EBioMedicine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Global Transcriptome Analysis of RNA Abundance Regulation by *ADAR* in Lung Adenocarcinoma

Michael F. Sharpnack [a], Bin Chen [b], Dvir Aran [b], Idit Kosti [b], Douglas D. Sharpnack [c], David P. Carbone [d], Parag Mallick [e,**], Kun Huang [a,f,*,1]

[a] Department of Biomedical Informatics, The Ohio State University, Columbus, OH, United States
[b] Institute for Computational Health Sciences, University of California San Francisco, San Francisco, CA, United States
[c] Vet Path Services, Inc., Cincinnati, OH, United States
[d] Comprehensive Cancer Center, The Ohio State University, Columbus, OH, United States
[e] Canary Center for Cancer Early Detection, Stanford University, Palo Alto, CA, United States
[f] Current Address: Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

ARTICLE INFO

Article history:
Received 3 October 2017
Received in revised form 27 November 2017
Accepted 5 December 2017
Available online xxxx

Keywords:
RNA editing
Lung adenocarcinoma
Cancer
Post-transcriptional regulation
Bioinformatics
ADAR

ABSTRACT

Despite tremendous advances in targeted therapies against lung adenocarcinoma, the majority of patients do not benefit from personalized treatments. A deeper understanding of potential therapeutic targets is crucial to increase the survival of patients. One promising target, *ADAR*, is amplified in 13% of lung adenocarcinomas and in-vitro studies have demonstrated the potential of its therapeutic inhibition to inhibit tumor growth. *ADAR* edits millions of adenosines to inosines within the transcriptome, and while previous studies of *ADAR* in cancer have solely focused on protein-coding edits, >99% of edits occur in non-protein coding regions. Here, we develop a pipeline to discover the regulatory potential of RNA editing sites across the entire transcriptome and apply it to lung adenocarcinoma tumors from The Cancer Genome Atlas. This method predicts that 1413 genes contain regulatory edits, predominantly in non-coding regions. Genes with the largest numbers of regulatory edits are enriched in both apoptotic and innate immune pathways, providing a link between these known functions of *ADAR* and its role in cancer. We further show that despite a positive association between *ADAR* RNA expression and apoptotic and immune pathways, *ADAR* copy number is negatively associated with apoptosis and several immune cell types' signatures.

1. Introduction

Each year, >500,000 people die from lung adenocarcinoma (LUAD) globally, and only a minority of these tumors harbor alterations targetable by existing personalized therapies (Imielinski et al., 2012; Collisson et al., 2014). Research into targeting new oncogenes is crucial to expanding the proportion of treatable tumors. The genomic locus containing the adenosine deaminase acting on RNA (*ADAR*) gene is amplified in 13% of LUAD and has been functionally confirmed as an oncogene in lung (Anadón et al., 2015), breast (Fumagalli et al., 2015), stomach (Chan et al., 2016), chronic myeloid leukemia (Jiang et al., 2013) and liver (Chen et al., 2013b) cancers. Knockdown of *ADAR* in LUAD is associated with reduced in-vitro cell viability, and decreased metastatic potential in xenograft mouse models (Anadón et al., 2015). Therefore, targeted therapies towards and a greater functional understanding of *ADAR* amplified tumors have the potential to greatly increase the number of treatable cases of LUAD.

Adenosine to inosine RNA editing is mediated predominantly by the *ADAR* gene and to a lesser extent, the *ADARB1* gene. This modification occurs in millions of sites across the transcriptome, mostly within Alu repeats (Bahn et al., 2015). Previous studies of the oncogenic role of *ADAR* in cancer have focused on the editing of coding sequences of tumor-associated proteins, such as *AZIN1* (Chen et al., 2013b) and *NEIL1* (Anadón et al., 2015). Although this role has been confirmed in functional studies of hepatocellular carcinoma, <1% of known RNA editing sites reside in coding sequences (Han et al., 2015). The function of RNA editing in non-protein coding regions is only known for a few edited genes. For example, it has been shown that *ADAR*-mediated RNA editing changes the accessibility to the HuR RNA binding protein (RBP) within the cathepsin S (*CTSS*) 3′ UTR, which enhances its mRNA's stability in endothelial cells (Stellos et al., 2016). In addition, Zhang et al. showed that RNA editing of the *MDM2* 3′ UTR segment can abolish mir-200b mediated repression of the *MDM2* mRNA. Recently, RNA editing of

* Correspondence to: K. Huang, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States.
** Correspondence to: P. Mallick, Canary Center for Cancer Early Detection, Stanford University, Palo Alto, CA, United States.
E-mail addresses: paragm@stanford.edu (P. Mallick), kunhuang@iu.edu (K. Huang).
[1] Lead author.

https://doi.org/10.1016/j.ebiom.2017.12.005

non-coding regions by *ADAR* has been implicated in immune regulation (Mannion et al., 2014; Liddicoat et al., 2015; Pestal et al., 2015; George et al., 2016) and apoptosis (Yang et al., 2017; Sakurai et al., 2017), but it is not clear how these functions are related to *ADAR*'s oncogenic potential.

Here, we investigate the role of mRNA abundance regulation by RNA editing in human LUAD. We create a pipeline that inputs RNA gene abundances and RNA editing frequencies, and outputs potentially regulatory pairs of RNA editing sites and mRNA target genes. This pipeline is applied to The Cancer Genome Atlas (TCGA) LUAD RNA sequencing data to obtain a global picture of the RNA regulatory editome. We identify enrichment of both apoptosis and immunoregulatory genes potentially regulated by *ADAR*-mediated RNA editing. *ADAR* is alternatively spliced into two separate isoforms: a short, nuclear, and constitutively expressed p110 isoform; and a long, cytosplasmic, and interferon inducible p150 isoform (Pestal et al., 2015). The p110 isoform is responsible for fine tuning apoptosis, at least in part via modulating accessibility to Staufen 1 binding sites in edited mRNA's (Yang et al., 2017; Sakurai et al., 2017). The p150 isoform edits double stranded dsRNA's and prevents them from activating the MDA5-MAVS interferon response. Neither of these functions of *ADAR* have been linked to cancer, although it was seen by Fumagalli, et al. that *ADAR* RNA expression is jointly explained by *ADAR* genomic copy number and *STAT1* expression, a marker of interferon activity (Fumagalli et al., 2015). We further establish that *ADAR* genomic copy number (CN) is negatively associated with immune and apoptosis pathways, as well as immune cell signatures, establishing potential oncogenic roles for *ADAR* in LUAD.

2. Materials and Methods

2.1. Data

RNAseq abundances, copy number data, and editing frequencies were downloaded for lung adenocarcinoma (LUAD) matched normal and tumor samples. RNAseq gene level expression (RNAseqV2 RSEM) and copy number data were downloaded from Broad Institute Firehose (https://gdac.broadinstitute.org). Gistic 2.0 copy number data was further processed to gene level as follows. If a gene is contained within a segment as defined in the gistic-processed data, the value of that segment is assigned to the gene. Editing frequency data was downloaded from synapse.org (https://www.synapse.org/#!Synapse:syn2374375/files/) (Han et al., 2015). Editing frequency is the number of edited reads covering a given editing site divided by the total number of reads covering that editing site. Editing sites are annotated in the RADAR (Ramaswami and Li, 2014) database, which collects and curates functionally validated A-I RNA editing sites.

2.2. Identification of Regulatory Edits

First, RNA editing sites were matched to the genes containing them. Often, there are multiple RNA editing sites within each gene. Then, for each gene, RNA editing frequencies were associated to RNA abundances with spearman correlation. Significant edits were determined using R's (v1.0.136) built in significance of correlation test and corrected for multiple hypothesis testing (Benjamini Hochberg corrected q-value < 0.1). All correlations performed in this study are spearman correlations.

2.3. RNA Binding Protein and microRNA Motif Analysis and Secondary Structure Visualization

Regulatory edits were grouped by gene and searched for continuously edited regions (CER). A CER was defined as >5 sequential significant RNA regulatory editing sites, with no two consecutive RNA editing sites being separated by >100 base pairs. These CERs were then input as a bed file into the RBPmap tool (http://rbpmap.technion.ac.il) (Paz et al., 2014), which searches for RNA binding protein (RBP) motifs within genomic regions. We used the following parameters in our searches: high stringency, hg19 reference, all Human/Mouse motifs, and no conservation filter. MicroRNA motif enrichment was calculated using the targetScan web tool (Grimson et al., 2007), and microRNA's with combined score < -0.3 were considered for further analysis. Secondary structure prediction was made with the Forna (Kerpedjiev et al., 2015) web app.

2.4. Pathway and Immune Cell Analysis

Pathway scores for tumor samples were calculated using single sample gene set enrichment analysis (SSGSEA) (Barbie et al., 2009). Immune exclusion and tumor purity scores were downloaded from Aran et al. (2015) and converted to immune infiltrate scores (1-SCORE). Immune cell subset analysis was performed using the TIMER web app (https://cistrome.shinyapps.io/timer/) (Li et al., 2016).

2.5. Histological Slide Evaluation of Immune Cell Subsets

Histological slides for TCGA LUAD tumors, were visualized using the digital slide archive (http://cancer.digitalslidearchive.net) (Gutman et al., 2013). We selected a subset of tumors such that there were roughly equal representation of *ADAR* amplified and normal copy numbers. In total, 97 LUAD tumors, 45 of which had no evidence of *ADAR* amplification, and 52 of which had high *ADAR* copy number, were selected for analysis and we conducted the experiment with no knowledge of the *ADAR* CN status of these tumors. We were able to visually estimate neutrophil, lymphocyte, and macrophage infiltrations, as well as necrosis on a scale of 0–3.

3. Results

3.1. Pipeline for Discovery of Regulatory A-I RNA Editing Sites

We created a pipeline for discovery of regulatory RNA editing sites as follows. We first matched each RNA editing site with its host RNA abundance (Fig. 1A), then tested all RNA edits for their association with host RNA abundance (Fig. 1B–C). All relevant code can be found at github.com/michaelsharpnack/RNA_edits.

3.2. Landscape of Regulatory Editing Sites in Lung Adenocarcinoma

Previous papers have focused on the potential for protein coding RNA editing sites to modify oncogene and tumor suppressor functions; however, the frequencies of many non-coding RNA editing sites are significantly increased in LUAD. In the TCGA LUAD cohort, 4115 non-protein-coding RNA editing sites were differentially edited (t-test, BH q-value < 0.1) while only 20 protein coding RNA editing sites were differentially edited (Supplementary Fig. 1). We therefore applied our regulatory RNA editing pipeline to the TCGA LUAD RNAseq dataset to discover alternative functions of the *ADAR* oncogene.

RNAseq data for 488 LUAD tumor and 57 matched normal samples were run through our regulatory RNA editing pipeline. This data includes 54,362 frequently edited sites, and 52,276 RNA editing site-RNA abundance combinations tested for their regulatory potential. 5468 (10%) of the edit-gene combinations were predicted to have regulatory potential across 1413 genes (Fig. 2A). The majority of the significant RNA editing sites had a positive association with RNA abundance (4976 or 91%, Fig. 2B). It is possible that the enrichment of positive RNA editing-mRNA abundance associations is due to a bias in detection; given that we would not be able to measure frequently edited transcripts that are degraded. Gene expression and length are both potential sources of bias in regulatory RNA editing site detection. The expression of each gene with detectable RNA editing sites is, although significantly so, very weakly associated with the ability to detect regulatory RNA editing sites ($\rho = 0.059$, $p = 5 \times 10^{-4}$). Gene length and the number

Please cite this article as: Sharpnack, M.F., et al., Global Transcriptome Analysis of RNA Abundance Regulation by *ADAR* in Lung Adenocarcinoma, EBioMedicine (2017), https://doi.org/10.1016/j.ebiom.2017.12.005