Characterisation of APOBEC3B-Mediated RNA Editing in Breast Cancer Cells Reveals Regulatory Roles of NEAT1 and MALAT1 lncRNAs

Paul Clarke (paul.clarke@icr.ac.uk)
Institute of Cancer Research https://orcid.org/0000-0001-9342-1290
Chi Zhang
Yujing Lu
Mei Wang
Bingjie Chen
Feifei Xiong
Costas Mitsopoulos
Olivia Rosanesse
Xiuling Li

Article

Keywords: APOBEC3B, RNA editing, cytidine deamination, NEAT1, MALAT1

Posted Date: November 21st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3619470/v1

Additional Declarations: **Yes** there is potential conflict of interest.

Abstract

RNA editing is a crucial post-transcriptional process that influences gene expression and increases the diversity of the proteome as a result of amino acid substitution. Recently, the APOBEC3 family has emerged as a significant player in this mechanism, with APOBEC3A (A3A) having prominent roles in base editing during immune and stress responses. APOBEC3B (A3B), another family member, has gained attention for its potential role in generating genomic DNA mutations in breast cancer. In this study, we coupled an inducible expression cell model with a novel methodology for identifying differential variants in RNA (DVRs) to map A3B-mediated RNA editing sites in a breast cancer cell model. Our findings indicate that A3B engages in selective RNA editing including targeting NEAT1 and MALAT1 long non-coding RNAs that are often highly expressed in tumour cells. Notably, the binding of these RNAs sequesters A3B and suppresses global A3B activity against RNA and DNA. Release of A3B from NEAT1/MALAT1 resulted in increased A3B activity at the expense of A3A activity suggesting a regulatory feedback loop between the two family members. This research substantially advances our understanding of A3B's role in RNA editing, its mechanistic underpinnings, and its potential relevance in the pathogenesis of breast cancer.

Introduction

RNA editing is a dynamic post-transcriptional process that modulates transcript sequences without altering the underlying genomic DNA sequence [1]. It plays a crucial role in expanding proteomic diversity and regulating gene expression in higher eukaryotes. Among the diverse mechanisms of RNA editing, two major types of modifications have been identified: deamination of adenine to inosine (A > I) or cytidine to uracil (C > U). These modifications impact on the cellular proteome as they are read as guanosine and uridine respectively during translation, thereby altering protein sequences encoded by genomic DNA [2, 3].

In mammals, the APOBEC (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like) protein family, alongside the activation-induced deaminase (AID) and cytidine deaminase (CDA), encompasses a group of cytidine deaminases [4]. While AID is recognised for its role in C > U mutation of DNA during antibody gene diversification, the APOBEC proteins have gained attention for their broader-spectrum roles in RNA editing [5, 6]. The human APOBEC family comprises ten members, APOBEC1, APOBEC2, and APOBEC4, that have restricted tissue expression, while the APOBEC3 genes are more widely expressed, though predominantly in immune cells [7]. Initially characterized for their ability to inhibit retroviruses and transposable elements by deaminating cytidines in single-stranded DNA, recent studies have revealed the involvement of APOBEC3 family members in RNA editing. Physiological conditions, such as interferon stimulation, hypoxia, and cellular crowding, can induce endogenous APOBEC3A (A3A)-mediated C > U RNA editing in monocytes and immune cells [8–10]. The induction of RNA editing by A3A under these circumstances suggests its involvement in cellular stress responses and immune modulation. Furthermore, A3A-mediated RNA editing of specific transcripts have been observed, indicating its potential impact on gene expression regulation and protein diversity [11, 12]. However, the prevalence and significance of other APOBEC3-mediated RNA editing in various biological contexts, including cancer is less well understood and remains an area of active investigation [3].

The last 10 years have witnessed the growing interest in the function of another member of the APOBEC3 family in cancer. APOBEC3B (A3B) has been implicated in driving genetic heterogeneity in cancers by inducing C > T transitions and C > G transversions at 5'-TCW motifs (W = A or T) [13, 14], a mutation pattern observed in more than 50% of primary breast cancers [15, 16]. Being the sole constitutively nuclear-localised member of the APOBEC3 family, A3B emerges as one of the primary sources underlying the genomic APOBEC3 mutational signature identified in breast cancer [15, 17]. The discovery that several members of the APOBEC3 family exhibit RNA-binding capabilities has also fuelled speculation about a role for A3B in RNA editing [18]. This is supported by the findings that the catalytic C-terminal domain (CTD) of A3B shares a high degree of similarity to A3A that binds RNA with high affinity [19]. However, the identification of transcriptome-wide A3B-mediated RNA editing sites and their cellular or mechanistic consequences in normal tissue or cancer cells remains unexplored. The challenges are twofold: firstly, even minimal genomic editing by A3B can be erroneously attributed to RNA editing due to the intricacies of the transcription process; secondly, RNA editing activities stemming from other members of the APOBEC3 family can confound the precise identification of A3B-specific RNA editing sites. Consequently, there is an imperative need for a more rigorous methodology to surmount these obstacles.

In this paper, we set out to comprehensively map the RNA sites edited by A3B using a breast cancer cell model with inducible A3B expression in conjunction with a newly-developed methodology to identify differential variants in RNA (DVRs). By further analysis of these sites, we sought to understand the mechanism and functional impact underlying the RNA-editing function of A3B and the transcriptomic hotspots of A3B-editing.

Results

Identification of A3B-mediated RNA editing sites.

To discern RNA editing sites resulting from A3B activity, we constructed a lentiviral expression system enabling the doxycycline-induced expression of a Flag-tagged A3B fusion protein (Fig. 1A). By comparing A3B-induced to un-induced cells, we could identify RNA editing sites solely attributable to the elevated A3B catalytic activity. To mitigate potential interference from other APOBEC3 family members, notably A3A and A3G, our lentiviral inducible system was established in T-47D cells. These cells are characterised by a notably lower expression of these proteins relative to many breast cancer cell lines [20]. Furthermore, they possess a loss-of-function mutation in the *TP53* gene, making them resistant to programmed cell death triggered by A3B induction [21]. We meticulously regulated the degree of induced expression of the Flag-tagged A3B protein (Fig. 1B, S1A) to ensure its levels were akin to the peak levels observed for endogenous A3B in a range of breast cancer cell lines, as previously documented [15]. For control purposes, we also constructed an inducible system for an enzymatically inactive A3B variant (E65Q/E225Q), where both the NTD and CTD catalytic domains are rendered inactive by mutation, henceforth referred to as A3B** [22]. This approach ensures the identification of A3B RNA editing sites that are physiologically pertinent within a breast cancer framework.

To identify A3B-editing sites in T-47D transcriptome and distinguish the RNA sites from edited sites resulting from A3B-catalysed modification of genomic DNA, we used next-generation sequencing and an analysis workflow that was devised to identify the DVRs attributed to the induced A3B. This method joins the most salient features from the previously established VaDiR [23] and rMATS-DVR [24] workflows where: i) False-positive RNA variants, arising from transcription of DNA's single nucleotide variants (SNV), are filtered using GATK Mutect2 [25], which utilises joint mutation calling from pairing RNA sequencing (RNA-seq) with control whole genome sequencing (WGS) data; and ii) The rMATS statistical approach which is employed on sample groups with multiple replicates (n = 4) to derive highly confident DVRs. Our approach contrasted with other previously delineated methods for detecting APOBEC3 RNA editing sites as it captures not only RNA editing events but also considers the consistency and magnitude of these events [23, 24]. A schematic representation of this analytical procedure is depicted in Fig. 1C and the outputs at different stages are shown in **Figure S1B**. We compared the results for A3B-induced RNA variant (RV) and DVR identification using our analytical pipeline outlined in Fig. 1C to the original rMATS-DVR method [24, 26] and found a reduced false-positive overlap with pre-existing DNA SNVs present in T-47D genomic DNA (GSE193225) (**Supplementary Table 1–2**).

To ascertain if our analytical workflow could pinpoint DVRs ascribed to augmented A3B activity, we processed T-47D cell samples with and without doxycycline exposure (Fig. 1D). We verified the induction of A3B at both the mRNA and protein levels following doxycycline treatment and extracted DNA and RNA samples for next generation sequencing and DVR assessment. The first round of analysis with our DVR detection pipeline revealed a marked rise in DVRs in poly-A enriched RNA following A3B induction, this included a build-up of A > G(I) DVRs and an even more pronounced accumulation of C > U DVRs (Fig. 1E-F). In addition, closer inspection of these DVRs finds the editing levels for most of the C > U DVRs were increased due to prolonged exposure to ectopically-expressed A3B, which was a phenomenon not found for A > G(I) DVRs (**Figure S1C**). These results suggest that the C > U editing sites detected were related to additional A3B catalytic activity by doxycycline induction.

The RNA preparation technique employed will dictate the RNA categories that will be sequenced. Initially, we employed poly-A enrichment, predominantly capturing mRNA. Next, to gain a more comprehensive perspective on RNA editing, we compared next generation sequencing of poly-A enriched RNA with a ribosomal RNA depletion protocol that facilitated the incorporation of non-polyadenylated RNA variants into our DVR analysis. Post 72 hours of A3B induction, we demonstrated that the ribosomal RNA depletion technique augmented DVR detection by ≈ 3-fold (Fig. 1E-F). The predominant DVRs discerned in the ribosomal-depleted RNA stemmed from A > G(I) editing that was increased by ≈ 6-fold compared to the poly-A enriched RNA. The ribosomal RNA depletion-enriched samples also manifested heightened C > U DVRs, by ≈ 0.5-fold relative to poly-A enriched RNA.

Employing the 72-hour induction and ribosome depletion approach, which optimally sensitised DVR detection, we examined DVRs produced by the A3B** construct as a negative control. We ascertained that the induction levels of both the mutant and native A3B were comparable (Fig. 1B). However, the expression of the enzymatically inert A3B** significantly diminished DVR detection compared to its active

counterpart (Fig. 1E-F). Notably, C > U variant production was profoundly affected relative to A > G(I) variants. The induction of the inactive A3B** yielded a mere two C > U DVRs, in contrast to the several hundred C > U DVRs observed with the wild-type A3B (Fig. 1E). By conducting differential expression analyses on the RNA-seq results, we also investigated whether the ectopic induction of A3B or A3B** affects the intrinsic APOBEC or ADAR families of RNA editors. As shown in **Figure S2**, for all the genes investigated none were significantly altered except for APOBEC3F which was up-regulated but to negligible levels compared to intrinsic and ectopic A3B. This result affirms that the C > U DVRs were unlikely to be attributed to activities from other members of APOBEC family, and that the A > G(I) DVRs were unlikely to be linked to potential transcriptional regulation of ADAR family due to A3B induction. Collectively, our findings underscore that the C > U editing sites discerned resulted from prolonged A3B induction, and the C > U DVRs identified through our refined detection method hinge on the catalytic efficacy of the doxycycline-induced A3B.

Validation and characterisation of DVRs.

To ensure that RNA C > U DVRs we identified were not inaccurately attributed to genomic DNA SNVs we used the RNAfold algorithm [27] to estimate the free energy linked to RNA folding of 100 nucleotides of sequence surrounding the DVRs. We then predicted the relationship between RNA structure and location of the base modification, specifically if the edited sites were located within double-stranded base-paired stems or single-stranded loop structures. Our analysis revealed that sequences adjacent to A > G(I) DVRs exhibited a comparable RNA folding energy distribution as curated RNA A > I edited sites [28], but differed from the genomic A > G SNVs listed in dbSNP [29], which exhibited less energetically favourable folding (Fig. 2A). Sequences adjacent to C > U DVRs also followed the A > G(I) pattern, with sequences adjacent to the RNA C > U edited sites exhibiting more energetically favourable folding than the known genomic C > T SNVs (dbSNP) or previously documented genomic A3B C > U edited sites (GSE193225, Fig. 2B) [26].

Detailed examination of the RNAfold-predicted structures revealed a marked enrichment of the A > G(I) and C > U DVRs in predicted single-stranded RNA loop structures (Fig. 2C). These single-stranded RNA structures are more susceptible to base deamination, which is in concordance with previous studies with ADAR and A3A [8]. Heightened editing within predicted loop structures was noted for UUCK > UUUK (K = G or U), VUCN > VUUN (V = A, C, or G, N = A, U, C or G) and VVCN > VVUN DVRs, suggesting a specific affinity for the nucleotides surrounding the C > U DVRs. To delve deeper into this phenomenon, we charted the distribution of nucleotides adjacent to the DVRs (Fig. 2D**)**. For A > G(I) DVRs, the patterns of surrounding nucleotides closely mirrored known RNA A > I sites [8], validating the DVRs' origin from RNA editing, presumably by the ADAR family of RNA editors. Conversely, for C > U DVRs, there was a clear predilection for the UUCV sequence motif, a contrast to the favoured TCW (W = A or T) sequence motif typical detected for genomic DNA sites edited by A3B (GSE193225; Fig. 2D) [26].

To further validate the role of A3B in the generation DVRs, we employed eCLIP-seq to map A3B-bound RNA sites in T-47D cells with high precision. We then scrutinised the data to determine whether A3B binding was enriched at DVR sites. By using this method we aimed to ascertain whether the C > U DVRs

result from RNA binding of A3B, where A3B binding sites were identified by next generation sequencing following isolation of RNA that was UV cross-linked to the induced Flag-tagged A3B (Fig. 3A). We processed the sequencing data using both the standard ENCODE protocol [30] and the PureCLIP programme [31]. The enriched eCLIP-seq signals were positioned immediately after C > U DVRs (Fig. 3B-C). This enrichment stems from the hindrance of reverse transcription by the crosslinked A3B:RNA complex during library preparation, affirming that A3B binds the RNA sites where C > U editing occurred. In contrast, there was no A3B eCLIP-seq signal enrichment around or adjacent to the A > G(I) DVRs, suggesting RNA binding or editing by A3B was not directly involved in generating these DVRs (Fig. 3C). Also, no enrichment of eCLIP-seq signal was found near previously identified A3B-mediated TC > TT genomic SNPs(Fig. 3C). While eCLIP-seq identified enrichment of A3B binding signal at C > U DVRs, not all C > U DVRs were overlapping with A3B binding sites identified by eCLIP-seq. We interpreted this as a reflection of the dynamic nature of A3B action, where A3B briefly binds, edits the RNA, then disengages and in the absence of repair mechanisms leaves a long-lived DVR. Further analysis of sequences around the A3B-associated eCLIP-seq clusters revealed a heightened frequency of the UUCV and UUCK motifs, previously pinpointed through RNA-seq and RNA-fold analysis, at A3B binding sites (**Figure S3**). Overall, the evidence from our eCLIP-seq analysis were consistent with the output from our RNA-seq analysis pipeline comparing catalytically active and inactive A3B. Together with the RNAfold analysis these data indicated that the C > U DVRs we detected were highly likely to result from the RNA editing activity of A3B, and not from editing of genomic DNA.

A3B edits RNA in a selective manner.

We subsequently examined the impact of the identified A3B-mediated DVRs on transcripts. Initially, we utilised the Variant Effect Predictor (VEP) [32] to annotate the RNA locations and the potential consequences of the DVRs within the T-47D transcriptome. Figure 4A depicts a significant enrichment of C > U DVRs in mRNA exons and 3'UTRs, when compared to un-induced T-47D control transcriptome positions matching the A3B C > U DVR sequence motif. Among the exonic DVRs, events in untranslated exons were predominant, with missense mutations in the resulting protein sequence being the next most common. Conversely, A > G(I) DVRs were primarily intronic, aligning with prior findings [8]. We also found A > G(I) DVRs were associated with transcripts showing significantly (FDR < 0.05) altered levels following A3B induction, contrasting with C > U DVRs that were generally associated with RNAs showing no or low levels of altered expression (**Figure S4**). This indicated that unlike A > G(I) editing which can alter mRNA splicing or stability, C > U RNA editing does not significantly impact RNA processing or degradation.

To discern any potential selectivity in A3B's activity, we quantified the distribution inequality of its C > U activity using the Gini coefficient [33]. We constructed Lorenz curves, plotting the cumulative proportion of C > U, indicative of A3B activity, against the rank of DVRs based on the depth of C > U mutations from RNA-seq experiments (Fig. 4B). In these plots, a straight diagonal line signifies an even distribution of A3B generated C > U DVRs across the transcriptome. Any deviation from this line suggests an unequal distribution of A3B activity, with the curve's increasing curvature highlighting a greater degree of inequality. The Gini coefficient, derived from the plot, ranges from 0 (even distribution) to 1 (maximum

inequality). For A3B C > U editing, the Gini coefficient was determined to be 0.375 and 0.300 for RNA samples processed using ribosomal depletion and poly-A enrichment methods, respectively. This points to a substrate bias in A3B activity, indicating a preference for editing specific RNAs. On closer analysis, approximately 40% of the A3B edited C > U DVRs were concentrated in the top 23 edited transcripts for the ribosomal-depleted samples and the top 44 genes for poly-A enriched samples (Fig. 4C). Specifically, for RNA processed via the ribosome depletion method, the editing of two non-coding RNAs, NEAT1 and MALAT1, constituted a significant fraction of the overall activity of A3B, accounting for 7.3% and 1.8%, respectively (Fig. 4C). These findings underscore A3B's substrate selectivity within the T-47D cell transcriptome and highlight specific hotspots for A3B-mediated C > U editing.

A3B binds and edit NEAT1 and MALAT1.

The pronounced enrichment of C > U editing by A3B on NEAT1 and MALAT1 RNAs prompted us to postulate a model of selective binding of A3B to these lncRNAs. The eCLIP-seq results, confirmed a marked enrichment of A3B-crosslinked RNA fragments to both *NEAT1* and *MALAT1* RNAs. These data revealed read clusters with heightened signals compared to the control size-matched input RNA, in this case suggesting stable A3B binding (**Figure S5**). Notably, some clusters displayed C > U DVRs, reinforcing the notion that A3B binding at these loci could result in RNA editing (Fig. 5A). Subsequent RNA immunoprecipitation (RIP) assays in 293T cells overexpressing flag-tagged A3A and A3B demonstrated co-precipitation of NEAT1 and MALAT1 RNA variants with flag-tagged A3B, but not A3A (Fig. 5B). We have previously employed proximity biotinylation-based labelling (BioID) to pinpoint A3B's binding partners across several breast cancer cell lines [26]. Mass spectrometry analysis from that study identified several A3B interactors that were also NEAT1 and MALAT1 RNA binding proteins, notably nuclear and/or paraspeckle proteins FUS, NONO, and SFPQ [26, 34]. We corroborated these findings by demonstrating co-immunoprecipitation of A3B with FUS, NONO, and SFPQ in T-47D cells. This association persisted after removal of nucleic acids by benzonase treatment, suggesting direct protein-protein interactions between A3B and these proteins. For further validation, we analysed sequences flanking all identified C > U DVRs in T-47D cell's transcriptome using the MEME package [35], which revealed high enrichment for NONO and FUS binding motifs in proximity to C > U DVRs, (Fig. 5C). Collectively, our results propose a model where A3B interacts and edits NEAT1 and MALAT1 lncRNAs that is mediated through their associated lncRNA protein partners.

NEAT1 and MALAT1 regulate A3B's catalytic activity.

NEAT1 and MALAT1 non-coding lncRNAs and their protein-binding partners are frequently overexpressed to similar level in different cancer types [36, 37]. We postulated that interactions between A3B and NEAT1/MALAT1 ribonucleoproteins (RNPs) might modulate A3B's catalytic activity, either by acting as suicide substrates or by sequestering A3B. To explore this, we utilised a previously established fluorescent reporter system to gauge the enzymatic activity of APOBEC3 enzymes in 293T cells [38]. We then assessed the implications of acute NEAT1 and MALAT1 depletion using antisense oligonucleotides (ASOs) that are frequently used to deplete non-coding RNAS (**Figure S6A)**. We introduced plasmid

constructs encoding Cas9n fusion proteins with A3A, A3B, and A3B's CTD into 293T cells, alongside an eGFP variant containing a T > C mutation at L202 (L202S, TTA > TCA) that ablates fluorescence and creates a potential A3B editing site. A3B editing of the L202 site generates a functional GFP, which can be quantified using fluorescence microscopy [38]. Notably, while acute depletion of NEAT1 and MALAT1 remained inconsequential for the editing activity of the A3A fusion protein, the activities of both full-length A3B and CTD-truncated A3B were significantly enhanced post depletion of NEAT1 or MALAT1 expression (Fig. 6A). This evidence supports our model that NEAT1 and MALAT1 lncRNAs act as negative modulators of A3B activity.

To determine the impact of prolonged NEAT1 and MALAT1 depletion on APOBEC3 enzymes' activity, we stably expressed shRNA targeting these lncRNAs and confirmed the efficacy of the shRNAs through RT-qPCR (**Figure S6B-C**). We incorporated five breast cancer cell lines with documented A3A and A3B expression profiles into this study [20], and gauged the activity of intrinsic A3B in these lines. A3B activity was measured by immunoprecipitating the A3B complex and measuring enzymatic function using a DNA deaminase biosensor incorporating BspHI restriction enzyme [39]. In the four A3B expressing cell lines, NEAT1 and MALAT1 depletion led to an enhanced enzymatic activity of A3B, underscoring that sustained suppression of their expression augments A3B activity (Fig. 6B). However, this was not observed in HCC2218 cells, a line with minimal A3B expression.

Recent findings have highlighted a negative feedback mechanism governing A3A and A3B activity [40], We explored whether the sustained increase in A3B activity, following loss of NEAT1 and MALAT1, might affect A3A function. We assessed level of C > U editing at the *DDOST1* C558 site (C558U, UCG > UUG) utilizing digital droplet PCR (ddPCR), a recognised technique designed for measuring A3A activity but not for other APOBEC3 family members since it was identified to be A3A-specific [41]. Our results revealed that while prolonged depletion of NEAT1 and MALAT1 did not alter A3A activity in the A3A-exclusive HCC2218 cells, but did diminish A3A activity in cells with high expression of both A3A and A3B (BT474 and HCC202) (Fig. 6C). The concomitant increase in A3B activity following NEAT1 and MALAT1 removal in these cells (Fig. 6B) suggests a model where release of active A3B in some cells was restrains of A3A activity.

Discussion

Two members of the APOBEC3 family (A3A and A3G) are the only members that have so far been reported to have *bona fide* RNA editing activity. The A3B member of this family is reported as key molecular driver of cancer mutations, however, its high sequence and structural similarity to A3A has also suggested this family member may have RNA editing activity [19, 42] [40]. Characterization of the RNA editing functions of A3B in human cancers has remained a challenge primarily due to interference from the genomic DNA cytidine deamination activity of A3B [7] and, depending on the cellular context, background RNA editing from other APOBEC3 family members - most notably A3A [8, 43]. Consequently, a better understanding of A3B-mediated RNA editing necessitated the development of an effective

cellular models, the implementation of more stringent controls, and the enhancement of analytical methodologies.

In the present study, we report the identification and mapping of A3B-mediated RNA editing sites by integrating a lentiviral inducible expression breast cancer model with next-generation sequencing and an analysis pipeline to detect and map RNA edits (DVRs). By combining WGS from controls with inducible expression of A3B, or a catalytically dead A3B, we developed a sensitive method that could: 1) detect low-level editing events and 2) isolate statistically significant DVRs that were solely attributed to elevated A3B expression, while managing interference from other APOBEC3 family members. Using this model, we were able to locate A3B RNA editing sites in the T-47D breast cancer transcriptome, strengthening the predictions that A3B can operate as an RNA-editing enzyme in cancer cells.

Unlike A3B's genomic editing sites which are typically repaired by base excision repair (BER) [44], the uracil bases resulting from A3B's RNA editing are retained. In a previous study with a BER-deficient cellular model we unveiled A3B's preference for genomic regulatory regions such as enhancers and promoters [26]. In contrast, here we discovered that the RNA edited sites of A3B had a higher incidence of coding region mutations in A3B's, suggesting a significant source of amino acid alterations that could contribute to A3B-mediated tumour adaptation and drug resistance [45, 46]. We also found editing sites in mRNA 5'- and 3'-UTRs, which could influence translation initiation, stability, and localization of RNAs. These observations and the additional detection of A > G(I) DVRs in T-47D cells, in conjunction with A3B induction, raises questions regarding the functional impact of RNA editing by A3B that need to be addressed through further targeted experimental designs. Elevated A3B expression did not significantly influence ADAR deaminase expression, nor did it result in increased binding of A3B at A > G(I) sites or shown proximity between C > U and A > G(I) coordinates. Therefore, the regulation of A > G(I) editing by A3B is most likely to be indirect.

We also showed that both poly-A enrichment and ribosome RNA depletion methods of library production successfully identified DVRs, but these techniques revealed preferences in the types of A3B-mediated DVRs detected. This comparison particularly illuminated a regulatory mechanism of A3B by long non-coding lncRNA, specifically by NEAT1 and MALAT1, which are abundant components of paraspeckles and nuclear speckles [36, 37, 47]. Our results align with structural and sequence homology data that predict the A3B NTD should bind to RNA [43]. Our findings suggest the involvement of NEAT1 and/or MALAT1 and A3B in an RNP complexes. Furthermore, although through homology A3B's NTD is predicted to bind RNA binding, our experiments using the CRISPR-Cas9n reporter system demonstrate that the C-terminal domain (CTD) alone can also manifest RNA editing function, and this can be influenced by cellular NEAT1/MALAT1 levels. These results indicate a potential complex mechanism of regulation of A3B that requires further exploration.

Among the many discoveries related to A3B's RNA editing sites, a significant finding pertains to A3B's sequence and RNA selectivity. A3B's uneven activity distribution has revealed a preference for lncRNAs NEAT1 and MALAT1 and suggested a sequestration mechanism repressing global A3B activity. This is

consistent with a previous report that the deamination activity of A3B is attenuated in an RNA-dependent manner and activated in cell lysates following RNAse A treatment [43]. These significant lncRNAs, previously involved in gene regulation and cellular processes including in cancer, also link to A3B's potential paraspeckle and/or nuclear speckle binding properties—an intriguing avenue for subsequent research.

Our data also introduces the potential of utilizing NEAT1/MALAT as biomarkers to predict A3A activity in cells and cancer patients [48, 49]. Recent comprehensive studies have demonstrated A3A's central role in driving genomic mutations in cancer, while A3B has weaker genome mutagenic capacity. We found that loss of NEAT1/MALAT1 can reduce A3A activity coincident with A3B activation, this complements a recent study that found A3B can function as a significant negative regulator of the genomic cytosine deaminase activity of A3A and A3A-driven mutagenesis [40]. These results call for a more nuanced investigations to evaluate the interrelationship between A3A and A3B activity against RNA or DNA and the role of incRNA binding and will contribute to a comprehensive understanding of how APOBEC3 enzymes promote cancer evolution by causing genome and transcriptome heterogeneity.

Materials and Methods

Cell culturing and cell line preparation

T-47D (ATCC HTB-133) was obtained from ATCC. Lenti-X 293T cells were obtained from Clonetech. T-47D cells were maintained in RPMI-1640 medium supplemented with bovine insulin, and Lenti-X 293T cells were cultured with DMEM medium. Both media were supplemented with 10% v/v FBS (PAA) and 0.5% v/v pen/strep. All cells were cultured at 37 ºC with 5% CO_2 and were regularly tested for mycoplasma contamination. To generate stable inducible cell lines, exponentially growing T-47D cells were transfected with lentivirus in the presence of 10 µg/ml polybrene. After a transduction period of 48 hours, selection was carried out using medium supplemented with 4 µg/mL puromycin (Gibco). Transduction efficiency was monitored by GFP expression as well as quantitative RT-PCR following induction of sample cells with doxycycline. After verification, lentivirus cassettes were maintained by culturing cells with 4 µg/mL puromycin. For the induction of lentiviral protein expression, cells were exposed to full RPMI-1640 culture medium supplemented with 100 ng/mL doxycycline (Sigma).

Whole Genome Sequencing

Exponentially growing T-47D cells were cultured in T175 flasks and were harvested using trypsin digestion. Genomic DNA was extracted using a Qiagen Genomic-tips 500/G kit. Sequencing libraries for whole genome sequencing with DNA nanoball (DNB) technology were constructed (Beijing Genomics Institute (BGI) Inc., Hong Kong). Sequencing was carried out using BGISEQ-500 sequencer with a mean sequencing coverage of greater than 30× for each of the samples using 2 × 150 bp configuration. Sequence reads from WGS was aligned to GRCh38 genome assembly using Burrow-Wheelers aligner [50].

RNA sequencing and transcript quantification

Total RNA from T-47D cells were extracted using the MagNA pure 96 platform (Roche), and RNA integrity number (RIN) was determined by BioAnalyzer 2100 (Agilent). RNA libraries were constructed and sequenced using poly-dT enrichment or ribosomal RNA depletion sample preparation methods in conjunction with DNB sequencing technology on a BGISEQ-500 instrument (BGI). Raw reads were aligned to GRCh38 genome assembly and GENCODE [51] GRCh38.p13 annotation using STAR programme [52], followed by processing of the resultant file with SAMtools [53]. Read counting on genomic features was carried out using Rsubread [54], and the R package DESeq2 [55] was used to perform statistic-based quantification. GENCODE genes with normalised read count of top 60% were considered being expressed in T-47D cells.

Further method details are provided in the supplementary text.

Declarations

Acknowledgements

We thank Dr. Alexia Hervieu Viches, and Dr. Alexandra Vasile form the Institute of Cancer Research, London, for helpful discussions and contributing to the idea of this work. We are grateful for the Scientific Computing facility at the Institute of Cancer Research, and NeoTrident Technology Ltd. for providing computational resources and relevant training. This work was financially supported by Cancer Research UK (C309/A11566 and C2739/A22897) and The Institute of Cancer Research (London, United Kingdom). C.Z. was sponsored by Shanghai Pujiang Programme (22PJD104), and Science and Technology Commission of Shanghai (23S11901100).

Author Contributions

C.Z., X.L., O.R. and P.A.C, conceived this study and provided grant support. C.Z., M.W., F.X. and X.L. conducted the experiments and analysis unless otherwise noted. C.Z., and B.C., conducted analysis on WGS data. K.M. provided guidance on data analysis using sequencing data. C.Z., X.L. and P.A.C. prepared the manuscript.

Competing Interests

C.Z., K.M., O.R. and P.A.C. are current or past employees of The Institute of Cancer Research, which has a commercial interest in the discovery and development of A3B inhibitors and operates a reward to inventors' scheme. Separately, C.Z., M.W., F.X. and X.L. are employees of Shanghai Institute of Biological Products, an entity presently engaged in the commercial development of therapeutic biologics. No conflicts of interest have been reported by the remaining authors.

Data availability

The high through-put sequencing data, together with processed data files, was available on GEO database with accession number GSE245700 and GSE245701.

References

1. Qiu L, Jing Q, Li Y, Han J. RNA modification: mechanisms and therapeutic targets. *Mol Biomed* 2023; 4: 25.
2. Slotkin W, Nishikura K. Adenosine-to-inosine RNA editing and human disease. *Genome Med* 2013; 5: 105.
3. Pecori R, Di Giorgio S, Paulo Lorenzo J, Nina Papavasiliou F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat Rev Genet* 2022; 23: 505-518.
4. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* 2008; 9: 229.
5. Fritz EL, Rosenberg BR, Lay K, Mihailovic A, Tuschl T, Papavasiliou FN. A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat Immunol* 2013; 14: 749-755.
6. Mertz TM, Collins CD, Dennis M, Coxon M, Roberts SA. APOBEC-Induced Mutagenesis in Cancer. *Annu Rev Genet* 2022; 56: 229-252.
7. Silvas TV, Schiffer CA. APOBEC3s: DNA-editing human cytidine deaminases. *Protein Sci* 2019; 28: 1552-1566.
8. Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, Baysal BE. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat Commun* 2015; 6: 6881.
9. Baysal BE, De Jong K, Liu B, Wang J, Patnaik SK, Wallace PK, Taggart RT. Hypoxia-inducible C-to-U coding RNA editing downregulates SDHB in monocytes. *PeerJ* 2013; 1: e152.
10. Alqassim EY, Sharma S, Khan A, Emmons TR, Cortes Gomez E, Alahmari A *et al*. RNA editing enzyme APOBEC3A promotes pro-inflammatory M1 macrophage polarization. *Commun Biol* 2021; 4: 102.
11. Farajollahi S, Maas S. Molecular diversity through RNA editing: a balancing act. *Trends Genet* 2010; 26: 221-230.
12. Salter JD, Bennett RP, Smith HC. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* 2016; 41: 578-594.
13. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K *et al*. Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012; 149: 979-993.
14. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X *et al*. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016; 534: 47-54.
15. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B *et al*. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 2013; 494: 366-370.
16. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* 2013; 45: 977-983.
17. Salamango DJ, McCann JL, Demir O, Brown WL, Amaro RE, Harris RS. APOBEC3B Nuclear Localization Requires Two Distinct N-Terminal Domain Surfaces. *J Mol Biol* 2018; 430: 2695-2708.

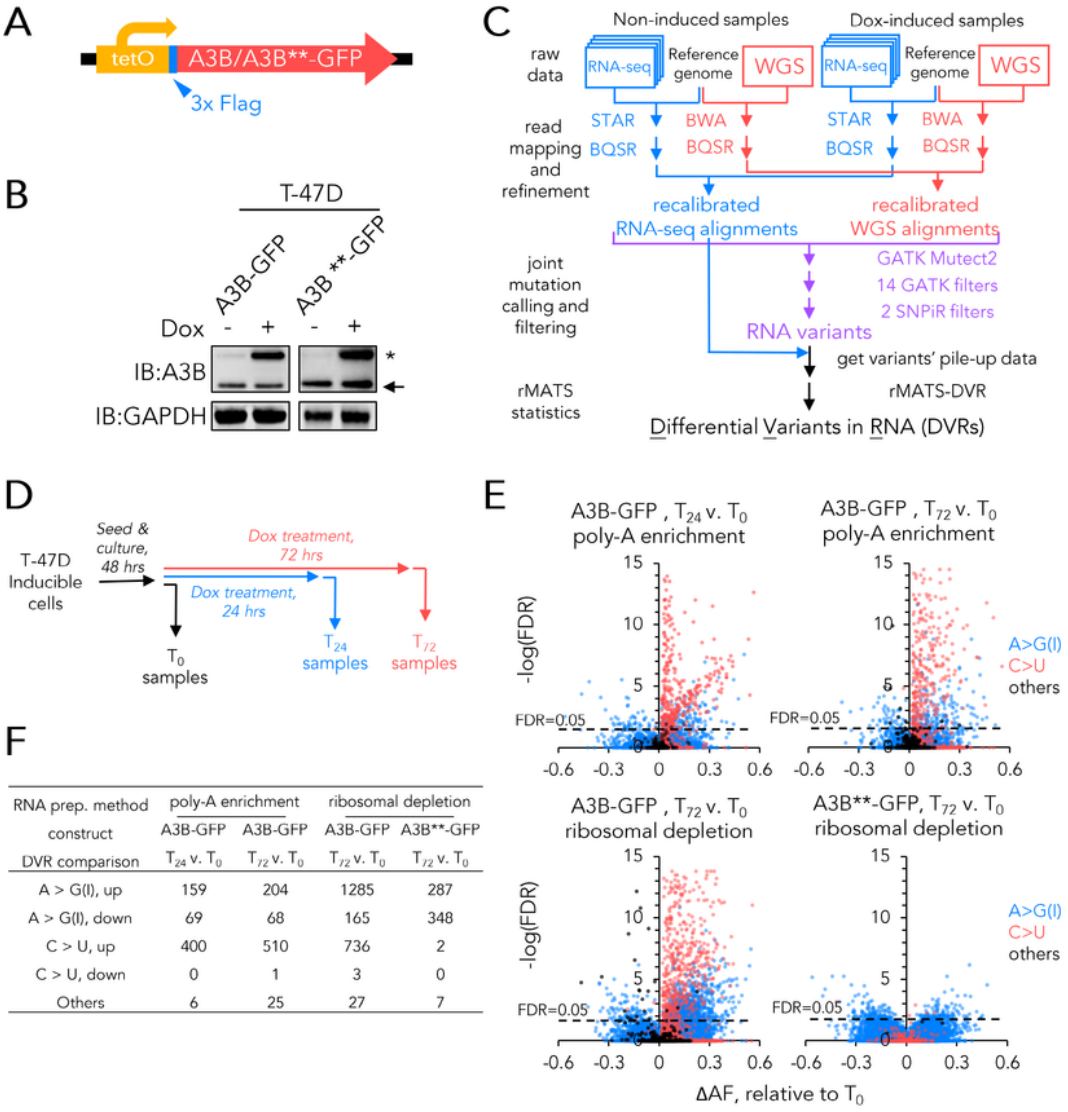
18. Prohaska KM, Bennett RP, Salter JD, Smith HC. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley Interdiscip Rev RNA* 2014; 5: 493-508.
19. Salter JD, Smith HC. Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands. *Trends Biochem Sci* 2018; 43: 606-622.
20. Periyasamy M, Patel H, Lai CF, Nguyen VTM, Nevedomskaya E, Harrod A *et al*. APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. *Cell Rep* 2015; 13: 108-121.
21. Serebrenik AA, Starrett GJ, Leenen S, Jarvis MC, Shaban NM, Salamango DJ *et al*. The deaminase APOBEC3B triggers the death of cells lacking uracil DNA glycosylase. *Proc Natl Acad Sci U S A* 2019; 116: 22158-22163.
22. Wang D, Li X, Li J, Lu Y, Zhao S, Tang X *et al*. APOBEC3B interaction with PRC2 modulates microenvironment to promote HCC progression. *Gut* 2019; 68: 1846-1857.
23. Neums L, Suenaga S, Beyerlein P, Anders S, Koestler D, Mariani A, Chien J. VaDiR: an integrated approach to Variant Detection in RNA. *Gigascience* 2018; 7: 1-13.
24. Wang J, Pan Y, Shen S, Lin L, Xing Y. rMATS-DVR: rMATS discovery of differential variants in RNA. *Bioinformatics* 2017; 33: 2216-2217.
25. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 2019.
26. Zhang C, Lu Y, Chen B, Bai Z, Hervieu A, Licciardello M *et al*. R-loop editing by DNA cytosine deaminase APOBEC3B determines the activity of estrogen receptor enhancers. *bioRxiv* 2022.
27. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res* 2008; 36: W70-74.
28. Mansi L, Tangaro MA, Lo Giudice C, Flati T, Kopel E, Schaffer AA *et al*. REDIportal: millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Res* 2021; 49: D1012-D1019.
29. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999; 9: 677-679.
30. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B *et al*. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 2016; 13: 508-514.
31. Krakau S, Richard H, Marsico A. PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol* 2017; 18: 240.
32. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A *et al*. The Ensembl Variant Effect Predictor. *Genome Biol* 2016; 17: 122.
33. Ursu A, Childs-Disney JL, Angelbello AJ, Costales MG, Meyer SM, Disney MD. Gini Coefficients as a Single Value Metric to Define Chemical Probe Selectivity. *ACS Chem Biol* 2020; 15: 2031-2040.

34. Roux KJ, Kim DI, Burke B, May DG. BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* 2018; 91: 19 23 11-19 23 15.
35. Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res* 2012; 40: e128.
36. Wang Z, Li K, Huang W. Long non-coding RNA NEAT1-centric gene regulation. *Cell Mol Life Sci* 2020; 77: 3769-3779.
37. Arun G, Aggarwal D, Spector DL. MALAT1 Long Non-Coding RNA: Functional Implications. *Noncoding RNA* 2020; 6.
38. Martin AS, Salamango DJ, Serebrenik AA, Shaban NM, Brown WL, Harris RS. A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Sci Rep* 2019; 9: 497.
39. Zhang YH, Guo XC, Zhong JB, Zhong DX, Huang XH, Fang ZY *et al*. Discovery of APOBEC Cytidine Deaminases Inhibitors Using a BspH1 Restriction Enzyme-Based Biosensor. *Chemistryselect* 2022; 7.
40. Petljak M, Dananberg A, Chu K, Bergstrom EN, Striepen J, von Morgen P *et al*. Mechanisms of APOBEC3 mutagenesis in human cancer cells. *Nature* 2022; 607: 799-807.
41. Jalili P, Bowen D, Langenbucher A, Park S, Aguirre K, Corcoran RB *et al*. Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. *Nat Commun* 2020; 11: 2971.
42. Maiti A, Hou S, Schiffer CA, Matsuo H. Interactions of APOBEC3s with DNA and RNA. *Curr Opin Struct Biol* 2021; 67: 195-204.
43. Xiao X, Yang H, Arutiunian V, Fang Y, Besse G, Morimoto C *et al*. Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. *Nucleic Acids Res* 2017; 45: 7540.
44. Krokan HE, Saetrom P, Aas PA, Pettersen HS, Kavli B, Slupphaug G. Error-free versus mutagenic processing of genomic uracil--relevance to cancer. *DNA Repair (Amst)* 2014; 19: 38-47.
45. Law EK, Sieuwerts AM, LaPara K, Leonard B, Starrett GJ, Molan AM *et al*. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci Adv* 2016; 2: e1601737.
46. Venkatesan S, Rosenthal R, Kanu N, McGranahan N, Bartek J, Quezada SA *et al*. Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann Oncol* 2018; 29: 563-572.
47. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI *et al*. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell* 2014; 55: 791-802.
48. Jakobsdottir GM, Brewer DS, Cooper C, Green C, Wedge DC. APOBEC3 mutational signatures are associated with extensive and diverse genomic instability across multiple tumour types. *BMC Biol* 2022; 20: 117.
49. Butler K, Banday AR. APOBEC3-mediated mutagenesis in cancer: causes, clinical significance and therapeutic potential. *J Hematol Oncol* 2023; 16: 31.

50. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-1760.
51. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F *et al*. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012; 22: 1760-1774.
52. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15-21.
53. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* 2021; 10.
54. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 2019; 47: e47.
55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15: 550.

Figures

Figure 1



F

RNA prep. method	poly-A enrichment		ribosomal depletion	
construct	A3B-GFP	A3B-GFP	A3B-GFP	A3B**-GFP
DVR comparison	T_{24} v. T_0	T_{72} v. T_0	T_{72} v. T_0	T_{72} v. T_0
A > G(I), up	159	204	1285	287
A > G(I), down	69	68	165	348
C > U, up	400	510	736	2
C > U, down	0	1	3	0
Others	6	25	27	7

Figure 1

Identification of A3B-mediated DVRs using a lentiviral inducible system. (A). Schematic of the A3B wildtype or catalytically dead (**) lentiviral inducible expression cassette used in this study. (B). Immunoblots showing the expression of endogenous (arrow) and inducible-exogenous (asterisk) A3Bs following induction of T-47D cells using 100 ng/ml doxycycline for 48 hours. GAPDH was used as loading control. (C). Schematic of the analysis pipeline for identifying A3B-mediated DVRs. For RNA-seq,

n = 4 biological repeats were used. For WGS, n = 1 was used. (D). Schematic of treatment processes for RNA-seq samples. (E). Volcano plot depicting fold change of alternative allele fraction for each DVRs (ΔAF) upon A3B induction against FDR derived from log ratio test, where AF = (read depth of alternative allele)/(total read depth), and $\Delta AF=AF_{dox\text{-}treated}-AF_{dox\text{-}untreated}$. In this study, FDR of ≤0.05 (dotted line) was applied as criteria for DVRs qualification. (F). Table showing break down of various types of DVRs in different sample groups relative to non-induced samples.

Figure 2

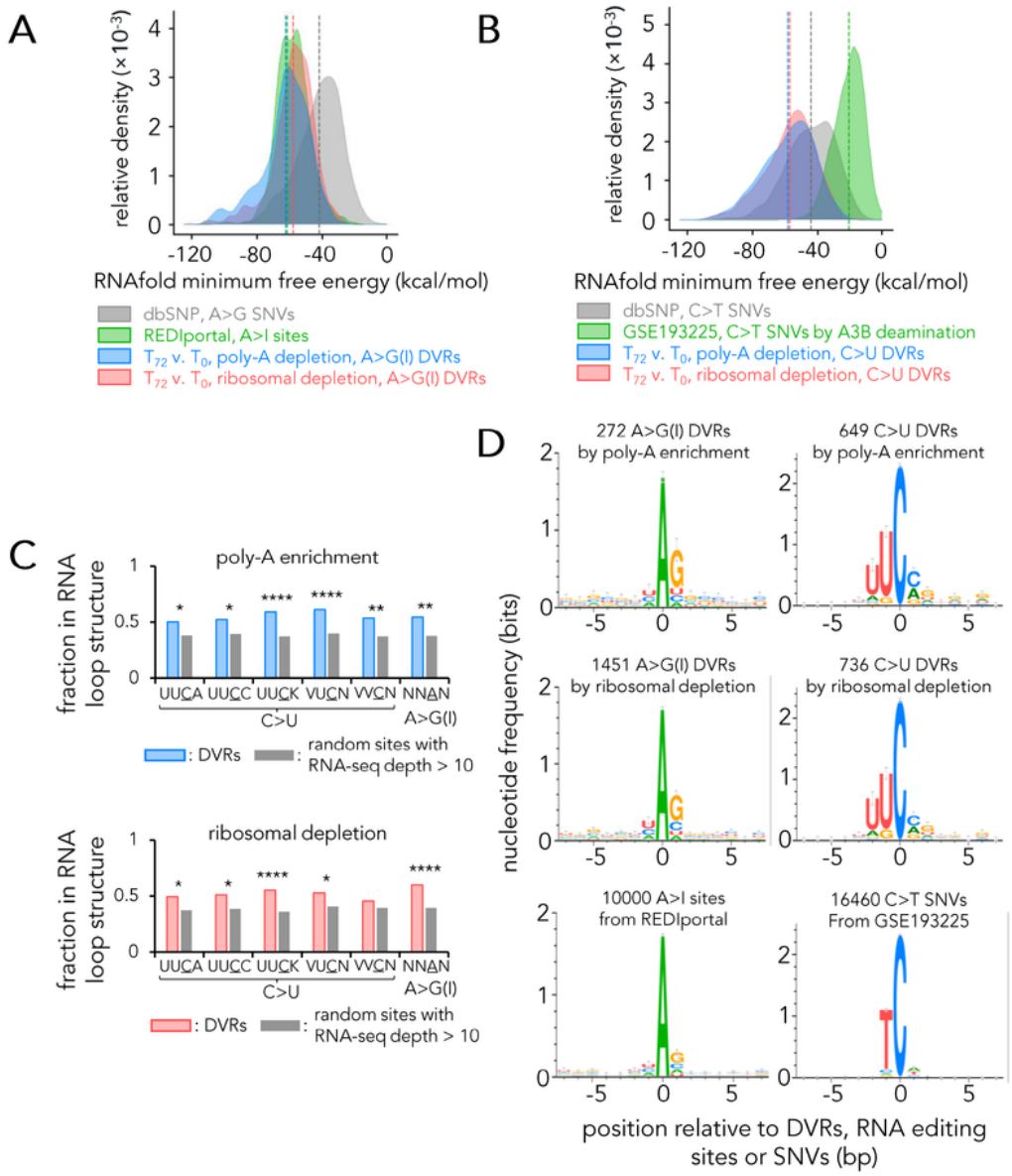


Figure 2

Characterisations of A3B-mediated DVRs. (A). RNA folding minimum free energy predicted by RNAfold for A>G genomic DNA SNVs, known RNA A>I sites and A>G(I) DVRs identified in this study. (B). RNA folding minimum free energy predicted for C>T genomic DNA SNVs, known A3B-mediated genomic DNA C>U editing sites and C>U DVRs identified in this study. Random sampling was applied for SNVs from dbSNP database (n = 100,000) and A>I sites (n = 10,000) from REDIportal. (C). Fraction of indicated types of DVRs or control sites within loop structure predicted by RNAfold. For random control, n = 10,000 random sites with RNA-seq read depth of greater than 10 in T-47D transcriptome were sampled. Error bars denote SD. *, ** and **** denote $p<0.05$, 0.01 and 0.0001 respectively using Z-test. (D). Consensus sequence logo plots for sequences centred at editing sites were generated by WebLogo3.

Figure 3

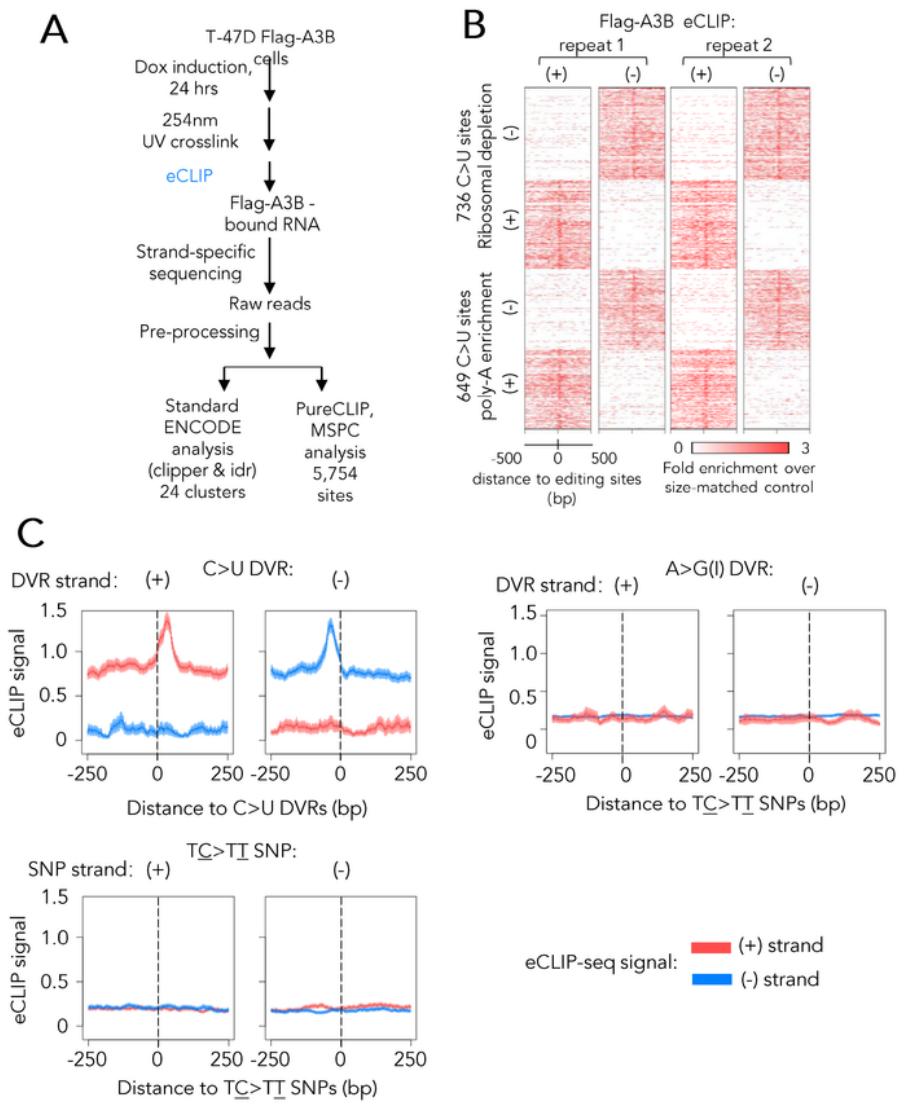


Figure 3

Validation of A3B-mediated C>U DVRs using eCLIP-seq. (A). Schematic of the eCLIP-seq and analysis pipeline in this study. (B). Heat maps showing eCLIP-seq signals in regions flanking all C>U DVRs identified upon A3B induction. Results from n = 2 eCLIP-seq signal from different strands are shown separately. (C). Profile plot showing eCLIP-seq signals in regions flanking all C>U DVRs identified upon A3B induction. For comparison, eCLIP-seq signals flanking A3B-mediated TC>TT SNPs identified from

GSE193225 are shown. For (B) and (C), data represent fold enrichment of eCLIP-seq signal over size-matched control.

Figure 4

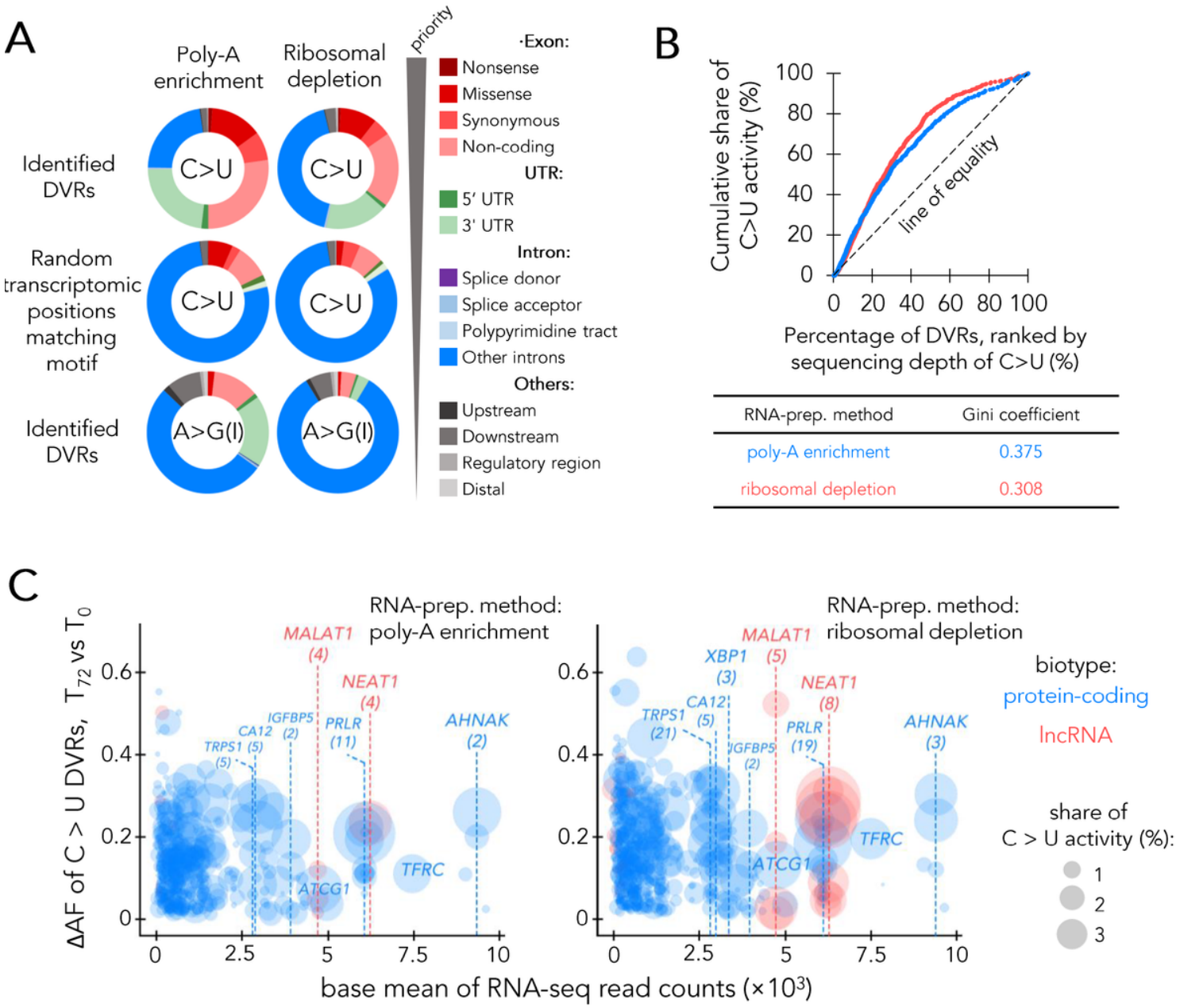


Figure 4

A3B edits RNA in a selective manner. (A). Distributions for editing sites across different types of RNA features by type of RNA editing. For variant with multiple editing consequences, only the one with highest impact predicted by VEP was counted. The rank of impact was shown on the right. (B). Lorenz curve

showing the level of inequality distribution of A3B's activity across identified C>U DVRs, and the calculated Gini coefficients for different sample group. (C) Bubble plot showing distribution of A3B activity across identified C>U DVRs. Vertical axis shows difference in faction of U bases, and horizontal axis shows level of gene expression for edited genes (as normalised read counts by DESeq2, also referred to as base mean), with size of the bubble denote share of total A3B's C>U editing activity. Numbers within parenthesis denote the number of DVRs within indicated genes.

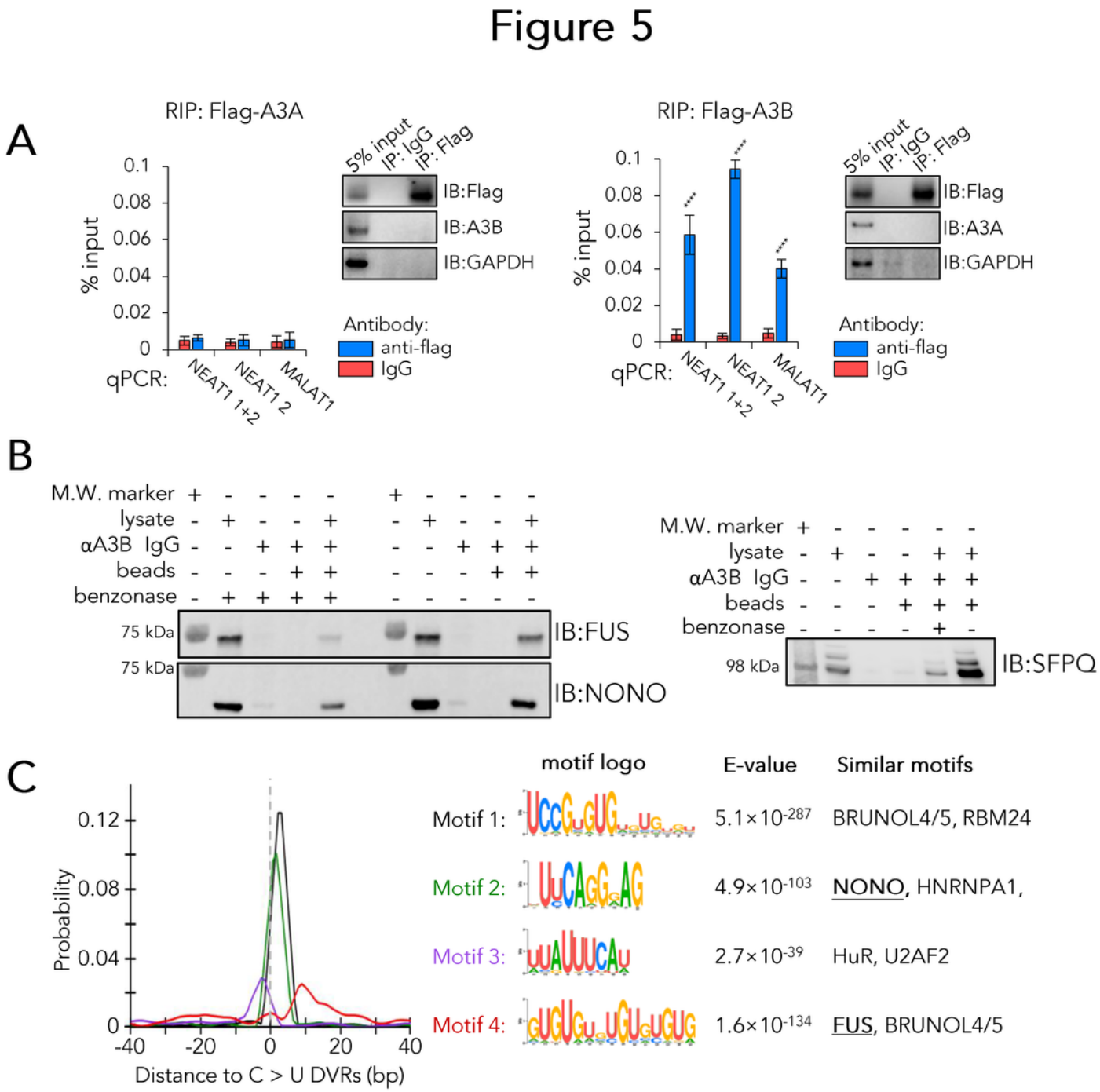


Figure 5

NEAT1 and MALAT1 binds to A3B. (A). Results from RIP-qPCR with co-immunoprecipitation of over-expressed flag-tagged A3A (left) or A3B(right) in T-47D cells. Data denotes mean of n=3 biological replicates, and error bars for SD. (B). Representative results for co-immunoprecipitation of A3B in T-47D cells. (C). Analysis of motif enrichment at ± 40 bp sequences centred at all C>U DVRs identified in this study. Profiles showing probability of top-ranking sequence motifs were shown on the left, and the details for the top-ranking sequence motifs were shown on the right.

Figure 6

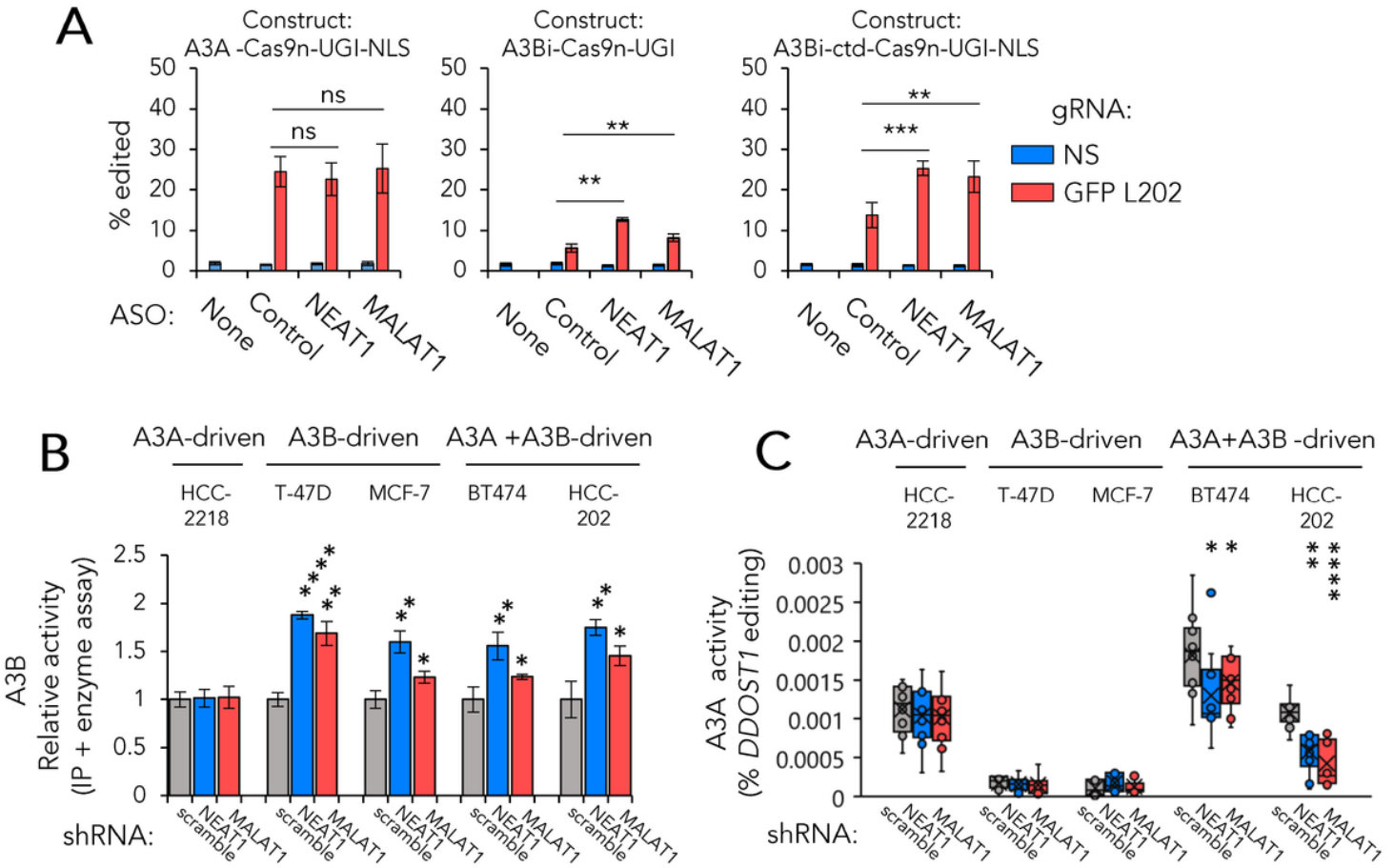


Figure 6

NEAT1 and MALAT1 regulates A3B's activity in cells. (A). Quantification of the effect of lncRNA depletion using ASOs on APOBEC3 enzymes' activity in cells with the APOBEC-Cas9n reporter system. 293T cells

were used in this study. Data denotes mean of n=3 biological repeats, and error bars for SD. (B). Quantification of A3B activity in cells using the BspH1 biosensor with immunoprecipitated A3B complex in breast cancer cell lines stably expressing shRNAs. Bar graph was shown with data normalised using scramble shRNA control. Data represent mean of n=3 biological repeats, and error bars for SD. (C). Quantification of A3A activity in cells by level of *DDOST1* 558 C>U editing, measured by ddPCR. Box plot representing n=9 biological repeats. *, **, ***, **** and ns denote $p<0.05$, 0.01, 0.001, 0.0001 and non-significant using two-tailed Student's T test, evaluating difference against scramble control shRNA-expressing samples.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalList1.xlsx
- SupplementalList2.xlsx
- Manuscript20231116SuppFigures.pdf
- Manuscript20231116SuppText.docx