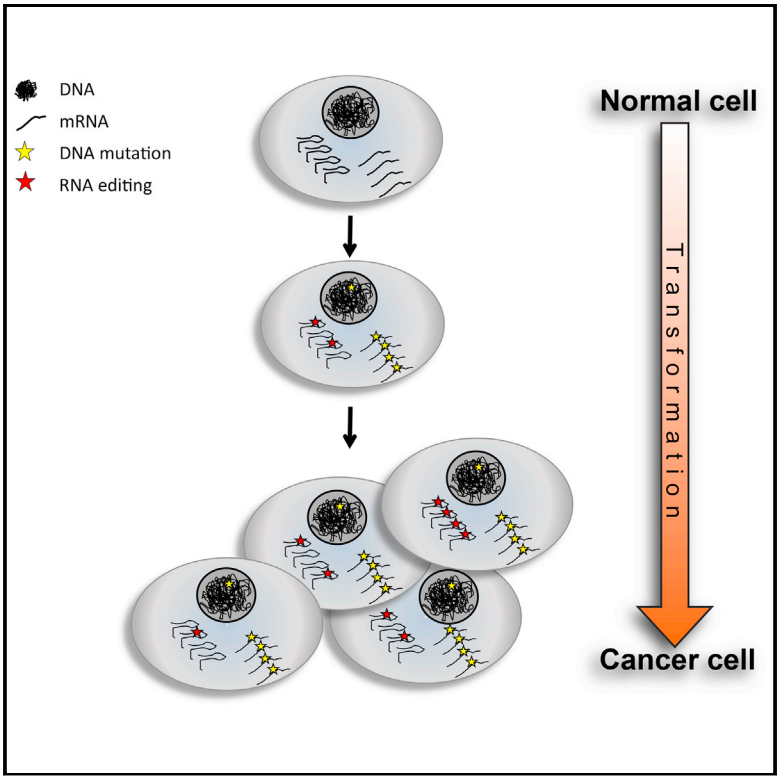
Report

Cell Reports

Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors

Graphical Abstract



Authors

Nurit Paz-Yaacov, Lily Bazak, Ilana Buchumenski, ..., Binyamin A. Knisbacher, Eli Eisenberg, Erez Y. Levanon

Correspondence

elieis@post.tau.ac.il (E.E.), erez.levanon@biu.ac.il (E.Y.L.)

In Brief

Paz-Yaacov et al. show that several types of cancer are accompanied by elevated activity of RNA editing, a process that changes the sequence of RNA from that encoded in the genome. Similar to genomic mutations, this mechanism results in multiple changes of the genetic information, which may be beneficial for cancer progression.

Highlights

- Level of A-to-I RNA editing by ADAR enzymes is elevated in various cancer types
- Extensive editing in cancer introduces RNA diversity or RNA mutations
- RNA modification events in tumors are as abundant as genomic DNA mutations
- Increased editing activity is associated with poor prognosis

Paz-Yaacov et al., 2015, Cell Reports *13*, 267–276
October 13, 2015
http://dx.doi.org/10.1016/j.celrep.2015.08.080

CellPress

Cell Reports
Report

Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors

Nurit Paz-Yaacov,[1] Lily Bazak,[1] Ilana Buchumenski,[1] Hagit T. Porath,[1] Miri Danan-Gotthold,[1] Binyamin A. Knisbacher,[1] Eli Eisenberg,[2,3,*] and Erez Y. Levanon[1,3,*]
[1]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 52900, Israel
[2]School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel
[3]Co-senior author
*Correspondence: elieis@post.tau.ac.il (E.E.), erez.levanon@biu.ac.il (E.Y.L.)
http://dx.doi.org/10.1016/j.celrep.2015.08.080

SUMMARY

Genomic mutations in key genes are known to drive tumorigenesis and have been the focus of much attention in recent years. However, genetic content also may change farther downstream. RNA editing alters the mRNA sequence from its genomic blueprint in a dynamic and flexible way. A few isolated cases of editing alterations in cancer have been reported previously. Here, we provide a transcriptome-wide characterization of RNA editing across hundreds of cancer samples from multiple cancer tissues, and we show that A-to-I editing and the enzymes mediating this modification are significantly altered, usually elevated, in most cancer types. Increased editing activity is found to be associated with patient survival. As is the case with somatic mutations in DNA, most of these newly introduced RNA mutations are likely passengers, but a few may serve as drivers that may be novel candidates for therapeutic and diagnostic purposes.

INTRODUCTION

Cancer is driven by alterations of the genomic information, mainly mutations in key genes that provide the cancerous cell a selective advantage for clonal multiplication. However, mutations in the DNA are not the only source for modifying the genomic content. RNA editing, a site-specific modification, alters the mRNA sequence from its genomic blueprint. RNA editing is catalyzed by the adenosine deaminase acting on RNA (ADAR) family of enzymes (Bass, 2002; Nishikura, 2010; Savva et al., 2012). This alteration results in dynamic RNA mutations, changes in the mRNA transcripts, which could ultimately lead to outcomes similar to those of genomic mutations. Unlike a genomic mutation, RNA editing affects varying fractions of the copies of the targeted transcript, leading to much higher flexibility. Alteration of editing was associated with cancer in some studies (Cenci et al., 2008; Chen et al., 2013; Han et al., 2014; Hu et al., 2015; Maas et al., 2001; Paz et al., 2007; Qin et al., 2014; Shoshan et al., 2015), but transcriptome-wide characterization of this modification across multiple cancer tissues has not been reported so far.

Here we show, using global measurements of RNA editing levels in hundreds of cancer samples, that adenosine to inosine conversion (A-to-I editing) and the enzymes mediating this modification are significantly altered in most cancer types screened, resulting in a sizable global effect on the transcriptome. In most tumor types, editing levels are elevated compared to their matched normal tissues, with the strongest signal detected in breast, thyroid, head and neck, and lung cancers. Overall, the number of RNA nucleotides modified by editing events in cancerous tissues is of the same order of magnitude as the genomic DNA mutation load. While the vast majority of editing events take place within *Alu* repeats (Athanasiadis et al., 2004; Bazak et al., 2014a; Blow et al., 2004; Kim et al., 2004; Levanon et al., 2004), we have found several non-*Alu* sites in coding sequences that were altered significantly in cancer. Our results suggest that editing may supplement genomic DNA alterations as a means to drive tumorigenesis. We hypothesize that classification of both DNA and RNA modifications is beneficial to determining a patient's profile and treatment. As is the case with cancer-associated somatic mutations, most RNA editing events are likely passengers and only a few may serve as drivers in each patient. Identifying these driver RNA editing sites may provide novel candidates for therapeutic and diagnostic purposes.

A-to-I RNA editing, mediated by the ADAR family of enzymes, is considered the most common RNA modification in mammals, with millions of editing sites detected so far in human (Bazak et al., 2014a). The majority of human editing sites lie in non-coding regions of the transcriptome, especially in *Alu* elements, which are likely to form long double-stranded RNA (dsRNA) structures, the optimal targets for ADARs (Lehmann and Bass, 1999). Deregulation of the RNA editing process in tumors, leading to an elevated level of modified RNA nucleotides, therefore may be analogous to genomic mutations.

Inosines are interpreted by the ribosome as guanosines. Thus, A-to-I editing of coding mRNAs by ADARs may lead to recoding, i.e., translation of a protein with amino acids different from those encoded in the genome. Recoding by RNA editing serves as a mechanism for creating structurally and functionally different isoforms of proteins from the same transcript. Hundreds of recoding sites have been identified in human (Bahn et al., 2012;

Li et al., 2009; Park et al., 2012; Peng et al., 2012; Ramaswami and Li, 2014; Ramaswami et al., 2013; Zhang and Xiao, 2015), but only a few dozen of them are conserved (Pinto et al., 2014; Xu and Zhang, 2014). However, the vast majority of editing activity in the human cell occurs at editing sites that do not result in recoding and reside in the primate-specific *Alu* elements. In a recent study (Bazak et al., 2014a), we quantified the relative ADAR activity in recoding sites and found that it consists of less than 1% of the global editing activity, which is mostly *Alu* editing. We therefore focus first on quantifying *Alu* editing in order to explore the global editing pattern in cancer tissues, as a means toward understanding the global differential editing in tumors.

RESULTS

Altered Editing in *Alu* Sites in Cancer and Matched Normal Tissues

We started by studying *Alu* editing as a proxy for estimating the global editing activity in a tissue. Employing an *Alu*-specific editing detection algorithm (see Experimental Procedures), we systematically screened A-to-I RNA editing alteration in 712 matched normal and cancerous RNA sequencing (RNA-seq) samples, originating from nine tissue types, using data from The Cancer Genome Atlas (TCGA) collection (Collisson et al., 2014; Davis et al., 2014; Cancer Genome Atlas Network, 2012a, 2012b; Cancer Genome Atlas Research Network, 2014a, 2014b). In six of the cancer types tested (all but kidney, prostate, and liver), we identified an elevated *Alu* editing activity in tumors compared to the matched normal tissues. Both the number of edited sites detected and the *Alu* editing index (AEI), representing the weighted average editing level across all expressed *Alu* sequences, were significantly increased (Figure 1A; Tables 1 and S1). There are typically thousands more editing sites detected in cancer tissues compared to their normal matched tissues. In comparison, the total number of genomic mutations ranges between hundreds and tens of thousands (Alexandrov et al., 2013). The same trend was seen using the complementary hyper-editing detection scheme (see Experimental Procedures and Figure 1C), specialized for identifying clusters of editing sites that are overlooked by standard alignment methods (Tables S1 and S2). As expected, the number of editing sites identified in the hyper-editing analysis correlated well with the AEI in each tissue (Figure 1D).

The accuracy of editing prediction is usually estimated by the ratio of A-to-G mismatches to the control mismatch (the most frequent mismatch identified excluding A-to-G), which represents the noise level. Both the *Alu* and the hyper-editing analyses showed a high signal-to-noise ratio (Figures 2A and 2C). In addition, one could observe the familiar ADAR motif signature for the A-to-G sites, but a random-like pattern in the control (Figures 2B and 2D). The fraction of detected sites that overlaps known SNPs and the distribution of sites between the read and transcript strands (Figure S1) also support the notion that the A-to-G mismatches are mainly due to A-to-I editing. Taken together, these results attest to the A-to-I editing origin of the observed A-to-G mismatches.

ADAR Expression Is Altered in Matched Normal and Tumor Tissues

The most obvious explanation for altered editing levels is a modified expression level of the ADAR enzymes. Indeed, the differential editing correlated well with the differential expression of ADARs in the TCGA tissue collection (Figure 1B; Table S3). ADAR1 levels were elevated in cancer for most tissues tested, except for colon and kidney for which, consistently, no elevation of editing was observed. On the other hand, expression of ADAR2, which is much less abundant in the tissues studied, showed a more complex pattern, consistent with the notion that ADAR1 is the main enzyme mediating *Alu* editing (Figure 1B; Table S3.

AEI Is Associated with Survival Rate in Head and Neck, Liver, and Breast Cancers

Importantly, editing alterations seem to be in association with clinical results. Kaplan-Meier curves were generated for all tissues, exhibiting that AEI (but not ADAR levels) could be used to classify patients for survival rate in liver cancer (Figure 3A), head and neck cancer (Figure 3B), and, to a lesser but significant extent, breast cancer (Figure S2A). Interestingly, the effect in liver cancer was gender dependent (Figure 3C): while the normal index level was similar in both genders, in the tumors males expressed high index values while females showed lower index values. Our dataset size did not allow us to examine whether AEI could be used to classify patients within genders. ADAR1 expression provided only a partial explanation for the gender differences (Figure 3D). Hepatocellular carcinoma is known to be gender biased, with females showing a longer survival rate (Buch et al., 2008; Dohmen et al., 2003). Elevated editing in AZIN1 recently was reported in hepatocellular carcinoma resulting in an oncogenic activity (Chen et al., 2013). In our limited set of data, we did observe a significant alteration of editing in the AZIN1 recoding site, but no correlation to the survival rate (Figure S2B).

Altered Editing in Recoding Sites

While most editing takes place in *Alu* repeats (Bazak et al., 2014a; Ramaswami and Li, 2014), the sites located in coding regions are more likely to have a functional role. Recently, elevated editing in AZIN1 recoding site was reported in hepatocellular carcinoma resulting in an oncogenic activity (Chen et al., 2013). Here we extend this analysis and search systematically for similar cases of a significant alteration of editing in recoding sites, resulting in dozens of additional candidate drivers.

We looked at all such sites documented in the RADAR database (Ramaswami and Li, 2014) and added all sites identified in our hyper-editing analysis. Focusing on sites in coding sequences and filtering out synonymous *Alu* sites, SNPs, and immunoglobulin sites resulted in a list of 9,484 sites (see Experimental Procedures).We looked for sites that exhibited statistically significant and sizable (>5% in absolute terms) difference in editing level between normal and tumor samples. Sixty such sites were identified in at least one cancer type (Table S4; Figure 4A). Note that, typically, the coverage provided by RNA-seq does not allow for effective and accurate quantification of editing levels, and thus the full scope of modified recoding sites

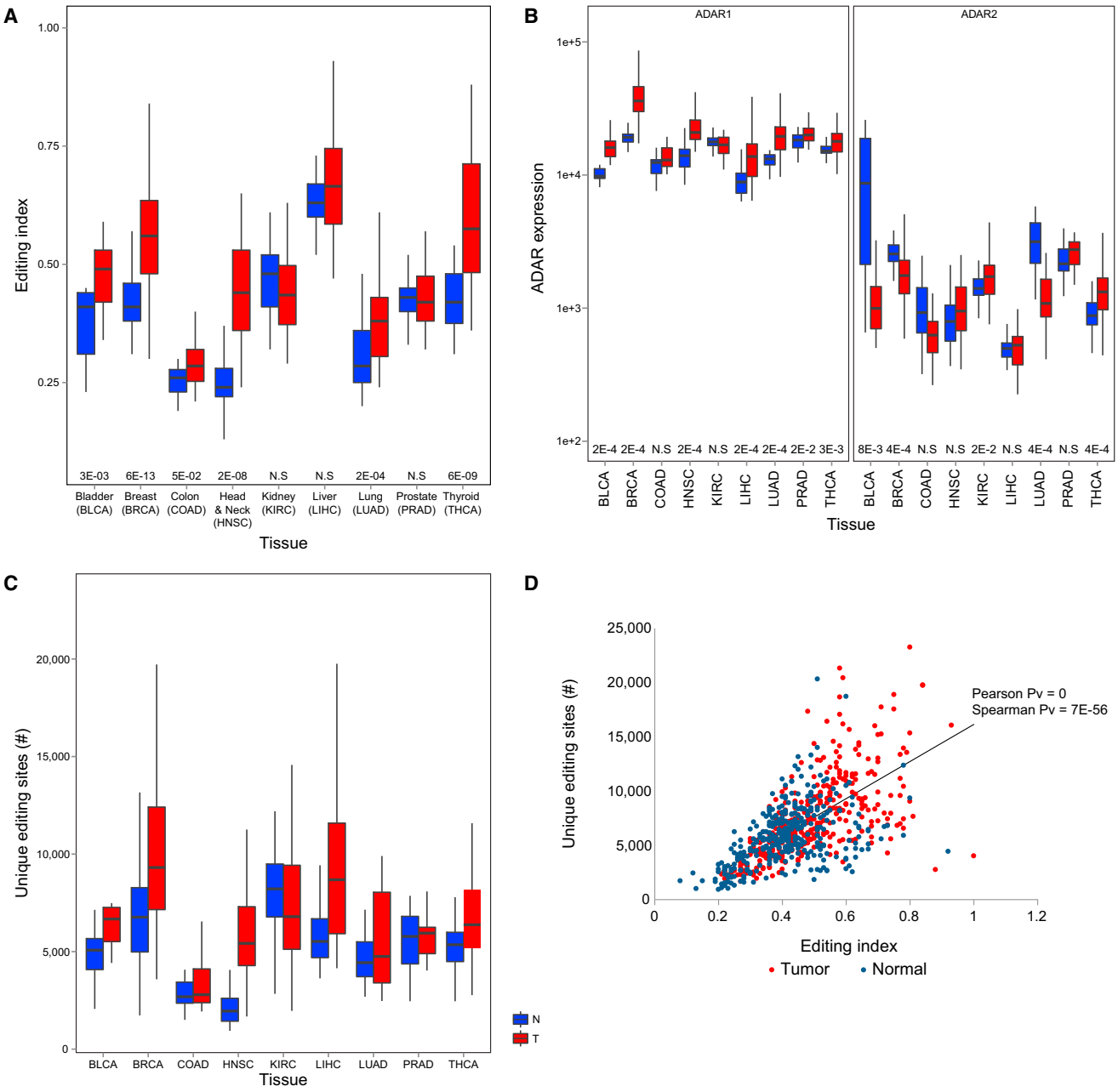


Figure 1. A-to-I RNA Editing Alteration in Cancer Versus Normal Tissues

(A) Editing is globally elevated in cancer. The distribution of *Alu* editing index (AEI) values in matched cancer and normal tissues is shown. The p values are shown (bottom) and were calculated using the one-sample Wilcoxon signed-rank test (for the index difference between matched samples) and FDR-corrected for the nine tests performed.

(B) ADAR1 expression levels are elevated in cancer. ADAR1 and ADAR2 expression level distributions in the matched normal and tumor tissues are shown. ADAR1 levels are about an order of magnitude higher than those of ADAR2 and are elevated in cancer for all tissues tested, except colon and kidney, for which, consistently, low or no elevation of editing was observed. In most tissues, but not in HNSC and THCA, the ADAR level may fully account for the observed AEI differences (Table S3).

(C) Hyper-editing screening of the TCGA data results in a higher number of editing sites in cancer tissues. Distributions of the number of sites detected by the hyper-editing screen for normal and tumor samples again show a higher number found in cancer, for most tissues.

(D) Correlation between the numbers of editing sites identified in the hyper-editing analysis and the AEI in each tissue. Red and blue colors refer to tumor and normal samples, respectively. Pearson and Spearman correlation p values are indicated. All sites that overlapped known SNPs were removed. The strong correlation between the independent editing measurements methods demonstrate that the two detection schemes work consistently.

Table 1. Editing Differences between Cancer and Normal Samples

	Detected Editing Sites		AEI		
Tissue Type	Average Normal	Average Tumor	Average Normal	Average Tumor	p Value (Index Difference)
BLCA (n = 13)	14,680	20,901	0.377	0.475	3.1E−03
BRCA (n = 95)	28,700	30,775	0.430	0.561	6.3E−13
COAD (n = 18)	10,143	12,847	0.257	0.293	0.046
HNSC (n = 29)	9,344	16,024	0.266	0.440	2.2E−08
KIRC (n = 62)	20,819	17,572	0.467	0.438	0.065
LIHC (n = 30)	13,587	20,619	0.631	0.663	0.24
LUAD (n = 36)	13,033	15,915	0.300	0.388	2.2E−04
PRAD (n = 31)	22,966	24,083	0.426	0.425	0.84
THCA (n = 42)	22,012	26,325	0.422	0.595	5.7E−09

The table presents, for editing sites in *Alu* sequences, the following: (1) the average number of detected editing sites identified in total, for normal and tumor tissues, (2) AEI value for normal and cancer tissues, and (3) statistical significance using the one-sample Wilcoxon signed-rank test to look at the distribution of AEI differences (AEI[tumor] − AEI[normal]) for the matched samples. FDR-corrected p values are presented. AEI, *Alu* editing index; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; COAD, colon adenocarcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; PRAD, prostate adenocarcinoma; THCA, thyroid carcinoma. See also Table S1.

might be larger than described here. Of the 60 sites, 14 exhibited consistent altered editing in at least three (of the nine) tissues. Interestingly, only 13 of the 52 non-synonymous sites (25%) were predicted to have a deleterious effect on the protein function (see Experimental Procedures and Table S4), compared to 53% deleterious events among all non-synonymous sites in our list (p = 0.0001; Table S4). This further supports the hypothesis of modified editing being a putative driver for tumorigenesis.

Altered Editing of MicroRNA Targets within *Alu* Elements

Many 3′ UTRs harbor microRNA (miRNA) targets, some of which reside within *Alu* sequences and might be modified by A-to-I editing. Alteration of miRNA regulation has been shown to be associated with cancer progression. We looked at all 222,778 editing sites within *Alu* repeats in 3′ UTRs that modify an miRNA target, and we applied the same procedure as was done for the sites in coding sequences. This resulted in 3,689 sites in 503 genes that showed significantly altered complementary seed sites (Table S5). Most of the sites, such as the targets at PPIA and METTL7A, were altered in multiple tissues.

DISCUSSION

Cancer is a complex and diverse disease affected by a multitude of genetic alterations in each patient. DNA mutations are recruited by cancer and passed on to all daughter cells, thus all tumor cells either have the mutation or not. In contrast, RNA editing levels span the whole range between 0% and 100%, and may change with time and cell state. These dynamic and flexible features of RNA editing may be utilized for the benefit of cancer progress, with varying levels of editing at each cell, possibly depending on tumor stage and genomic mutation background. We showed here that, in most cancers, editing is significantly elevated compared to the matched normal tissue. In a few cases, we also showed low editing index to be correlated with better survival. Although these changes mostly reflect *Alu* editing, they might turn out functional in several ways.

First, *Alu* editing might be relevant as a probe for the global editing activity. Note that direct detection of modified editing levels in specific recoding sites using RNA-seq is challenging. Typical coverage of a few dozen reads allows only for a rough estimate of the editing level, masking much of the alterations that might occur between different biological conditions. Thus, while we report here 60 significantly modified recoding sites, many more may be undetectable due to limited coverage. Moreover, as is the case with DNA mutations, additional sites might be seen in sub-populations, thus undetectable in the current study. *Alu* editing level is a much more sensitive and robust probe for the global editing machinery activity. Its elevation, probably by activation of the editing enzymes, indicates that the number of recoding sites and their editing levels may be elevated as well.

Second, altered editing in non-coding *Alu* repeats may result in a profound effect on the gene stability, partnering, interaction, and localization (Prasanth et al., 2005; Vitali and Scadden, 2010; Zhang and Carmichael, 2001). For example, editing of an *Alu* element was shown to downregulate the formation of circular RNAs (Ivanov et al., 2015), which in turn may increase the level of certain miRNAs, e.g., mir-7 that is implicated in many cancers (Hansen et al., 2013). Furthermore, many thousands of miRNA targets are located within *Alu* sequences (Liang and Landweber, 2007) or can be created due to editing (Borchert et al., 2009), and we have specifically verified that they also are affected by *Alu* editing alteration. Many of these targets are within known cancer-related genes and therapeutic targets such as DHFR and CTSB.

Finally, *Alu* editing recently was suggested to induce selective editing in recoding sites up to several hundred bases apart (Daniel et al., 2014). Thus, the major increase in *Alu* editing may result in induced recoding events in coding sequences located close to the *Alu* elements.

Taken together, the results presented here, as well as current results by other groups (Fumagalli et al., 2015; Han et al.,

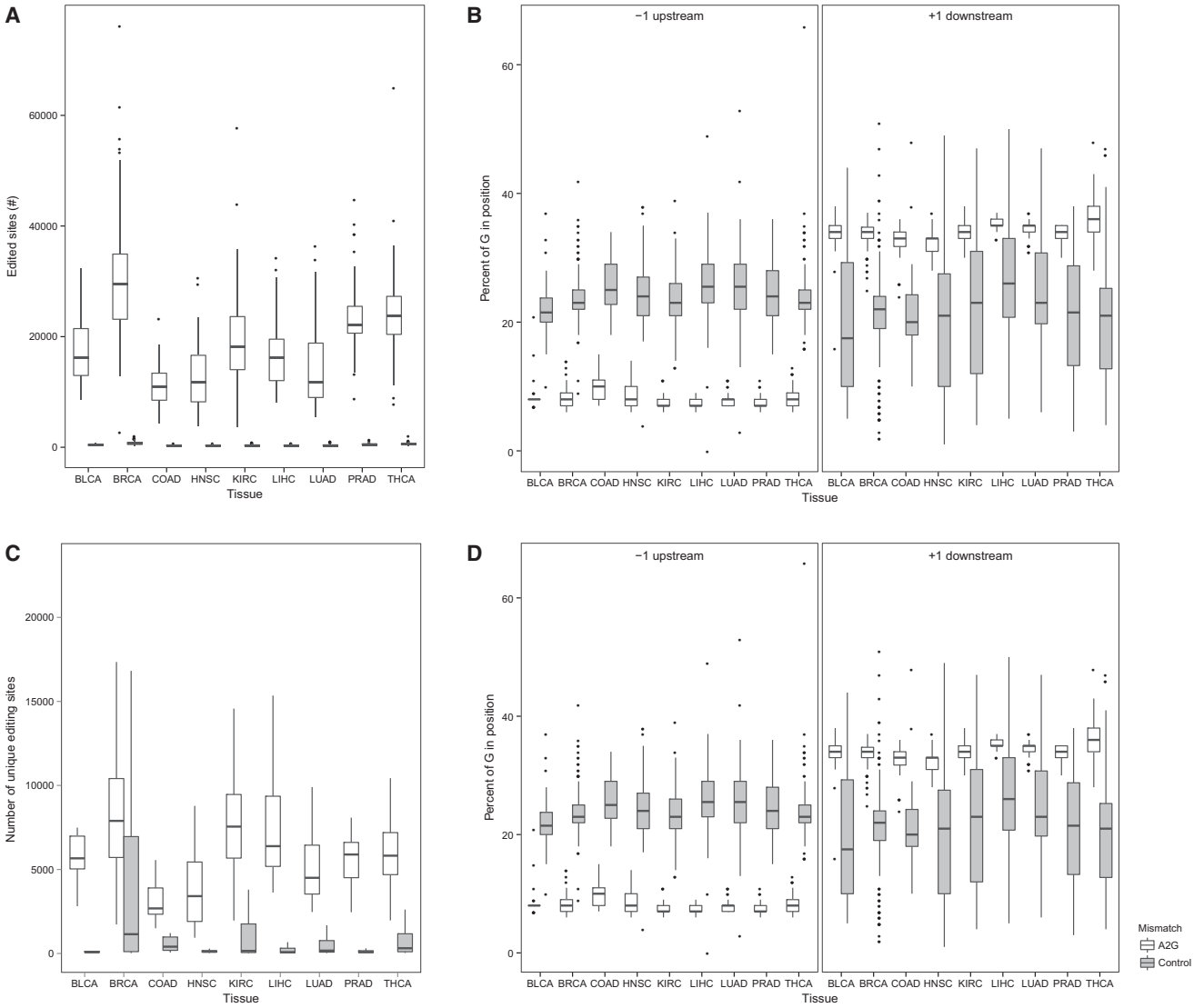


Figure 2. Editing Detection Scheme Yields a Clean Signal

(A) Distribution of the number of A-to-G mismatches within clusters in *Alu* repeats, over specific samples, compared to control (same for the second most frequent mismatch identified, G-to-A). The ratio of the number of control mismatches to that of A-to-G mismatches is an estimator of the noise level. Clearly, A-to-G mismatches are vastly over-represented, attesting to them being due to A-to-I editing rather than sequencing or alignment errors, SNPs, mutations, and so forth, and the noise level (estimated by the number of G-to-A mismatches) is rather low.

(B) ADAR-mediated editing is characterized by under-representation of G base in the nucleotide upstream to the edited site (−1) and over-representation of G base in +1 position from the edited site. The detected A-to-G sites, but not the control G-to-A mismatches, exhibit the familiar motif signature of A-to-I editing, as expected.

(C) Number of unique editing sites detected when applying the hyper-editing detection approach. Similar to the results in *Alu,* A-to-G mismatches are vastly over-represented.

(D) The detected sites in the hyper-editing analysis also show the familiar ADAR motif signature of A-to-I editing, while the control sites do not (see also Figure S1).

2015), demonstrate that, in addition to genomic mutations, reprogramming by RNA modifications is another source for transcriptome diversification in cancerous tissues. These correlative observations reflect ADAR1 overexpression in cancer, and it remains an open question as to whether a few of the resulting RNA mutations play a causative role in tumor progression. Future work, including larger numbers of sample tissues and functional assays, is required to explore this question.

RNA editing by adenosine deaminases is not the first reported case of recruitment of deaminase proteins for the creation of cancer mutations. Recently it was shown that APOBEC3B, a cytidine deaminase protein, can increase the rate of tumor genome evolution by introducing loads of C-to-U mutations into the DNA of cancer cells (Burns et al., 2013; Nik-Zainal et al., 2012). The amount of excess RNA editing events in cancer seems to be much higher than that of the reported cases of

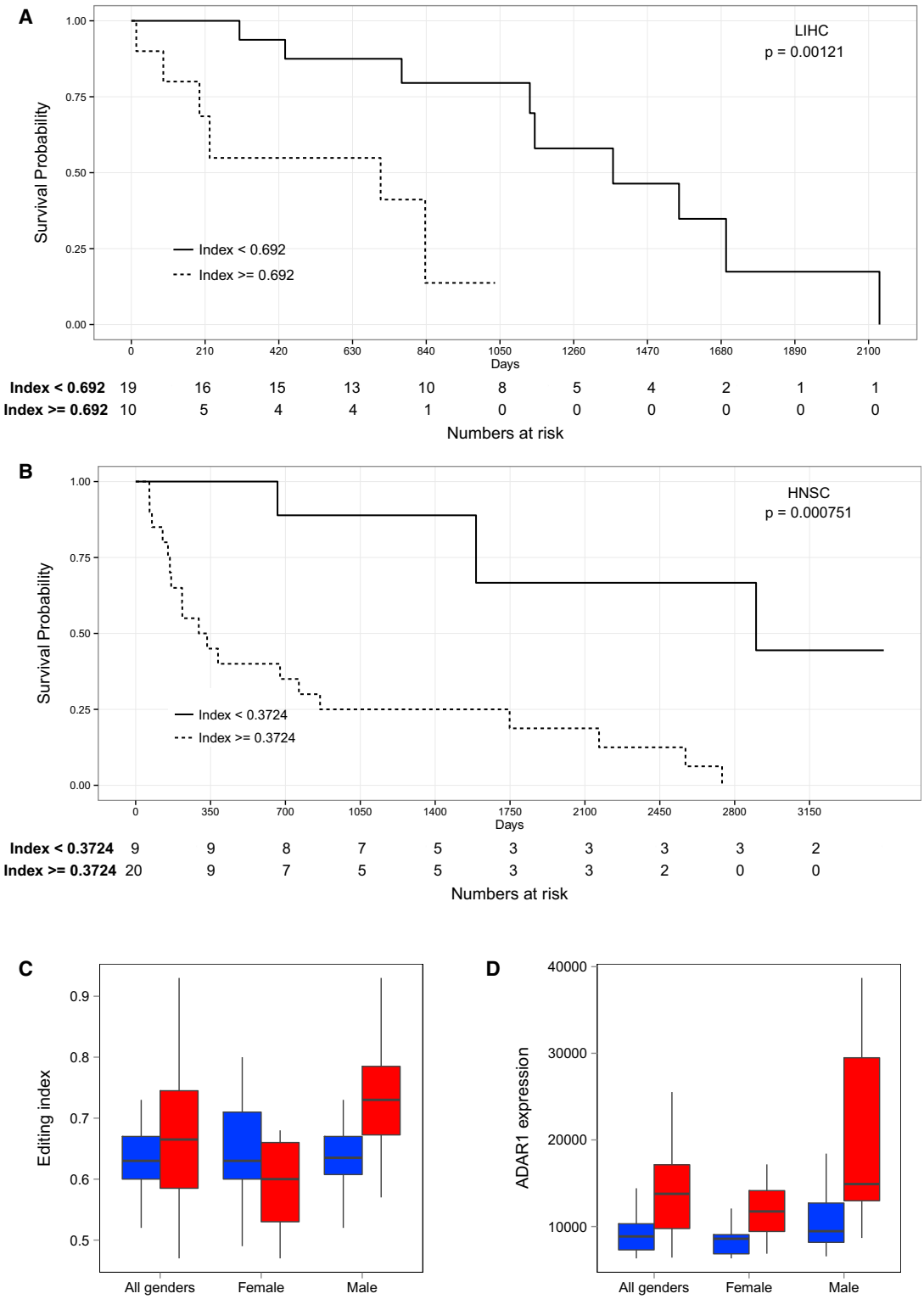


Figure 3. Kaplan-Meier Analysis of AEI

(A and B) Kaplan-Meier analysis demonstrates AEI is a good prognostic predictor in liver hepatocellular carcinoma (A) and head and neck tumors (B).

(C and D) AEI (C) and ADAR1 expression (D) are gender dependent in liver cancer tissues. Differences in ADAR expression do not fully account for the gender differences. Looking at each gender separately, a significant (opposite direction) effect is visible for both genders (all n = 30, female = 13, male = 16). The merged data for both genders do not show a clear trend in the editing index. Normal and tumor samples are indicated in blue and red, respectively (see also Figure S2).

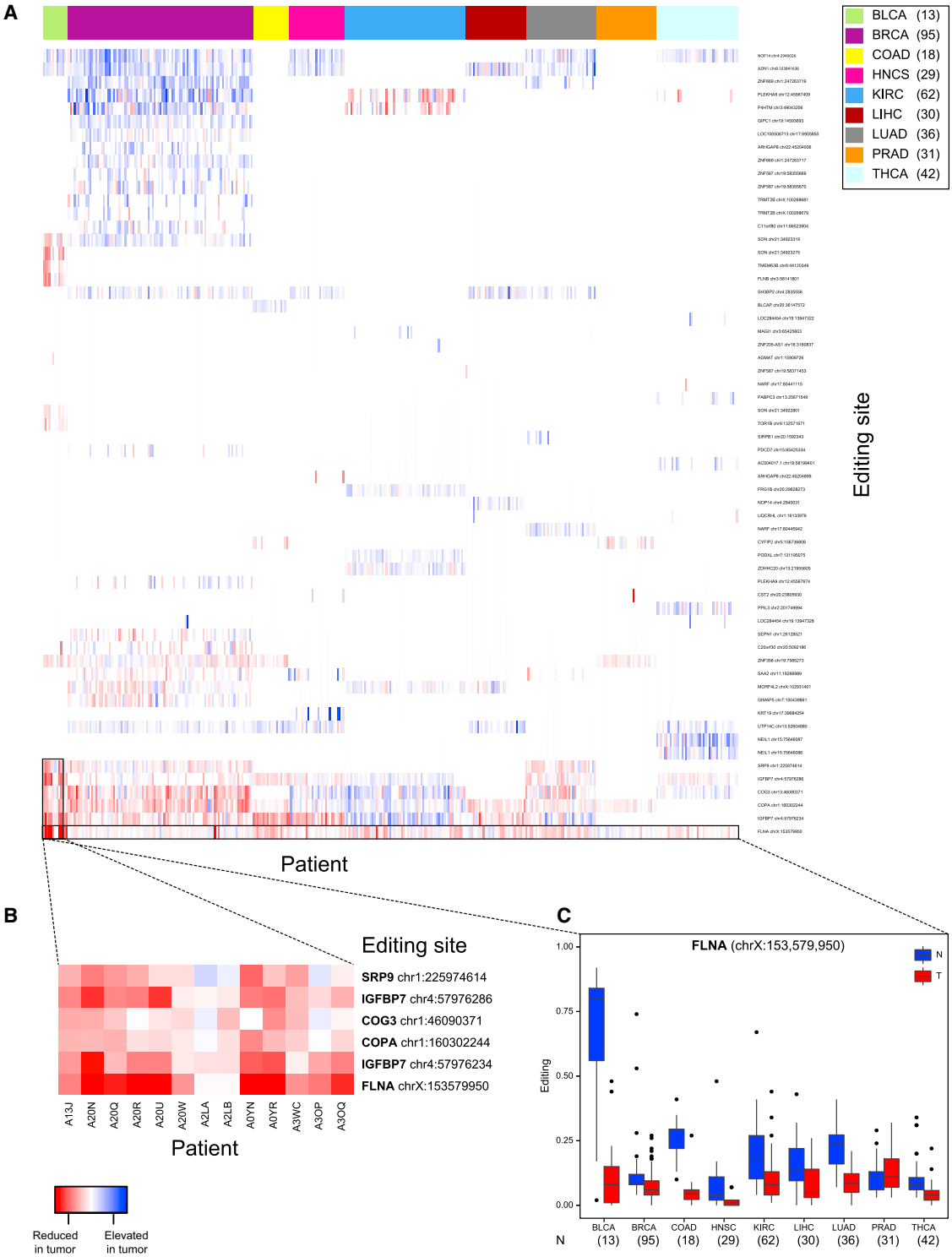


Figure 4. Pan-cancer Editing Alteration in Coding Sequences

(A) Each column represents one of the 356 patients, and each row one of the 60 sites that were significantly modified in at least one cancer type. The color in each position represents the difference in editing level per site per patient (blue, elevated editing in tumor; red, reduced in tumor). For example, the editing levels of the P4HTM site are reduced in most kidney tumors while elevated in most breast cancer samples. Patients are grouped by tissue type and the top is color-coded to indicate the tissue.

(B) Enlarged view of the bottom-left corner of (A). All bladder cancer samples and six editing sites are presented.

(C) Boxplot presentation of the distributions of editing levels in the FLNA editing site, for normal and cancer samples in each of the nine tissue types.

APOBEC3B. Actually, the number of additional RNA editing events in cancerous tissue, the RNA mutations, is comparable to the total number of somatic mutation events in the cancer's DNA. The heterogeneity introduced by this phenomenon, overlooked so far, may be recruited by tumor cells to play a role in cancer progression.

EXPERIMENTAL PROCEDURES

Detection of Editing in RNA-Seq Data

RNA-seq data were downloaded from TCGA collection (https://tcga-data.nci.nih.gov/tcga/) for tissues containing RNA-seq for both normal and tumor samples from the same patient. In total, 356 patients were analyzed (712 RNA-seq) for nine different cancer types as follows (number indicates the number of matched samples that were analyzed after passing quality test from each tissue): bladder urothelial carcinoma (BLCA, 13 patients), breast invasive carcinoma (BRCA, 95 patients), colon adenocarcinoma (COAD, 18 patients), head and neck squamous cell carcinoma (HNSC, 29 patients), kidney renal clear cell carcinoma (KIRC, 62 patients), liver hepatocellular carcinoma (LIHC, 30 patients), lung adenocarcinoma (LUAD, 36 patients), prostate adenocarcinoma (PRAD, 31 patients), and thyroid carcinoma (THCA, 42 patients).

Sequence reads were aligned using STAR (Dobin et al., 2013) to hg19 with parameters that accept only uniquely aligned reads (outFilterMultimapNmax = 1) and limit the number of mismatches to 0.05 of the mapped length (outFilterMismatchNoverLmax = 0.05). Then, three editing detection schemes were applied, as described subsequently.

Detection of Editing in *Alu* Elements

Over 99% of all editing activity in humans takes place in *Alu*. We thus used the *Alu* editing as a robust probe for the global editing activity. The magnitude of the effect and the distinctive features of editing in *Alu* repeats (e.g., clusters of sites) result in an extremely clean signal, with minimal false-positive rate (Bazak et al., 2014a).

To quantify *Alu* editing, we followed the approach described previously (Bazak et al., 2014b). We collected all mismatches between the aligned reads and the reference genome that occur within *Alu* elements, discarding mismatches in read positions with quality Phred score <30 and those located at sites reported as genomic SNPs in dbSNP (SNP build 135) (Eisenberg et al., 2005). We filtered out mismatches' positions that might be explained by sequencing errors, based on a probabilistic model assuming an a priori sequencing error rate of 0.001 (corresponding to Phred score Q = 30). Benjamini-Hochberg multiple testing correction was applied (for all *Alu* adenosines) with a false discovery rate (FDR) of 0.05, resulting in a list of putative reliable mismatches. We then looked for *Alu* elements for which a single type of mismatch dominated, i.e., *Alu* elements in which the number of mismatches of the most common mismatch type was higher than the number of mismatches of all other types combined (see Table S1). Virtually all such *Alu* elements were dominated by A-to-G mismatches and thus considered to be edited (Figure 2; Figure S2). Editing sites of the dominant type within these *Alu*s were considered to be editing sites. Editing sites that we detected showed the familiar ADAR sequence preference (Figure 2B; Eggington et al., 2011), tended to be more deeply covered than non-A-to-G sites, and the edited version was supported by more reads. Detected sites were uniformly distributed across read positions and (when occurring within RefSeq transcripts) conformed to the expressed strand (Figure S1), supporting the validity of the detection scheme.

AEI

Comparing editing levels in specific *Alu* sites requires ultra-high coverage, as virtually all adenosines within most *Alu* repeats are being edited to some extent, mostly to a low degree (<1%) (Bazak et al., 2014a, 2014b). Sites that happen to be detected in a specific sample are not necessarily stronger; most of them are a random sample of a small random fraction of the weakly edited sites. A more robust measure of the global editing level in a given sample is provided by the AEI. This index measures the averaged editing level across all *Alu* adenosines, weighted by their expression. It may be quantified by the ratio of the number of A-to-G mismatches (presumably due to inosines) to the total number of reads—nucleotides aligned to a genomic adenosine within an *Alu* repeat (representing edited and non-edited transcript adenosines). As previously described (Bazak et al., 2014b), this index averages over millions of adenosines and is, therefore, rather robust to statistical noise. Whenever segmental duplication data were available in TCGA, we verified that the differences in the AEI were not accounted for by duplications in the cancerous cells.

Hyper-editing in RNA-Seq Data

Traditionally, RNA editing is detected by comparing RNA sequences to their source DNA and searching for high-confidence A-to-G mismatches. We have shown recently that this approach misses many heavily edited reads (Carmi et al., 2011), which differ so widely from the corresponding DNA to the extent that standard schemes fail to align them properly. A newly devised pipeline allows for picking up the contribution of these heavily edited reads (Porath et al., 2014). In this approach, one transforms all As to Gs in both the unmapped reads and the reference genome and re-aligns them. Here, too, the sites we found showed the familiar ADAR sequence preference (Figure 2D), tended not to overlap known SNPs (unlike detected sites of non-A-to-G type) (Figure S1B), were uniformly distributed across read positions (data not shown), and (when occurring within RefSeq transcripts) conformed to the expressed strand (Figure S1C). Also, the results were even more clean than those observed with the AEI. We applied this pipeline to the TCGA dataset, resulting in further support of the trend seen using the AEI.

A-to-I Editing Sites in Coding Sequence

To analyze editing in the coding sequences, we compiled a list of editing sites to be tested for alteration in cancer. We took all sites documented in the RADAR database (Ramaswami and Li, 2014) in coding regions (UCSC annotation, 47,025 sites), supplemented by novel sites detected using the hyper-editing scheme within coding sequence (8,299 sites). We then excluded synonymous sites within *Alu* repeats, sites located in immunoglobulins (likely to be due to somatic hyper-mutations) and known SNPs (dbSNP 135). We did keep the following three sites: chr20, 36147572; chr5,156736808; and chrX,153579950 in the BLCAP, CYFIP2, and FLNA sites, respectively, which are known editing sites (Levanon et al., 2005). We further removed all sites annotated by wAnnovar (Chang and Wang, 2012) as residing in non-coding regions, resulting in 9,484 sites. We then quantified the editing levels in these sites using the REDItools script (Picardi and Pesole, 2013), trimming six bases at both ends of the reads.

For each of the sites, differences in editing level between normal and cancer tissues were evaluated using the χ^2 test followed by 5% FDR multiple-testing correction. Finally, we discarded sites where the absolute difference in the average editing level between normal and cancer samples was less than 5%.

Putative deleterious sites are defined as those sites for which at least two of wAnnovar analysis tools indicate a deleterious outcome. We found 3,520 such events among the 6,723 non-synonymous sites in our list.

ADAR Expression

The DESeq package (Anders and Huber, 2010) in R was used for differential gene expression analysis on normal and tumor samples (using UCSC gene annotation tables [Karolchik et al., 2003]). The number of reads aligned to each gene was calculated using featureCounts (Liao et al., 2014) with the same alignment used in the *Alu*-specific approach (detailed in Table S3).

Survival Analysis

Kaplan-Meier analysis was executed in R to identify differential survival rates of patients, classified by their tumor's editing index. Index values most significantly separating between high- and low-survival groups were selected for each tissue. Only patients with paired tumor and normal samples were included in the analysis. The p values were corrected for multiple testing over the nine cancer types considered, setting the FDR to 0.05.

miRNA Target Alterations in *Alu* within 3′UTRs

We have downloaded the human miRNA seed sequences from miRbase, and we compiled the list of perfect complementary matches to these seeds within *Alu* repeats in 3′ UTRs. We then looked for editing sites that modify these

targets, and we applied the same pipeline and filters as described for the sites in coding sequences to detect differentially edited sites.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.08.080.

AUTHOR CONTRIBUTIONS

N.P.-Y. designed the analysis, participated in each of the individual analyses, and supervised them. L.B. performed the *Alu* editing part, H.T.P. was in charge of the hyper-editing analysis, I.B. conducted the known editing sites screen, and M.D.-G. took care of ADAR expression level analysis. B.A.K. assisted with the analysis and the manuscript preparation. E.E. and E.Y.L. conceived of the project and designed the analysis. N.P.-Y., E.E., and E.Y.L. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Gilad Finkelshtein and Michal Barak for downloading the data, Gidi Rechavi for fruitful discussion, Moran Gal and Jasmine Jacob-Hirsch for assisting with functional analysis, and Orshay Gabay for graphical help. The results shown here are based on data generated by the TCGA Research Network (http://cancergenome.nih.gov/). This work was supported by the European Research Council (grant 311257), the Leon and Maria Taubenblatt Prize for Excellence in Medical Research, the I-CORE Program of the Planning and Budgeting Committee in Israel (grants 41/11 and 1796/12), and the Israel Science Foundation (1380/14 [E.Y.L.] and 379/12 [E.E.]).

Received: April 27, 2015
Revised: July 2, 2015
Accepted: August 28, 2015
Published: October 1, 2015

REFERENCES

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. *2*, e391.

Bahn, J.H., Lee, J.-H., Li, G., Greer, C., Peng, G., and Xiao, X. (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res. *22*, 142–150.

Bass, B.L. (2002). RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem. *71*, 817–846.

Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., Isaacs, F.J., Rechavi, G., Li, J.B., Eisenberg, E., and Levanon, E.Y. (2014a). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. *24*, 365–376.

Bazak, L., Levanon, E.Y., and Eisenberg, E. (2014b). Genome-wide analysis of Alu editability. Nucleic Acids Res. *42*, 6876–6884.

Blow, M., Futreal, P.A., Wooster, R., and Stratton, M.R. (2004). A survey of RNA editing in human brain. Genome Res. *14*, 2379–2387.

Borchert, G.M., Gilmore, B.L., Spengler, R.M., Xing, Y., Lanier, W., Bhattacharya, D., and Davidson, B.L. (2009). Adenosine deamination in human transcripts generates novel microRNA binding sites. Hum. Mol. Genet. *18*, 4801–4807.

Buch, S.C., Kondragunta, V., Branch, R.A., and Carr, B.I. (2008). Gender-based outcomes differences in unresectable hepatocellular carcinoma. Hepatol. Int. *2*, 95–101.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366–370.

Cancer Genome Atlas Network (2012a). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Cancer Genome Atlas Network (2012b). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Cancer Genome Atlas Research Network (2014a). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature *507*, 315–322.

Cancer Genome Atlas Research Network (2014b). Integrated genomic characterization of papillary thyroid carcinoma. Cell *159*, 676–690.

Carmi, S., Borukhov, I., and Levanon, E.Y. (2011). Identification of widespread ultra-edited human RNAs. PLoS Genet. *7*, e1002317.

Cenci, C., Barzotti, R., Galeano, F., Corbelli, S., Rota, R., Massimi, L., Di Rocco, C., O'Connell, M.A., and Gallo, A. (2008). Down-regulation of RNA editing in pediatric astrocytomas: ADAR2 editing activity inhibits cell migration and proliferation. J. Biol. Chem. *283*, 7251–7260.

Chang, X., and Wang, K. (2012). wANNOVAR: annotating genetic variants for personal genomes via the web. J. Med. Genet. *49*, 433–436.

Chen, L., Li, Y., Lin, C.H., Chan, T.H.M., Chow, R.K.K., Song, Y., Liu, M., Yuan, Y.-F., Fu, L., Kong, K.L., et al. (2013). Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. Nat. Med. *19*, 209–216.

Collisson, E.A., Campbell, J.D., Brooks, A.N., Berger, A.H., Lee, W., Chmielecki, J., Beer, D.G., Cope, L., Creighton, C.J., Danilova, L., et al.; Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550.

Daniel, C., Silberberg, G., Behm, M., and Öhman, M. (2014). Alu elements shape the primate transcriptome by cis-regulation of RNA editing. Genome Biol. *15*, R28.

Davis, C.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D., Shen, H., Buhay, C., Kang, H., Kim, S.C., Fahey, C.C., et al.; Cancer Genome Atlas Research Network (2014). The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell *26*, 319–330.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dohmen, K., Shigematsu, H., Irie, K., and Ishibashi, H. (2003). Longer survival in female than male with hepatocellular carcinoma. J. Gastroenterol. Hepatol. *18*, 267–272.

Eggington, J.M., Greene, T., and Bass, B.L. (2011). Predicting sites of ADAR editing in double-stranded RNA. Nat. Commun. *2*, 319.

Eisenberg, E., Adamsky, K., Cohen, L., Amariglio, N., Hirshberg, A., Rechavi, G., and Levanon, E.Y. (2005). Identification of RNA editing sites in the SNP database. Nucleic Acids Res. *33*, 4612–4617.

Fumagalli, D., Gacquer, D., Rothé, F., Lefort, A., Libert, F., Brown, D., Kheddoumi, N., Shlien, A., Konopka, T., Salgado, R., et al. (2015). Principles governing A-to-I RNA editing in the breast cancer transcriptome. Cell Rep. *13*, this issue, 277–289.

Han, S.-W., Kim, H.-P., Shin, J.-Y., Jeong, E.-G., Lee, W.-C., Kim, K.Y., Park, S.Y., Lee, D.-W., Won, J.-K., Jeong, S.-Y., et al. (2014). RNA editing in RHOQ promotes invasion potential in colorectal cancer. J. Exp. Med. *211*, 613–621.

Han, L., Diao, L., Yu, S., Xu, X., Li, J., Zhang, R., Yang, Y., Werner, H.M.J., Eterovic, A.K., Yuan, Y., et al. (2015). The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. Cancer Cell *28*, this issue, 515–528.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Hu, X., Wan, S., Ou, Y., Zhou, B., Zhu, J., Yi, X., Guan, Y., Jia, W., Liu, X., Wang, Q., et al. (2015). RNA over-editing of BLCAP contributes to

hepatocarcinogenesis identified by whole-genome and transcriptome sequencing. Cancer Lett. *357*, 510–519.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., and Rajewsky, N. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep. *10*, 170–177.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al.; University of California Santa Cruz (2003). The UCSC Genome Browser Database. Nucleic Acids Res. *31*, 51–54.

Kim, D.D.Y., Kim, T.T.Y., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res. *14*, 1719–1725.

Lehmann, K.A., and Bass, B.L. (1999). The importance of internal loops within RNA substrates of ADAR1. J. Mol. Biol. *291*, 1–13.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., et al. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. *22*, 1001–1005.

Levanon, E.Y., Hallegger, M., Kinar, Y., Shemesh, R., Djinovic-Carugo, K., Rechavi, G., Jantsch, M.F., and Eisenberg, E. (2005). Evolutionarily conserved human targets of adenosine to inosine RNA editing. Nucleic Acids Res. *33*, 1162–1168.

Li, J.B., Levanon, E.Y., Yoon, J.-K., Aach, J., Xie, B., Leproust, E., Zhang, K., Gao, Y., and Church, G.M. (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science *324*, 1210–1213.

Liang, H., and Landweber, L.F. (2007). Hypothesis: RNA editing of microRNA target sites in humans? RNA *13*, 463–467.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Maas, S., Patt, S., Schrey, M., and Rich, A. (2001). Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Proc. Natl. Acad. Sci. USA *98*, 14687–14692.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. *79*, 321–349.

Park, E., Williams, B., Wold, B.J., and Mortazavi, A. (2012). RNA editing in the human ENCODE RNA-seq data. Genome Res. *22*, 1626–1633.

Paz, N., Levanon, E.Y., Amariglio, N., Heimberger, A.B., Ram, Z., Constantini, S., Barbash, Z.S., Adamsky, K., Safran, M., Hirschberg, A., et al. (2007). Altered adenosine-to-inosine RNA editing in human cancer. Genome Res. *17*, 1586–1595.

Peng, Z., Cheng, Y., Tan, B.C.-M., Kang, L., Tian, Z., Zhu, Y., Zhang, W., Liang, Y., Hu, X., Tan, X., et al. (2012). Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat. Biotechnol. *30*, 253–260.

Picardi, E., and Pesole, G. (2013). REDItools: high-throughput RNA editing detection made easy. Bioinformatics *29*, 1813–1814.

Pinto, Y., Cohen, H.Y., and Levanon, E.Y. (2014). Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. Genome Biol. *15*, R5.

Porath, H.T., Carmi, S., and Levanon, E.Y. (2014). A genome-wide map of hyper-edited RNA reveals numerous new sites. Nat. Commun. *5*, 4726.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. Cell *123*, 249–263.

Qin, Y.-R., Qiao, J.-J., Chan, T.H.M., Zhu, Y.-H., Li, F.-F., Liu, H., Fei, J., Li, Y., Guan, X.-Y., and Chen, L. (2014). Adenosine-to-inosine RNA editing mediated by ADARs in esophageal squamous cell carcinoma. Cancer Res. *74*, 840–851.

Ramaswami, G., and Li, J.B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic Acids Res. *42*, D109–D113.

Ramaswami, G., Zhang, R., Piskol, R., Keegan, L.P., Deng, P., O'Connell, M.A., and Li, J.B. (2013). Identifying RNA editing sites using RNA sequencing data alone. Nat. Methods *10*, 128–132.

Savva, Y.A., Rieder, L.E., and Reenan, R.A. (2012). The ADAR protein family. Genome Biol. *13*, 252.

Shoshan, E., Mobley, A.K., Braeuer, R.R., Kamiya, T., Huang, L., Vasquez, M.E., Salameh, A., Lee, H.J., Kim, S.J., Ivan, C., et al. (2015). Reduced adenosine-to-inosine miR-455-5p editing promotes melanoma growth and metastasis. Nat. Cell Biol. *17*, 311–321.

Vitali, P., and Scadden, A.D.J. (2010). Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. Nat. Struct. Mol. Biol. *17*, 1043–1050.

Xu, G., and Zhang, J. (2014). Human coding RNA editing is generally nonadaptive. Proc. Natl. Acad. Sci. USA *111*, 3769–3774.

Zhang, Z., and Carmichael, G.G. (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell *106*, 465–475.

Zhang, Q., and Xiao, X. (2015). Genome sequence-independent identification of RNA editing sites. Nat. Methods *12*, 347–350.