Principles governing A-to-I RNA editing in the breast cancer transcriptome

Debora Fumagalli,[1,*] David Gacquer,[3,*] Françoise Rothé,[1,*] Anne Lefort,[3] Frederick Libert,[3] David Brown,[1] Naima Kheddoumi,[1] Adam Shlien,[8] Tomasz Konopka,[3] Roberto Salgado,[1] Denis Larsimont,[2] Kornelia Polyak,[6] Karen Willard-Gallo,[7] Christine Desmedt,[1] Martine Piccart,[1] Marc Abramowicz,[5] Peter J Campbell,[8] Christos Sotiriou,[1,*] Vincent Detours[3,4,*]

[1]Breast Cancer Translational Research Laboratory, Jules Bordet Institute, Université Libre de Bruxelles (ULB), Boulevard de Waterloo, 125 - 1000 Brussels, Belgium

[2]Department of Pathology, Jules Bordet Institute, Université Libre de Bruxelles (ULB), Boulevard de Waterloo, 125 - 1000 Brussels, Belgium

[3]IRIBHM, Université Libre de Bruxelles (ULB), [4]WELBIO, Route de Lennik, 808 - 1070 Brussels, Belgium

[5]Department of Genetics, Hôpital Erasme, Route de Lennik, 808 - 1070 Brussels, Belgium

[6]Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA

[7]Molecular Immunology Unit, Jules Bordet Institute, Université Libre de Bruxelles (ULB), Boulevard de Waterloo, 125 - 1000 Brussels, Belgium

[8]Cancer Genetics and Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK

*Contributed equally to this work.

Running title: A-to-I editing in the cancer transcriptome

Corresponding author:
Vincent Detours
IRIBHM, Université Libre de Bruxelles (ULB), WELBIO
Route de Lennik, 808
1070 Brussels, Belgium
Email Address: vdetours@ulb.ac.be
Phone: +32 2 555 4220
Fax: +32 2 555 4655

List of online supplementary material:

- Supplementary Methods, provided as a pdf file
- Supplementary Figures, provided as one pdf file
- Supplementary Tables, provided as a multitab xls file

ABSTRACT

A-to-I editing substitutes inosines for adenosines at specific positions in mRNAs and can substantially alter a cell's transcriptome. Currently, little is known about how this type of editing functions in cancer. This study demonstrates that A-to-I edited mRNA loci are conserved not only across matched normal and tumor breast tissues but also between patients; however, the editing frequency is higher in tumor compared to normal breast tissue. We developed a model that shows the editing enzyme ADAR uniformly controls the editing frequency across all loci, proportional to each individual locus' editability, which is a property of its surrounding nucleotide sequence. We also demonstrate that the type I interferon response and *ADAR* DNA copy number together determine ADAR expression in breast and potentially the vast majority of human cancers. ADAR silencing using shRNA lentivirus transduction in breast cancer cell lines led to a decrease in cell proliferation and an increase in apoptosis. Our results reveal that A-to-I editing is a pervasive but well controlled phenomenon in cancer that can profoundly influence the transcriptome and cellular functions in breast cancer.

INTRODUCTION

While intense effort is currently being dedicated to cancer genome sequencing, comparatively little attention has been devoted at understanding how faithful RNA sequences are to the DNA sequences from which they were derived. Messenger RNA (mRNA) is the target of a series of post-transcriptional modifications that can affect its structure and stability, one of the most relevant being RNA editing (1–3). The most common form of RNA editing in humans, the A-to-I type, is catalyzed by the adenosine deaminases that act on RNA (ADARs) family of enzymes, which bind double stranded RNA (dsRNA) and turn adenosines into inosines at precise positions (1, 3). Inosines are subsequently interpreted as guanosines by the cellular transcription machinery. ADAR enzymes are essential in mammals (4, 5) and exist in three forms: ADAR (also known as ADAR1), which is ubiquitous and has two isoforms—p110 is constitutive and p150 is inducible; ADARB1 (also known as ADAR2), principally expressed in the brain; and ADARB2 (also known as ADAR3), which contrary to ADAR and ADARB1 seems to be enzymatically inactive (6, 7).

A-to-I edits can profoundly influence cellular functions and regulations by altering mRNA splicing, localization and translation, and interfering with the binding of regulatory RNAs (8, 9). In addition to mRNA, ADAR can target non-coding RNAs such as micro RNAs (miRNAs), small-interfering RNAs (siRNAs) and long non-coding RNAs (lncRNAs), affecting both their structure and activities (10–13). A-to-I editing has been shown to occur predominantly in highly repetitive *Alu* sequences, likely because their frequency ($>10^6$) in the human genome make their arrangement in quasi-palindrome configurations prone to RNA duplex formation highly probable (2, 9, 14, 15). High-throughput sequencing studies suggest that tens of thousands of positions are targeted by A-to-I editing in the human transcriptome (16–19), and recent investigations report that potentially all adenosines in specific *Alu* repeats undergo A-to-I editing (15).

Currently, a limited number of studies on A-to-I RNA editing in cancer have been published with the findings pointing to a diversity of effects. For example, in brain cancer editing inhibits cell growth and is reduced in glioma (20, 21) and pediatric astrocytoma (22). In contrast, A-to-I editing increases during chronic myeloid leukemia progression (23). In hepatocellular carcinoma, A-to-I editing of the antizyme inhibitor 1 (AZIN1) increases and neutralizes a key inhibitor of the polyamine synthesis pathway, thereby promoting proliferation *in vitro* and increasing tumor initiation and volume in a mouse xenograft model (24). Both the magnitude and mechanisms regulating A-to-I editing in the majority of cancers, including breast cancer (BC), remain largely unknown.

The main objective of this study was to investigate and characterize the extent of A-to-I RNA editing in BC and define the principles governing the editing process in this as well as other types of cancer. We demonstrate that A-to-I edited mRNA loci are conserved both in matched normal and tumor breast tissues and across different patients. The editing frequency is, however, markedly increased at all editable loci in tumor compared to normal tissues. The number of edited loci and their editing frequency were highly correlated with ADAR expression. In breast and the majority of other cancers, ADAR expression appears to be regulated by two main factors: the type 1 interferon response and *ADAR* copy gains. Furthermore, ADAR silencing using shRNA lentivirus

transduction in breast cancer cell lines led to a decrease in cell proliferation and an increase in apoptosis. Our results show that A-to-I editing is a pervasive phenomenon in cancer that can drive aberrant transcriptome expression associated with a pro-proliferative and an anti-apoptotic phenotype in breast cancer suggesting its potential role during carcinogenesis.

METHODS

The methods are fully detailed in the Supplementary Methods available online. Briefly, the exome and transcriptome of 58 well characterized BC samples representing the four main known subtypes based on immunohistochemistry, namely TN, HER2+, luminal A and luminal B, and 10 matched normal samples were profiled using exome sequencing and RNA-Seq in paired-end mode on the Illumina HiSeq 2000 platform. Gene expression and SNPs profiles were obtained with Affymetrix® HG-U133 Plus 2.0 Array chips and Affymetrix® Genome-Wide Human SNP Arrays 6.0 for 57 and 49 tumor samples, respectively. RNA reads obtained from RNA-Seq were aligned simultaneously on the human genome and all known exonic junctions. Variant calls were submitted to a series of filters limiting artifact associated with RNA-seq. The identified RNA-DNA differences (RDDs) were validated in an independent cohort of 15 BC samples; moreover, few events as well as their edit frequency were validated using an independent technology (Roche FLX sequencer). The effect of interferon (IFN) on ADAR expression and editing was evaluated on 6 BC cell lines and one immortalized, non-transformed mammary epithelial cell line, MCF-10A. Cell lines were treated for 1, 2 or 5 days with IFN α, β, or γ. The effect of treatment on ADAR p110 and p150 protein and gene expression levels were evaluated quantifying the immunoblot signals and qRT-PCR, data respectively, while the effect of IFN treatment on editing distribution and frequency was investigated using amplicon sequencing (Roche FLX sequencer). In each sample, the mean editing frequency was correlated with clinico-pathological parameters and the expression of ADAR. The intracellular localization of ADAR was defined using IHC. The association between editing and *ADAR* amplification and/or a surrogate of interferon response (STAT1 expression) was evaluated in breast and 19 additional cancer types obtained from TCGA. Experimental assays and computational analyses are detailed in the Supplementary Methods.

RESULTS

Detection and validation of A-to-I editing sites in breast tissue

The extent of A-to-I RNA editing in BC was investigated by paired exome and transcriptome sequencing of a broad series of BC samples representing the major intrinsic subtypes including 17 triple negative (TN), 14 HER2-positive (HER2), 16 luminal A (LA) and 11 luminal B (LB) tumors (Supplementary Table S1). Paired exome and transcriptome sequencing of matched, tumor-adjacent normal tissue was performed on ten cases from this series. RNA-DNA single nucleotide differences (RDDs) were called as outlined in Supplementary Fig. S1 (details in Supplementary Methods).

Overall, we detected 16,027 RDDs in one or more samples, with all possible base changes represented (Fig. 1A). Among these, 560 RDDs were located in *Alu* regions and all were of the A-to-I type (Fig. 1A and Supplementary Table S2), consistent with the notion that A-to-I editing occurs predominantly in forward-facing *Alu* forming dsRNA duplexes processed by ADAR. Forty seven percent of the A-to-I *Alu* RDDs were present in the DARNED RNA editing site database (25). In contrast, only 2.5% of A-to-I, non-*Alu* RDDs and 0.6% of non A-to-I RDDs were found in the DARNED database (Fig. 1B).

Breast tissue is not well represented in the studies covered by the DARNED database. Since gene expression and RNA editing frequency (defined for each sample as the ratio of the number of RNA-seq reads documenting the non-reference base relative to the total number of reads covering the site) could be regulated in a tissue specific manner, we further validated our findings on an independent breast series. This independent validation series included 15 BC samples with paired transcriptome and full genome sequencing data from the Sanger Institute. The genomic coordinates of our putative RDDs and the coordinates of 1,000 random *Alu* positions were sent to the Sanger Institute without any additional information. This blind test confirmed 90% of the *Alu* RDDs, while only one of the 1,000 random *Alu* sites was detected in the validation series. Beyond *Alu*, overlap with the validation series was below 40% (Fig. 1B). Given the low confirmation rate of RDDs located outside of *Alu* regions in both the DARNED database and the independent validation series and that the majority of human editing events are A-to-I detected in *Alu* repeats (2, 9, 14, 15), our subsequent analyses focused exclusively on the subset of A-to-I RDDs located in *Alu* sequences. Because antizyme Inhibitor 1 (*AZIN1*) has also repeatedly been shown to be edited, this target was also included in our analyses (16–18, 24, 26–29).

To evaluate the accuracy of edited transcript frequencies measured in our full transcriptome data, we generated amplicons of the *AZIN1* editing site region for 36 samples that were then analyzed by an independent sequencing technology (Roche FLX sequencer). The edit frequencies measured from full transcriptome and amplicon sequencing were remarkably consistent (Fig. 1C and Supplementary Tables S2 and S3) and thereby validated the accuracy of these estimations.

Sequencing depth is a key factor in detecting single nucleotide variations (15), leading us to ask whether exome and RNA sequencing depths could influence the number of detectable *Alu* edit sites. As anticipated, while the number was not dependent on exome sequencing coverage, it did greatly increase with transcriptome coverage (Fig. 1D and E; Supplementary Table S4). A plateau was not reached in our dataset, which had a maximum coverage of $\sim 3\times10^7$ reads/sample. This suggests that with higher transcriptome coverage additional A-to-I editing sites should be detectable in the breast transcriptome.

More A-to-I editing was found in tumor compared to normal matched breast tissue

To determine whether A-to-I editing is specifically altered in BC, the mean editing frequencies across all edited sites were compared between matched normal and breast tissues for ten cases where paired exome and transcriptome sequencing data was available for the normal tissue. *AZIN1* has repeatedly been reported to be edited with a high editing frequency for *AZIN1* shown to promote carcinogenesis in hepatocellular cancer (24). Therefore, we compared the specific edit frequency of the *AZIN1* transcript determined by high-depth amplicon sequencing (Roche FLX sequencer) between tumor and matched normal breast tissues. The global mean editing frequency and the *AZIN1* specific editing frequency were higher in tumor compared to matched-normal breast tissues (Fig. 2A and B).

Lower epithelial cell content in normal compared to tumor breast tissue is a potential explanation for these observations. Thus, the site-averaged editing frequencies across all 560 *Alu* sites from the independent validation series (15 BCs) were compared to eight normal breast organoids (i.e. freshly isolated uncultured intact breast milk ducts). Editing was higher in tumor compared to normal cells (Fig. 2C), which validates our findings.

Global A-to-I editing is governed by simple principles

The general principles governing A-to-I editing in BC were investigated in multiple, matched exome-transcriptome data pairs. The ADAR family of enzymes catalyzes A-to-I editing, leading us to first determine their expression levels in normal and tumor breast tissues as well as their association with editing frequency using transcriptome sequencing data. ADAR was expressed 9-fold more than ADARB1 and >1,000-fold more than ADARB2 ($p<10^{-16}$, Supplementary Fig. S2A), which was anticipated because these last two isoforms are principally expressed in the brain. Moreover, while ADAR expression was higher in tumor compared to patient-matched normal breast tissues ($p=0.005$, Supplementary Fig. S2B), an inverse borderline-significant trend was observed for ADARB1 ($p=0.1$, Supplementary Fig. S2C).

The mean editing frequency (defined as the average editing frequency of all 560 *Alu* sites) was significantly positively correlated with ADAR mRNA expression levels (Fig. 3A; Supplementary Table S4), while it was weakly anti-correlated with ADARB1 expression levels (Supplementary Fig. S2D), as previously reported (24). The global

association detected between ADAR mRNA expression and the mean editing frequency was also observed at individual editing sites (Supplementary Fig. S3 and Supplementary Table S2). Considering both the high levels of ADAR mRNA expression and its strong correlation with the mean editing frequency found, our further analyses were focused on ADAR.

Editing site distribution across normal and BC tissues was investigated by plotting the maximum edit frequency for all editing sites against the number of samples where editing of these sites was detected (Fig. 3B). These two variables were highly correlated indicating that if a site was highly edited in one sample it was very likely to be edited in many other samples. This also suggested that the editing sites are conserved between normal and tumor tissues and among BC patients.

The editing frequencies of all 560 A-to-I *Alu* editing sites were ordered by site and sample using an increasing average editing frequency (Fig. 3C; negative controls in Supplementary Fig. S3), which revealed that high editing frequencies were present in the samples with more editing sites and high ADAR expression. Conversely, samples with lower ADAR expression had fewer edited sites, which were edited at lower frequencies. Taken together, these data suggest a simple quantitative model of A-to-I editing (Fig. 3D). In this model, turning-up the ADAR expression 'knob' leads to detectable editing at more sites and a proportionally increased editing frequency of all the editable sites. Conversely, when ADAR expression is low, editing is detectable at fewer sites and at a lower frequency.

Validation of the A-to-I editing model

We challenged this A-to-I editing model by inducing ADAR expression in four breast cell lines (three tumor and one normal tissue derived cell lines) with interferon alpha, a known ADAR inducer (30). The effect of inducing ADAR overexpression on the editing frequency of *AZIN1* and four of the most edited *Alu* regions in the discovery series was analyzed by amplicon sequencing (Roche FLX sequencer). These experiments demonstrated: First, that the same sites were edited in all cell lines (Supplementary Fig. S4 and Supplementary Table S5), including 90 of the 91 sites detected by whole transcriptome sequencing *in vivo*. Second, the editing frequency profiles were conserved across all cell lines (Supplementary Fig. S4). Third, increasing ADAR induction time increased editing frequencies at all edited positions (Fig. 4A and Supplementary Fig. S4). Fourth, increasing ADAR induction time and/or the depth of coverage increased the number of detected editing sites (Fig 4B). Due to deeper coverage (typically >1,000X for the Roche FLX sequencer) of the cell line amplicons, we identified 137 new sites in addition to the 90 in the discovery dataset, which suggests there are likely even more sites to identify in breast tissue.

We took advantage of the long reads (>300bp) and high coverage in the Roche FLX data to further validate our model by applying it to thousands of individual mRNA molecules transcribed from the same DNA region in the same individual. Focusing on one 256bp *Alu* region in one cell line, 65 of 68 adenosines potentially targeted by ADAR (Fig. 4C) were edited in at least one of the 2,842 mRNA molecules analyzed. The number

of edited positions per transcript was highly variable, ranging from 0 to 26 (38% of all adenosines). As expected, the sets of edited positions in "low-edited" mRNA molecules tended to be subsets of those edited in "high-edited" mRNA molecules. These findings further validate our A-to-I editing model. Nevertheless, the editing process had a strong stochastic component at the level of individual molecules. Interpreting editability as a probability of edition by ADAR reconciles the molecule-level randomness with the predictable editability observed at the level of transcripts populations.

Modeling editability

Beyond a dependency on ADAR mRNA levels, increases in editing frequencies at any given position may be proportional to an intrinsic, sequence-related property called 'editability'. Editability depends upon biophysical interactions between an individual site with its surrounding RNA sequence and partnering as a duplex with ADAR. Because editability is zero at the vast majority of positions, the activity of ADAR appears to be site-specific.

In principle, editability should be partially predictable from sequencing data, which led us to develop and validate a simple proof-of-principle statistical model for editability. The model relies on the notion that: first an edited site must be part of an RNA duplex (i.e. in a sequence with a nearby palindromic match) and second ADAR activity is dependent on a specific nucleotide sequence in the vicinity of the edited base (details in Supplementary Fig. S5 and Supplementary Methods). To build the model, we analyzed the edit frequencies of 51,621 edited *Alu* sites with ≥20X coverage from an independent sample sequenced at very high coverage (26). These sites were then ordered by genomic position. Half were used to fit a statistical model of edit frequencies based on DNA data alone (Supplementary Fig. S5). Editability scores were then computed for the other half, revealing that they are strongly associated with the observed edit frequencies (Fig. 4D). Our validated statistical model supports the notion that the edit frequency of an individual site in a given sample is controlled both by a site-specific factor, its editability, and a global factor, ADAR expression.

Association of ADAR expression, A-to-I editing and clinico-pathological variables

The relevance of ADAR expression to the A-to-I editing process led us to analyze its tissue and cellular localization by immunohistochemistry (IHC). Uniform ADAR expression was detected in cancer cells (Fig. 5A) but to a lesser extent in normal cells and tumor-infiltrating lymphocytes (TILs; see Fig. 5B). Moreover, ADAR staining was markedly stronger in nucleoli (Fig. 5C), in agreement with previous findings (31, 32).

To investigate the potential clinical impact of A-to-I editing, we determined whether the mean editing frequency was associated with tumor cell content (i.e. the proportion of malignant epithelial cells, adipose, stroma, normal epithelial cells and TILs) and/or well-established clinico-pathological parameters, including estrogen receptor, progesterone receptor, the proliferation marker Ki67, HER2 status, tumor size, nodal

status and histological grade. The mean editing frequency was positively correlated with the percentage of TILs (Spearman's correlation $\rho=0.3$, $p=0.02$), tumor size ($\rho=0.3$, $p=0.01$) and HER2 IHC staining ($\rho=0.3$, $p=0.01$; see Supplementary Fig. S6A-C and Supplementary Tables S1 and S4). Multivariate analysis of this dataset suggests that TILs and HER2 IHC are dependent variables in their association with editing frequency (Supplementary Fig. S6G).

To circumvent our limited sample size, correlations between these variables and ADAR expression were assessed in a large cohort of 787 BC patients with HER2 analyzed by IHC (33). TILs were not scored in this series so the level of Signal Transducer and Activator of Transcription 1 (*STAT1*) expression, a proxy for type I interferon response, was used instead. This independent BC series confirmed an association between ADAR and STAT1 expression but not for HER2 status or tumor size (Supplementary Fig. S6D-F). The lack of an association with estrogen receptor, Ki67 and HER2 indicates that ADAR expression is not correlated with a specific BC subtype beyond their link with adaptive immune response.

The interferon response and gains in *ADAR* copy number independently control A-to-I editing in cancer

The biological processes potentially associated with RNA editing were investigated by searching for genes whose expression had a strong positive correlation with the mean editing frequency (details in Supplementary Methods). Remarkably, 62 of the 85 genes identified were located on chromosome 1q ($p=10^{-66}$). Since *ADAR* is located on chromosome 1q, we next used SNP data to determine *ADAR* copy numbers in our samples. *ADAR* amplification was frequent in our series (44%) and correlated with high mean edit frequencies (Fig. 6A).

Chromosome 1q contains hundreds of genes and therefore its amplification could have a systemic impact on the BC transcriptome (33). Therefore, we further characterized the genes correlated with editing that were independent from 1q amplification. First, the microarray expression data were adjusted for 1q copy number to remove any potential confounding effects of *ADAR* amplification and then gene set analysis was performed (34) to identify canonical pathways associated with the mean edit frequency. The 13 significant pathway gene sets revealed by this analysis were all involved in interferon responses, interferon-related DNA and RNA sensing and lymphocyte biology (Supplementary Fig. S7A). We also investigated gene sets with shared transcription factor binding motifs between their promoters. The 7 significant gene sets identified were overwhelmingly related to *NFκB* and the interferon response, including the Interferon Response Factors *IRF1*, *IRF2* and *IRF7* (Supplementary Fig. S7B). To further investigate the relationship between interferon-related genes and ADAR expression, the median expression levels of STAT1 (Fig. 6B) and 389 type I interferon-inducible genes (Supplementary Fig. S7C) derived from 10 microarray studies (35) were measured. The expression of STAT1 and the 389 genes were positively associated with ADAR expression, suggesting that increased editing was part of a broader type I interferon response related to the chronic inflammatory state in cancer.

The respective roles of *ADAR* copy number and STAT1 expression (as a proxy for interferon response) in the A-to-I editing process were further defined using multivariate analysis to demonstrate that they are independently associated with *ADAR* expression (Fig. 6B-D). When STAT1 was a strong predictor of ADAR expression (Fig. 6B), it could be strengthened by adjusting ADAR expression for *ADAR* DNA copy number (Fig. 6C). The independent effect of type I interferon response and *ADAR* amplification was also supported by measuring the constitutive p110 and interferon-inducible p150 ADAR isoforms (Supplementary Fig. S8). STAT1 expression was more strongly correlated with p150 than p110 and conversely *ADAR* copy number more with p110 than p150.

While *ADAR* amplification is likely limited to malignant epithelial cells, the type-I interferon effect could be principally mediated by TILs. Seven breast cell lines (derived from the four principal BC molecular subtypes and normal breast) were treated with individual interferons (α, β, and γ) to determine if editing can be directly increased by interferon. ADAR p150 protein expression increased with all three interferons in all cell lines at each time point (Fig. 6E, Supplementary Fig. S9A), while p110 induction was weaker and less consistent. The moderate but significant correlation between p110 and STAT1 mRNA detected in primary tumors suggests that a small amount of p110 was induced (Supplementary Fig. S8F). The same four cell lines used to validate our A-to-I editing model were analyzed for p150 and p110 ADAR mRNA isoform expression levels, the editing proportion of AZIN1 and the four most edited *Alu* regions previously selected. The mRNA levels for p110 and p150 isoforms paralleled their protein expression (Supplementary Fig. S9B). Moreover, editing increased at all editable sites with all interferons in the four cell lines (Fig. 6F), although higher editing levels were observed at 2 or 5 days compared to untreated or 1 day. The induction of ADAR and editing was lowest for IFN-γ. These experiments confirm that type-I interferon response affect A-to-I mRNA editing in epithelial cells.

ADAR is involved in cell proliferation and apoptosis in breast cancer

As we have shown that ADAR expression and mean editing frequency were higher in breast tumors compared to matched normal tissues, we aimed to further investigate its role on cell proliferation, migration and apoptosis. To that aim, ADAR expression was stably knocked down in 3 different representative breast cancer cell lines (MDA-MB-231, MCF7 and BT474) using shRNA lentiviral particles (shRNA ADAR). The 3 cell lines were transduced with scramble shRNA lentiviral particles (shRNA control) as a negative control for the functional experiments. ADAR silencing was confirmed by Western blot analysis (Fig. 7A).

To assess the role of ADAR in cell proliferation, MTT assays were performed. These experiments showed that ADAR silencing led to a statistically significant decrease in cell proliferation (shRNA ADAR) compared to the control cells (shRNA control) in all cell lines (Fig. 7B). These results suggest that ADAR promotes cell proliferation. No significant effect of ADAR silencing was found regarding cell migration. The role of ADAR in apoptosis was investigated using Annexin V assays. ADAR silencing led to a statistically significant increase in cell apoptosis (shRNA ADAR) compared to the

control cells (shRNA control) in all cell lines (Fig. 7C) suggesting that ADAR may act as an anti-apoptotic factor.

The role of ADAR gains and interferon responses in other cancers

ADAR amplification is frequent in human cancers (Fig.8) and inflammatory responses are pervasive in this disease. This information led us to investigate whether these two factors were related to ADAR expression in 4,480 cancers from the The Cancer Genome Atlas (TCGA) (36), where sample-matched expression and copy number profiles are available. The representative analyses shown in Fig. 6B and 6C were reproducible across the TCGA dataset, which spans 20 types of cancer from 16 organs (Fig. 7). Overall, ADAR expression was consistently associated with both *ADAR* copy number and STAT1 expression. Similar to BC, adjusting ADAR expression for *ADAR* copy number increased the correlation between ADAR and STAT1 for all except pancreatic, kidney and thyroid tumors. The frequency of *ADAR* amplification was low in kidney and thyroid tumors, therefore correcting for *ADAR* copy number had a limited effect. These data suggest that ADAR expression could be principally driven by interferon in these two types of cancer. In most cancers, however, the editing process is driven by both type I interferon and *ADAR* copy number amplification. A correlation between ADAR copy number and ADAR expression has also been recently reported in esophageal cancer (29).

DISCUSSION

Next generation sequencing has fostered greater interest in RNA editing as a mechanism that governs additional and unexpected layers of genomic complexity. Recent studies have focused on defining the optimal bioinformatic strategies for identifying reliable RNA editing sites (16, 18, 26, 37, 38). Studies of the mammalian genome and transcriptome have shown that A-to-I editing, catalyzed by the ADAR family of enzymes, is the main form of RNA editing (1–3, 15). At first this phenomenon appeared to be restricted to a few targets located in coding regions; however, it is now clear that most A-to-I editing sites are located in non-coding regions. In particular, these sites tend to cluster in *Alu* regions, which due to their highly repetitive nature, are prone to form double-stranded RNA structures recognized by ADAR family enzymes (2, 9, 14, 15).

Initial studies found that the altered editing of specific targets were a determinant of tumor behavior, spawning efforts to define a potential role for A-to-I RNA editing in cancer initiation and development (20, 39, 40). Solid and hematologic cancers have been investigated with a few targets identified (24, 29); however, most of these studies performed an in depth analysis on small cohorts of patients. The magnitude of A-to-I editing in cancer as well as the mechanisms controlling and regulating the A-to-I editing machinery are currently unknown.

To address both points, we preformed the largest survey on RNA editing in cancer to date by profiling dozens of BCs with their matched healthy breast tissue. This was accomplished by using exome and transcriptome sequencing together with SNPs and

gene expression arrays in combination with state-of-the-art bioinformatics analyses. The increased sample size of this study opened a window to principles governing A-to-I editing that were previously out of reach. A significant finding from our study is the demonstration that the same sites are edited in normal and tumor breast tissues as well as in several BC cell lines. We further show that while the editing frequency profiles are conserved across tissues and BC cell lines, the frequency of editing was significantly higher in tumors compared to their matched normal breast tissues. High editing frequencies were also detected in samples with more detectable editing sites and/or where ADAR expression was high. These data provided the basis for our A-to-I editing model, where increases in ADAR expression acts on all editable positions in the transcriptome proportionally to their editability. This model was successfully validated on BC cell lines. ADAR appears to have site-specific activity, which is partly influenced by the biophysics of interactions between nucleotides in the surrounding RNA sequences and their duplex partnering with ADAR, a property that we define as editability.

Longer ADAR induction times and/or deeper sequencing coverage increased the number of editing sites detected. Interestingly, no plateau was reached at the depths we investigated, with up to 3×10^7 aligned reads per sample. A previous study made a similar observation using coverage up to 5×10^8 mRNA reads/sample, where no plateau was reached despite >140,000 A-to-I sites detected in the *Alu*'s (26). Differences in number of edited sites between the cited work and the present study could also be due to the cell type analyzed (a lymphoblastoid cell line vs. breast tissue/cell lines, respectively) and the DNA sequencing strategies (whole genome vs. coding sequences and their neighborhood). It is anticipated that a large number of additional A-to-I editing sites beyond those identified here remain to be discovered in BC. The data presented here clearly demonstrate that A-to-I editing is a pervasive phenomenon in cancer, and suggest that it is a major source of mRNA sequence variability in breast and potentially other types of cancer. The importance of this observation is that editing has the potential to significantly impact transcriptional regulation and cellular functions in tumor cells. Indeed, our in vitro studies have shown that ADAR silencing decreases cell proliferation and promotes apoptosis supporting the potential carcinogenic role of ADAR and consequently A-to-I editing in breast cancer.

This study is the first to identify two key factors independently determining ADAR expression and thereby A-to-I RNA editing in breast and other human cancers: first the type-I interferon response in tumors, and second *ADAR* copy number gains. The mean editing frequency was found to be significantly correlated with the expression of STAT1 and type I interferon target genes, both in our patient series and a large pool of BC datasets. This observation suggests that the immune response profoundly impacts the transcriptome sequence of tumor cells and thereby influences the internal mechanisms that govern their behavior.

Recent work has revealed that aberrant expression of ADAR and APOBEC family proteins occurs in many human diseases, including cancer. In breast and other tumor types, mutational signatures are associated with APOBEC family proteins (41) with evidence that APOBEC-mediated mutagenesis is highly active in human cancers (42, 43). Although, the role of ADARs and RNA editing in human disease is just beginning to emerge (44), the link between A-to-I editing by ADAR and the type I

interferon response shown here suggests that editing can not only affect tumor cell activities, but also those of immune and other cells present in the tumor microenvironment. A significant role for ADAR is further supported by our demonstration that its expression is significantly upregulated by *ADAR* copy number gains in breast (up to 75%) and other cancers (up to 70%). Overall, these data highlight the potential magnitude of A-to-I RNA editing in tumors and thereby the possibility for large-scale clinical consequences. RNA editing and/or APOBEC-mediated mutagenesis could shape the immunogenicity of the tumor and thereby directly affect anti- and/or pro-tumor immune responses. RNA editing itself, the processes it regulates and its potential to differentially direct activities in response to the chronic inflammatory tumor microenvironment, may have important implications for clinical progression in breast and other cancers.

This study identifies the principles governing A-to-I editing in breast and potentially all types of cancer. These data further identify the type I interferon pathway as an important mechanism regulating this phenomenon. The wide-spread editing we observed, in combination with the conservation of editing sites detected across tissues and patients, suggests there are may be clinical and therapeutic implications for a wide range of cancer patients. However, modulating editing at an individual site is entangled with many processes. The model we established for A-to-I editing implies that modulation of ADAR will also affect all editable sites in expressed transcripts. In addition, ADAR has recently been shown to enhance miRNA processing (45). Thus, variation in the editing frequency at an individual site is likely to be correlated with global variation of miRNA expression. Finally, variation of ADAR expression *in vivo* will possibly be associated with modification of the hundreds of genes located on 1q and/or controlled by interferon. Determining whether increasing A-to-I editing limits or enhances cancer progression will need to take into account all of these potential variables. More research is needed to identify the critical editing sites, establish their potential as markers of cancer evolution and investigate them as a new class of therapeutic targets.

DATA ACCESS

Sequencing and SNPs array data obtained from the enrolled patients are archived at European Genome-phenome Archive, https://www.ebi.ac.uk/ega, under accession number EGAS00001000495; amplicon sequencing data obtained from cell lines are archived at European Nucleotide Archive, http://www.ebi.ac.uk/ena/home, under study accession number ERP004253; gene expression array data are archived at the Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo, under accession number GSE43358. Clinical information, results of sequence and arrays preprocessing and biological assays are available in the online Supplementary Tables.

FUNDING

This work was supported by a grant of the Belgian National Cancer Plan PNC29. D. Gacquer and T. Konopka have been supported by a WELBIO grant. M. Abramowicz, D. Brown and C. Sotiriou were supported by the FNRS. A. Shlien is funded by the H.L. Holmes Award from the National Research Council Canada and an EMBO fellowship. C. Desmedt has been supported by the Brussels Region.

ACKNOWLEDGMENTS

We thank Raphael Leplae and the ULB Computing Center for their support, Roland De Wind for pathology support, Cédric Blanpain, Sabine Costagliola, Jacques E. Dumont, Pierre Vanderhaeghen and Gilbert Vassart for the helpful discussions.

DISCLOSURE DECLARATION

No conflict of interest.

FIGURES LEGEND

Figure 1. Detection of A-to-I editing. **(A)**, Substitution frequencies of RDDs. **(B)**, Percentage of RDDs confirmed in the validation dataset, N=15 BCs (in red), and the DARNED database (in blue). The negative control set is composed of 1,000 sites selected at random positions in randomly selected *Alu*s. **(C)**, Each dot represents a sample for which the frequency of edited AZIN1 transcripts has been measured with Illumina full transcriptome sequencing (x-axis) and Roche FLX amplicon sequencing (y-axis). ρ denotes the Spearman's correlation. **(D)**, **(E)**, Number of detected *Alu* A-to-I sites as a function of transcriptome and exome coverages.

Figure 2. A-to-I editing and ADAR expression in normal and tumor breast tissue. **(A)**, Each dot represents a patient with the mean editing frequency in her normal (*x*-axis) and her matched tumor breast tissue (*y*-axis). **(B)**, Same as **(A)**, except that the AZIN1 editing frequency measured by Roche FLX amplicon sequencing is depicted. **(C)**, The mean editing frequency of 8 breast organoid cultures is compared to that of 15 breast tumors.

Figure 3. Model of A-to-I editing. **(A)**, Each dot represents a sample with its RNA-seq-estimated ADAR expression on the *x*-axis (in $\log_2$ of Fragments Per Kilobase per Million mapped reads), and its mean editing frequency across all 560 *Alu* sites on the *y*-axis. The RNA-seq expression of ADAR is highly correlated with microarrays and qRT-PCR expression (Supplementary Fig. S3). **(B)**, Each dot represents an *Alu* A-to-I editing site with the maximal edit frequency across all samples on the *x*-axis and the number of samples in which it was detectably edited on the *y*-axis. **(C)**, Heatmap of editing frequencies across all *Alu* A-to-I edit sites in all samples. Both are ordered by increasing (down-to-up, left-to-right) mean editing frequencies. Smoothed contour lines labels give the percentage of edited transcripts. The bottom panel shows corresponding ADAR expression. Negative controls are presented in Supplementary Fig. S3. **(D)**, Model of A-to-I editing. Turning the ADAR 'expression knob' clockwise increases ADAR expression. As a result, more transcripts are edited (red dots) and the editing frequency of all editable sites increases accordingly (compare green vs. red bars). Moreover, the detection limit at some sites for which editing was previously undetectable was passed. The detection limit depends on sequencing coverage, which is lower on the right-most exon. Importantly, the ranking of editing frequencies of the different sites is unaltered by *ADAR* expression.

Figure 4. Validation of the A-to-I editing model. **(A)**, Effect of increasing ADAR expression in the cell line MCF7 on editing in 4 representative *Alu* regions and AZIN1. The full-length of sequenced regions are shown in Supplementary Fig. S4 for MCF7 and 3 more cell lines. Complete ADAR Western blots quantifications underlying the color scale are provided in Fig. 6E and in Supplementary Fig. S9A (see baseline t=0 and IFN-α, $t \in \{1, 2, 5\}$ days tracks). Increasing ADAR expression increases the editing

frequency at all editable positions, as predicted by the model of Fig. 3D. Similar results were obtained for IFN-β and IFN-γ (global, position-less, view Fig. 6F). Arrows point at editing sites detectable only at higher ADAR expression in our assay. **(B)**, Increasing sequencing coverage (*x*-axis) or ADAR expression (color scale) increases the number of detectable editing sites (*y*-axis). Coverage variation was implemented by down-sampling the total pool of sequencing reads, starting from 2,000x, down to 100x, and re-running the variant detection pipeline for each down-sampled alignment. Each data point is the mean of 10 down-sampling experiments. Error bars were omitted for the sake of clarity. **(C)**, Editing of individual mRNA molecules. Each black dot depicts an edited base (*y*-axis) in a sequencing read (*x*-axis) spanning the entire 256bp region. Reads and sites were ordered by decreasing editing frequencies. Hundred and eighty-five non-edited reads were omitted from the figure. **(D)**, Each dot represents an *Alu* A-to-I site with coverage >20x from high coverage sample GM12878. Measured editing frequencies, *x*-axis, are compared to the editability score computed from the DNA sequence alone, *y*-axis. The calculation was run on 25,681 sites not used to develop the model.

Figure 5. (**A**) Representative ADAR staining of a luminal A tumor. **(B)**, Zooming in panel **(A)** reveals that tumor staining (black arrows) is higher that in normal epithelium (green arrows) and lymphocytes (red arrows). **(C)**, Zooming further in panel **(B)** reveals a higher staining of nucleoli (black arrows).

Figure 6. *ADAR* amplification and the interferon response are independent predictors of ADAR expression in cancer. **(A)**, The top panel shows the frequencies of amplifications/deletions along chromosome 1 in our series. The middle panel shows the genes whose expression is highly associated with that of ADAR. Nineteen genes not located on chr1 are omitted. The bottom panel shows the Spearman's correlation coefficient and associated *p*-values of non-segmented copy-number array probes with the sample-wise mean editing frequencies. **(B)**, Dots represent tumor samples, with STAT1 expression on the *x*-axis and ADAR expression on the *y*-axis. **(C)**, Same as **(B)** with *ADAR* expression adjusted for *ADAR* copy number. **(D)**, Association p-values of *ADAR* copy number and STAT1 expression with ADAR expression increase in a multivariate analysis, demonstrating that ADAR expression is independently associated with these two variables. **(E)**, Seven breast cancer cell lines were exposed to interferon α, β and γ for 1, 2 and 5 days. Western blots quantifications are depicted for each cell line, interferon and time. Because expression dynamic ranges vary among cell lines, each line has its own color scale extending from low expression in green to high expression in red. The underlying gels are presented in Supplementary Fig. S9A and blot quantification in Supplementary Table S6. **(F)**, Editing frequencies in the absence of treatment (*x*-axis) vs. interferon treatment (*y*-axis). Points depict the editing sites in AZIN1 and the 4 *Alu* regions of Supplementary Fig. S4. Points are above the identity line $x=y$ (black diagonals), i.e. interferon increase editing frequencies at all sites. Library preparation failed for MCF7/IFN-γ at 5 days. Limited sequencing coverage precluded detection of some editing events for MDA-MB-231, t=0 and t=1 days.

Figure 7. ADAR involvement in cell proliferation and apoptosis. **(A)**, Western blot analysis of ADAR silencing after shRNA lentiviral transduction in MDA-MB-231, MCF7 and BT474 breast cancer cell lines. **(B)**, ADAR silencing statistically decreases cell proliferation. Cell growth curves for ADAR-knock down cells (shRNA ADAR) and control cells (shRNA control) in MDA-MB-231, MCF7 and BT474 BC cell lines. **(C)**, ADAR silencing statistically increases cell apoptosis. Illustration of the percentage of apoptotic cells in ADAR-knock down cells (shRNA ADAR) and control cells (shRNA control) in MDA-MB-231, MCF7 and BT474 BC cell lines. Error bars depict standard deviation of three independent experiments.

Figure 8. *ADAR* amplification and the interferon response predict ADAR expression in human cancers. We included all TCGA datasets and tumors (see 'N' column) for which both copy number and RNA-seq expression data (pipeline v3) were available. Datasets are ordered by decreasing median ADAR expression (top to bottom). The three leftmost plots depict the distributions of ADAR expression, *ADAR* DNA copy number and STAT1 expression across each dataset. The two rightmost bar plots extend to TCGA data the calculation presented for our data in Fig. 6B and 6C. In most cancers, adjusting ADAR expression for *ADAR* copy number increases the Spearman correlation, ρ, with STAT1 (compare the orange bars to the blue bars).

REFERENCES

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

2. 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.

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

4. Wang,Q., Khillan,J., Gadue,P. and Nishikura,K. (2000) Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science*, **290**, 1765–1768.

5. Higuchi,M., Maas,S., Single,F.N., Hartner,J., Rozov,A., Burnashev,N., Feldmeyer,D., Sprengel,R. and Seeburg,P.H. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature*, **406**, 78–81.

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

7. Chen,C.X., Cho,D.S., Wang,Q., Lai,F., Carter,K.C. and Nishikura,K. (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA*, **6**, 755–767.

8. Rueter,S.M., Dawson,T.R. and Emeson,R.B. (1999) Regulation of alternative splicing by RNA editing. *Nature*, **399**, 75–80.

9. 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.

10. Blow,M.J., Grocock,R.J., van Dongen,S., Enright,A.J., Dicks,E., Futreal,P.A., Wooster,R. and Stratton,M.R. (2006) RNA editing of human microRNAs. *Genome Biol.*, **7**, R27.

11. Kawahara,Y., Zinshteyn,B., Chendrimada,T.P., Shiekhattar,R. and Nishikura,K. (2007) RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep.*, **8**, 763–769.

12. Hundley,H.A. and Bass,B.L. (2010) ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem. Sci.*, **35**, 377–383.

13. Kapusta,A., Kronenberg,Z., Lynch,V.J., Zhuo,X., Ramsay,L., Bourque,G., Yandell,M. and Feschotte,C. (2013) Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.*, **9**, e1003470.

14. 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.

15. Bazak,L., Haviv,A., Barak,M., Jacob-Hirsch,J., Deng,P., Zhang,R., Isaacs,F.J., Rechavi,G., Li,J.B., Eisenberg,E., *et al.* (2014) 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.

16. 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.

17. Ju,Y.S., Kim,J.-I., Kim,S., Hong,D., Park,H., Shin,J.-Y., Lee,S., Lee,W.-C., Kim,S., Yu,S.-B., *et al.* (2011) Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals. *Nat. Genet.*, **43**, 745–752.

18. 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.

19. 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.

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

21. 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.

22. 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.

23. Jiang,Q., Crews,L.A., Barrett,C.L., Chun,H.-J., Court,A.C., Isquith,J.M., Zipeto,M.A., Goff,D.J., Minden,M., Sadarangani,A., *et al.* (2013) ADAR1 promotes malignant progenitor reprogramming in chronic myeloid leukemia. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 1041–1046.

24. 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.

25. Kiran,A. and Baranov,P.V. (2010) DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics*, **26**, 1772–1776.

26. Ramaswami,G., Lin,W., Piskol,R., Tan,M.H., Davis,C. and Li,J.B. (2012) Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods*, **9**, 579–581.

27. Li,M., Wang,I.X., Li,Y., Bruzel,A., Richards,A.L., Toung,J.M. and Cheung,V.G. (2011) Widespread RNA and DNA sequence differences in the human transcriptome. *Science*, **333**, 53–58.

28. Shah,S.P., Morin,R.D., Khattra,J., Prentice,L., Pugh,T., Burleigh,A., Delaney,A., Gelmon,K., Guliany,R., Senz,J., *et al*. (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, **461**, 809–813.

29. 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.

30. Patterson,J.B. and Samuel,C.E. (1995) Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol. Cell. Biol.*, **15**, 5376–5388.

31. Desterro,J.M.P., Keegan,L.P., Lafarga,M., Berciano,M.T., O'Connell,M. and Carmo-Fonseca,M. (2003) Dynamic association of RNA-editing enzymes with the nucleolus. *J. Cell. Sci.*, **116**, 1805–1818.

32. Sansam,C.L., Wells,K.S. and Emeson,R.B. (2003) Modulation of RNA editing by functional nucleolar sequestration of ADAR2. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 14018–14023.

33. Curtis,C., Shah,S.P., Chin,S.-F., Turashvili,G., Rueda,O.M., Dunning,M.J., Speed,D., Lynch,A.G., Samarajiwa,S., Yuan,Y., *et al*. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352.

34. Efron,B. and Tibshirani,R. (2007) On testing the significance of sets of genes. *Ann Appl Stat*.

35. Schoggins,J.W., Wilson,S.J., Panis,M., Murphy,M.Y., Jones,C.T., Bieniasz,P. and Rice,C.M. (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, **472**, 481–485.

36. The Cancer Genome Atlas - Cancer Genome - TCGA.

37. 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.

38. 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.

39. Beghini,A., Ripamonti,C.B., Peterlongo,P., Roversi,G., Cairoli,R., Morra,E. and Larizza,L. (2000) RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia. *Hum. Mol. Genet.*, **9**, 2297–2304.

40. Ishiuchi,S., Tsuzuki,K., Yoshida,Y., Yamada,N., Hagimura,N., Okado,H., Miwa,A., Kurihara,H., Nakazato,Y., Tamura,M., *et al.* (2002) Blockage of Ca(2+)-permeable AMPA receptors suppresses migration and induces apoptosis in human glioblastoma cells. *Nat. Med.*, **8**, 971–978.

41. 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.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.

42. Roberts,S.A., Lawrence,M.S., Klimczak,L.J., Grimm,S.A., Fargo,D., Stojanov,P., Kiezun,A., Kryukov,G.V., Carter,S.L., Saksena,G., *et al.* (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.*, **45**, 970–976.

43. Burns,M.B., Temiz,N.A. and Harris,R.S. (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.*, **45**, 977–983.

44. Gallo,A. and Locatelli,F. (2012) ADARs: allies or enemies? The importance of A-to-I RNA editing in human disease: from cancer to HIV-1. *Biol Rev Camb Philos Soc*, **87**, 95–110.

45. Ota,H., Sakurai,M., Gupta,R., Valente,L., Wulff,B.-E., Ariyoshi,K., Iizasa,H., Davuluri,R.V. and Nishikura,K. (2013) ADAR1 forms a complex with Dicer to promote microRNA processing and RNA-induced gene silencing. *Cell*, **153**, 575–589.

Figure 1

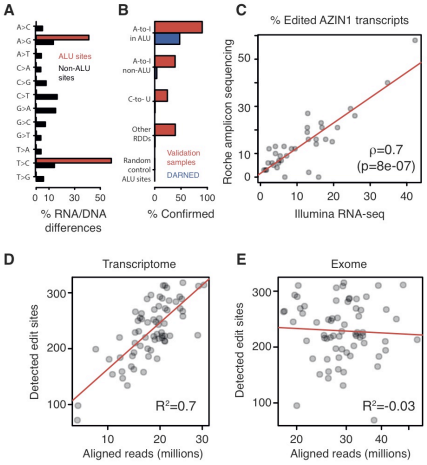

A
A>C
A>G
A>T
C>A
C>G
C>T
G>A
G>C
G>T
T>A
T>C
T>G
ALU sites
Non-ALU sites
0 10 30 50
% RNA/DNA differences
B
A-to-I in ALU
A-to-I non-ALU
C-to-U
Other RDDs
Random control ALU sites
Validation samples
DARNED
0 50 100
% Confirmed
C
% Edited AZIN1 transcripts
Roche amplicon sequencing
0 10 30 50
ρ=0.7
(p=8e-07)
0 10 20 30 40
Illumina RNA-seq
D
Transcriptome
Detected edit sites
100 200 300
R²=0.7
10 20 30
Aligned reads (millions)
E
Exome
Detected edit sites
100 200 300
R²=-0.03
20 30 40
Aligned reads (millions)

Figure 2

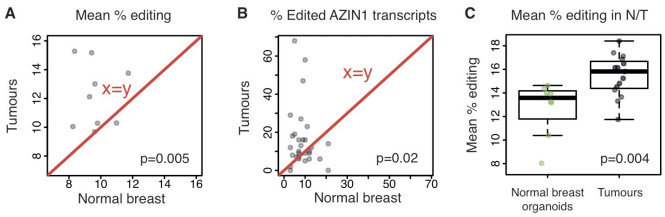

A
Mean % editing
Tumours
Normal breast
x=y
p=0.005
B
% Edited AZIN1 transcripts
Tumours
Normal breast
x=y
p=0.02
C
Mean % editing in N/T
Mean % editing
Normal breast organoids
Tumours
p=0.004

Figure 3

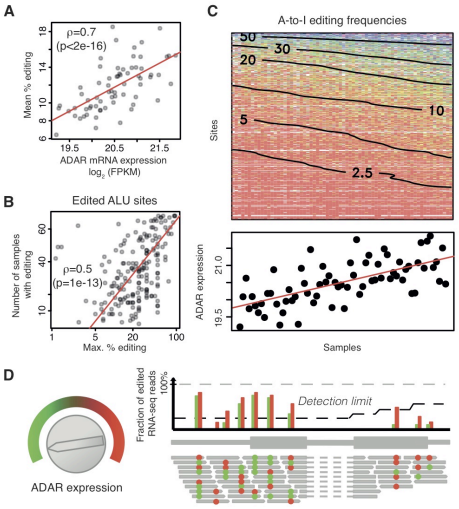

A
ρ=0.7
(p<2e-16)
Mean % editing
ADAR mRNA expression
log2 (FPKM)
B
Edited ALU sites
ρ=0.5
(p=1e-13)
Number of samples with editing
Max. % editing
C
A-to-I editing frequencies
Sites
50
30
20
10
5
2.5
ADAR expression
Samples
D
ADAR expression
Fraction of edited RNA-seq reads
100%
Detection limit

Figure 4

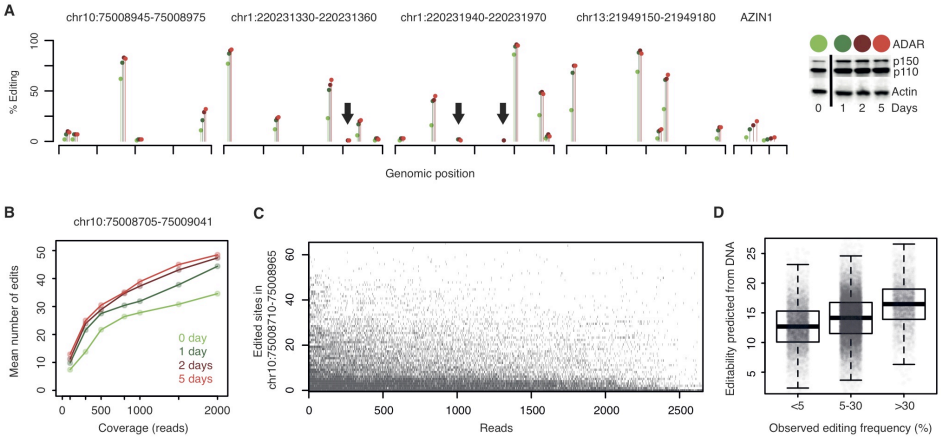

A
chr10:75008945-75008975
chr1:220231330-220231360
chr1:220231940-220231970
chr13:21949150-21949180
AZIN1
% Editing
Genomic position
ADAR
p150
p110
Actin
0 1 2 5 Days
B
chr10:75008705-75009041
Mean number of edits
Coverage (reads)
0 day
1 day
2 days
5 days
C
Edited sites in chr10:75008710-75008965
Reads
D
Editability predicted from DNA
<5
5-30
>30
Observed editing frequency (%)

Figure 5

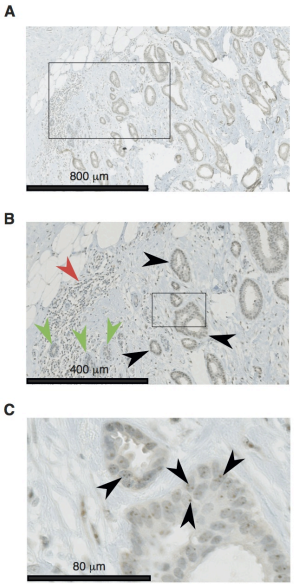


Figure 6

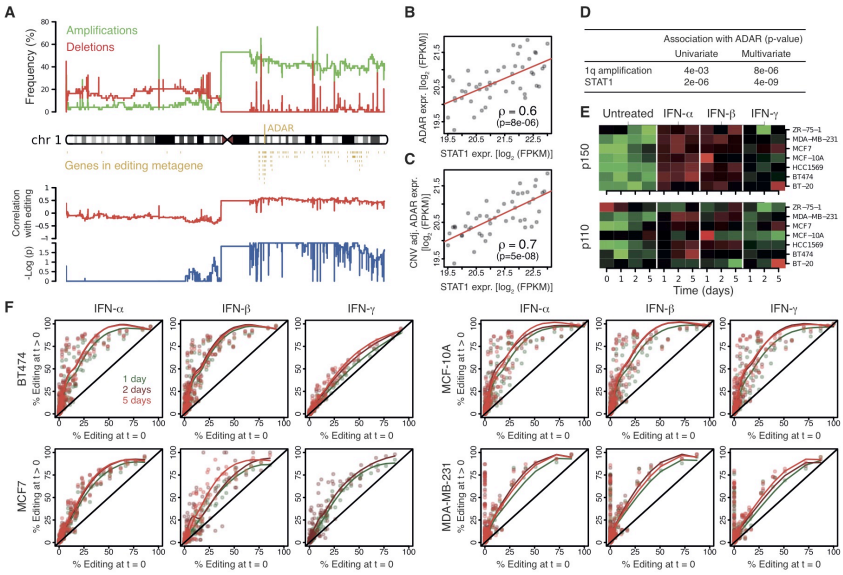

A
Amplifications
Deletions
Frequency (%)
chr 1
ADAR
Genes in editing metagene
Correlation with editing
-Log (p)
B
ADAR expr. [log2 (FPKM)]
ρ = 0.6
(p=8e-06)
STAT1 expr. [log2 (FPKM)]
C
CNV adj. ADAR expr. [log2 (FPKM)]
ρ = 0.7
(p=5e-08)
STAT1 expr. [log2 (FPKM)]
D
Association with ADAR (p-value)
Univariate
Multivariate
1q amplification 4e-03 8e-06
STAT1 2e-06 4e-09
E
Untreated
IFN-α
IFN-β
IFN-γ
p150
p110
ZR-75-1
MDA-MB-231
MCF7
MCF-10A
HCC1569
BT474
BT-20
Time (days)
F
IFN-α
IFN-β
IFN-γ
BT474
MCF7
MCF-10A
MDA-MB-231
% Editing at t > 0
% Editing at t = 0
1 day
2 days
5 days

Figure 7

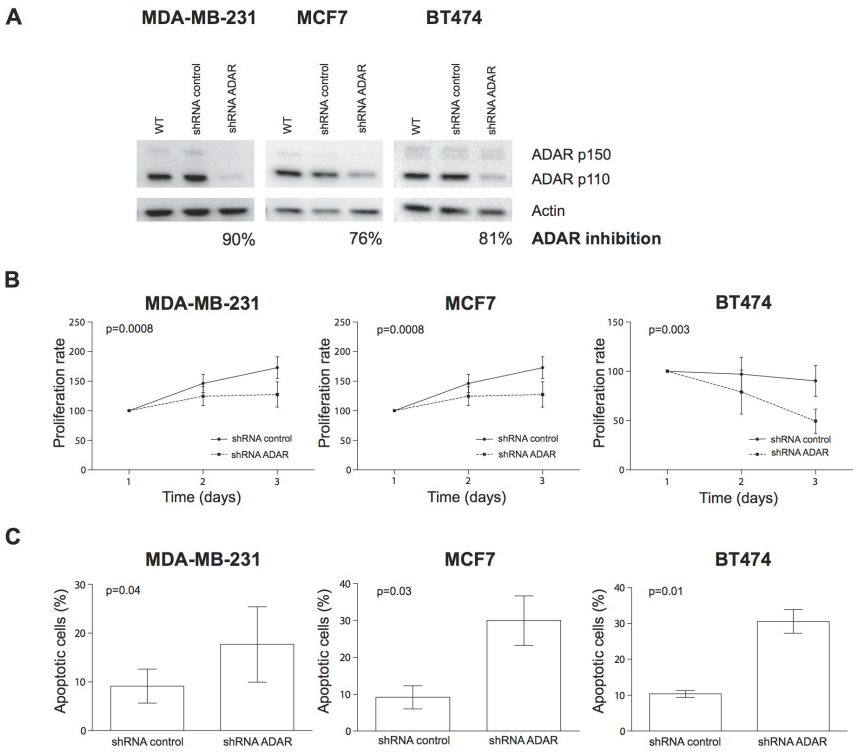
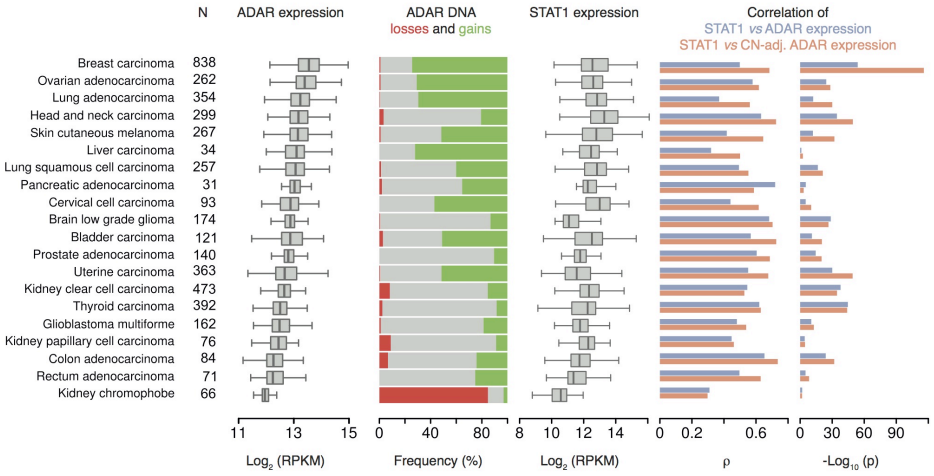

A
MDA-MB-231
MCF7
BT474
WT
shRNA control
shRNA ADAR
ADAR p150
ADAR p110
Actin
90%
76%
81%
ADAR inhibition
B
MDA-MB-231
p=0.0008
MCF7
p=0.0008
BT474
p=0.003
Proliferation rate
Time (days)
shRNA control
shRNA ADAR
C
MDA-MB-231
p=0.04
MCF7
p=0.03
BT474
p=0.01
Apoptotic cells (%)
shRNA control
shRNA ADAR

Figure 8


N
ADAR expression
ADAR DNA
losses and gains
STAT1 expression
Correlation of
STAT1 vs ADAR expression
STAT1 vs CN-adj. ADAR expression
Breast carcinoma 838
Ovarian adenocarcinoma 262
Lung adenocarcinoma 354
Head and neck carcinoma 299
Skin cutaneous melanoma 267
Liver carcinoma 34
Lung squamous cell carcinoma 257
Pancreatic adenocarcinoma 31
Cervical cell carcinoma 93
Brain low grade glioma 174
Bladder carcinoma 121
Prostate adenocarcinoma 140
Uterine carcinoma 363
Kidney clear cell carcinoma 473
Thyroid carcinoma 392
Glioblastoma multiforme 162
Kidney papillary cell carcinoma 76
Colon adenocarcinoma 84
Rectum adenocarcinoma 71
Kidney chromophobe 66
11 13 15
Log2 (RPKM)
0 40 80
Frequency (%)
8 10 12 14
Log2 (RPKM)
0 0.2 0.6
ρ
0 30 60 90
-Log10 (p)