Genome-wide identification and analysis of A-to-I RNA editing events in the malignantly transformed cell lines from BEP2D induced by $\alpha$-particles radiation

Qiaowei Liu$^{1,2,3,}$†, Hao Li$^{1,}$†, Lukuan You$^{1,2}$, Tao Li$^{1,2}$, Lingling Li$^{1,2}$, Pingkun Zhou$^{3}$, Xiaochen Bo$^{3}$, Hebing Chen$^{3,*}$, Xiaohua Chen$^{3,*}$, Yi Hu$^{1,2,*}$

1 Medical School of Chinese PLA, Beijing 100853, P.R. China;
2 Department of Medical Oncology, Chinese PLA General Hospital, Beijing 100853, P.R. China;
3 Beijing Institute of Radiation Medicine, Beijing 100850, P.R. China;
Correspondence: huyi0401@aliyun.com (Y.H.); chenxh0676@sina.com (X.C.); chb-1012@163.com (H.C.);
†These authors contributed equally to this work.
Abstract

Adenosine (A) to inosine (I) RNA editing is the most prevalent RNA editing mechanism in humans and play critical roles in tumorigenesis. However, the effects of radiation on RNA editing and the mechanisms of radiation-induced cancer were poorly understood. Here, we analyzed human bronchial epithelial BEP2D cells and radiation-induced malignantly transformed cells with next generation sequencing. By performing an integrated analysis of A-to-I RNA editing, we found that genome-encoded single-nucleotide polymorphisms (SNPs) might induce the downregulation of ADAR2 enzymes, and further caused the abnormal occurrence of RNA editing in malignantly transformed cells. These editing events were significantly enriched in differentially expressed genes between normal cells and cancer cells. In addition, oncogenes CTNNB1 and FN1 were highly edited and significantly overexpressed in cancer cells, thus may be responsible for the lung cancer progression. Our work provides a systematic analysis of RNA editing from lung tumor specimens with high-throughput RNA sequencing and DNA sequencing. Moreover, these results demonstrate further evidence for RNA editing as an important tumorigenesis mechanism.

Keywords: RNA editing, radiation-induced cancer, BEP2D, oncogenes

Introduction

Lung cancer remains the leading cause of cancer death in both men and women, and radon exposure is the second most common cause of lung cancer after smoking [1]. However, the molecular mechanisms of radon-induced lung cancer remain unclear.

Adenosine (A) to inosine (I) RNA editing is the most prevalent RNA editing mechanism in humans, where ADAR enzymes convert A to I at specific nucleotide sites of select transcripts without affecting the DNA sequence identity [2]. Intriguingly, RNA editing have been reported to introduce many nucleotide changes in tumorigenesis, such as AZIN1 editing in liver cancer [3], PODXL editing in gastric cancer [4], RHOQ editing in colorectal cancer [5], and GABRA3 editing in breast cancer [6]. Recent study has shown that increased recoding RNA editing of DNA repair enzyme NEIL1 (K242R) is identified in non-small cell lung cancer samples as a result of ADAR gene amplification [7]. However, there were limited studies to date in further exploring the characteristics of RNA editing in lung cancer. What’s more, the effect of radiation on RNA editing is poorly understood.
Here, we investigate A-to-I RNA editing in human bronchial epithelial cells (BEP2D) and malignantly transformed cells (BERP35T1 and BERP35T4), which are important models to characterize the radiation-mediated carcinogenesis of lung [8-10]. By performing high-throughput RNA sequencing, we identified A-to-I editing sites with three robust bioinformatics methods. We then systemically compared editing events in normal cells and cancer cells. Further, by performing genome-wide DNA sequencing, we revealed that the genomic variants in \textit{ADAR2} gene was responsible for the abnormal editing events in cancer cells. Finally, we reported two potential editing genes, \textit{CTNNB1} and \textit{FN1}, in lung cancer.

**Results**

**Identification of A-to-I RNA editing**

The prevalence and importance of A-to-I RNA editing have been illuminated in recent years largely owing to the rapid adoption of high-throughput sequencing technologies [11,12]. To analyze A-to-I RNA editing in BEP2D cells and malignantly transformed cell lines, we performed high-throughput RNA sequencing (RNA-Seq) on BEP2D cell and transformed BEP2D cells, which were irradiated with 1.5Gy dose of $\alpha$-particles emitted by $^{238}\text{PuO}_2$. Two transformed cell lines, BERP35T1 and BERP35T4, were investigated (Fig. 1A). Three biological replicates were sequenced and analyzed for each cell line. We then calculated gene expression level with Cufflinks program [13], for each cell line, biological replicates of RNA-seq showed highly reproducible results (Fig. 1B). Thus, our sequencing data were of high quality for downstream analysis.

Recent studies have reported that the most challenging part of identifying RNA editing is the discrimination of RNA editing sites from genome-encoded single-nucleotide polymorphisms (SNPs) and technical artifacts caused by sequencing or read-mapping errors [14-16]. To accurately identify RNA editing sites, we performed three famous methods including GIREMI [17], RNAEditor [18] and Separate method from Jin Billy Li [14] (See Methods). The GIREMI method combines statistical inference of mutual information (MI) between pairs of single-nucleotide variants (SNVs) in RNA-seq reads with machine learning to predict RNA editing sites. RNAEditor identify RNA editing by detecting ‘editing islands’. Separate method from Jin Billy Li identify RNA editing sites by strict filtering processes. For each sample, we only used RNA editing sites that can be detected in all three methods. For each cell line, we combined RNA editing sites from three biological replicates. As the most prevalent editing type in humans is adenosine-to-inosine (A-to-I) editing and most noncanonical editing are false positives [19],
we only analyzed A-to-I RNA editing in this study. Final, 5659, 3820, 2446 A-to-I RNA editing sites were identified in BEP2D cell line and transformed cell lines BERP35T1 and BERP35T4, respectively (Table 1 and Supplemental Table 1).

**A-to-I RNA editing and associated genes in BEP2D and malignantly transformed cell**

We next investigated the difference of A-to-I RNA editing between BEP2D cell line and malignantly transformed cell lines. First, 3,683~4,217 editing events were disappeared and 1,004~1,844 new editing events occurred from normal BEP2D cell to malignantly transformed cells, indicating a dramatic changes of RNA editing when BEP2D cell was irradiated (Fig. 2A). Generally, A-to-I editing is pervasive in Alu repeats because of the double-stranded RNA structures formed by inverted Alu repeats in many genes [20,21]. We found that although RNA editing is quite different in BEP2D and malignantly transformed cell lines, editing sites are still conserved in Alu repeats (Fig. 2B) and ~60% of RNA editing events occurred in intergenic regions (Fig. 2C), thus, the basic distribution of A-to-I RNA editing did not change.

Next, we examined genes targeted by A-to-I RNA editing sites (editing genes for short). In general, 484, 426 and 305 genes were edited in BEP2D, BERP35T1 and BERP35T4 cell lines (Supplemental Table 2). We found that, in BEP2D cell line, 88% of editing genes were targeted by BEP2D-specific editing sites, but in BERP35T1 and BERP35T4, only 70% and 53% of editing genes were targeted by cell-specific editing sites (Fig. 2D). This result suggested that the editing rate of genes decreased when cell was irradiated and malignantly transformed. In addition, we performed gene ontology (GO) analysis to reveal the biological function of editing genes. We found that editing genes were enriched in different biological processes. For BEP2D, editing genes were enriched in protein processes and translation process, for BERP3T1, nervous system development and hemophilic cell adhesion process were highlighted and editing genes were enriched in DNA replication in BERP35T4 cell line (Fig. 2E).

To examine whether RNA editing affects transcription activity, we identified differentially expressed genes (DEGs) by performing Cuffdiff program [13] (Supplemental Table 3). We found that DEGs between normal BEP2D cell and malignantly transformed cells were significantly enriched in genes with RNA editing events ($p$ value < 1E-3, hypergeometric test, Fig. 2F). This observation indicated that the dynamics of RNA editing sites was related to gene dysregulation and may further induce tumorigenesis.
**ADAR2 down-regulation by genome SNPs**

We next investigated the mechanism responsible for the differences observed in RNA editing between normal BEP2D cell and malignantly transformed cells. In human, A-to-I editing is performed by the ADAR family, which contains 3 genes: **ADAR1**, **ADAR2** and **ADAR3** [22-24]. We thus examined the transcript levels of **ADAR** genes. The expression level of **ADAR1** in BEP2D cell was comparable to that in BERP35T1 and BERP35T4 cell and **ADAR3** was silence in both BEP2D cell and malignantly transformed cells (Fig. 3A). However, **ADAR2** expression level significantly reduced in BERP35T1 and BERP35T4 (Fig. 3B). Previous studies have confirmed that **ADAR2** is lowly expressed in cancer e.g. glioblastoma [25], gastric cancer[4]. Thus the tumor progression of BEP2D seems to mainly be induced by **ADAR2** downregulation.

We further explored the possible mechanism of **ADAR2** downregulation in malignantly transformed cells. Radiation induced DNA alterations change gene expression and further increase cancer risk [26,27]. We next performed DNA sequencing (DNA-seq) on BEP2D cell line and malignantly transformed cells. We identified genome-encoded single-nucleotide polymorphisms (SNPs) using GATK pipeline [28] (See Methods). Surprisingly, nearly 2-fold SNPs in **ADAR2** gene were detected in malignantly transformed cells compared to normal BEP2D (Fig. 3C, Supplemental Table 4) and more SNPs in **ADAR2** exon were observed (Fig. 3D). For example, known SNP rs11701974, a genetic variant of HLA-DQB1 associated with human longevity [29], was detected in 3’ UTR of **ADAR2** and specific in BERP35T1 cell and BERP35T4 cell (Fig. 3E). Moreover, we identified 32 and 24 novel SNPs in BERP35T1 cell and BERP35T4 cell, respectively (Supplemental Table 4). For example, chr21:46643782 was altered in malignantly transformed cells (Fig. 3F). These results indicate that genomic variants induced by radiation may lead **ADAR2** downregulation and further decrease the editing rate of malignantly transformed cells.

**Oncogene CTNNB1 and FN1 are highly edited and significantly overexpressed in malignantly transformed cell**

To gain insights into the biological relevance of RNA editing in malignantly transformed cell, we investigated 285 oncogenes from previous studies [30] (Supplemental Table 5). We found that oncogenes **CTNNB1**, **PABPC1** and **VHL** were edited in BERP35T1 cell. Notably, the expression level of **CTNNB1** in BERP35T1 cell was significantly higher than that in BEP2D cell (p-value = 0.00185, Fig. 4A). Previous studies reported that activating mutations in **CTNNB1** have oncogenic activity resulting in tumor development and somatic mutations are found in various tumor types [31-34]. We found two A-
to-I editing events (chr3:41262966 and chr3:41262974) occurred in BERP35T1 cell and CTNNB1 was overexpression in BERP35T1 cell (Fig. 4B). Similarly, we found three oncogenes FNI, METTL14 and VHL were edited in BERP35T4 cell. Notably, the expression level of FNI in BERP35T4 cell was significantly higher than that in BEP2D cell (p-value = 5E-5, Fig. 4C). Previous studies have reported that transcriptional activation of FNI and gene fusions of FNI promote the malignant behavior of multiple cancers [35-37]. A strong A-to-I editing event (chr2:216236508) and a weak A-to-I editing event (chr2:216236482) were observed in BERP35T4 and FNI was overexpression in BERP35T4 cell (Fig. 4D). These results suggest that RNA editing are associated with oncogene overexpression and may further induce cancer progression.

Materials and Methods

Cell culture

The BEP2D cell line is a human papilloma-virus (HPV18)-immortalized human bronchial epithelial cell line and was established by Dr Curtis C. Harris (National Cancer Institute, MD, USA) [38]. These cells are anchorage dependent and non-tumorigenic in late passages. We got the authorization for research use only from Dr Curtis C. Harris and the Passage 20 of the BEP2D cell line was kindly provided by Tom K Hei (Center for Radiological Research, College of Physician and Surgeons, Columbia University, New York, USA) in the summer of 1993. The authentication of BEP2D cell line was tested by using short tandem repeats (STRs) analysis in June 2018. Although the information of BEP2D cells was not found in DSMZ and ATCC, our STRs results did match BBM cell line (ATCC Cell No. CRL-9482), BZR cell line (ATCC Cell No. CRL-9483) and BEAS-2B cell line (ATCC Cell No. CRL-9609), which are three tranformants derived from human bronchial epithelial cell (supplementary S1). The BERP35T1 and BERP35T4 malignant transformant cell lines were derived from BEP2D cells irradiated with 1.5 Gy of α-particle emitted from $^{238}$Pu source and were described in detail in a previous paper[10]. The cells were cultured in serum-free LHC-8 medium (Gibco, USA) at 37°C under a 95% air/5% CO$_2$ atmosphere.

RNA sequencing

Total RNAs were extracted from cells with RNAiso Reagent (TaKaRa, Dalian, China) following the manufacturer’s instruction. RNA degradation and contamination was monitored on 1% agarose gels. RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies,
CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total amount of 1 µg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer’s recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3’ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer’s instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq platform and 150 bp paired-end reads were generated.

**DNA sequencing**

Total DNAs were extracted from cells with DNAiso Reagent (TaKaRa, Dalian, China) following the manufacturer’s instruction. The quality of isolated genomic DNA was verified by using these two methods in combination: (1) DNA degradation and contamination were monitored on 1% agarose gels. (2) DNA concentration was measured by Qubit® DNA Assay Kit in Qubit® 2.0 Flurometer(Life Technologies, CA, USA). A total amount of 1µg DNA per sample was used as input material for the DNA library preparations. Sequencing library was generated using Truseq Nano DNA HT Sample Prep Kit (Illumina, USA) following manufacturer’s recommendations and index codes were added to each sample. Briefly, genomic DNA sample was fragmented by sonication to a size of 350 bp. Then DNA fragments were endpolished, A-tailed, and ligated with the full-length adapter for Illumina sequencing, followed by further PCR amplification. After PCR products were purified (AMPure XP system), libraries
were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified by real-time PCR (3nM). The clustering of the index-coded samples was performed on a cBot Cluster Generation System using Hiseq PE Cluster Kit (Illumina) according to the manufacturer’s instructions. After cluster generation, the DNA libraries were sequenced on Illumina Hiseq platform and 150 bp paired-end reads were generated.

**RNA editing identification**

We adopted three previously published methods to accurately identify A-to-I RNA editing sites.

For Jinbilly’s method [14], we used the Burrows-Wheeler algorithm (BWA)[39] to align RNA-seq reads to a combination of the reference genome (hg19) and exonic sequences surrounding known splice junctions from available gene models. We chose the length of the splice junction regions to be slightly shorter than the RNA-seq reads to prevent redundant hits. Picard (http://picard.sourceforge.net/) was then used to remove identical reads (PCR duplicates) that mapped to the same location. GATK tools were used to perform local realignment around insertion and/or deletion polymorphisms and to recalibrate base quality scores. Variant calling was performed using GATK UnifiedGenotyper tool with options stand_call_conf of 0 and stand_emit_conf of 0. Further filtering were performed as described [40].

For GIREMI method [17], RNA-seq mapping and preprocessing was same as Jinbilly’s method. For each mismatch position, a total read coverage of \( \geq 5 \) was required and the variant allele was required to be present in at least three reads. We then the following types of mismatches: those located in simple repeats regions or homopolymer runs of \( \geq 5 \) nt, those associated with reads substantially biased toward one strand, those with extreme variant allele frequencies (>95% or <10%) and those located within 4 nt of a known spliced junction. Finally, we perform GIREMI tool to call RNA editing.

For RNAEditor [18], fastq format files from RNA-seq data were directly used as input for RNAEditor tools to call RNA editing sites.

**SNP identification**

We used the Bowtie2 [41] to align DNA-seq reads to reference genome hg19, Picard (http://picard.sourceforge.net/) was then used to remove identical reads (PCR duplicates) that mapped to the same location. GATK tools were used to perform local realignment around insertion and/or deletion polymorphisms and to recalibrate base quality scores. Variant calling was performed using GATK UnifiedGenotyper tool.
Statistical analysis

Gene expression was calculated using Cufflinks program default parameters. Differentially expressed genes (DEGs) were identified by Cuffdiff program, three biological replicates for each cell line were combined as the input of Cuffdiff and a p-value was reported to show the significance of DEGs. RNA editing targeted genes were assigned with ‘bedops’ program. GO analysis were performed by using DAVID [42].

Discussion

Radon is an important environmental factor that is associated with the development of human bronchogenic Carcinoma. In order to elucidate the cellular and molecular mechanisms of radon-induced lung cancer, we established a model system of α-particle transformed human cells in our previous work [43], a number of alterations in cytogenetics [44,45], genomic instability [46], DNA repair [10], and gene expression [47] were found in these cell models. However, a genome-wide systematic analysis of this model based on next generation of sequencing is absent. In this work, we performed high-throughput RNA sequencing and genome-wide DNA sequencing to this model and discovered a new mechanism probably for tumorigenesis.

We all known the radon-oncogenic effect is DNA damage, but there were few articles about the effects of radiation on RNA editing. In this work, we provided genome-wide identification and analysis of A-to-I RNA editing events in the malignantly transformed cell line BEP2D induced by α-particles radiation, the results show that RNA editing sites changed greatly and the total amount decreased after radiation.

Although 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. 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 [47,48], but little is known about how RNA editing operates in cancer [49]. Our work demonstrated that, in these cell models, genome-encoded single-nucleotide polymorphisms (SNPs) induced the downregulation of ADAR2 enzymes, and further caused the abnormal occurrence of RNA editing in malignantly transformed cells. Then, the abnormal occurrence of RNA editing led to abnormal expression of oncogenes, such as, CTNNB1 and FNI, thus may be responsible for the lung cancer progression. These results demonstrate further evidence for RNA editing as an important tumorigenesis mechanism, and RNA editing sites...
might represent a new class of therapeutic targets.

Acknowledgements

We gratefully acknowledge funding from the Innovation Project of People's Liberation Army (grant no.16CXZ042).

Conflicts of Interest: The authors declare no conflict of interest.

References


rhabdomyosarcoma cell lines. J Cancer Res Clin Oncol.


particle transformed human bronchial epithelial cells. Radiation & Environmental Biophysics 43: 257-263.


Tables

<table>
<thead>
<tr>
<th>Sample</th>
<th>GIREMI</th>
<th>JinBilly</th>
<th>Editor</th>
<th>All methods</th>
<th>RNA editing</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEP2D REP1</td>
<td>3721</td>
<td>13139</td>
<td>20585</td>
<td>2822</td>
<td>5659</td>
</tr>
<tr>
<td>BEP2D REP2</td>
<td>3712</td>
<td>13776</td>
<td>20322</td>
<td>2892</td>
<td></td>
</tr>
<tr>
<td>BEP2D REP3</td>
<td>3982</td>
<td>15356</td>
<td>22106</td>
<td>3078</td>
<td></td>
</tr>
<tr>
<td>BERP35T1 REP1</td>
<td>1929</td>
<td>7693</td>
<td>10877</td>
<td>1429</td>
<td>3820</td>
</tr>
<tr>
<td>BERP35T1 REP2</td>
<td>3355</td>
<td>13258</td>
<td>18762</td>
<td>2496</td>
<td></td>
</tr>
<tr>
<td>BERP35T1 REP3</td>
<td>1990</td>
<td>9081</td>
<td>12728</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>BERP35T4 REP1</td>
<td>1746</td>
<td>7587</td>
<td>10895</td>
<td>1302</td>
<td>2446</td>
</tr>
<tr>
<td>BERP35T4 REP2</td>
<td>1537</td>
<td>7088</td>
<td>9948</td>
<td>1122</td>
<td></td>
</tr>
<tr>
<td>BERP35T4 REP3</td>
<td>1603</td>
<td>6597</td>
<td>9317</td>
<td>1202</td>
<td></td>
</tr>
</tbody>
</table>
Figure Legends

**Figure. 1. Cell culture and RNA sequencing**

A. Photomicrographs of BEP2D (Left), BERP35T1 (Middle) and BERP35T4 (Right) showing the cellular atypia of the malignant transformed cells (100×).

B. Scatter plots showing the consistency of normalized gene expression between biological replicates for each cell line.

**Figure. 2. Identification and characterization of A-to-I RNA editing**

A. Venn plot showing the overlaps of RNA editing sites in BEP2D, BERP35T1 and BERP35T4 cell lines.

B, C. Bar plots showing the proportion of A-to-I RNA editing in Alu/non-Alu regions (B) and Genebody/Intergenic regions (C).

D. Bar plot showing the proportion of tissue-specific editing genes in each cell line.

E. GO enrichment for editing genes in each cell line, value was negative log10 of p-value.

F. We compared the proportion of DEGs in total genes (blue bar) and DEGs in editing genes (red bar), p-value was calculated by hypergeometric test.

**Figure. 3. Down-regulation of ADAR2 induce RNA editing**

A, B. Bar plots showing the normalized gene expression of ADAR1, ADAR2 and ADAR3 in each cell line.

C, D. Bar plots showing the number of genomic variants in ADAR2 (C) and exons of ADAR2 (D).

E, F. IGV plots showing the sequencing reads information of SNP rs11701974 and novel SNP chr21:46643782.

**Figure. 4. Oncogene CTNNB1 and FN1 are highly edited and significantly overexpressed in malignantly transformed cell**

A, C. Bar plots showing the normalized gene expression of oncogenes CTNNB1 and FN1 in normal BEP2D cell and malignantly transformed BERP35T1 and BERP35T4 cells, respectively.

B, D. IGV plots showing the editing events in oncogenes CTNNB1 and FN1.
Supplementary Tables

Table S1. RNA editing sites in each cell line.

Table S2. Genes with RNA editing sites (sheet 1) and genes with tissue-specific RNA editing sites (sheet 2).

Table S3. Differentially expressed genes.

Table S4. SNPs in gene ADAR2 for each cell line.

Table S5. Oncogenes used in this study.

Supplementary Materials

Figure 1. Cell culture and RNA sequencing

A

BEP2D  |  BERP35T1  |  BERP35T4

B

Normalized gene expression (log2(FPKM+1))
Figure. 3. Down-regulation of ADAR2 induce RNA editing

A

B

C

D

E

F

rs11701974

cln21-46643782 (novel SNP)

ADAR2 (3' UTR)

ADAR2 (exon11)
Figure 4. Oncogene CTNNB1 and FN1 are highly edited and significantly overexpressed in malignantly transformed cell