



## A human-specific switch of alternatively spliced *AFMID* isoforms contributes to *TP53* mutations and tumor recurrence in hepatocellular carcinoma

Kuan-Ting Lin, Wai Kit Ma, Juergen Scharner, et al.

*Genome Res.* 2018 28: 275-284 originally published online February 15, 2018  
Access the most recent version at doi:[10.1101/gr.227181.117](https://doi.org/10.1101/gr.227181.117)

---

**References** This article cites 40 articles, 7 of which can be accessed free at:  
<http://genome.cshlp.org/content/28/3/275.full.html#ref-list-1>

**Creative Commons License** This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white-bordered box containing the words "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, and the Cellecta logo, which consists of a cluster of green dots.

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

© 2018 Lin et al.; Published by Cold Spring Harbor Laboratory Press

## Research

# A human-specific switch of alternatively spliced *AFMID* isoforms contributes to *TP53* mutations and tumor recurrence in hepatocellular carcinoma

Kuan-Ting Lin,<sup>1</sup> Wai Kit Ma,<sup>1</sup> Juergen Scharner,<sup>1</sup> Yun-Ru Liu,<sup>2</sup> and Adrian R. Krainer<sup>1</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; <sup>2</sup>Joint Biobank, Office of Human Research, Taipei Medical University, Taipei, Taiwan 11031

Pre-mRNA splicing can contribute to the switch of cell identity that occurs in carcinogenesis. Here, we analyze a large collection of RNA-seq data sets and report that splicing changes in hepatocyte-specific enzymes, such as *AFMID* and *KHK*, are associated with HCC patients' survival and relapse. The switch of *AFMID* isoforms is an early event in HCC development and is associated with driver mutations in *TP53* and *ARID1A*. The switch of *AFMID* isoforms is human-specific and not detectable in other species, including primates. Finally, we show that overexpression of the full-length *AFMID* isoform leads to a higher NAD<sup>+</sup> level, lower DNA-damage response, and slower cell growth in HepG2 cells. The integrative analysis uncovered a mechanistic link between splicing switches, de novo NAD<sup>+</sup> biosynthesis, driver mutations, and HCC recurrence.

[Supplemental material is available for this article.]

Liver cancer is the second leading cause of cancer death worldwide and has very poor prognosis, with an incidence rate almost equal to the mortality rate (ratio=0.95) (Ferlay et al. 2015). The global incidence of liver cancer has increased in the past 20 yr, resulting in a doubling in disease-specific mortality (Llovet et al. 2015). Hepatocellular carcinoma (HCC) is the primary malignancy of the liver. The only approved drug for HCC is the protein tyrosine kinase inhibitor sorafenib, which can only prolong survival by about 3 mo. Surgical resection has the best prognosis for long-time survival, but only a minority (~15%) of HCC patients have enough normal liver remaining at the time of diagnosis. Even if surgical resection is successful, most HCC patients (~90%) die within 5 yr, because of intrahepatic recurrent HCC tumors (HCCs). The 5-yr survival rate is ~17% in the United States. Unfortunately, recent clinical trials of experimental HCC drugs have all failed (Llovet et al. 2015). Accordingly, there is an urgent unmet clinical need in prevention, diagnosis, prognosis, and treatment for this deadly cancer.

HCC cells are highly heterogeneous: Different areas within the same tumor often have different patterns of morphology, immunohistochemical staining, and driver mutations (Friemel et al. 2015). The negative results from recent HCC clinical trials also highlight the intrinsic resistance of HCC to therapies (Villanueva and Llovet 2014). One important aspect of how cell identity is determined is through alternative pre-mRNA splicing patterns, which are regulated in a cell-type-specific manner. Recent studies identified some recurrent splicing events in HCC, but they did not establish associations with clinical information and were limited to small patient cohorts (Sebestyen et al. 2015, 2016; Zhang et al. 2016b). Also, the detection methods used in the previous studies were limited to analyzing only two isoforms at a time.

To provide an integrative analysis of splicing patterns during the transition from hepatocytes to HCC cells, the present study analyzed ~6000 samples of RNA-seq data comprising human hepato-

cytes, Kupffer cells, adult and fetal livers, dysplastic lesions, early HCCs, HCCs, and cancer cell lines from various tissue types. We sought to identify robust splicing events associated with survival, recurrence, and driver mutations in HCC, by using a modified percent spliced-in (PSI) index (see Methods). In particular, we describe an *AFMID* alternative splicing event and propose that it plays a critical role in early HCC development and progression.

## Results

### Concordant splicing events in HCC tumors and liver-cancer cell lines

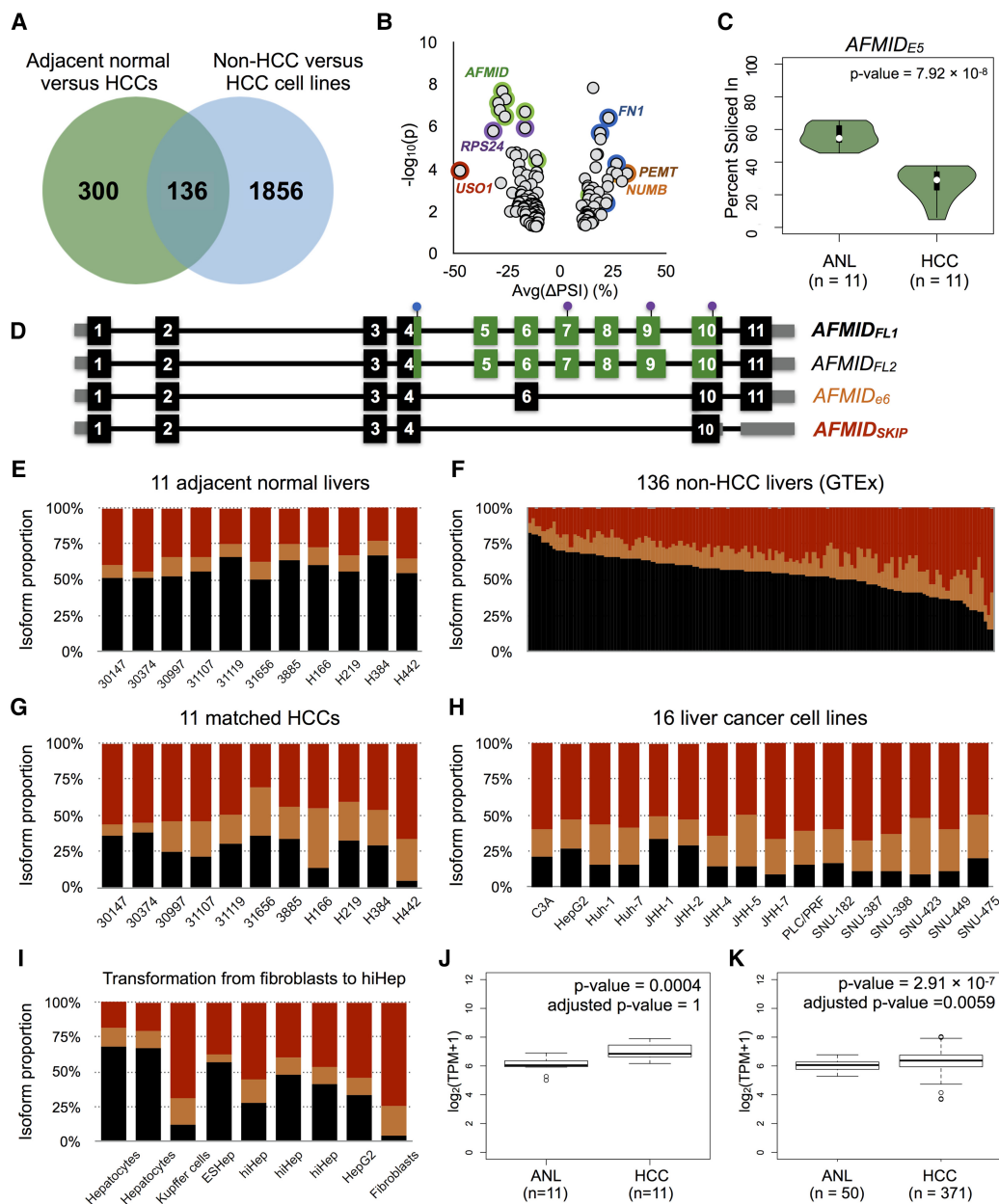
Liver-cancer cell lines are malignant clones derived from heterogeneous liver tumors. Concordant splicing events that coexist in both HCCs and liver-cancer cell lines could be among the main characteristics preserved in liver-cancer evolution. Concordance means that the splicing event involves the same exon/exons and increases or decreases in the same direction. To identify splicing events, we used a new PSI index to analyze RNA-seq data sets (Methods; Supplemental Fig. S1). We started with an RNA-seq data set of 11 primary HCCs and matched adjacent normal liver (ANL) tissues from a recent study (Jhunjunwala et al. 2014). Also, we analyzed 136 non-HCC liver samples from the Genotype-Tissue Expression (GTEx) database and 16 liver-cancer cell lines (The GTEx Consortium 2015; Klijn et al. 2015). We identified 436 and 1992 robust splicing events in HCCs and liver-cancer cell lines, respectively (Supplemental Tables S1, S2). To identify highly reproducible splicing events, we required at least 20 supporting reads for the event in at least 80% of the samples.

Among the splicing events, we identified 136 overlapping events involving 135 alternative exons with concordant splicing changes ( $P = 5.16 \times 10^{-133}$ ) (Fig. 1A). The top 10 genes with the

**Corresponding author:** [krainer@cshl.edu](mailto:krainer@cshl.edu)

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.227181.117>.

© 2018 Lin et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.



**Figure 1.** Concordant splicing events in HCC patient samples and liver-cancer cell lines. (A) The Venn diagram shows the number of overlapping splicing events between two independent comparisons: 11 adjacent normal livers (ANLs) versus 11 matched HCCs (green circle); and 136 non-HCC liver samples versus 16 liver-cancer cell lines (blue circle). (B) Each dot in the dot plot represents an alternative exon. The color coding of the circles indicates different exons from the same gene. The x-axis shows the average PSI change of an alternative exon between ANL and HCCs. The y-axis is the statistical significance, denoted as  $-\log_{10}(p)$ , of the splicing event in 11 ANLs versus 11 HCCs. (C) The violin plot shows the PSI distributions of exon 5 of *AFMID* in 11 ANLs and 11 matched HCCs. (D) Exon/intron structure of *AFMID* isoforms. Each dark block is a coding exon; the gray boxes represent the UTR exons or portions of exons; the green boxes denote domain regions. Three purple dots represent the active-site residues (Ser164, Asp247, and His279). The blue dot on exon 4 represents the HGGXW motif (the oxyanion hole). (E–H) The stacked bar charts in E and G represent the estimated proportion of *AFMID* isoforms in the 11 pairs of ANLs and matched HCCs, based on RNA-seq junction reads; F shows the 136 non-HCC liver samples, and H shows the 16 liver-cancer cell lines. (I) The stacked bar chart shows the PSIs of *AFMID* isoforms in human hepatocytes, Kupffer cells, hiHep cells (ESHep and hiHep), HepG2, and skin fibroblasts. (J,K) The two box plots show the overall gene-expression level of *AFMID* in 11 matched HCC patient samples and 371 HCC patient samples (50 with matched ANLs) from TCGA, respectively. Expression levels are shown as transcripts per million (TPM).

largest PSI changes were *PENT*, *KIAA1551*, *NUMB*, *FN1*, *MYO1B*, *USO1*, *RPS24*, *AFMID*, *KHK*, and *ARHGEF10L* (Fig. 1B). *PENT* (phosphatidylethanolamine N-methyltransferase) mainly expresses the *PENT<sub>B</sub>* isoform (NM\_007169) in ANLs, but we found that it switches to greater expression of the *PENT<sub>A1</sub>* isoform

(NM\_148172) in HCCs (Supplemental Fig. S2; Shields et al. 2001). *NUMB* is a cell-fate determinant in cell development; the *NUMB<sub>L</sub>* isoform is known to be up-regulated in HCCs (Lu et al. 2015). *FN1* (fibronectin 1) had increased inclusion of its EDB exon. The *FN1<sub>EDB</sub>* isoform is a known hallmark of tumor

angiogenesis and is regulated by the stiffness of the extracellular matrix (Bordeleau et al. 2015). *USO1* (vesicle transport factor p115) showed a reduction in exon 13 inclusion (Chr 4: 76716488–76716509) and was recently proposed as a potential splicing marker in HCC (Danan-Gotthold et al. 2015). *KHK* (ketohexokinase) switched from *KHK<sub>C</sub>* to *KHK<sub>A</sub>* isoforms. A recent study showed that switching *KHK<sub>A</sub>* to *KHK<sub>C</sub>* can induce heart disease (Mirtschink et al. 2015). Conversely, switching *KHK<sub>C</sub>* to *KHK<sub>A</sub>* drives HCC development (Li et al. 2016). *AFMID* (arylformamidase) showed a decrease in the full-length isoform and a higher proportion of the other two alternative isoforms in liver-cancer cells; one isoform skips five exons (exon 5 to exon 9) and the other skips four exons (exons 5, 7, 8, and 9). Among these top 10 genes, *PEMT*, *KHK*, and *AFMID* encode liver-enriched enzymes. The events involving the alternative exons of *AFMID* had the most significant *P*-values (Fig. 1C). Exon 6 of *AFMID* has both increased and decreased PSI values in liver-cancer cells because it is present in the two full-length isoforms (*AFMID<sub>FL1</sub>* and *AFMID<sub>FL2</sub>*) and in an alternative isoform (*AFMID<sub>e6</sub>*) that skips exons 5, 7, 8, and 9 (Fig. 1D). In liver-cancer cells, the exon 6 PSI values were increased for the *AFMID<sub>e6</sub>* isoform and decreased for the *AFMID<sub>FL1</sub>* and *AFMID<sub>FL2</sub>* isoforms.

### The switch of *AFMID* isoforms corresponds to loss of normal hepatocyte identity

*AFMID* is located on Chromosome 17q25.3 and encodes arylformamidase, a controlling enzyme in tryptophan metabolism. The *AFMID* expression level is evolutionarily constrained across multiple species (Pervouchine et al. 2015). *AFMID* generally expresses four isoforms, including *AFMID<sub>FL1</sub>*, *AFMID<sub>FL2</sub>*, *AFMID<sub>e6</sub>*, and *AFMID<sub>SKIP</sub>* (Fig. 1D). *AFMID<sub>FL1</sub>* is the major isoform and has a shorter exon 9 than *AFMID<sub>FL2</sub>*. *AFMID* has a HGGXW motif (in exon 4), an alpha/beta hydrolase-fold domain (in exons 4–10), and an active site triad (in exons 7, 9, and 10, respectively) (Fig. 1D; Pabarcus and Casida 2002, 2005). *AFMID<sub>SKIP</sub>* splices out exons 5–9, where the core-domain region resides. Because the *AFMID<sub>SKIP</sub>* isoform is incompletely annotated in genome browsers, we lacked direct information showing whether this isoform includes constitutive exons 1–4. To determine the complete structure of *AFMID<sub>SKIP</sub>*, we analyzed Pacific Biosciences (PacBio) long reads derived from several cancer cell lines (Tilgner et al. 2014). We found that most of the isoforms that lack exons 5–9 have exons 1 through 4 in GM12878, GM12891, GM12892, and K562 cell lines (Supplemental Fig. S3).

In the present study, we used the PSI of exon 5 to represent the PSI of *AFMID<sub>FL</sub>*, as both *AFMID<sub>FL1</sub>* and *AFMID<sub>FL2</sub>* share the same exon 5. In normal or non-HCC livers, *AFMID* primarily expresses *AFMID<sub>FL</sub>* (Fig. 1E,F). In HCCs or liver-cancer cell lines, *AFMID* expresses mostly *AFMID<sub>SKIP</sub>* and *AFMID<sub>e6</sub>* (Fig. 1G,H). Analyzing the RNA-seq data set from a recent study of human induced hepatocytes (hiHeps), we found that human hepatocytes had the highest PSI values of *AFMID<sub>FL</sub>*, and hiHep cells had higher PSI values of *AFMID<sub>FL</sub>* than HepG2 and fibroblasts (Fig. 1I; Huang et al. 2014). From the RNA-seq data set of human Kupffer cells (specialized macrophages located in the liver), we also found that *AFMID<sub>SKIP</sub>* and *AFMID<sub>e6</sub>* isoforms were the dominant isoforms (Fig. 1I; Costa-Silva et al. 2015). Both sets of data showed that the high-*AFMID<sub>FL</sub>* pattern is characteristic of human hepatocytes.

Interestingly, although the PSI values of *AFMID<sub>FL</sub>* were significantly decreased in HCCs, the overall gene-expression level of *AFMID* was maintained at similar levels in the two RNA-seq data

sets: (1) 11 ANLs versus 11 HCCs (~1.97-fold increase, adjusted *P*-value = 1) (Fig. 1J); and (2) 50 ANLs vs. 371 HCCs (~0.25-fold increase, adjusted *P*-value = 0.0059) (Fig. 1K). Real-time RT-PCR experiments with RNA from 20 ANLs and 19 HCCs showed that the overall *AFMID* level did not significantly change (~25% reduction, *P* = 0.2963), but the *AFMID<sub>FL</sub>* isoform level was significantly down-regulated in HCCs by about twofold (~55% reduction, *P* = 0.0042) (Supplemental Fig. S4). Both sets of data suggest that *AFMID* is not a differentially expressed gene—given the small magnitude of the change—but rather the proportions of its isoforms are switched in HCCs.

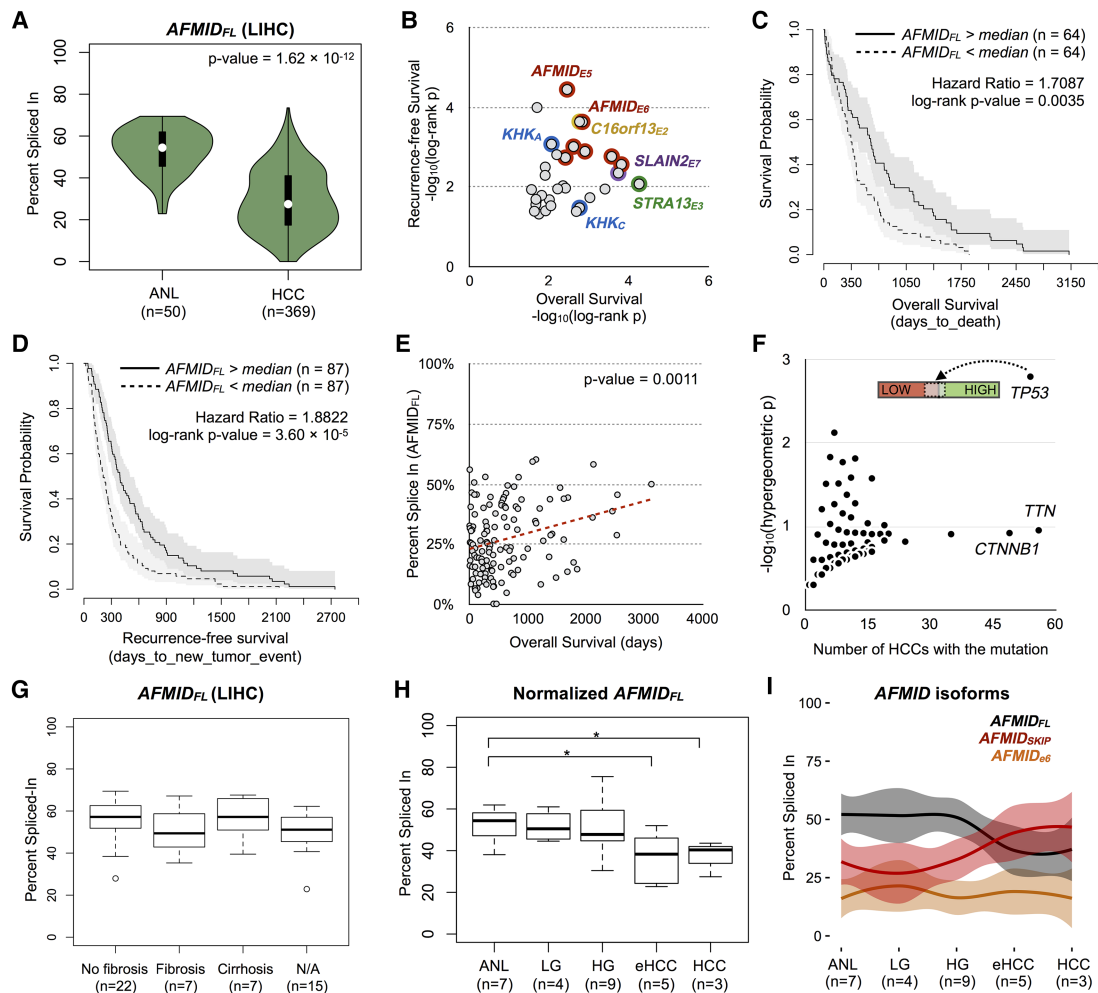
### The decrease of the *AFMID<sub>FL</sub>* isoform is associated with HCC patients' survival

Since losing cell identity can have biomarker potentials, we tested whether the PSI values of the alternative exons are correlated with the clinical information from the liver hepatocellular carcinoma (LIHC) data set of The Cancer Genome Atlas (TCGA) (The Cancer Genome Atlas Research Network et al. 2013). We used “days\_to\_death” as the overall survival, and “days\_to\_new\_tumor\_event\_after\_initial\_treatment” as the recurrence-free survival. We excluded all right-censoring data, which refers to patients for whom there is only the “days\_to\_last\_followup” information. This left us with 130 and 174 HCC patients with overall and recurrence-free survival information, respectively. We used the median of the PSI values to separate the HCC patients into high and low groups. For example, a high-*AFMID<sub>FL</sub>* HCC means an HCC sample whose PSI value for *AFMID<sub>FL</sub>* is higher than the median for the overall HCC samples.

From the 50 ANLs and 369 HCCs from the LIHC data set, we identified 250 splicing events (Supplemental Table S3). Seventy-nine of the 135 alternative exons identified above were present again in the set of 250 splicing events. *AFMID<sub>FL</sub>* also had decreased PSI values in HCCs from the LIHC data set (Fig. 2A). Survival analysis showed that 32 alternative exons had significant log-rank *P*-values in both overall and recurrence-free data sets (Fig. 2B; Supplemental Table S4). The top five genes were *AFMID*, *C16ORF13*, *SLAIN2*, *STRA13*, and *KHK*. The PSI values of exons 5 and 6 of *AFMID* had the strongest prognostic values in the RNA-seq data of TCGA (Fig. 2B). Patients with lower PSI values of *AFMID<sub>FL</sub>* died sooner (hazard ratio = 1.7087, *P* = 0.0035) and tended to have a recurrence earlier (hazard ratio = 1.8822, *P* = 3.60 × 10<sup>-5</sup>) (Fig. 2C,D). For predicting HCC recurrence, *AFMID* is similar to *MKI67* (encoding the proliferation marker Ki-67), which had a log-rank *P*-value of 3.80 × 10<sup>-5</sup>. Among the 369 HCCs from the LIHC data set, low-*MKI67* samples have significantly higher PSI values of *AFMID<sub>FL</sub>* (*P* = 3.71 × 10<sup>-7</sup>) (Supplemental Fig. S5). The overall survival was positively correlated with the PSI values of *AFMID<sub>FL</sub>* in HCC (*P* = 0.0011) (Fig. 2E). Sixty-three of the 64 low-*AFMID<sub>FL</sub>* patients died within 5 yr. The median survival for low-*AFMID<sub>FL</sub>* patients was ~11.77 mo (30 d per month), whereas for high-*AFMID<sub>FL</sub>* patients it was ~19.95 mo.

### The decrease of the *AFMID<sub>FL</sub>* isoform is associated with driver mutations

The major source of nicotinamide adenine dinucleotide (NAD<sup>+</sup>) production in the hepatocyte is through tryptophan metabolism. A recent study showed that inhibition of the de novo NAD<sup>+</sup> biosynthesis pathway leads to NAD<sup>+</sup> depletion, DNA-damage responses, and HCC development in mice (Tummala et al. 2014). Feeding the mutant mice with nicotinamide riboside (NR), the



**Figure 2.** The switch of *AFMID* isoforms is associated with patient outcome and is already evident in early-stage HCC. (A) The violin plot shows the PSI distribution of *AFMID<sub>FL</sub>* isoform in adjacent normal livers (ANLs) and HCCs in the LIHC data set of TCGA. (B) The dot plot summarizes the log-rank *P*-values of each alternative exon from overall (*x*-axis) and recurrence-free (*y*-axis) survival analysis, based on the LIHC data set of TCGA. Each circle represents one alternative exon. Alternative exons from the same gene are labeled with the same color. (C,D) The plots show the survival curves of *AFMID*-high and *AFMID*-low patients, based on the PSIs of the *AFMID<sub>FL</sub>* isoform in overall and recurrence-free survival analysis, respectively. The PSI of the *AFMID<sub>FL</sub>* isoform was determined by using the PSI of exon 5. (E) The dot plot shows the correlation between the PSI of the *AFMID<sub>FL</sub>* isoform and overall survival days. (F) The plot shows the enrichment (*y*-axis) of mutated genes in high-*AFMID<sub>FL</sub>* or low-*AFMID<sub>FL</sub>* HCCs, based on the LIHC data set of TCGA. The *x*-axis shows the number of HCCs carrying the mutated gene. Each dot represents one gene mutated in at least one of the HCC samples. The bar labeled *TP53* demonstrates that there are more *TP53*-mutated HCCs (light gray bar) in low-*AFMID<sub>FL</sub>* HCCs (red bar). (G) The box plots show the PSIs of the *AFMID<sub>FL</sub>* isoform in ANLs from patients with no fibrosis, fibrosis, or cirrhosis, in the LIHC data set. N/A means that the annotation of the liver sample was not available. (H) The box plot shows the PSI distribution of the *AFMID<sub>FL</sub>* isoform in ANLs, low-/high-grade dysplastic lesions (LG/HG), early HCCs (eHCC), and HCCs. (I) The plot shows the PSI patterns of *AFMID<sub>FL</sub>* (black), *AFMID<sub>SKIP</sub>* (red), and *AFMID<sub>e6</sub>* (orange) isoforms in different groups of liver samples.

precursor of the salvage pathway for generating NAD<sup>+</sup>, compensates for the loss of de novo NAD<sup>+</sup> biosynthesis and prevents HCC development (Tummala et al. 2014). The study also showed that *AFMID* protein is down-regulated or not detected in human HCCs by Western blotting, and depletion of *Afmid* in nontumorigenic mouse liver cells (AML-12) resulted in aggressive tumors (Tummala et al. 2014). If the switch of *AFMID* isoforms increases DNA-damage responses in normal hepatocytes, low-*AFMID<sub>FL</sub>* HCCs would have a higher chance of accumulating driver mutations, such as in *TP53* and *ARID1A* (Villanueva and Llovet 2014). To test this hypothesis, we used the nonsilent mutations. Among 369 HCC samples from TCGA, we found that 37 out of the 54 *TP53*-mutated HCC samples were enriched in low-*AFMID<sub>FL</sub>* HCC samples (hypergeometric *P* = 0.0016) (Fig. 2F;

Supplemental Table S5). In other words, low-*AFMID<sub>FL</sub>* HCC samples appear to have a twofold higher chance of gaining *TP53* mutations. Among 9762 genes mutated in at least one of the 369 HCC samples, only *TP53* had a *P*-value lower than 0.01. *TTN* and *CTNNB1* were mutated in a similar number of HCC samples, but the *P*-values were not significant. Incorporating silent mutations yielded the same enrichment for *TP53*. Forty of 61 HCC samples with *TP53* mutations were enriched in the low-*AFMID<sub>FL</sub>* group (*P* = 0.0016) (Supplemental Table S6). In addition, we tested whether the 1st quartile (Q1) and the 4th quartile (Q4) of HCCs are associated with nonsilent mutations in terms of PSI values of *AFMID<sub>FL</sub>*. Ranked by the PSI values of *AFMID<sub>FL</sub>*, 93 HCCs were in Q1 (PSI > 41%, high *AFMID<sub>FL</sub>*) and the other 93 HCCs were in Q4 (PSI < 17%, low *AFMID<sub>FL</sub>*). Twenty-six of the 186 HCCs had

*TP53* mutations, and 20 of them were enriched in Q4 ( $P = 0.0020$ ). Among the 6370 mutated genes in the 186 HCCs, *ARID1A* also showed significant enrichment in Q4 ( $P = 0.0289$ ) (Supplemental Table S7). Seven out of eight HCCs with *ARID1A* nonsilent mutations were enriched in Q4. Overall, there is a consistent enrichment of *TP53* mutations in low-*AFMID<sub>FL</sub>* HCCs.

### The switch of *AFMID* isoforms occurs in early-stage HCC

To establish when the switch of *AFMID* isoforms is likely to occur, we investigated the PSI distributions of *AFMID<sub>FL</sub>* in two data sets: (1) the LIHC data sets from TCGA; and (2) an RNA-seq data set that covers several stages in HCC development. First, among the 50 HCC patients with matched HCC samples from TCGA, 22 patients showed no fibrosis, seven showed fibrosis, and six showed cirrhosis. The PSI distributions of *AFMID<sub>FL</sub>* were not statistically different in the ANL samples (Fig. 2G), indicating that the switch of *AFMID* isoforms is not associated with fibrosis or cirrhosis. Next, we investigated the other RNA-seq data set from a recent study (Marquardt et al. 2014). This data set is composed of seven ANL samples, four low-grade dysplastic lesions, nine high-grade dysplastic lesions, five early HCCs, and three late HCCs. The sequencing depth for the samples in the RNA-seq data set ranges from 7 million to 339 million reads. This RNA-seq data set has particularly strong enrichment for reads in the 3' end. In general, exons near the 5' end tend to have much lower coverage. For example, in the 419 TCGA patient samples (50 ANLs and 369 HCCs), the PSI values of exon 5 and exon 9 had high correlation (correlation = 0.85) (Supplemental Fig. S6A). However, in the 28 patient samples, the PSI values of the two exons were weakly correlated (correlation = -0.16) (Supplemental Fig. S6B). The PSI values of exon 5 are generally much lower than the PSI values of exon 9. The imbalance in PSI values leads to inaccurate estimations of the *AFMID* isoform proportions if we use the PSI values of exon 5 alone to represent *AFMID<sub>FL</sub>*. Accordingly, we normalized the PSI of *AFMID<sub>FL</sub>* by using the average PSI of exon 5 and exon 9 in the second data set. After normalization, the PSI distributions of *AFMID<sub>FL</sub>* were not statistically different in ANL, low-grade, and high-grade dysplastic lesions. In contrast, the PSI values of *AFMID<sub>FL</sub>* were significantly lower in early HCCs ( $P = 0.0318$ ) and HCCs ( $P = 0.0356$ ) than in ANLs (Fig. 2H). Combining the three isoforms in one figure, we could show that the PSI values of *AFMID<sub>FL</sub>* and *AFMID<sub>SKIP</sub>* start to intersect at the early stage of HCC (Fig. 2I).

### The decrease of the *AFMID<sub>FL</sub>* isoform in other cancers

*AFMID* has highest expression in the liver, because of this organ's high demand for NAD<sup>+</sup> (400–800 μmol/kg protein) (Houtkooper et al. 2010). To investigate if cancers from other organs have the same switch, we analyzed RNA-seq data from 5213 noncancer samples from the GTEx portal and from 675 cancer cell lines originated from 31 tissue types (Fig. 3A; The GTEx Consortium 2015; Klijn et al. 2015). We found that all of the cancer cell lines expressed higher proportions of *AFMID<sub>SKIP</sub>* and *AFMID<sub>e6</sub>* isoforms (Fig. 3B). The 5213 noncancer samples generally had higher PSI values for the *AFMID<sub>FL</sub>* isoform, compared to their cancer cell line counterpart (Fig. 3C). Among the 15 matched tissue types from GTEx, liver and kidney had the largest decrease in PSI values, on average, and lung had the most significant  $P$ -value (Fig. 3D). On the other hand, brain had very little change in PSI values. We note that the non-HCC livers with a lower RNA integrity (RIN) score in the GTEx data set tend to have lower PSI values of *AFMID<sub>FL</sub>*. For example, 82 of the 136 non-HCC livers have RIN < 7 and signifi-

cantly lower PSIs ( $P = 9.0 \times 10^{-6}$ ). RNA integrity is probably one factor that influences the PSI estimation of *AFMID<sub>FL</sub>*, considering that *AFMID<sub>FL</sub>* is twice as long as the other alternative isoforms.

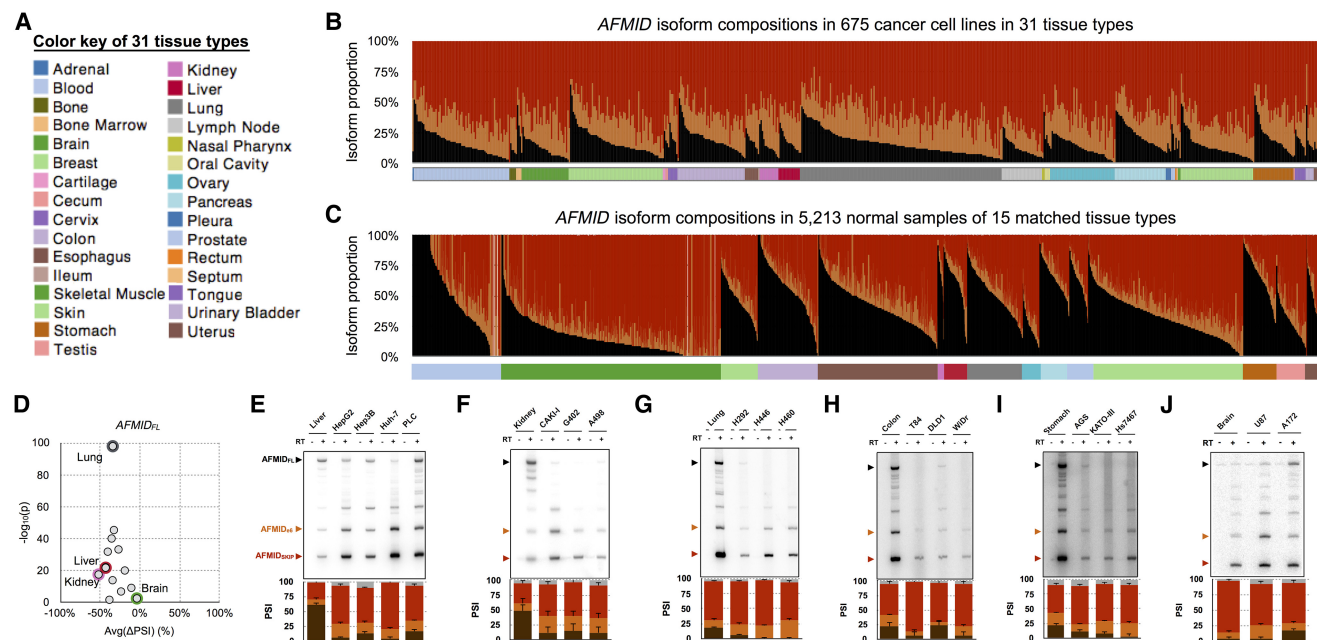
To experimentally validate the event in multiple cancer types, we performed radioactive RT-PCR of RNA from normal liver, kidney, lung, colon, stomach, and brain tissue samples versus their cancer cell line counterparts (Fig. 3E–J). We confirmed that *AFMID<sub>FL</sub>* is the dominant isoform in normal liver and kidney. We also confirmed that *AFMID<sub>FL</sub>* decreases and *AFMID<sub>SKIP</sub>* increases in most of the cancer cell lines tested, except for the brain cell lines (Fig. 3E–J). Real-time RT-PCR gave similar results. The overall *AFMID* expression levels were not significantly different between normal liver tissues and HepG2 and Hep3B cells, whereas Huh-7 and PLC/PRF/5 cell lines had significantly lower overall *AFMID* levels (Supplemental Fig. S7A). *AFMID<sub>FL</sub>* was generally lower in the four liver-cancer cell lines, whereas *AFMID<sub>e6</sub>* and *AFMID<sub>SKIP</sub>* were mostly unchanged (Supplemental Fig. S7A). Further RT-PCR analysis showed that the *AFMID<sub>SKIP</sub>* isoform is also the predominant isoform in human fetal liver (Supplemental Fig. S7B). The pattern is consistent with recent RNA-seq data from human fetal liver (Gerrard et al. 2016).

### The switch of *AFMID* isoforms is human-specific

*AFMID* is highly conserved (82 orthologs) and often paired with *TK1* in the genome of many species. It is not known if *AFMID* in other species is regulated by alternative splicing. Using an ultra-deep RNA-seq data set, we found zero supporting junction reads for *Afmid<sub>SKIP</sub>* and *Afmid<sub>e6</sub>* isoforms in fetal (E18), post-natal day 14 or day 28 (PN14 and PN28), and 3-mo-old adult (A3M) mouse liver samples (Bhate et al. 2015). Likewise, we did not find supporting junction reads for *Afmid<sub>SKIP</sub>* and *Afmid<sub>e6</sub>* isoforms in RNA-seq data from the *Mst1<sup>-/-</sup>; Mst2<sup>Flox/Flox</sup>* mouse HCC model (Fitamant et al. 2015). Using RNA-seq data from chimpanzee (*Pan troglodytes* and *Pan paniscus*), *Pongo pygmaeus*, *Macaca mulatta*, gorilla, mouse, and chicken, we again found that neither *Afmid<sub>SKIP</sub>* nor *Afmid<sub>e6</sub>* isoforms are expressed in liver, kidney, heart, muscle, and brain (Brawand et al. 2011; Barbosa-Morais et al. 2012). We conclude that *AFMID* splicing regulation is human-specific (Fig. 4A). Further radioactive RT-PCR showed no bands for the alternative isoforms in mouse liver and tumor samples (Fig. 4B). This analysis again confirmed that the alternative splicing regulation of *AFMID* isoforms is specific to human cells.

### Overexpression of *AFMID<sub>FL</sub>* isoform increases NAD<sup>+</sup> level, reduces DNA-damage response, and slows down cell growth in HepG2 cells

The mechanism of inhibiting de novo NAD<sup>+</sup> biosynthesis in the mouse liver was described in a recent study (Tummala et al. 2014). However, it was not known if increasing the *AFMID<sub>FL</sub>* isoform in human liver-cancer cells would affect the NAD<sup>+</sup> level and the response to DNA damage. Moreover, it was not known if this perturbation affects cell growth. To test the hypothesis that *AFMID<sub>FL</sub>* regulates NAD<sup>+</sup> levels in human cells, we expressed *AFMID<sub>FL</sub>* in HepG2 cells (Fig. 4C). Overexpression of *AFMID<sub>FL</sub>* reduced the total TP53 protein level by ~40% (Fig. 4D), suggesting a reduction in the DNA-damage response. The higher *AFMID<sub>FL</sub>* level increased NAD<sup>+</sup> levels by ~1.74-fold ( $P = 0.0458$ ) (Fig. 4E). Finally, overexpression of *AFMID<sub>FL</sub>* slowed down cell growth in HepG2 cells at day 7 ( $P = 0.0454$ ) and day 8 ( $P = 0.038$ ) (Fig. 4F).



**Figure 3.** The switch of *AFMID* isoforms in cancers. (A) The color key of 31 tissue types. (B) The stacked bar chart shows the proportions of *AFMID* isoforms (*AFMID*<sub>FL</sub>: black, *AFMID*<sub>SKIP</sub>: red, and *AFMID*<sub>eg</sub>: orange) in 675 cancer cell lines, based on RNA-seq analysis. (C) The stacked bar chart shows the proportion of *AFMID* isoforms in noncancer samples of 15 matched tissue types. (D) The dot plot shows the average  $\Delta$ PSI between noncancer samples and cancer cell lines from the same tissue type. The x-axis is the average  $\Delta$ PSI, and the y-axis shows the statistical significance ( $-\log_{10}[P\text{-value}]$ ) of the  $\Delta$ PSI. (E–J) The six panels show radioactive RT-PCR results of *AFMID* isoforms in normal tissues and cancer cell lines from liver, kidney, lung, colon, stomach, and brain, respectively (24 amplification cycles). The three arrowheads on the left indicate the expected sizes of *AFMID*<sub>FL</sub>, *AFMID*<sub>eg</sub>, and *AFMID*<sub>SKIP</sub> isoforms, respectively. The stacked bar chart below each RT-PCR plot shows the average PSI values and standard deviations from triplicate experiments. The PSI bars of *AFMID*<sub>FL</sub>, *AFMID*<sub>eg</sub>, *AFMID*<sub>SKIP</sub>, and unknown bands are colored in black, orange, red, and gray, respectively.

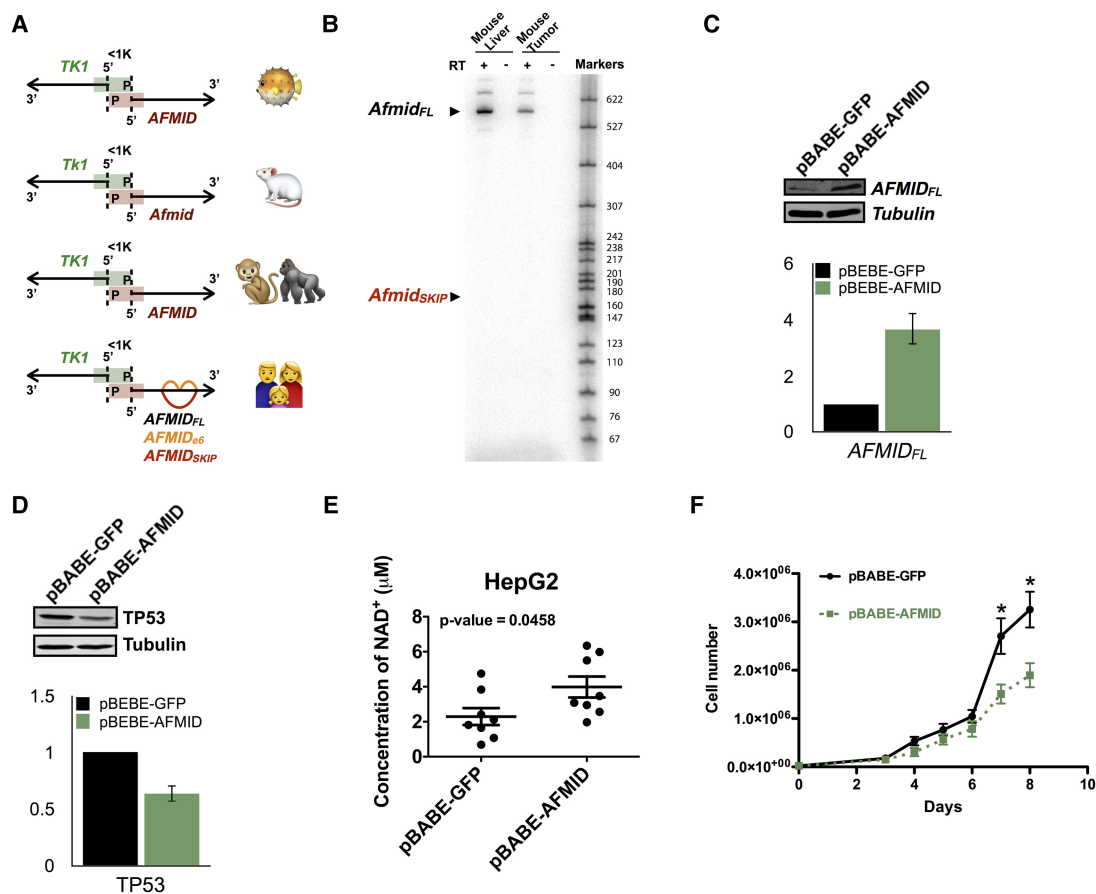
## Discussion

HCC's heterogeneity is a challenge for developing advances in prognosis and treatment (Friemel et al. 2015; Llovet et al. 2015). We tried to overcome this challenge by characterizing the splicing events in liver-cancer cells. We report that hepatocyte-specific splicing patterns have outstanding power in predicting HCC recurrence. Especially, the *AFMID* splicing event is associated with the presence of early driver mutations, such as mutated *TP53* and *ARID1A*. The switch of *AFMID* isoforms represents a new regulatory step in tryptophan/kynurenine metabolism and revealed the disruption of de novo  $\text{NAD}^+$  biosynthesis in hepatocytes in the early stages of HCC development. Low-*AFMID*<sub>FL</sub> HCCs tend to have a higher chance of carrying *TP53* mutations but not *CTNNB1* mutations. This is consistent with the current understanding that mutated *CTNNB1* is a later event (Friemel et al. 2015). Only the link between mutated *TP53* and the switch of *AFMID* isoforms was preserved during HCC evolution. Indeed, seven of 16 liver-cancer cell lines we analyzed lacked *TP53* mutations (Klijn et al. 2015). Therefore, the switch of *AFMID* isoforms can occur without *TP53* mutations. The switch may play an important role in early HCC evolution, because it increases the chance of accumulating driver mutations in HCC-initiating cells (Fig. 5A).

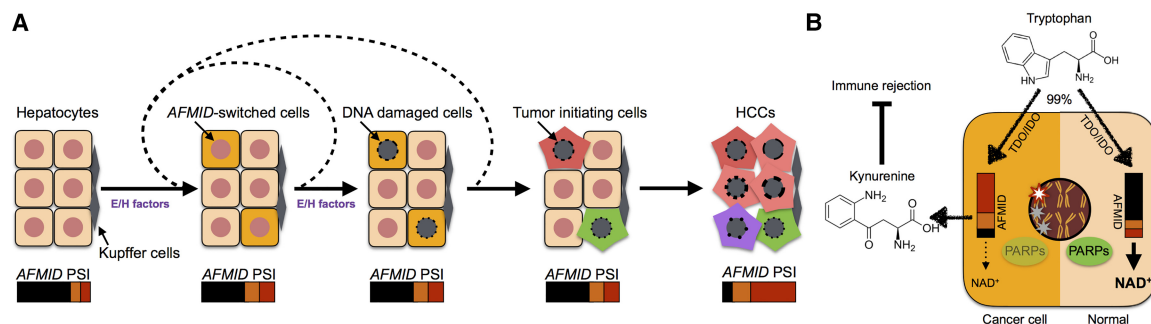
$\text{NAD}^+$  is a vital coenzyme in energy metabolism in eukaryotic cells (Houtkooper et al. 2010; Canto et al. 2015).  $\text{NAD}^+$  repletion increases life span in mice (Zhang et al. 2016a). However, the  $\text{NAD}^+/\text{NADH}$  ratio is very low in cancer cells; they maintain sufficient  $\text{NAD}^+$  for a high rate of glycolysis by converting pyruvate to lactate, while turning off other sources of  $\text{NAD}^+$  production (Vander Heiden and DeBerardinis 2017). For example, the switch

of *AFMID* isoforms impairs the major source of  $\text{NAD}^+$  production in hepatocytes. The switch may facilitate proliferation, but it also increases DNA-damage responses. For example, poly-(ADP-ribose) polymerase (PARP) and sirtuin are both  $\text{NAD}^+$ -dependent enzymes. PARP enzymes consume  $\text{NAD}^+$  to generate PAR polymers for repairing DNA. Sirtuin enzymes are associated with longevity, aging, and cancer (Herranz et al. 2010; Canto et al. 2015). Accordingly, the dysregulation of the de novo  $\text{NAD}^+$  pathway is a key event in HCC development. The switch of *AFMID* isoforms contributes to the accumulation of driver mutations and increases cancer susceptibility (Fig. 5B). Our discovery of the two human-specific isoforms (*AFMID*<sub>SKIP</sub> and *AFMID*<sub>eg</sub>) may lead to uncovering new mechanisms in tryptophan metabolism, as these are the predominant isoforms in cancer cells. Their roles in kynurenine secretion need to be further investigated. Switching the *AFMID*<sub>SKIP</sub> and *AFMID*<sub>eg</sub> isoforms back to *AFMID*<sub>FL</sub> may impact the secretion of kynurenine by redirecting the flux of tryptophan back to de novo  $\text{NAD}^+$  biosynthesis. This, in turn, may enhance  $\text{NAD}^+$  production and reduce immune escape of cancer cells (Fig. 5B). Also, modulating the splicing switch has potential implications for neurodegenerative diseases (Vecsei et al. 2013).

The new PSI index used in the present study is flexible and can compute PSI values of individual exons in complex splicing events. Previous single-exon PSI approaches were designed for simple splicing events with only one alternative exon, but they can be ambiguous in the case of mutually exclusive exons, multi-exon skipping, and more complex events. For example, we identified exon 5-6-7-8-9 and exon 5-7-8-9 skipping splicing events of *AFMID*. MISO cannot detect such splicing events because its database does not annotate multi-exon skipping events (Katz et al.



**Figure 4.** Phylogenetic comparisons of *AFMID* splicing patterns and functional experiments of the *AFMID<sub>FL</sub>* isoform. (A) The bidirectional gene pair, *TK1* and *AFMID*, is a conserved structure in Fugu, mouse, monkey, gorilla, and human (from top to bottom). The black arrows indicate the direction of transcription for each gene. The promoter region is indicated by green and red bars for *TK1* and *AFMID*, respectively. The distance between the two transcription start sites is <1 kb. In the human diagram, the red arc indicates the *AFMID<sub>SKIP</sub>* isoform, and the orange arcs represent the *AFMID<sub>e6</sub>* isoform. (B) Radioactive RT-PCR analysis of *AFMID* isoforms in mouse liver and tumor samples. Expected sizes of isoforms are labeled on the left. Size markers are shown on the right. (C, D) Representative Western blots at the top show (C) *AFMID<sub>FL</sub>* and (D) TP53 levels in HepG2 cells, respectively. HepG2 cells with a GFP plasmid are labeled pBABE-GFP, and HepG2 cells with an *AFMID<sub>FL</sub>* plasmid are labeled pBABE-AFMID. The bar charts below show the quantifications of Western blots from triplicate experiments. (E) The plot shows the NAD<sup>+</sup> levels in HepG2 cells with pBABE-GFP or pBABE-AFMID from eight biological replicates. (F) Cell growth curves of HepG2 cells with pBABE-GFP (black) or pBABE-AFMID (green) are shown. Error bars indicate SEM. (\*)  $P < 0.05$ .



**Figure 5.** Model of the *AFMID* isoform switch in HCC. (A) The flow chart represents HCC progression (left to right). At the start, six representative hepatocytes are shown (nuclei in red). Later, because of environmental or hereditary factors (E/H factors), a subset of hepatocytes switches *AFMID* isoforms. The environmental factors include WNT signals in peri-central hepatocytes in daily liver regeneration, cytokines released during inflammation, chemical damage, and virus infection. The hereditary factors include driver mutations, such as in *TP53* and *ARID1A*. The E/H factors temporarily disrupt the identity of hepatocytes and reduce the NAD<sup>+</sup> level in the hepatocytes. The reduced NAD<sup>+</sup> level gives rise to increased DNA damage in the nucleus. After recursively accumulating driver mutations, the DNA-damaged hepatocytes become HCC initiating cells. The switch of *AFMID* isoforms can then be observed by bulk RNA-seq. (B) The diagram shows two states of hepatocyte cells. On the right, a normal hepatocyte expresses high levels of *AFMID<sub>FL</sub>* isoform, and tryptophan can be converted to NAD<sup>+</sup> for PARPs to fix DNA damage. On the left, the HCC cell expresses low levels of *AFMID<sub>FL</sub>* and has low NAD<sup>+</sup>, so DNA damage is increased and kynurenine is secreted to inhibit immune rejection.

2010). Even if the exon 5-6-7-8-9 skipping event were annotated, MISO would not report the exon 5-7-8-9 skipping event because the splicing change can only be detected when all isoforms are considered together (Supplemental Fig. S1). Likewise, a recent paper using SUPPA did not report the splicing events of *AFMID* or *KHK* in the TCGA-LIHC data set (Sebestyen et al. 2016). In addition, our new PSI index reported an alternative first exon event of *PEMT*, which was missed by previous studies (Sebestyen et al. 2015, 2016; Zhang et al. 2016b). Because the new PSI index takes into account all the junction reads involving the 3' splice site of exon 4, the switch from *PEMT<sub>B</sub>* to *PEMT<sub>A1</sub>* can be detected (Supplemental Fig. S2). Related approaches, such as MAJIQ, have similar advantages and can detect these splicing changes (Vaquero-Garcia et al. 2016). Additional advantages of our new PSI index are that it can: (1) report PSI values and exon coordinates of individual exons in a multiple-exon skipping event, (2) amend the GC-content and 3'-end biases from RNA-seq by using the junction reads from both ends, and (3) detect the splicing changes from alternative exons that share the same upstream intron. Thus, the new PSI index provides more precise and suitable information for downstream analysis.

In summary, the present study provides the first integrative analysis of splicing events in liver cancer. We identified new splicing-based biomarkers in hepatocyte-enriched enzymes, such as *PEMT*, *KHK*, and *AFMID*. We found that *AFMID* alternative splicing constitutes a key event in liver carcinogenesis and a new switch in tryptophan/kynurenine metabolism.

## Methods

### A new PSI index

Traditionally, the PSI index is denoted as  $(a + b)/(a + b + 2c)$ , where  $a$  and  $b$  stand for the number of splice-junction reads connecting the alternative exon to the upstream and downstream constitutive exons, respectively (Barbosa-Morais et al. 2012), and  $c$  stands for the number of junction reads connecting the two constitutive exons. We modified the PSI index as follows:

$$\Psi = \frac{a + b}{\sum_{i=1}^n C_1 S_i + \sum_{j=1}^m C_2 S_j} \times 100\%$$

where  $C_1$  and  $C_2$  stand for the upstream and downstream constitutive exons, respectively.  $C_1 S_i$  stands for the total number of junction reads whose 5' splice site is connected to the upstream constitutive exon in a given splicing event. Similarly,  $C_2 S_j$  stands for the junction reads whose 3' splice site is connected to the downstream constitutive exon. Because the denominator is the sum of junction reads connecting to the flanking constitutive exons, the equation does not have the ambiguity for mutually exclusive exon events in which  $c$  might not exist. Also, in the new PSI equation,  $a$  and  $b$  stand for the number of junction reads connecting the alternative exon to its upstream and downstream exons, respectively;  $a$  and  $b$  do not necessarily reflect connections to  $C_1$  and  $C_2$  exons. For alternative splice site events, only  $C_1$  or  $C_2$  is used in the denominator, because the event only involves one constitutive exon.

In the present study, the alternative exons were identified based on the Ensembl 75 gene annotation. For a given alternative exon, each sample was required to have more than 20 supporting junction reads in the denominator of the PSI index. If fewer than 80% of the samples met the criteria, the splicing events were not considered as candidates for highly reproducible splicing events.

In addition, splicing events with lower than 10% PSI change or whose  $P$ -value was larger than 0.05 were also excluded.

### RNA-seq data processing

RNA-seq data sets were downloaded from several sources, such as Sequence Read Archive (SRA), European Genome-phenome Archive (EGA), TCGA, and GTEx. The present study used three different approaches to obtain the counts of junction reads (Supplemental Fig. S1). For data sets without alignment files or junction read information, we used STAR (2.4.1c) to obtain the counts of junction reads (Dobin et al. 2013). For data sets that already have alignment files or junction reads, we extracted the counts of junction reads either directly from the alignment files or from the file with junction read information. For example, the RNA-seq data set of the 11 primary HCCs and matched normal livers was downloaded from SRA and aligned by STAR. The RNA-seq data set of 675 cancer cell lines was downloaded from EGA and aligned by STAR. For TCGA's LIHC data set, we downloaded the alignment files from The Cancer Genomics Hub (<http://cghub.ucsc.edu>) and extracted the counts of junction reads from the alignment files. Recurrent HCCs (02A or 02B) in the LIHC data set were excluded. The counts of junction reads of the 5213 non-cancer samples were downloaded from the GTEx portal (<http://www.gtexportal.org/>). GRCh37 (hg19) and Ensembl 75 were the reference genome and gene annotation for human data sets, respectively. mm10 was the reference genome for mouse data sets. PPGY2 was the reference genome for *Pongo pygmaeus*. CHIMP2.1.4 was the reference genome for *Pan troglodytes* and *Pan paniscus*. MMUL1.0 was the reference genome for *Macaca mulatta*. gorGor3.1 was the reference genome for *Gorilla gorilla*. galGal4 was the reference genome for *Gallus gallus*. The pre-existing alignments and junction-count annotations from TCGA and GTEx were based on GRCh37 (hg19). To enable others to easily reproduce our findings, we used the same settings as in these two major data sources in our study. TCGA adopted the new reference genome GRCh38 (hg38) in 2016, and we verified that the PSI changes are reproducible.

### Statistical analysis

The two-sample  $t$ -test was used to elaborate the significance of PSI differences between noncancer and cancer cells. The log-rank test was used for survival analysis. The hypergeometric test was used for enrichment analysis of somatic mutations. Correlation testing was based on Pearson's product moment correlation coefficient. The adjustment method for  $P$ -values used the Bonferroni correction.

### RT-PCR

Total RNA was extracted from cell lines using TRIzol (Invitrogen). Genomic DNA was removed by treatment with DNase I (Promega). Reverse transcription of 0.5–1  $\mu$ g of total RNA was carried out using ImProm-II reverse transcriptase (Promega). Semiquantitative PCR in the presence of [ $\alpha$ -<sup>32</sup>P]-dCTP was performed with AmpliTaq polymerase (Applied Biosystems). The human-specific primer set (Forward: 5'-GGCCACCAGGAAGAGCCTGC-3', Reverse: 5'-CCTTCTGGGTCAGATTCTCAAC-3') was used to amplify endogenous *AFMID* transcripts; these primers anneal to exons 3 and 10. After 24 amplification cycles, the products were resolved using a 5% native polyacrylamide gel, and the resolved bands were visualized on a Typhoon 9410 phosphorimager (GE Healthcare). The signal intensities were quantified using ImageJ software (Schneider et al. 2012). Primer sequences are listed in Supplemental Table S8.

## Real-time PCR

One-half microgram of total RNA was extracted and reverse-transcribed as for RT-PCR. Complementary DNA (cDNA) was analyzed on a 7900HT Fast Real-Time PCR system (ThermoFisher Scientific). Fold changes were calculated using the  $\Delta\Delta C_q$  method and are reported as three biological replicates with three technical repeats each with  $\pm$ SEM. Real-time PCR results for HCC patient samples were obtained using a Bio-Rad system. For empirical validations, 20 ANLs and 19 HCCs from the Taiwan Liver Cancer Network were selected based on gender and cirrhosis status. These samples were used in accordance with the IRB procedures of Taipei Medical University. Primer sequences are listed in Supplemental Table S8.

## Overexpression of *AFMID<sub>FL</sub>*

cDNA for *AFMID<sub>FL</sub>* or *GFP* were synthesized and subcloned into the pBABE-Puro vector. HepG2 cells were grown in DMEM. HepG2 cells expressing *AFMID<sub>FL</sub>* or *GFP* were generated by retroviral transduction and puromycin selection (1.5  $\mu$ g/mL).

## Western blot analysis

Cells were washed with PBS and lysed in RIPA buffer (50 mM Tris pH 8, 150 mM NaCl, 1% [v/v] NP40, 0.5% [w/v] deoxycholate, 0.1% [w/v] SDS) supplemented with Complete Protease Inhibitor Cocktail tablets (Roche) and PhosSTOP (Roche). Thirty micrograms of total protein were loaded on a 12% SDS-polyacrylamide gel and transferred onto a nitrocellulose membrane (Millipore). For AFMID detection, the membrane was blocked in 1% (w/v) BSA in Tween 20-TBST (50 mM Tris pH 7.5, 150 mM NaCl, 0.05% [v/v] Tween 20). For TP53 detection, the membrane was blocked in 5% (w/v) milk in Tween 20-TBST. Blots were incubated with AFMID (Proteintech, 19522-1-AP), Tubulin (Sigma, T9026), or TP53 (Santa Cruz, sc-126) primary antibodies. IR-Dye 680 anti-mouse or IR-Dye 800 anti-rabbit immunoglobulin G (IgG) secondary antibodies (Licor) were used for infrared detection and quantification with an Odyssey Imaging System (Licor).

## NAD<sup>+</sup> assay

The concentration of NAD<sup>+</sup> was measured using an EnzyChrom NAD<sup>+</sup>/NADH Assay kit (BioAssay Systems) in accordance with the manufacturer's instructions. Briefly, 10<sup>6</sup> cells were homogenized using 100  $\mu$ L NAD extraction buffer at 60°C for 5 min. The extract was then neutralized with 100  $\mu$ L NADH extraction buffer and 20  $\mu$ L assay buffer, followed by centrifugation at 15,000 rpm for 5 min. Forty microliters of supernatant were then mixed with 80  $\mu$ L of working reagent solution (60  $\mu$ L assay buffer, 1  $\mu$ L Enzyme A, 1  $\mu$ L Enzyme B, 14  $\mu$ L lactate, and 14  $\mu$ L MTT). During the reaction, optical density at 565 nm was measured using a SpectraMax i3 plate reader (Molecular Devices) at 0- and 15-min time points, at room temperature. The concentration of NAD<sup>+</sup> was determined using a calibration curve with NAD standards, as per the manufacturer's instructions.

## Cell growth assay

Cells were seeded in 6-well plates at a density of 2  $\times$  10<sup>4</sup> cells/well. The cell number at the indicated time points in triplicate was measured with a Bright-Line hemocytometer (Hausser Scientific).

## Acknowledgments

K.-T.L., W.-K.M., J.S., and A.R.K. acknowledge support from National Cancer Institute grant CA13106. We thank M. Wigler for sharing cancer cell lines, and D. Tuveson and D. Fearon for

valuable suggestions. We thank the Taiwan Liver Cancer Network (TLCN) for providing the liver tissue samples.

**Author contributions:** K.-T.L. designed the study and performed the bioinformatics analysis. W.-K.M. performed functional assays. W.-K.M. and J.S. performed radioactive RT-PCR experiments. Y.-R.L. performed quantitative RT-PCR experiments for human HCC patient samples. K.-T.L. and A.R.K. wrote the manuscript with input from all co-authors.

## References

- Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**: 1587–1593.
- Bhate A, Parker DJ, Bebee TW, Ahn J, Arif W, Rashan EH, Chorghade S, Chau A, Lee JH, Anakk S, et al. 2015. ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. *Nat Commun* **6**: 8768.
- Bordeleau F, Califano JP, Negron Abril YL, Mason BN, LaValley DJ, Shin SJ, Weiss RS, Reinhart-King CA. 2015. Tissue stiffness regulates serine/arginine-rich protein-mediated splicing of the extra domain B-fibronectin isoform in tumors. *Proc Natl Acad Sci* **112**: 8314–8319.
- Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**: 343–348.
- The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**: 1113–1120.
- Canto C, Menzies KJ, Auwerx J. 2015. NAD<sup>+</sup> metabolism and the control of energy homeostasis: a balancing act between mitochondria and the nucleus. *Cell Metab* **22**: 31–53.
- Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, Becker A, Hoshino A, Mark MT, Molina H, et al. 2015. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* **17**: 816–826.
- Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R, Levanon EY. 2015. Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res* **43**: 5130–5144.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**: E359–E386.
- Fitamant J, Kottakis F, Benhamouche S, Tian HS, Chuvin N, Parachoniak CA, Nagle JM, Perera RM, Lapouge M, Deshpande V, et al. 2015. YAP inhibition restores hepatocyte differentiation in advanced HCC, leading to tumor regression. *Cell Rep* doi: 10.1016/j.celrep.2015.02.027.
- Friemel J, Rechsteiner M, Frick L, Bohm F, Struckmann K, Egger M, Moch H, Heikenwalder M, Weber A. 2015. Intratumor heterogeneity in hepatocellular carcinoma. *Clin Cancer Res* **21**: 1951–1961.
- Gerrard DT, Berry AA, Jennings RE, Piper Hanley K, Bobola N, Hanley NA. 2016. An integrative transcriptomic atlas of organogenesis in human embryos. *eLife* **5**: e15657.
- The GTEx Consortium. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**: 648–660.
- Herranz D, Munoz-Martin M, Canamero M, Mulero F, Martinez-Pastor B, Fernandez-Capetillo O, Serrano M. 2010. Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer. *Nat Commun* **1**: 3.
- Houtkooper RH, Canto C, Wanders RJ, Auwerx J. 2010. The secret life of NAD<sup>+</sup>: an old metabolite controlling new metabolic signaling pathways. *Endocr Rev* **31**: 194–223.
- Huang P, Zhang L, Gao Y, He Z, Yao D, Wu Z, Cen J, Chen X, Liu C, Hu Y, et al. 2014. Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. *Cell Stem Cell* **14**: 370–384.
- Jhunjunwala S, Jiang Z, Stawiski EW, Gnad F, Liu J, Mayba O, Du P, Diao J, Johnson S, Wong KF, et al. 2014. Diverse modes of genomic alteration in hepatocellular carcinoma. *Genome Biol* **15**: 436.
- Katz Y, Wang ET, Airolidi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**: 1009–1015.

- Klijin C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, Degenhardt J, Mayba O, Gnad F, Liu J, et al. 2015. A comprehensive transcriptional portrait of human cancer cell lines. *Nat Biotechnol* **33**: 306–312.
- Li X, Qian X, Peng LX, Jiang Y, Hawke DH, Zheng Y, Xia Y, Lee JH, Cote G, Wang H, et al. 2016. A splicing switch from ketohexokinase-C to ketohexokinase-A drives hepatocellular carcinoma formation. *Nat Cell Biol* **18**: 561–571.
- Llovet JM, Villanueva A, Lachenmayer A, Finn RS. 2015. Advances in targeted therapies for hepatocellular carcinoma in the genomic era. *Nat Rev Clin Oncol* **12**: 408–424.
- Lu Y, Xu W, Ji J, Feng D, Sourbier C, Yang Y, Qu J, Zeng Z, Wang C, Chang X, et al. 2015. Alternative splicing of the cell fate determinant Numb in hepatocellular carcinoma. *Hepatology* **62**: 1122–1131.
- Marquardt JU, Seo D, Andersen JB, Gillen MC, Kim MS, Conner EA, Galle PR, Factor VM, Park YN, Thorgeirsson SS. 2014. Sequential transcriptome analysis of human liver cancer indicates late stage acquisition of malignant traits. *J Hepatol* **60**: 346–353.
- Mirtschink P, Krishnan J, Grimm F, Sarre A, Horl M, Kayikci M, Fankhauser N, Christinat Y, Cortijo C, Feehan O, et al. 2015. HIF-driven *SF3B1* induces KHK-C to enforce fructolysis and heart disease. *Nature* **522**: 444–449.
- Pabarcus MK, Casida JE. 2002. Kynurenine formamidase: determination of primary structure and modeling-based prediction of tertiary structure and catalytic triad. *Biochim Biophys Acta* **1596**: 201–211.
- Pabarcus MK, Casida JE. 2005. Cloning, expression, and catalytic triad of recombinant arylformamidase. *Protein Expr Purif* **44**: 39–44.
- Pervouchine DD, Djebali S, Breschi A, Davis CA, Barja PP, Dobin A, Tanzer A, Lagarde J, Zaleski C, See LH, et al. 2015. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. *Nat Commun* **6**: 5903.
- Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671–675.
- Sebestyen E, Zawisza M, Eyraas E. 2015. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res* **43**: 1345–1356.
- Sebestyen E, Singh B, Minana B, Pages A, Mateo F, Pujana MA, Valcarcel J, Eyraas E. 2016. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res* **26**: 732–744.
- Shields DJ, Agellon LB, Vance DE. 2001. Structure, expression profile and alternative processing of the human phosphatidylethanolamine N-methyltransferase (*PEMT*) gene. *Biochim Biophys Acta* **1532**: 105–114.
- Tilgner H, Grubert F, Sharon D, Snyder MP. 2014. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc Natl Acad Sci* **111**: 9869–9874.
- Tummala KS, Gomes AL, Yilmaz M, Grana O, Bakiri L, Ruppen I, Ximenez-Embun P, Sheshappanavar V, Rodriguez-Justo M, Pisano DG, et al. 2014. Inhibition of de novo NAD<sup>+</sup> synthesis by oncogenic URI causes liver tumorigenesis through DNA damage. *Cancer Cell* **26**: 826–839.
- VanderHeiden MG, DeBerardinis RJ. 2017. Understanding the intersections between metabolism and cancer biology. *Cell* **168**: 657–669.
- Vaquero-Garcia J, Barrera A, Gazzara MR, Gonzalez-Vallinas J, Lahens NF, Hogenesch JB, Lynch KW, Barash Y. 2016. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife* **5**: e11752.
- Vecsei L, Szalardy L, Fulop F, Toldi J. 2013. Kynurenines in the CNS: recent advances and new questions. *Nat Rev Drug Discov* **12**: 64–82.
- Villanueva A, Llovet JM. 2014. Liver cancer in 2013: mutational landscape of HCC—the end of the beginning. *Nat Rev Clin Oncol* **11**: 73–74.
- Zhang H, Ryu D, Wu Y, Gariani K, Wang X, Luan P, D'Amico D, Ropelle ER, Lutolf MP, Aebersold R, et al. 2016a. NAD<sup>+</sup> repletion improves mitochondrial and stem cell function and enhances life span in mice. *Science* **352**: 1436–1443.
- Zhang L, Liu X, Zhang X, Chen R. 2016b. Identification of important long non-coding RNAs and highly recurrent aberrant alternative splicing events in hepatocellular carcinoma through integrative analysis of multiple RNA-Seq datasets. *Mol Genet Genomics* **291**: 1035–1051.

Received July 26, 2017; accepted in revised form January 24, 2018.