FLNB editing level in tumour tissues when compared with NT liver tissues as the cut-off value for *FLNB* hyper-editing to subdivide HCC patients. Similarly, a decrease of not less than 10% of the *COPA* editing level in tumour tissues when compared with NT liver samples was used as the cut-off value for *COPA* hypo-editing. In the GZ cohort, approximately 52% (65/125) and 67% (84/125) of the HCC specimens displayed *FLNB* hyper-editing or *COPA* hypo-editing, respectively compared with the matched NT liver specimens (*FLNB*: $p<0.0001$; *COPA*: $p<0.0001$; paired Student's t test) (figure 4C,D). Consistently, *FLNB* hyper-editing and *COPA* hypo-editing were found in approximately 49% (23/47) and 74% (35/47) of the HCC cases in the SH cohort, respectively (*FLNB*: $p<0.0001$; *COPA*: $p<0.0001$; paired Student's t test) (figure 4C,D).

We conclude that as a consequence of the differentially expressed ADARs (ADAR1 and ADAR2), HCC displays a disrupted A to I editing balance, which was reflected by the altered gene specific editing patterns.

ADAR1 has oncogenic ability, while *ADAR2* functions as a tumour suppressor gene

As upstream regulators of A to I RNA editing, ADAR proteins have a number of reported target transcripts, such as serotonin receptor subunit 2C (*5-HT$_{2C}$R*), glutamate receptor (*GluRs*), filamin A (*FLNA*) and bladder cancer associated protein (*BLCAP*).[17 21 22] This prompted us to study the functional roles of ADAR1 and ADAR2 during hepatocarcinogenesis. For this purpose, we introduced *ADAR1* (p110) or *ADAR2* expression constructs into two HCC cell lines (SNU-423 and SNU-449) expressing the relative low endogenous ADAR1 (p110) and ADAR2 among six HCC cell lines using a lentiviral system (figure 5A,B). As detected by in vitro functional assays, cells transduced with the ADAR1 lentivirus (423-AR1 and 449-AR1) had accelerated growth rates and higher frequency of focus formation than cells transduced with the control LacZ lentivirus (423-LacZ and 449-LacZ) (figure 5C and see online supplementary figure S4A). However, introduction of ADAR2 into SNU-423 and SNU-449 cells (423-AR2 and 449-AR2) could effectively inhibit tumorigenic abilities, including significant inhibition of cell growth rate and reduction in frequency of focus formation (figure 5C and see online supplementary figure S4A). Consistent with the clinical correlation between tumour recurrence and the differentially expressed ADAR1 and ADAR2 in tumours, 449-AR1 and 449-AR2 cells demonstrated increased and decreased migrative and invasive capabilities, respectively, compared with 449-LacZ cells (figure 5D,E). As expected, this phenotype could also be observed in 423-AR1 and 423-AR2 cells compared with control cells (see online supplementary figure S4B,C). Xenograft studies in mice demonstrated that the growth rate of tumours induced by 449-AR1 or 449-AR2 cells was markedly higher or lower, respectively, than those induced by 449-LacZ cells (figure 5F,G). All of these data demonstrate that ADAR1 and ADAR2 have the opposite effects on tumorigenicity. ADAR1 has oncogenic ability while ADAR2 functions as a tumour suppressor gene, suggesting that there is a tight link between the unbalanced A to I editing mediated by the differentially expressed ADARs and HCC initiation and progression.

DISCUSSION

The data from our transcriptome measurement demonstrate differentially expressed gene profiling and reveal multiple types of SNVs. These variants are expected to consist of (in descending order of frequency) inherited polymorphisms, somatic mutations, sequencing errors and actual differences between RNA and DNA (eg, RNA editing and polyadenylation). RNA editing is broadly defined as the post-transcriptional process that alters the sequence of primary RNA transcripts. Of the various types of RNA editing, A to I (G) modification is most widespread in higher eukaryotes.[18 23 24] The most frequent RNA editing mechanism in mammals involves the conversion of specific adenosines into inosines by the ADAR family of enzymes. The majority of A to I substitutions occur not within coding portions of mRNA but largely in non-coding RNAs.[25] Early RNA editing studies have revealed that the editing occurs in many tissues and organs. In humans, this process is thought to occur predominantly in the brain and may be a key regulator of neural development. Recent studies have demonstrated that altered RNA editing is associated with numerous human pathologies, particularly cancers. Accumulating evidence has indicated that a hypo-editing phenotype is found in brain tumours and tumour tissues, such as prostate, lung, kidney and testis; additionally, the hypo-editing phenotype is linked to several cancer phenotypes in paediatric astrocytomas and malignant gliomas.[17 26 27]

Recent studies have revealed that the developmental and cell type specific modulation of A to I RNA editing is linked to ADAR expression and localisation.[19] It has been reported that all of the three editing enzymes, ADAR1, ADAR2 and ADAR3, were found to be downregulated in brain tumours. Consistently, overexpression of ADAR1 and ADAR2 in the U87 glioblastoma multiforme cell line resulted in a decreased proliferation rate, suggesting that the reduced A to I editing in brain tumours is involved in the pathogenesis of cancer.[17] In contrast, ADAR1 and/or ADAR2 were found to be upregulated in tumour tissues, such as prostate cancer and breast cancer tissues.[28] Similar to many other solid tumours, HCC development is believed to be a multistep process involving the accumulation of genetic and epigenetic alterations.[7 8] However, the role of RNA editing in HCC progression remains unknown. In this study, we identified an average of 20 007 A to I RNA editing events in transcripts. Intriguingly, unlike most types of cancers that are associated with a general decrease or increase in RNA editing activity, HCC is neither a hypo- nor hyper-editing cancer and displays a disrupted A to I editing balance in coding regions and non-coding *Alu* repetitive elements in human HCC. Moreover, the connection between the differential expression of ADARs and an altered gene specific editing pattern was investigated for the first time to illustrate how the A to I RNA editing balance was deregulated in HCC. Based on RNA-Seq transcript quantitation, the highest transcript abundance of *ADAR1* was found in liver tissue whereas *ADAR2* was expressed at extremely low levels, and *ADAR3* was undetectable in all samples. Most ADAR proteins localise to the nucleus, with the exception of the ADAR1 p150 isoform which is shuttled between the nucleus and cytoplasm and is thought to be responsible for the A to I editing of the viral RNA produced by viruses in the cytoplasm of infected cells.[18 20] In this study, we revealed that both of the ADAR1 p110 and p150 variants were overexpressed in approximately 70% of the primary HCC samples whereas ADAR2 was downregulated in approximately 50% of HCC cases. Clinically, the differentially expressed ADAR1 and ADAR2 in HCC, as shown by ADAR1 overexpression and ADAR2 downregulation in tumours, predicts a poor prognosis for HCC patients.

In addition, A to I editing can be very specific, leading to deamination of select adenosine residues, or it can be almost random and lead to non-selective conversion of many inosines. For long dsRNA (>100 bp) within 3′UTR regions, many

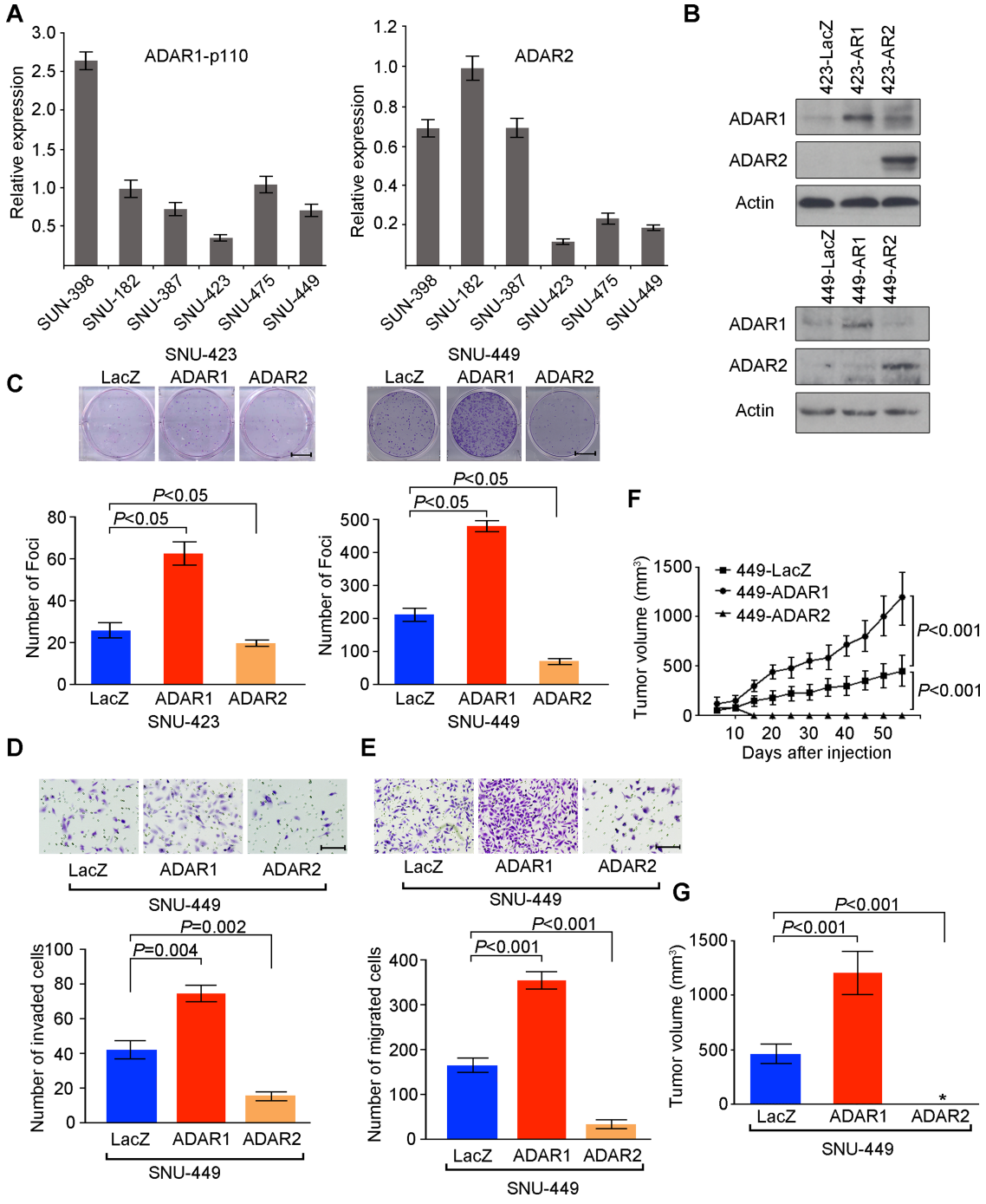


Figure 5 *ADAR1* (*Adenosine DeAminase that acts on RNA 1*) has oncogenic ability while *ADAR2* functions as a tumour suppressor gene. (A) Relative *ADAR1* p110 and *ADAR2* expression levels in six hepatocellular carcinoma cell lines, as detected by quantitative real time PCR (mean±SD of three independent experiments). (B) Western blotting showing expression of ADAR1 p110 and ADAR2 proteins in the indicated cell lines. β-Actin was the loading control. (C) Quantification of foci formation induced by the indicated stable cell lines. Triplicate independent experiments were performed and the data were expressed as the mean±SD of triplicate wells within the same experiment (unpaired two tailed Student's t test). Scale bar 1 cm. (D, E) Quantification of cells from the indicated cells that invaded through the Matrigel coated membrane (D) or migrated through the polyethylene terephthalate membrane (E) (unpaired two tailed Student's t test). Scale bar 200 μm. (F) Growth curves of tumours derived from the indicated cell lines over a period of 8 weeks. Data are presented as mean±SD (unpaired two tailed Student's t test). (G) Volumes of tumours derived from the indicated cell lines at the end point. Data are presented as mean±SD (unpaired two tailed Student's t test). *undetectable.

adenosine residues are edited promiscuously, leading to approximately 50% of adenosines being converted to inosines. However, in terms of A to I editing of protein coding sequences, it is highly selective, and an imperfect fold back dsRNA structure is formed between the exon sequence surrounding the editing site(s) and a downstream intronic complementary sequence termed editing site complementary sequence.[24] As described in our recent study, the *AZIN1* transcript undergoes A to I editing by a similar mechanism involving the fold back dsRNA structure configured from complementary edited exon 11 and the downstream 100 bp intronic sequences.[12] In addition, ADAR1 has a 5′ nearest neighbour preference of

A=U>C>G but no reported 3′ nearest neighbour preference.[29] On the other hand, ADAR2 has a 5′ nearest neighbour preference of A≈U>C=G and 3′ nearest neighbour preference of U=G>C=A.[30] In this study, two representative recoding editing events at codon 164 (Ile→Val) of the *COPA* gene and codon 2269 (Met→Val) of the *FLNB* gene were selected for further study. The overexpression and knockdown/rescue experiments demonstrated that *FLNB* editing was catalysed by both ADAR1 and ADAR2 and that *COPA* editing was specifically catalysed by ADAR2. Intriguingly, two individual cohorts of HCC samples displayed significantly higher and lower editing levels of *FLNB* and *COPA*, respectively, compared with matched NT liver tissues. Moreover, the altered gene specific editing activities were closely associated with HCC pathogenesis from normal to adjacent non-tumour to clinically verified HCC. To our knowledge, unlike the editing events that occur within 3′UTR regions, where the editing can affect transcript stability via affecting microRNA targeting or the nuclear retention of transcripts, those within coding regions will cause amino acid change and affect protein function rather than protein level. Similar to the *FLNB* transcript, *AZIN1* is one of the recoding editing targets that are placed in the 'common editing events' category. As a result of the A to I editing of *AZIN1* transcripts, the serine (S)→glycine (G) substitution at residue 367, located in β strand 15 (β15) and predicted to cause a conformational change, induced a cytoplasmic to nuclear translocation and conferred gain of function phenotype. Moreover, *AZIN1* editing was catalysed by ADAR1, and an 18% increase in *AZIN1* editing frequency was sufficient to promote tumorigenic phenotypes,[12] strongly suggesting there exists a causative relationship between the altered RNA editing activity and HCC progression. ADAR1 and ADAR2 proteins have numerous editing substrates; although it is tempting to investigate if the editing alterations in the specific genes analysed may be relevant to the malignant phenotype, studying the functional connection between the decreased/increased expression of ADARs and carcinogenesis is of extreme biological importance. Here, our functional studies have indicated that *ADAR1* has oncogenic ability while *ADAR2* functions as a tumour suppressor gene. Therefore, we propose a model in which the precise regulation of the expression levels of ADARs is essential for accurate editing, and the altered expression of ADARs could be at the origin of cell transformation.

Investigating the connection between RNA editing and cancer progression is only the initial step in this research. Recent efforts to identify RNA editing events in the human transcriptome using deep sequencing approaches have indicated that many of the identified RNA–DNA differences could be explained by errors in sequencing or mapping errors in the assignment of RNA-Seq reads to the reference transcriptome.[31–33] Moreover, most recoding sites may be modified only to levels of less than a few percent.[19] In this study, we reported two recoding editing events with high level modification rates. Specifically, *FLNB* editing could be detected in nearly all of the primary HCC samples. More importantly, as a result of the differentially expressed ADARs (ADAR1 and ADAR2) in tumours, alterations in the gene specific editing activities are closely associated with HCC pathogenesis. Therefore, we speculate that monitoring expression levels of ADARs or the global activity of RNA editing represents a useful early biomarker for the detection of disorders in HCC before clinical symptoms become apparent. Generation of in vivo models for gene specific editing deficiency or hyper-editing should better elucidate the physiological significance of particular editing events in the context of liver cancer.

Acknowledgements We wish to thank those patients who donated tumour tissues to our tissue bank.

Contributors THMC and CHL were involved in study conception and design, data acquisition, analysis and interpretation, manuscript drafting, and finalisation and submission. CHL performed illumina mRNA library preparation and all bioinformatics analysis of transcriptome sequencing. LQ contributed to data acquisition, analysis and interpretation. YL, KJY, ML, YS, JF, VHEN and RKKC provided technical expertise and support. Y-FY performed patient identification, enrolment and follow-up, and clinical database compilation and analysis. LC developed the study concept, and critically revised and finalised the manuscript with input from X-YG and DGT.

Funding This work was supported by the Singapore Translational Research (STaR) Award (NMRC/STaR/0001/2008), a grant from the Singapore Ministry of Education and National Research Foundation, the Hong Kong Research Grant Council Central Allocation (HKU5/CRF/08), the Hong Kong RGC Collaborative Research Grant (HKU 7/CRG09), the 'Hundred Talents Program' at Sun Yat-Sen University (85000-3171311) and the National Key Sci-Tech Special Project of Infectious Diseases (2012ZX10002-013).

Competing interests None.

Ethics approval The study was approved by the committees for ethics review for research involving human subjects at Sun Yat-Sen University, University of Hong Kong and National University of Singapore.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing The RNA-Seq data were deposited in the following repository: Repository/DataBank Accession: GEO; Accession ID: GSE33294. Databank URL: http://www.ncbi.nlm.nih.gov/geo/.

REFERENCES

1 Blow M, Futreal PA, Wooster R, *et al*. A survey of RNA editing in human brain. *Genome Res* 2004;14:2379–87.
2 Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* 2004;2:e391.
3 Maas S, Kawahara Y, Tamburro KM, *et al*. A-to-I RNA editing and human disease. *RNA Biol* 2006;3:1–9.
4 Li JB, Levanon EY, Yoon JK, *et al*. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* 2009;324:1210–13.
5 Parkin DM, Bray F, Ferlay J, *et al*. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
6 El-Serag HB. Hepatocellular carcinoma: recent trends in the United States. *Gastroenterology* 2004;127:S27–34.
7 Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 2002;31:339–46.
8 Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer* 2006;6:674–87.
9 Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet;* 2010;11:685–96.
10 Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 2008;92:255–64.
11 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57–63.
12 Chen L, Li Y, Lin CH, *et al*. Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nature Med* 2013;19:209–16.
13 Koboldt DC, Chen K, Wylie T, *et al*. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 2009;25:2283–5.
14 Kiran A, Baranov PV. DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics* 2010;26:1772–6.
15 Li M, Wang IX, Li Y, *et al*. Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 2011;333:53–8.
16 Gallo A, Galardi S. A-to-I RNA editing and cancer: from pathology to basic science. *RNA Biol* 2008;5:135–9.
17 Paz N, Levanon EY, Amariglio N, *et al*. Altered adenosine-to-inosine RNA editing in human cancer. *Genome Res* 2007;17:1586–95.
18 Bass BL. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 2002;71:817–46.
19 Farajollahi S, Maas S. Molecular diversity through RNA editing: a balancing act. *Trends Genet* 2010;26:221–30.
20 Samuel CE. Antiviral actions of interferons. *Clin Microbiol Rev* 2001;14:778–809.

21 Burns CM, Chu H, Rueter SM, *et al*. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* 1997;387:303–8.
22 Higuchi M, Single FN, Kohler M, *et al*. RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* 1993;75:1361–70.
23 Gott JM, Emeson RB. Functions and mechanisms of RNA editing. *Annu Rev Genet* 2000;34:499–531.
24 Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 2010;79:321–49.
25 Levanon EY, Eisenberg E, Yelin R, *et al*. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* 2004;22:1001–5.
26 Cenci C, Barzotti R, Galeano F, *et al*. Down-regulation of RNA editing in pediatric astrocytomas: ADAR2 editing activity inhibits cell migration and proliferation. *J Biol Chem* 2008;283:7251–60.
27 Maas S, Patt S, Schrey M, *et al*. Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. *Proc Natl Acad Sci USA* 2001;98:14687–92.
28 Shah SP, Morin RD, Khattra J, *et al*. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009;461:809–13.
29 Polson AG, Bass BL. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J* 1994;13:5701–11.
30 Lehmann KA, Bass BL. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* 2000;39:12875–84.
31 Kleinman CL, Majewski J. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* 2012;335:1302; author reply.
32 Lin W, Piskol R, Tan MH, *et al*. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* 2012;335:1302; author reply.
33 Pickrell JK, Gilad Y, Pritchard JK. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science* 2012;335:1302; author reply.

A Disrupted RNA Editing Balance Mediated by ADARs in Human Hepatocellular Carcinoma

Tim Hon Man Chan, Chi Ho Lin, Lihua Qi, Jing Fei, Yan Li, Kol Jia Yong, Ming Liu, Yangyang Song, Raymond Kwok Kei Chow, Vanessa Hui En Ng, Yuan-Fei Yuan, Daniel G. Tenen, Xin-Yuan Guan, Leilei Chen

Supplementary Figures 1-4, Tables 1-8

MATERIALS AND METHODS

RNA extraction and Illumina mRNA library preparation

Tumor and their adjacent non-tumorous (NT) tissues of 3 HCC patients (Case No. 448, 473 and 510) in GZ cohort were selected for RNA-Seq. All 3 patients are HBV-positive and HCV-negative. Total RNA was isolated using the mirVana^{TM} miRNA isolation kit (Ambion, Austin, TX, USA), and the total RNA was treated with the DNA-free kit (Ambion) for the removal of contaminated genomic DNA. PolyA^{+} RNA was purified using Dynabeads mRNA purification kit (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. Approximately 100 ng of mRNA was fragmented by incubation for 5 min at 94°C in 5 × Array Fragmentation Buffer (Ambion). Double stranded cDNA was synthesized using the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen) using random hexamers. The reaction was purified using a QiaQuick PCR column (Qiagen, Valencia, CA).

Double-stranded cDNA fragments were repaired using the DNA Terminator End Repair Kit (Lucigen, Middleton, WI) and purified using a QiaQuick PCR column. The Klenow 3' to 5' exo-polymerase (NEB, Ipswich, MA) was used to add a single 'A' base to the 3'end of blunt phosphorylated DNA fragments. Following purification, the Illumina PE Adapter (Illumina,San Diego, CA) was ligated to the end of the DNA fragments using the Quick Ligation™ Kit (NEB, Ipswich, MA). DNA fragments ranging from approximately 280 to 300 bp were excised from a 2% low-melting agarose gel. The fragments were enriched by 10 thermocycles using AccuPrime™ Pfx DNA Polymerase (Invitrogen). The PCR product

was run on a Novex 8% TBE polyacrylamide gel (Invitrogen) and stained with SYBR Gold (Invitrogen). Gel slice containing the 340- to 360-bp fragments were excised and purified using the QiaQuick Gel Extraction Kit (Qiagen). The concentration of the gel-purified DNA fragments was measured using a ND-1000 UV/Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Solexa sequencing and read mapping

Cluster generation and sequencing were conducted using the Standard Cluster Generation kit v4, and 36-Cycle Sequencing kit v3 on the Illumina Cluster Station and GAIIx following the manufacturer's instructions. cDNA libraries from 3 paired HCC tumors and the corresponding NT counterparts (HCC448N/T, HCC473N/T, and HCC510N/T) were sequenced with 58-base single-reads. Raw data from the GAIIx were analyzed using the Illumina Real Time Analysis (RTA) v1.6 software. A phi-X 174 control lane was included in each Solexa run for matrix, phasing, and error rate estimations as recommended by the manufacturer. The error rate of the Phi-X control error rate was < 0.28% for all of the sequencing runs.

Ribosomal RNA sequences were first removed from GA reads by aligning them to 28S (NCBI RefSeq accession NR_003287.2), 18S (NCBI RefSeq accession NR_003286.2), human ribosomal DNA complete repeating unit (HSU13369) and mitochondrial ribosomal RNA (Ensembl transcript ID ENST00000387347 and ENST00000389680) using Bowtie[1] with default parameters. The high-quality reads were then aligned against the human genome assembly (NCBI Build 37.1/hg19) using TopHat v1.0.14[2] with the RefSeq refGene annotation,

which was downloaded from the UCSC Genome Browser.[3] Finally, mapping results were processed with custom scripts and visualized on the UCSC Genome Browser as a custom track.

Plasmid construction

We amplified the full-length cDNA encoding ADAR1 p110 isoform (ADAR1 p110) by PCR with primers (F: 5'-CACCGAAAGAGGCAGGAACACCC-3'; R: 5'-CTATACTGGGCAGAGATAAAAGTTC-3'). The full-length cDNA encoding ADAR2 was amplified by PCR with primers (F: 5'-CACCATGGATATAGAAGATGAAGAAAAC-3'; R: 5'-TCAGGGCGTGAGTGAGAACTGG-3'). Subsequently, the purified *ADAR1 p110* or *ADAR2* PCR products were ligated to pLenti6/V5-TOPO®vector (*pLenti6-ADAR1 p110* or *pLenti6-ADAR2*) (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

Establishment of ADAR1 or ADAR2 overexpression cells using a lentiviral expression system

Either the *pLenti6-ADAR1 p110* or *pLenti6-ADAR2* expression construct or the empty pLenti6/V5 vector was transfected into the 293FT cell line. Virus-containing supernatants were collected for subsequent transduction into SNU-423 cells. At 48 hours after virus transduction, half of cells were collected for the relevant assays. The remaining cells were cultured in complete medium containing blasticidin (3 μg/mL; Sigma-Aldrich) for

establishing stably transduced cell lines.

ADAR1 and ADAR2 knockdown/rescue experiments

The *ADAR1* and *ADAR2*-specific shRNA expression vectors (pRS-shADAR1 or pRS-shADAR2) and scrambled non-effective shRNA cassette in the pGFP-V-RS plasmid (pRS-scramble) were purchased from OriGene Technologies, Inc (Rockville, MD). The sequences of the shRNAs directed against *ADAR1* or *ADAR2* were as follows: shADAR1: CCTGTGGAATCCAGTGACATTGTGCCTAC and shADAR2: ACTCAAGTATGACTTCCTCTCCGAGAGCG. The pRS-shADAR1 or pRS-shADAR2 construct was transfected into SUN-423 cells stably expressing ADAR1 p110 or ADAR2 (423-AR1 or 423-AR2), respectively, using Lipofectamine 2000 (Invitrogen). The pRS-scramble construct was transfected into cells as a negative control. To generate an *ADAR1 p110* or *ADAR2* mutant construct preserving the native amino acid sequence, we introduced six mutations into the shADAR1 or shADAR2 targeting sequence (29nt), respectively. PCR-directed mutagenesis was performed using an inner forward or reverse primer containing six nucleotide alterations (AR1-rescue-forward: 5'-GAGAACGGAGAAGGCACAATCCCAGTAGAGTCAAGCGATATT -3';

AR1-rescue-reverse: 5'-AATATCGCTTGAGTCTACTGGGATTGTGCCTTCTCCGTTCTC-3';

AR2-rescue-forward: 5'-TTGAACGAACTGCGCCCAGGACTGAAATACGATTTTCTGTCC-3';

AR2-rescue-reverse: 5'-GGACAGAAAATCGTATTTCAGTCCTGGGCGCAGTTCGTTCAA-3') with the corresponding external primers described above.

Twenty-four hours after transfection, the transfected cells were cultured for 3 days with 0.4 μg/mL puromycin (OriGene). Pooled populations of knockdown/rescue cells were subjected to *in vitro* experiments.

Complementary DNA (cDNA) synthesis and quantitative real-time PCR (QPCR)

Total RNA was isolated as described above. To quantify the *ADAR1 and ADAR2* expression levels in clinical samples, equal amounts of cDNA were synthesized using the Advantage RT-for-PCR kit (Clontech, Mountain View, CA) and used for QPCR analysis. QPCR was performed using the SYBR Green PCR master mix (Applied Biosystems) and the following primers:

qADAR1-F(5'-CCCTTCAGCCACATCCTTC-3'),qADAR1-R(5'GCCATCTGCTTTGCCACTT-3'), qADAR2-F(5'- CTGACACGCTCTTCAATGGTT-3') and qADAR2-R(5'-GGCGCAGTTCGTTCAAGAT-3'). 18S was amplified as an internal control using the following primers: q18S-F: 5'-CTCTTAGCTGAGTGTCCCGC-3'; q18S-R: 5'-CTGATCGTCTTCGAACCTCC-3'. PCR was performed using an ABI Prism 7900 System (Applied Biosystems), and data processing was performed using the ABI SDS v2.3 software (Applied Biosystems). For *ADAR1* and *ADAR2* expression in HCC and matched NT liver specimens, the relative target gene expression is indicated by $2^{-\Delta\Delta CT}$ ($\Delta\Delta C_T = \Delta C_T$

^{Tumor} - ΔC_T $^{Non\text{-}tumor}$) and normalized to the average relative expression level in all of the NT tissues, which was defined as 1.0.

Gene Ontology Analysis

DAVID[4] was used for functional enrichment analysis of the edited genes obtained from the RNA-Seq.

Focus Formation Assay

Briefly, 1×10^3 cells were seeded in a 6-well plate. After culture for 7 days, surviving colonies (>50 cells per colony) 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.

Cell migration assay

The transwell cell migration assay was performed using Bio-coat cell migration chambers (BD Biosciences) containing polyethylene terephthalate membranes (PET) of 8-μm pore size according to the manufacturer's instructions. Briefly, $5x10^4$ cells in FBS-free RPMI were added. RPMI supplemented with 20% FBS was added to the bottom chamber as a chemoattractant. After 24 hours, the number of cells that had migrated through the filter pores was counted in 10 fields under a 20× objective lens and imaged using SPOT imaging software (Nikon, Japan).

Matrigel Invasion Assay

We performed invasion assays using 24-well BioCoat Matrigel Invasion Chambers (BD Biosciences) according to the manufacturer's instructions. Briefly, 2×10^5 cells FBS-free RPMI were added to the top chamber, and 20% 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 in 10 fields under a 20× objective lens and imaged using SPOT imaging software (Nikon, Japan).

***In Vivo* Tumorigenicity Assay**

We subcutaneously injected approxiamtely 1×10^7 cells into the right flank of 4- to 5-week-old male severe combined immunodeficient (SCID) mice. We monitored tumor formation in the SCID mice over an 8-week period and calculated the tumor volume weekly by the formula V (volume) = 0.5 × L (length) × W (width) × W. All animal experiments were approved by and performed in accordance with the Institutional Animal Care and Use Committees of National University of Singapore.

Antibodies and Western blot analysis

Mouse anti-ADAR1 and anti-β-actin antibodies were purchased from Abcam (Cambridge,

MA). The mouse anti-ADAR2 and anti-GAPDH antibodies were purchased from Sigma-Aldrich (Missouri, USA) and Santa Cruz Biotechnology. Protein lysates were quantified and resolved on a SDS-PAGE gel, transferred onto a polyvinylidene difluoride (PVDF) 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).

Immunohistochemical staining (IHC)

The tissue microarray (TMA) blocks were sectioned (5 mM thick) for IHC staining. Briefly, sections were deparaffinized and rehydrated. The endogenous peroxidase activity was blocked with 3% hydrogen peroxide (H_2O_2) for 10min. For antigen retrieval, the slides were immersed in 10 mM citrate buffer (pH 6.0) and boiled for 15 min in a microwave oven. Non-specific binding was blocked with 5% normal goat serum for 10min. The slides were incubated in a 1:100 dilution of anti-ADAR1 (Abcam) at 4°C overnight in a humidified chamber. The slides were then sequentially incubated with biotinylated goat anti-mouse IgG (1:100 dilution, Santa Cruz) for 30 min at room temperature, streptavidin-peroxidase conjugate for 30 min at room temperature. Isotope-matched human IgG was used in each case as a negative control. Finally, the 3, 5-diaminobenzidine (DAB) Substrate Kit (Dako Ltd., Carpenteria, CA) was used for color development followed by Mayer's hematoxylin counterstaining. Based on staining intensities, the ADAR1 immunoreactivity was scored as negative (0) (total absence of staining), weak expression (1) (faint staining in <50%, or

moderate staining in <25% of tumor cells), moderate expression (2) (moderate staining in >=25% to <75%, or strong staining in <25% of tumor cells), and strong expression (3) (moderate staining in >=75%, or strong staining in >=25% of tumor cells).

REFERENCES

1. **Langmead B**, Trapnell C, Pop M, *et al*. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;**10**:R25.

2. **Trapnell C**, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;**25**:1105-11.

3. **Zweig AS**, Karolchik D, Kuhn RM, *et al*. UCSC genome browser tutorial. *Genomics*. 2008;**92**:75-84.

4. **Huang da W,** Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 2009;**4**:44-57.

FIGURE LEGENDS

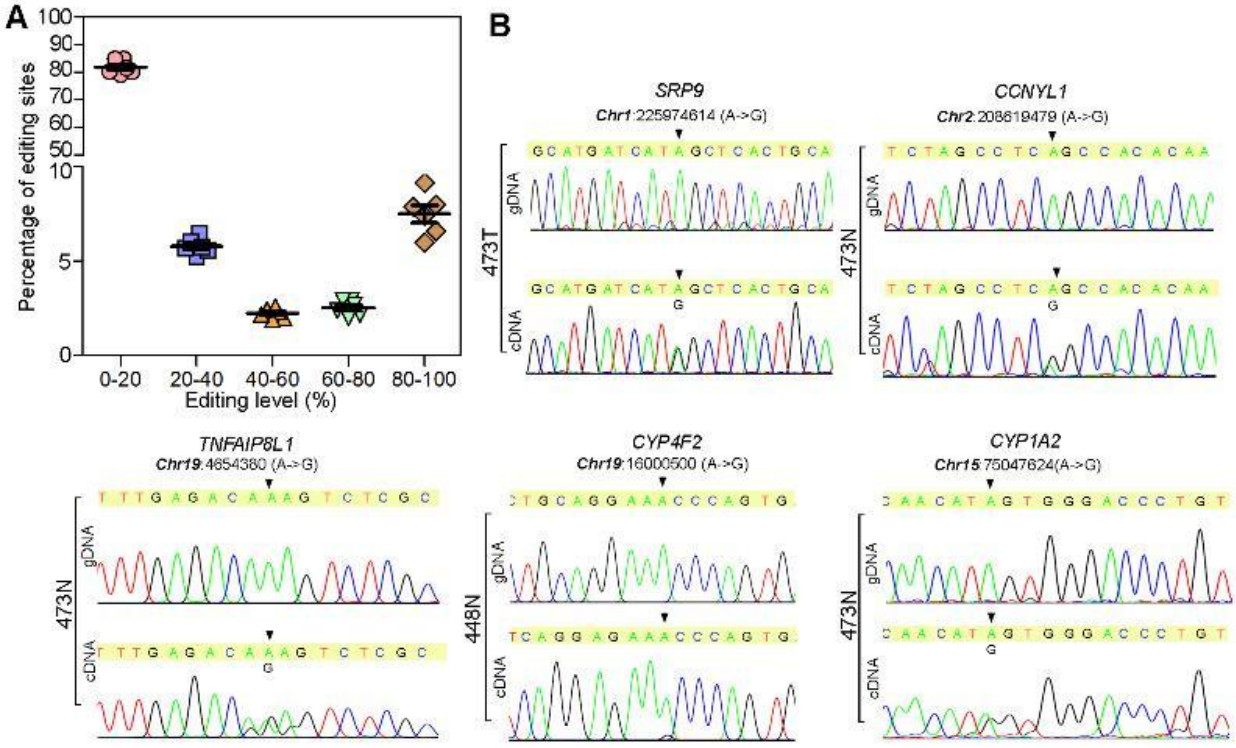
Supplementary Figure 1. **(A)** Distribution of the editing levels of potential A-to-I (G) editing sites. The data are presented as dot plots and indicate the mean ±SD of all 6 samples (3 pairs of primary HCC and matched NT liver specimens). **(B)** Validation of inferred editing sites from RNA-Seq by Sanger sequencing. The sequencing chromatograms of 5 representative gene loci are shown. The editing positions are indicated by arrows. The top trace is genomic DNA (gDNA), and the bottom trace is cDNA.

Supplementary Figure 2. The sequence chromatograms of the *AZIN1*, *FLNB*, *COPA* and *UTP14C* gDNA sequences in the indicated tumor and matched NT liver samples. An arrow indicates the editing position.

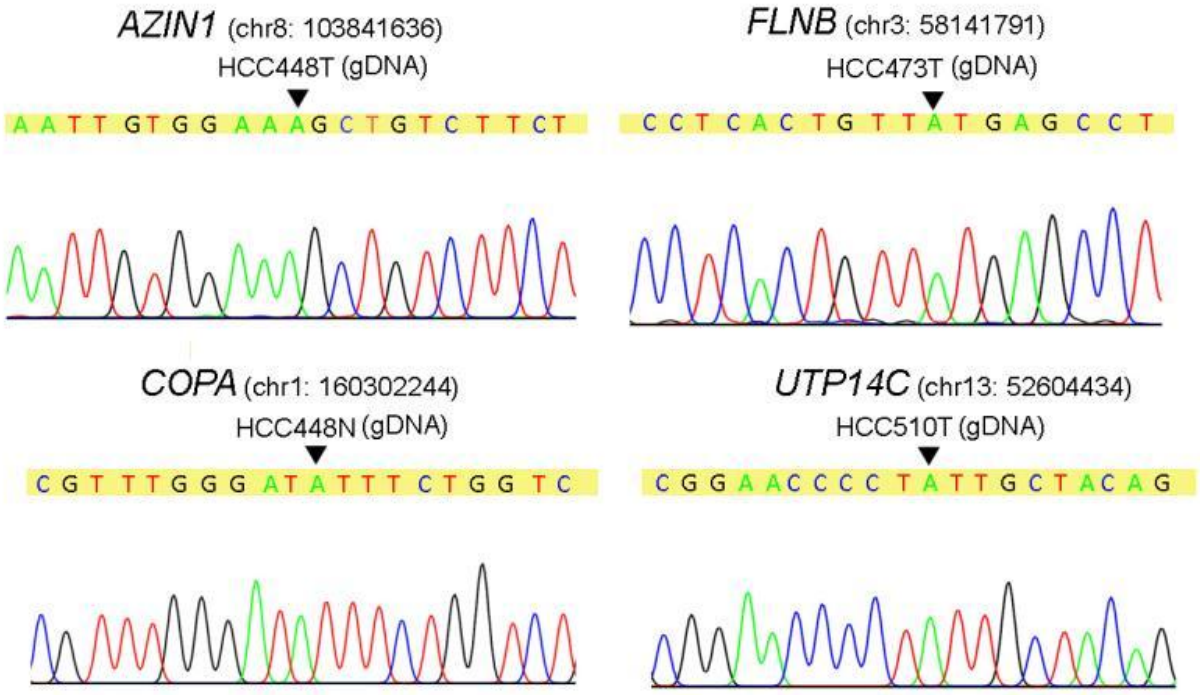
Supplementary Figure 3. **(A)** The numbers of potential editing sites within *Alu* sequences in 3 pairs of primary HCC and matched NT liver specimens. **(B)** Validation of inferred editing sites from RNA-Seq by Sanger sequencing. The editing positions are indicated by arrows.

Supplementary Figure 4. **(A)** XTT proliferation assays showing growth rates of the indicated stable cell lines. Triplicate independent experiments were performed and the data

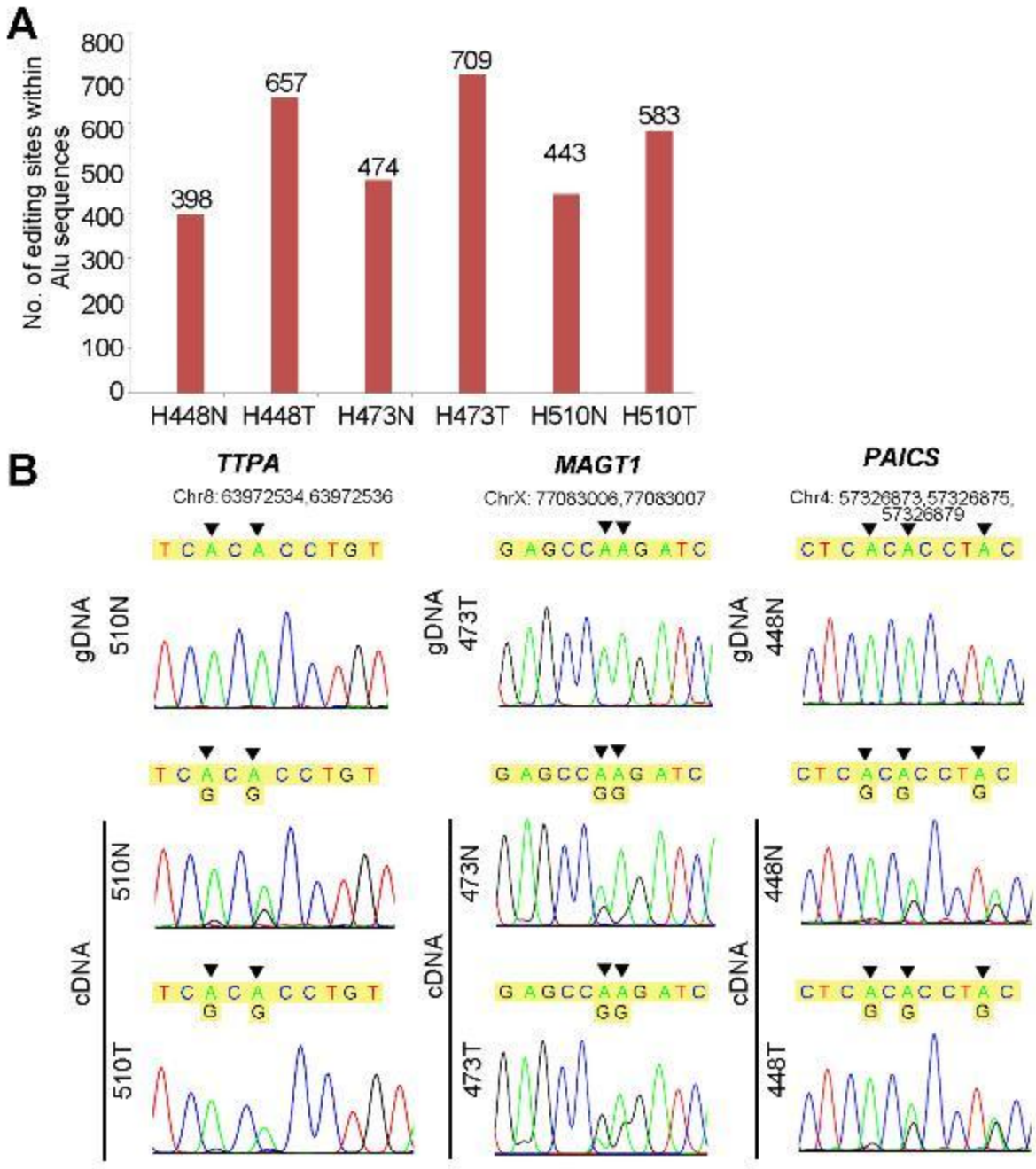
were expressed as the mean ±SD of triplicate wells within the same experiment (Unpaired, two-tailed Student's *t* test). **(B, C)** Quantification of cells from the indicated cells that invaded through Matrigel-coated membrane **(B)** or migrated through the polyethylene terephthalate (PET)-membrane **(C)** (Unpaired, two-tailed Student's *t* test). Scale bar, 200 μm.



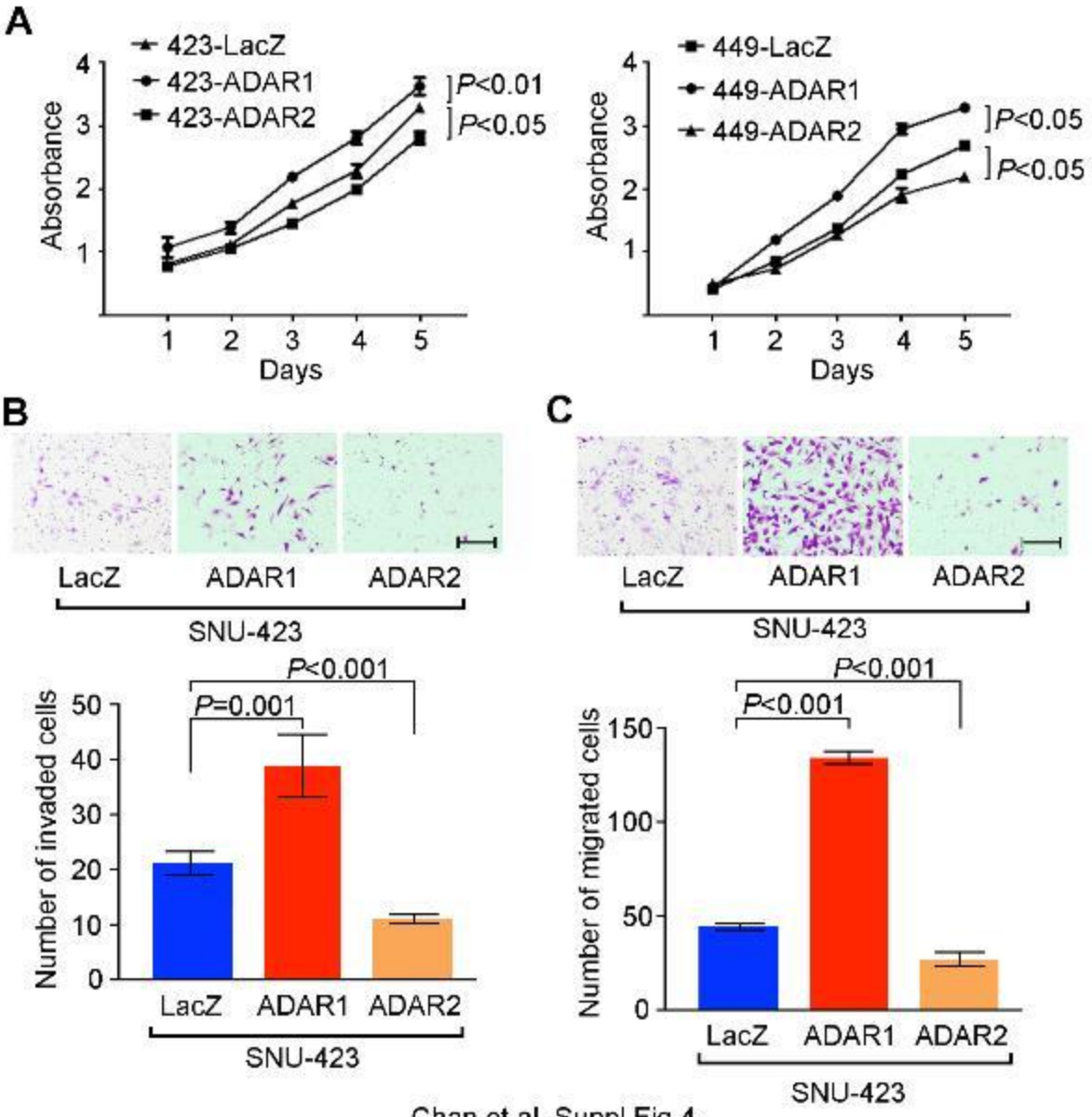
Chan et al. Suppl Fig. 1



Chan et al. Supplementary Figure 2



Chan et al. Suppl. Fig. 3



Chan et al. Suppl Fig.4

Supplementary Table 1. Distribution of potential "A-to-I (G)" editing events by transcript regions.

Region	HCC448N	HCC448T	HCC473N	HCC473T	HCC510N	HCC510T
CDS[1]	3,679	5,035	3,991	4,564	3,551	5,017
Intron	1,147	1,770	1,245	1,559	1,302	1,445
UTR[2]	5,045	5,803	5,391	5,931	4,430	6,396
Splicing site	611	990	638	950	737	988
Intergenic	1,406	1,734	1,505	1,874	1,185	2,067
Pseudo/ncRNA[3]	741	962	728	1,039	630	1,010
Total	12,629	16,294	13,498	15,917	11,835	16,923

[1]CDS, coding sequence;

[2]UTR, untranslational region;

[3]ncRNA, non-coding RNA.

Supplementary Table 2. Gene ontology (GO) analysis of the edited genes identified in all 3 non-tumor liver specimens

GO Term	No. of Genes	%	P-value	Benjamin	FDR*
acute inflammatory response	63	3.2	7.70E-21	2.80E-17	1.40E-17
response to wounding	161	8.2	7.70E-21	1.40E-17	1.40E-17
oxidation reduction	176	9	3.20E-19	3.80E-16	5.80E-16
cofactor metabolic process	78	4	2.60E-15	2.30E-12	4.70E-12
response to organic substance	170	8.7	1.90E-13	1.40E-10	3.60E-10
complement activation	33	1.7	3.50E-13	2.10E-10	6.50E-10
inflammatory response	99	5.1	4.60E-13	2.40E-10	8.40E-10
coenzyme metabolic process	63	3.2	6.10E-13	2.80E-10	1.10E-09
activation of plasma proteins involved in acute inflammatory response	33	1.7	6.40E-13	2.60E-10	1.20E-09
wound healing	71	3.6	1.60E-12	5.60E-10	2.90E-09
complement activation, classical pathway	27	1.4	2.10E-12	6.80E-10	3.80E-09
organic acid catabolic process	51	2.6	5.00E-12	1.50E-09	9.20E-09
carboxylic acid catabolic process	51	2.6	5.00E-12	1.50E-09	9.20E-09
humoral immune response mediated by circulating immunoglobulin	27	1.4	9.10E-12	2.50E-09	1.70E-08
immunoglobulin mediated immune response	35	1.8	1.70E-11	4.40E-09	3.10E-08
protein maturation	53	2.7	1.80E-11	4.30E-09	3.20E-08
protein processing	50	2.5	3.40E-11	7.70E-09	6.20E-08
B cell mediated immunity	35	1.8	4.00E-11	8.50E-09	7.30E-08
translational elongation	47	2.4	4.80E-11	9.60E-09	8.80E-08
protein maturation by peptide bond cleavage	42	2.2	1.10E-10	2.10E-08	2.00E-07
lymphocyte mediated immunity	38	1.9	1.50E-10	2.70E-08	2.80E-07
immune effector process	53	2.7	2.90E-10	5.00E-08	5.30E-07
coagulation	45	2.3	3.20E-10	5.30E-08	5.90E-07
blood coagulation	45	2.3	3.20E-10	5.30E-08	5.90E-07
innate immune response	53	2.7	6.70E-10	1.10E-07	1.20E-06
hemostasis	45	2.3	1.40E-09	2.10E-07	2.60E-06
translation	89	4.5	1.90E-09	2.80E-07	3.50E-06
cellular amino acid catabolic process	35	1.8	3.10E-09	4.30E-07	5.60E-06
leukocyte mediated immunity	39	2	3.20E-09	4.30E-07	5.90E-06
negative regulation of apoptosis	92	4.7	3.70E-09	4.80E-07	6.80E-06
negative regulation of programmed cell death	92	4.7	6.40E-09	8.10E-07	1.20E-05
sulfur metabolic process	45	2.3	6.80E-09	8.30E-07	1.30E-05
negative regulation of cell death	92	4.7	7.20E-09	8.40E-07	1.30E-05
adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	36	1.9	7.70E-09	8.70E-07	1.40E-05
adaptive immune response	36	1.9	7.70E-09	8.70E-07	1.40E-05

cellular amino acid derivative metabolic process	56	2.9	8.70E-09	9.60E-07	1.60E-05
amine catabolic process	36	1.9	1.00E-08	1.10E-06	1.90E-05
defense response	132	6.8	1.30E-08	1.30E-06	2.40E-05
organic acid biosynthetic process	53	2.7	1.70E-08	1.70E-06	3.10E-05
carboxylic acid biosynthetic process	53	2.7	1.70E-08	1.70E-06	3.10E-05
lipid localization	53	2.7	2.30E-08	2.30E-06	4.30E-05
proteasomal ubiquitin-dependent protein catabolic process	41	2.1	3.20E-08	3.10E-06	5.90E-05
proteasomal protein catabolic process	41	2.1	3.20E-08	3.10E-06	5.90E-05
regulation of body fluid levels	48	2.5	6.70E-08	6.20E-06	1.20E-04
acute-phase response	24	1.2	1.10E-07	9.90E-06	2.00E-04
cellular amino acid biosynthetic process	27	1.4	1.20E-07	1.00E-05	2.10E-04
lipid transport	48	2.5	1.30E-07	1.10E-05	2.40E-04
regulation of apoptosis	155	7.9	2.00E-07	1.70E-05	3.70E-04
generation of precursor metabolites and energy	78	4	2.00E-07	1.70E-05	3.70E-04
negative regulation of protein metabolic process	56	2.9	2.20E-07	1.70E-05	4.00E-04
negative regulation of cellular protein metabolic process	54	2.8	2.50E-07	2.00E-05	4.50E-04
complement activation, alternative pathway	15	0.8	2.90E-07	2.20E-05	5.30E-04
humoral immune response	33	1.7	3.10E-07	2.40E-05	5.80E-04
regulation of cellular protein metabolic process	104	5.3	3.20E-07	2.30E-05	5.80E-04
regulation of programmed cell death	155	7.9	3.30E-07	2.40E-05	6.00E-04
regulation of response to external stimulus	50	2.5	3.60E-07	2.50E-05	6.60E-04
regulation of cell death	155	7.9	3.90E-07	2.70E-05	7.10E-04
response to drug	60	3.1	3.90E-07	2.70E-05	7.20E-04
activation of immune response	36	1.9	4.40E-07	3.00E-05	8.10E-04
lipoprotein particle clearance	15	0.8	6.20E-07	4.10E-05	1.10E-03
response to inorganic substance	57	2.9	7.80E-07	5.10E-05	1.40E-03
hydrogen peroxide metabolic process	18	0.9	8.30E-07	5.30E-05	1.50E-03
homeostatic process	143	7.3	1.10E-06	7.10E-05	2.10E-03
peptide metabolic process	26	1.3	1.30E-06	7.70E-05	2.30E-03
organic ether metabolic process	26	1.3	1.30E-06	7.70E-05	2.30E-03
serine family amino acid metabolic process	18	0.9	1.30E-06	8.20E-05	2.50E-03
amine biosynthetic process	32	1.6	2.10E-06	1.30E-04	3.90E-03
anti-apoptosis	56	2.9	2.40E-06	1.40E-04	4.40E-03
response to hormone stimulus	83	4.2	2.40E-06	1.40E-04	4.50E-03
response to endogenous stimulus	89	4.5	2.50E-06	1.40E-04	4.60E-03
cellular lipid catabolic process	30	1.5	3.10E-06	1.70E-04	5.70E-03
glutathione metabolic process	18	0.9	3.20E-06	1.80E-04	6.00E-03

response to glucocorticoid stimulus	30	1.5	4.70E-06	2.60E-04	8.60E-03
chemical homeostasis	104	5.3	5.10E-06	2.70E-04	9.40E-03
secondary metabolic process	30	1.5	5.70E-06	3.00E-04	1.10E-02
sterol metabolic process	35	1.8	6.20E-06	3.20E-04	1.10E-02
response to oxidative stress	47	2.4	6.40E-06	3.30E-04	1.20E-02
cellular homeostasis	96	4.9	6.50E-06	3.30E-04	1.20E-02
fatty acid metabolic process	53	2.7	6.80E-06	3.40E-04	1.30E-02
regulation of lipid metabolic process	36	1.9	1.10E-05	5.40E-04	2.00E-02
acylglycerol metabolic process	23	1.2	1.20E-05	5.70E-04	2.20E-02
cholesterol transport	20	1	1.20E-05	5.70E-04	2.20E-02
sterol transport	20	1	1.20E-05	5.70E-04	2.20E-02
triglyceride metabolic process	21	1.1	1.20E-05	5.80E-04	2.20E-02
regulation of inflammatory response	29	1.5	1.30E-05	5.90E-04	2.30E-02
positive regulation of immune response	42	2.2	1.30E-05	6.10E-04	2.40E-02
regulation of fibrinolysis	11	0.5	1.40E-05	6.30E-04	2.50E-02
regulation of protein processing	9	0.5	1.50E-05	6.60E-04	2.70E-02
regulation of protein maturation by peptide bond cleavage	9	0.5	1.50E-05	6.60E-04	2.70E-02
neutral lipid metabolic process	23	1.2	1.50E-05	6.70E-04	2.80E-02
response to nutrient levels	51	2.6	1.60E-05	6.90E-04	2.90E-02
negative regulation of molecular function	74	3.8	1.70E-05	7.20E-04	3.10E-02
cholesterol metabolic process	32	1.6	1.70E-05	7.20E-04	3.10E-02
response to corticosteroid stimulus	30	1.5	1.80E-05	7.50E-04	3.30E-02
negative regulation of response to stimulus	33	1.7	1.80E-05	7.60E-04	3.30E-02
lipid homeostasis	23	1.2	1.90E-05	8.10E-04	3.60E-02
glycerol ether metabolic process	23	1.2	1.90E-05	8.10E-04	3.60E-02
protein oligomerization	47	2.4	2.10E-05	8.80E-04	3.90E-02
positive regulation of molecular function	111	5.7	2.20E-05	9.00E-04	4.10E-02
anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	26	1.3	2.30E-05	9.10E-04	4.20E-02
negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle	26	1.3	2.30E-05	9.10E-04	4.20E-02
positive regulation of response to stimulus	57	2.9	2.30E-05	9.00E-04	4.20E-02
antigen processing and presentation of peptide antigen	17	0.8	2.50E-05	9.60E-04	4.50E-02
oxidoreduction coenzyme metabolic process	23	1.2	2.50E-05	9.60E-04	4.60E-02
steroid metabolic process	51	2.6	2.70E-05	1.00E-03	4.90E-02

*FDR: false discovery rate

Supplementary Table 3. Gene ontology (GO) analysis of the edited genes identified in all 3 HCC specimens

GO Term	No. of Genes	%	P-value	Benjamin	FDR*
translation	135	4.6	5.80E-15	2.30E-11	1.10E-11
acute inflammatory response	63	2.1	1.90E-14	3.80E-11	3.60E-11
translational elongation	60	2	1.80E-12	2.40E-09	3.30E-09
cellular macromolecule catabolic process	212	7.2	2.50E-10	2.50E-07	4.60E-07
macromolecular complex assembly	197	6.7	5.00E-10	4.00E-07	9.20E-07
macromolecule catabolic process	222	7.5	5.50E-10	3.70E-07	1.00E-06
macromolecular complex subunit organization	206	7	7.90E-10	4.50E-07	1.50E-06
protein localization	243	8.3	8.40E-10	4.20E-07	1.60E-06
protein folding	77	2.6	8.70E-10	3.90E-07	1.60E-06
mRNA metabolic process	126	4.3	1.20E-09	4.70E-07	2.20E-06
protein transport	215	7.3	2.30E-09	8.30E-07	4.30E-06
establishment of protein localization	215	7.3	4.20E-09	1.40E-06	7.80E-06
intracellular transport	189	6.4	6.10E-09	1.90E-06	1.10E-05
proteolysis involved in cellular protein catabolic process	176	6	8.70E-09	2.50E-06	1.60E-05
protein catabolic process	180	6.1	1.00E-08	2.80E-06	1.90E-05
cellular protein catabolic process	176	6	1.20E-08	2.90E-06	2.20E-05
modification-dependent protein catabolic process	168	5.7	1.80E-08	4.30E-06	3.40E-05
modification-dependent macromolecule catabolic process	168	5.7	1.80E-08	4.30E-06	3.40E-05
ubiquitin-dependent protein catabolic process	89	3	3.30E-08	7.40E-06	6.20E-05
complement activation	30	1	4.30E-08	9.10E-06	8.10E-05
response to wounding	156	5.3	4.70E-08	9.40E-06	8.70E-05
regulation of cellular protein metabolic process	143	4.8	6.80E-08	1.30E-05	1.30E-04
activation of plasma proteins involved in acute inflammatory response	30	1	6.90E-08	1.30E-05	1.30E-04
RNA splicing	98	3.3	8.00E-08	1.40E-05	1.50E-04
complement activation, classical pathway	24	0.8	1.30E-07	2.10E-05	2.30E-04
proteolysis	264	9	1.90E-07	3.10E-05	3.60E-04
protein complex assembly	147	5	2.10E-07	3.20E-05	3.90E-04
protein complex biogenesis	147	5	2.10E-07	3.20E-05	3.90E-04
proteasomal ubiquitin-dependent protein catabolic process	48	1.6	2.30E-07	3.40E-05	4.30E-04
proteasomal protein catabolic process	48	1.6	2.30E-07	3.40E-05	4.30E-04
humoral immune response mediated by circulating immunoglobulin	24	0.8	3.90E-07	5.60E-05	7.30E-04

nuclear mRNA splicing, via spliceosome	62	2.1	4.10E-07	5.60E-05	7.50E-04
RNA splicing, via transesterification reactions	62	2.1	4.10E-07	5.60E-05	7.50E-04
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	62	2.1	4.10E-07	5.60E-05	7.50E-04
RNA processing	155	5.3	4.40E-07	5.90E-05	8.20E-04
lymphocyte mediated immunity	38	1.3	4.50E-07	5.80E-05	8.40E-04
acute-phase response	27	0.9	6.70E-07	8.40E-05	1.20E-03
mRNA processing	102	3.5	8.10E-07	9.90E-05	1.50E-03
sulfur metabolic process	50	1.7	1.30E-06	1.60E-04	2.50E-03
generation of precursor metabolites and energy	99	3.4	1.40E-06	1.60E-04	2.60E-03
regulation of catabolic process	44	1.5	2.20E-06	2.40E-04	4.00E-03
protein maturation by peptide bond cleavage	41	1.4	2.40E-06	2.60E-04	4.40E-03
inflammatory response	101	3.4	2.70E-06	2.80E-04	5.00E-03
immunoglobulin mediated immune response	30	1	4.60E-06	4.70E-04	8.50E-03
negative regulation of cellular protein metabolic process	65	2.2	5.30E-06	5.30E-04	9.80E-03
protein maturation	50	1.7	5.30E-06	5.20E-04	9.80E-03
posttranscriptional regulation of gene expression	72	2.4	5.80E-06	5.50E-04	1.10E-02
immune effector process	53	1.8	6.00E-06	5.60E-04	1.10E-02
regulation of lipid metabolic process	47	1.6	6.50E-06	6.00E-04	1.20E-02
protein processing	47	1.6	6.50E-06	6.00E-04	1.20E-02
leukocyte mediated immunity	39	1.3	8.00E-06	7.10E-04	1.50E-02
B cell mediated immunity	30	1	8.40E-06	7.30E-04	1.60E-02
response to hydrogen peroxide	30	1	8.40E-06	7.30E-04	1.60E-02
regulation of cellular catabolic process	32	1.1	8.80E-06	7.50E-04	1.60E-02
response to organic substance	183	6.2	1.00E-05	8.40E-04	1.90E-02
cellular macromolecule localization	117	4	1.30E-05	1.00E-03	2.40E-02
response to inorganic substance	69	2.3	1.30E-05	1.10E-03	2.50E-02
oxidation reduction	165	5.6	1.40E-05	1.10E-03	2.50E-02
negative regulation of protein metabolic process	65	2.2	1.40E-05	1.10E-03	2.70E-02
positive regulation of catabolic process	27	0.9	1.80E-05	1.30E-03	3.30E-02
cellular protein localization	116	3.9	1.80E-05	1.30E-03	3.30E-02
cofactor metabolic process	66	2.2	1.80E-05	1.30E-03	3.40E-02
positive regulation of cellular catabolic process	23	0.8	2.20E-05	1.60E-03	4.10E-02
response to reactive oxygen species	35	1.2	2.40E-05	1.70E-03	4.40E-02

*FDR: false discovery rate

Supplementary Table 4. List of potential non-tumor-Specific editing sites within coding regions in 3 pairs of HCC and matched NT liver samples.

Chr	Position	Ref.[1]	Var.[2]	No of Ref. reads	No. of Var. reads	Frequency (%)	Depth	Type[3]	Annotation Details	448 N	473 N	510 N
1	196881973	A	G	19	4	13.79	23	syn	CFHR4:NM_006684:exon3:c.A360G:p.G120G,CFHR4:NM_001201551:exon7:c.A1098G:p.G366G,CFHR4:NM_001201550:exon7:c.A1101G:p.G367G	Y	Y	Y
6	160962143	T	C	22	16	42.11	38	nonsyn	LPA:NM_005577:exon36:c.A5590G:p.K1864E	Y	Y	Y
9	37429814	A	G	1	610	95.91	611	syn	GRHPR:NM_012203:exon6:c.A579G:p.A193A	Y	Y	Y
14	24569425	A	G	6	13	61.9	19	nonsyn	PCK2:NM_001018073:exon7:c.A1237G:p.M413V	Y	Y	Y
1	16133912	T	C	1	97	97	98	nonsyn	UQCRHL:NM_001089591:exon1:c.A233G:p.K78R	Y	Y	N
1	160302244	T	C	9	6	40	15	nonsyn	COPA:NM_001098398:exon6:c.A490G:p.I164V,COPA:NM_004371:exon6:c.A490G:p.I164V	Y	Y	N
3	57570123	T	C	17	2	10.53	19	nonsyn	ARF4:NM_001660:exon2:c.A136G:p.I46V	Y	Y	N
3	101404666	T	C	35	4	10	39	nonsyn	RPL24:NM_000986:exon3:c.A188G:p.Q63R	Y	Y	N
3	145820566	T	C	18	2	10	20	syn	PLOD2:NM_000935:exon7:c.A753G:p.A251A	Y	Y	N
4	100235106	T	C	129	18	11.76	147	nonsyn	ADH1B:NM_000668:exon6:c.A700G:p.K234E	Y	Y	N
20	47708626	A	G	17	2	10	19	syn	CSE1L:NM_001316:exon22:c.A2409G:p.A803A	Y	Y	N
20	60883509	A	G	20	6	18.18	26	nonsyn	ADRM1:NM_007002:exon9:c.A1100G:p.E367G,ADRM1:NM_175573:exon9:c.A1100G:p.E367G	Y	Y	N
22	43036239	T	C	1	27	96.43	28	syn	ATP5L2:NM_001165877:exon1:c.A42G:p.A14A	Y	Y	N
22	43036242	T	C	1	27	96.43	28	syn	ATP5L2:NM_001165877:exon1:c.A39G:p.P13P	Y	Y	N
11	89429882	A	G	11	4	25	15	syn	FOLH1B:NM_153696:exon13:c.A1128G:p.R376R	Y	N	Y
17	40970834	T	C	16	2	11.11	18	nonsyn	BECN1:NM_003766:exon5:c.A322G:p.T108A	Y	N	Y

17	47782564	T	C	9	3	16.67	12	syn	SLC35B1:NM_005827:exon5:c.A450G:p.G150G	Y	N	Y
20	32693186	T	C	15	2	11.76	17	nonsyn	EIF2S2:NM_003908:exon2:c.A181G:p.T61A	Y	N	Y
X	120182330	A	G	14	3	14.29	17	syn	GLUD2:NM_012084:exon1:c.A792G:p.G264G	Y	N	Y
2	87244692	T	C	3	12	80	15	nonsyn	PLGLB2:NM_002665:exon2:c.A169G:p.K57E,PLGLB1:NM_001032392:exon2:c.A169G:p.K57E	N	Y	Y
2	178095529	T	C	32	4	11.11	36	nonsyn	NFE2L2:NM_006164:exon5:c.A1802G:p.D601A,NFE2L2:NM_001145412:exon5:c.A1754G:p.D585A,NFE2L2:NM_001145413:exon5:c.A1733G:p.D578A	N	Y	Y
8	97285615	A	G	16	2	11.11	18	nonsyn	PTDSS1:NM_014754:exon2:c.A268G:p.N90D	N	Y	Y
9	108097933	A	G	14	3	17.65	17	nonsyn	SLC44A1:NM_080546:exon4:c.A359G:p.Q120R	N	Y	Y
10	96702029	A	G	31	6	13.04	37	nonsyn	CYP2C9:NM_000771:exon3:c.A412G:p.K138E	N	Y	Y
17	1648692	A	G	3	32	78.05	35	syn	SERPINF2:NM_001165921:exon4:c.A168G:p.V56V	N	Y	Y
18	47317843	T	C	28	5	13.51	33	nonsyn	ACAA2:NM_006111:exon7:c.A880G:p.I294V	N	Y	Y

[1] Ref.=Reference;

[2] Var.=Variation

[3] Nonsyn, nonsynonymous; Syn, synononymous

Supplementary Table 5. List of potential tumor-specific editing sites within coding regions in 3 pairs of HCC and matched NT liver samples.

Chr	Position	Ref.[1]	Var.[2]	No of Ref. reads	No. of Var. reads	Frequency (%)	Depth	Type[3]	Annotation Details	448 T	473 T	510 T
2	216236742	T	C	403	54	11.34	457	nonsyn SNV	FN1:NM_212476:exon38:c.A6061G:p.T2021A,FN1:NM_212482:exon40:c.A6604G:p.T2202A	Y	Y	Y
15	76578730	T	C	72	10	10.2	82	nonsyn SNV	ETFA:NM_001127716:exon5:c.A397G:p.S133G,ETFA:NM_000126:exon6:c.A544G:p.S182G	Y	Y	Y
17	80445942	A	G	14	5	21.74	19	nonsyn SNV	NARF:NM_031968:exon12:c.A1418G:p.E473G,NARF:NM_001038618:exon11:c.A1103G:p.E368G,NARF:NM_001083608:exon10:c.A1136G:p.E379G,NARF:NM_012336:exon11:c.A1280G:p.E427G	Y	Y	Y
1	156182847	A	G	11	2	11.76	13	nonsyn SNV	PMF1:NM_001199653:exon1:c.A41G:p.E14G,PMF1-BGLAP:NM_001199661:exon1:c.A41G:p.E14G,PMF1-BGLAP:NM_001199663:exon1:c.A41G:p.E14G,PMF1:NM_001199654:exon1:c.A41G:p.E14G	Y	Y	N
1	225974614	A	G	29	23	42.59	52	nonsyn SNV	SRP9:NM_001130440:exon3:c.A192G:p.I64M	Y	Y	N
5	108713940	T	C	15	2	11.11	17	syn SNV	PJA2:NM_014819:exon4:c.A1248G:p.G416G	Y	Y	N
5	138661339	A	G	26	5	15.62	31	nonsyn SNV	MATR3:NM_001194955:exon13:c.A2359G:p.N787D,MATR3:NM_199189:exon16:c.A2359G:p.N787D,MATR3:NM_001194954:exon15:c.A2359G:p.N787D,MATR3:NM_018834:exon13:c.A2359G:p.N787D,MATR3:NM_001194956:exon12:c.A1495G:p.N499D	Y	Y	N

6	31752021	T	C	56	7	11.11	63	syn SNV	VARS:NM_006295:exon13:c.A1641G:p.V547V	Y	Y	N
6	44219221	A	G	20	6	20.69	26	nonsyn SNV	HSP90AB1:NM_007355:exon8:c.A1190G:p.Q397R	Y	Y	N
12	67701215	A	G	19	4	16	23	nonsyn SNV	CAND1:NM_018448:exon11:c.A2968G:p.K990E	Y	Y	N
1	36921393	T	C	10	3	23.08	13	nonsyn SNV	MRPS15:NM_031280:exon8:c.A770G:p.Q257R	Y	N	Y
2	198353049	T	C	13	2	12.5	15	nonsyn SNV	HSPD1:NM_002156:exon10:c.A1382G:p.Q461R,HSPD1:NM_199440:exon10:c.A1382G:p.Q461R	Y	N	Y
3	50112710	A	G	10	3	16.67	13	nonsyn SNV	RBM6:NM_005777:exon20:c.A3193G:p.R1065G,RBM6:NM_001167582:exon16:c.A1627G:p.R543G	Y	N	Y
8	97243292	T	C	364	50	10.94	414	syn SNV	UQCRB:NM_006294:exon4:c.A327G:p.A109A,UQCRB:NM_001199975:exon5:c.A231G:p.A77A	Y	N	Y
10	5010523	A	G	67	11	12.09	78	nonsyn SNV	AKR1C1:NM_001353:exon4:c.A392G:p.K131R	Y	N	Y
11	18424543	A	G	26	4	12.5	30	nonsyn SNV	LDHA:NM_001135239:exon4:c.A401G:p.E134G,LDHA:NM_001165414:exon5:c.A662G:p.E221G,LDHA:NM_001165415:exon5:c.A575G:p.E192G,LDHA:NM_001165416:exon5:c.A575G:p.E192G	Y	N	Y
11	62289306	T	C	28	4	11.76	32	nonsyn SNV	AHNAK:NM_001620:exon5:c.A12583G:p.K4195E	Y	N	Y
11	94801227	A	G	35	5	12.2	40	syn SNV	SRSF8:NM_032102:exon1:c.A837G:p.R279R	Y	N	Y
13	25670574	A	G	28	7	19.44	35	nonsyn SNV	PABPC3:NM_030979:exon1:c.A238G:p.K80E	Y	N	Y
13	41515296	T	C	22	5	11.63	27	syn SNV	ELF1:NM_172373:exon8:c.A1017G:p.G339G,ELF1:NM_001145353:exon7:c.A945G:p.G315G	Y	N	Y

13	52604985	A	G	6	7	53.85	13	nonsyn SNV	UTP14C:NM_021645:exon2:c.A2045G:p.Q682R	Y	N	Y
16	67231545	A	G	48	7	12.28	55	syn SNV	E2F4:NM_001950:exon8:c.A1077G:p.T359T	Y	N	Y
18	20572889	A	G	27	3	10	30	nonsyn SNV	RBBP8:NM_203291:exon11:c.A1099G:p.T367A,RBBP8:NM_203292:exon11:c.A1099G:p.T367A,RBBP8:NM_002894:exon11:c.A1099G:p.T367A	Y	N	Y
X	27766155	A	G	14	16	50	30	syn SNV	DCAF8L2:NM_001136533:exon1:c.A1143G:p.Q381Q	Y	N	Y
X	27766172	A	G	14	26	59.09	40	nonsyn SNV	DCAF8L2:NM_001136533:exon1:c.A1160G:p.Q387R	Y	N	Y
1	35925949	T	C	16	2	10.53	18	nonsyn SNV	KIAA0319L:NM_024874:exon9:c.A1384G:p.I462V	N	Y	Y
1	222895848	A	G	13	2	10.53	15	syn SNV	BROX:NM_144695:exon5:c.A393G:p.G131G	N	Y	Y
2	99234730	A	G	27	3	10	30	nonsyn SNV	UNC50:NM_014044:exon6:c.A743G:p.H248R	N	Y	Y
3	194149599	T	C	25	3	10.34	28	syn SNV	ATP13A3:NM_024524:exon27:c.A2922G:p.A974A	N	Y	Y
7	56122076	A	G	33	6	13.33	39	syn SNV	CCT6A:NM_001762:exon3:c.A216G:p.P72P	N	Y	Y
9	6013849	T	C	13	11	44	24	nonsyn SNV	RANBP6:NM_012416:exon1:c.A1759G:p.N587D	N	Y	Y
9	6013859	T	C	12	11	45.83	23	syn SNV	RANBP6:NM_012416:exon1:c.A1749G:p.Q583Q	N	Y	Y
11	86055746	A	G	43	5	10.2	48	syn SNV	C11orf73:NM_016401:exon4:c.A522G:p.A174A	N	Y	Y
12	9230302	T	C	14	4	21.05	18	nonsyn SNV	A2M:NM_000014:exon26:c.A3271G:p.I1091V	N	Y	Y
12	49315778	T	C	27	4	12.5	31	nonsyn SNV	FKBP11:NM_016594:exon6:c.A595G:p.K199E,FKBP11:NM_001143781:exon5:c.A289G:p.K97E	N	Y	Y
13	46090371	A	G	19	3	13.04	22	nonsyn SNV	COG3:NM_031431:exon17:c.A1903G:p.I635V	N	Y	Y

13	52604264	A	G	17	5	22.73	22	nonsyn SNV	UTP14C:NM_021645:exon2:c.A1324G:p.S442G	N	Y	Y
13	52604288	A	G	18	5	21.74	23	nonsyn SNV	UTP14C:NM_021645:exon2:c.A1348G:p.S450G	N	Y	Y
21	34948722	A	G	10	4	28.57	14	nonsyn SNV	SON:NM_138927:exon12:c.A7273G:p.R2425G	N	Y	Y
13	52604434	A	G	17	5	22.73	22	syn SNV	UTP14C:NM_021645:exon2:c.A1494G:p.L498L	N	Y	Y

[1] Ref.=Reference;

[2] Var.=Variation

[3] Nonsyn, nonsynonymous; Syn, synononymous

Supplementary Table 6. List of potential common editing sites within coding regions in 3 pairs of HCC and matched NT liver samples.

Chr	Position	Ref.[1]	Var.[2]	No of Ref. reads	No. of Var. reads	Frequency (%)	Depth	Type[3]	Annotation Details	448 N/T	473 N/T	510 N/T
8	103841636	T	C	36	16	28.57	52	nonsyn	AZIN1:NM_015878:exon12:c.A1099G:p.S367G,AZIN1:NM_148174:exon11:c.A1099G:p.S367G	Y	Y	Y
10	43882155	T	C	15	13	35.14	28	nonsyn	HNRNPF:NM_001098208:exon3:c.A1178G:p.E393G,HNRNPF:NM_001098206:exon4:c.A1178G:p.E393G,HNRNPF:NM_004966:exon4:c.A1178G:p.E393G,HNRNPF:NM_001098204:exon4:c.A1178G:p.E393G,HNRNPF:NM_001098207:exon4:c.A1178G:p.E393G,HNRNPF:NM_001098205:exon4:c.A1178G:p.E393G	Y	Y	Y
11	17097117	T	C	177	46	19.66	223	nonsyn	RPS13:NM_001017:exon4:c.A205G:p.N69D	Y	Y	Y
X	100667807	A	G	29	42	58.33	71	syn	HNRNPH2:NM_019597:exon2:c.A831G:p.G277G,HNRNPH2:NM_001032393:exon2:c.A831G:p.G277G	Y	Y	Y
X	120182218	A	G	6	38	84.44	44	nonsyn	GLUD2:NM_012084:exon1:c.A680G:p.N227S	Y	Y	Y
X	120182436	A	G	4	73	87.95	77	nonsyn	GLUD2:NM_012084:exon1:c.A898G:p.R300G	Y	Y	Y
X	120183106	A	G	3	118	84.89	121	nonsyn	GLUD2:NM_012084:exon1:c.A1568G:p.H523R	Y	Y	Y
2	198365911	A	G	34	7	16.28	41	syn	HSPE1-MOBKL3:NM_001202485:exon2:c.A117G:p.G39G,HSPE1:NM_002157:exon2:c.A117G:p.G39G	Y	N	Y

19	16000500	T	C	14	6	26.09	20	syn	CYP4F2:NM_001082:exon7:c.A651G:p.K217K	Y	N	Y
3	58141791	A	G	14	10	38.46	24	nonsyn	FLNB:NM_001457:exon41:c.A6877G:p.M2293V, FLNB:NM_001164318:exon41:c.A6844G:p.M2282V,FLNB:NM_001164317:exon42:c.A6970G:p.M2324V,FLNB:NM_001164319:exon40:c.A6805G:p.M2269V	Y	N	Y
13	52604880	A	G	10	5	33.33	15	nonsyn	UTP14C:NM_021645:exon2:c.A1940G:p.Q647R	Y	N	Y
X	51488018	A	G	3	10	66.67	13	syn	GSPT2:NM_018094:exon1:c.A1296G:p.R432R,	Y	N	Y
X	51488042	A	G	3	14	73.68	17	syn	GSPT2:NM_018094:exon1:c.A1320G:p.K440K,	Y	N	Y
2	70315692	A	G	212	72	25.09	284	nonsyn	PCBP1:NM_006196:exon1:c.A817G:p.S273G,	N	Y	Y
12	6646277	A	G	51	15	19.48	66	nonsyn	GAPDH:NM_002046:exon6:c.A338G:p.Q113R	N	Y	Y
13	25671320	A	G	2	231	96.25	233	syn	PABPC3:NM_030979:exon1:c.A984G:p.E328E	N	Y	Y
X	120182480	A	G	7	31	79.49	38	syn	GLUD2:NM_012084:exon1:c.A942G:p.L314L	N	Y	Y

[1] Ref.=Reference;

[2] Var.=Variation

[3] Nonsyn, nonsynonymous; Syn, synononymous

Supplementary Table 7. IHC scores of ADAR1 in 92 pairs of primary HCCs and matched non-tumor liver tissues.

		IHC scores of T (no. of sections)				Total
		0	**1**	**2**	**3**	
IHC scores of NT (No. of sections)	**0**	8	**10**	**11**	**4**	33 (35.9%)
	1	2	8	**25**	**13**	48 (52.2%)
	2	2	2	4	**3**	11 (11.9%)
Total		12 (13.0%)	20 (21.7%)	40 (43.5%)	20 (21.7%)	92

Supplementary Table 8. Cox proportional hazard regression analyses for Disease-free survival.

Clinical-pathological Features	Univariable analysis		Multivariable analysis	
	HR (95% CI)	***P***	**HR (95% CI)**	***P***
Age (>60 vs <=60)	0.598(0.255-1.406)	0.239		
Gender (Male vs Female)	0.868(0.420-1.794)	0.703		
AFP (>400 vs <=400ng/mL)	1.170(0.669-2.046)	0.582		
HBsAg (Positive vs Negative)	1.961(0.778-4.945)	0.154		
Cirrhosis (Present vs absent)	4.208(1.667-10.624)	**0.002**	5.201(1.588-17.036)	**0.006**
Tumor size (>5 vs <=5cm)	1.330(0.704-2.551)	0.380		
Tumor staging (AJCC)	1.246(0.913-1.863)	**0.046**	1.288(0.820-2.024)	0.273
Differentiation	0.657(0.405-1.066)	0.089		
ADAR1 OE & ADAR2 DR in tumors	2.028(1.298-3.168)	**0.002**	1.725(1.071-2.777)	**0.025**

CI = confidence interval; HR = hazard ratio; OE: overexpression; DR: downregulation