A survey on cellular RNA editing activity in response to Candida albicans infections

Yaowei Huang1, Yingying Cao1, Jiarui Li2,3, Yuanhua Liu1, Wu Zhong4, Xuan Li5*, Chen Chen2,3* and Pei Hao1*

From 16th International Conference on Bioinformatics (InCoB 2017)
Shenzhen, China. 20-22 September 2017

Abstract

Background: Adenosine-to-Inosine (A-to-I) RNA editing is catalyzed by the adenosine deaminase acting on RNA (ADAR) family of enzymes, which induces alterations in mRNA sequence. It has been shown that A-to-I RNA editing events are of significance in the cell's innate immunity and cellular response to viral infections. However, whether RNA editing plays a role in cellular response to microorganism/fungi infection has not been determined. Candida albicans, one of the most prevalent human pathogenic fungi, usually act as a commensal on skin and superficial mucosal, but has been found to cause candidiasis in immunosuppression patients. Previously, we have revealed the up-regulation of A-to-I RNA editing activity in response to different types of influenza virus infections. The current work is designed to study the effect of microorganism/fungi infection on the activity of A-to-I RNA editing in infected hosts.

Results: We first detected and characterized the A-to-I RNA editing events in oral epithelial cells (OKF6) and primary human umbilical vein endothelial cells (HUVEC), under normal growth condition or with C. albicans infection. Eighty nine thousand six hundred forty eight and 60,872 A-to-I editing sites were detected in normal OKF6 and HUVEC cells, respectively. They were validated against the RNA editing databases, DARNED, RADAR, and REDIportal with 50, 80, and 80% success rates, respectively. While over 95% editing sites were detected in Alu regions, among the rest of the editing sites in non repetitive regions, the majority was located in introns and UTRs. The distributions of A-to-I editing activity and editing depth were analyzed during the course of C. albicans infection. While the normalized editing levels of common editing sites exhibited a significant increase, especially in Alu regions, no significant change in the expression of ADAR1 or ADAR2 was observed. Second, we performed further analysis on data from in vivo mouse study with C. albicans infection. One thousand one hundred thirty three and 955 A-to-I editing sites were identified in mouse tongue and kidney tissues, respectively. The number of A-to-I editing events was much smaller than in human epithelial or endothelial cells, due to the lack of Alu elements in mouse genome. Furthermore, during the course of C. albicans infection we observed stable level of A-to-I editing activity in 131 and 190 common editing sites in the mouse tongue and kidney tissues, and found no significant change in ADAR1 or ADAR2 expression (with the exception of ADAR2 displaying a significant increase at 12 h after infection in mouse kidney tissue before returning to normal).

(Continued on next page)
ADAR3 are three members of the ADAR gene family found altered sequences in protein products. ADAR1, ADAR2, and ADAR3 are three members of the ADAR gene family found in human [3]. They consisted of two to three double-stranded RNA-binding domains (dsRBD) and a deaminase domain in the C-terminal region. ADAR1 has two isoforms, a constitutively expressed ADAR1 p110 and an IFN-induced ADAR1 p150 [4]. As shown in previous studies, dysregulated RNA editing is associated with numerous human diseases, including cancers like acute myeloid leukemia (AML), Astrocytoma, or hepatocellular carcinoma [5–7], and neurological or psychiatric disorders like Amyotrophic lateral sclerosis (ALS), epilepsy, depression, and suicide [8]. Mutant ADARs were also related to Aicardi-Goutières syndrome (AGS) [9, 10] and dyschromatosis symmetrica hereditaria (DSH) [11, 12].

Roles of A-to-I RNA editing were shown to be involved in innate immune response in virus-infected hosts, which displayed either antiviral or proviral activities by inducing changes that affect how viruses interact with their hosts [13, 14]. The RNA editing machinery may contribute to the hyper-mutation and diversification of RNA virus like noroviruses (NoVs) [15]. ADAR-mediated editing of the viral genome during replication was found to be partially responsible for the high mutation rates of NoVs, in addition to the low replication fidelity of RNA virus polymerases. Ebolavirus (EBOV) and Marburgvirus (MARV) were also found to contain novel filovirus-host interactions in infected hosts by addition of non-template-encoded residues within the EBOV glycoprotein (GP) or MARV nucleoprotein (NP), which was induced by apparent RNA-editing activities [16]. Although a lot of studies have been focused on the function of RNA editing both in hosts and microorganisms [17–19], no systematic study has been found on the function of RNA editing in mammalian hosts during the infection of fungi. Since innate immunity is known to respond to bacteria/fungi infections, it is important to investigate whether the RNA editing mechanism is involved in cellular response to microorganism infection.

**Candida albicans** is one of the most prevalent human pathogenic fungi. It grows as yeast or filamentous cells and can cause the infection candidiasis in humans, which was reported as the 4th leading cause [20] of bloodstream infections acquired during hospital stay and incurred a very high mortality and a severe morbidity in the U.S. [21, 22]. Many nutrient factors, iron and zinc sequestration, pH, temperature, and osmotic pressure are crucial for the growth of commensal *C. albicans* in human hosts [22]. Numerous studies have revealed the complex interplay between fungal virulence factors utilized by *C. albicans* and the host immune response [23–27]. The state of *C. albicans* growing and coexisting with a human host without causing any symptoms of disease, is very much dependent on the immunological cross-talk between *C. albicans* and the human innate immune system. Alterations in the host can replace commensal factors with virulence attributes once the pathogenicity is induced. As a result, fungal cells adhering to the host epithelial cells begins invasion by host endocytosis, then active penetration and directs damage follows. Whether RNA editing plays a role in *C. albicans*-host immune system cross talk as a part of the innate immune response system remains unknown. In the current study, to investigate the A-to-I RNA editing events in *C. albicans* infections, a big-data analysis was performed using community RNA-seq data sets generated from human endothelial cells and oral epithelial cells during in vitro infection with 2 *C. albicans* stains, WO1 or SC5314 and from tongue and kidney tissues from mouse model of hematogenously disseminated candidiasis (HDC) [28]. This work represents the first comprehensive analysis of A-to-I RNA editome in human epithelial and endothelial cells, and a possible role of A-to-I RNA editing in *C. albicans* infection was implicated with in vitro infection data.

**Results**

**Global profiles of A-to-I RNA editing events in human epithelial and endothelial cells**

RNA-seq data were obtained from oral epithelial cells (OKF6 cell line) and primary human umbilical vein endothelial cells (HUVEC cell line) infected with SC5314 or...
Changes in RNA editing in human epithelial and endothelial cells infected with C. albicans

To characterize the effects of C. albicans infection on A-to-I RNA editing in human epithelial and endothelial cells, we investigated the changes in editing activities during the course of infection. When mean editing level was used to measure the changes of editing activity (Fig. 2a, and see Additional file 10), it ranged from 0.36 to 0.44 in epithelial cells, and from 0.38 to 0.47 in endothelial cells. Upward changes were observed at later time points in both during the course of infection. We then looked more carefully at the pattern of A-to-I editing sites located in either stably expressed genes or differentially expressed ones. For stably expressed genes (Fig. 2b, and see Additional file 7), the mean editing level ranged from 0.28 to 0.31 in epithelial cells, and from 0.27 to 0.34 in endothelial cells. For differentially expressed genes (Fig. 2c, and see Additional file 8), the mean editing level spanned from 0.23 to 0.25 in epithelial cells compared to from 0.22 to 0.27 in endothelial cells. So for the subsets of stably and differentially expressed genes, no significant changes in editing pattern can be detected in epithelial and endothelial cells with WO1 or SC5314 infections.

As shown previously in human cancers [5, 39, 40], the normalized editing levels were found to be correlated to ADAR gene expression. Thus, we investigated the correlation between ADARs expression (Fig. 2d left panels) and the normalized editing levels of common editing sites (Fig. 2d right panels). Expression of ADAR1 and ADAR2 genes varied, whereas ADAR3 gene expression was not detectable in human epithelial and endothelial cells. While the expression of ADAR1 showed a decline in SC5314 infected HUVEC cell lines, no significant changes in either ADAR1 or ADAR2 expression were detected during the infection. In comparison, the normalized editing levels of the common editing sites (Fig. 2d right panel) exhibited an significantly upward changing patterns in all infection conditions, especially in Alu regions. The increased activities in A-to-I RNA editing over the course of infection can not be explained by the expression level of ADAR1 or ADAR2 (see Additional file 9). This observation differs from those of previous studies [5, 39, 40] and our new research data submitted with this manuscript in parallel, which suggests that the A-to-I editing activities were up-regulated by a different mechanism in C. albicans infection in human epithelial and endothelial cells.

Profiles of A-to-I RNA editing events in mouse model in vivo

We used C. albicans infected mouse model to further analyze RNA editing changes in epithelial and endothelial cells during infection. The RNA-seq data from mouse tongue and kidney tissues were available from a recent study [28]. 584 and 605 million mapped reads were mapped to the mouse genome, respectively (see Additional file 10). We identified 1133 and 955 discrepancy editing sites in the mouse tongue and kidney RNA-seq data, respectively (Fig. 3a, and see Additional file 11), for which A-to-G
mismatches were the majority. Compared to human epithelial and endothelial A-to-I editing sites, the mouse editing sites had a lower overlap between tongue and kidney tissues (Fig. 3a). We then annotate these editing sites (Fig. 3b), and found introns, intergenic regions of genes, and 3′-UTR were the most enriched regions for A-to-I editing sites for these tissues. Next, we analyzed the upstream and downstream 15-bases sequence flanking editing sites (exhibited by WebLogo3) (Fig. 3c). The flanking sequences of RNA editing sites in mouse tissues were similar to those of human with a depletion of G in 5′ and an increase in 3′ of editing sites, but the G depletion was less in mouse than in human. These data represents the first A-to-I RNA editome from mouse tongue and kidney tissues.
Changes of A-to-I RNA editing in mouse model infected with *C. albicans* in vivo

To illustrate the changes of RNA editing in mouse model infected with *C. albicans* in vivo, we analyzed the distribution of editing level across different mouse tissues. The mean editing level of sites in tongue tissues were between 0.40 to 0.52, and that of editing sites in kidney tissues were between 0.41 to 0.50 (Fig. 4a, and see Additional files 12 and 13). We then looked at the pattern of editing levels in stably expressed and differentially expressed genes (Fig 4b). The editing sites located in stably expressed genes showed no significant changes in tongue tissues. For those sites in stably expressed genes in kidney tissue, the editing level decreased initially throughout the course of infection and returned to pre-infection level at 48 h post infection. For the editing sites located in differentially expressed genes, 212 and 194 sites were found in up- and down-regulated genes of kidney tissues, respectively (Fig. 4d), whereas only 17 sites are found in tongue tissues. In kidney tissues the editing level of sites in up-regulated genes increased significantly at 6 h post infection and then decreased at 12- and 24- h post infection before returning to normal at 48 h after infection. For editing sites in down-regulated genes in kidney tissues, the editing level remained un-changed along the different time points after infection.

Accordingly, we examined the relationship between ADAR gene expression and the editing level of A-to-I RNA editing sites (Fig. 4d, and see Additional file 14).
We obtained 131 and 190 common editing sites in mouse tongue and kidney tissue, respectively. The ADAR genes showed no significant change in expression among tongue tissues. The ADAR2 gene, which was found to be highly expressed in brain tissues reported previously [8, 41], exhibited a high expression level in tongue. The ADAR2 gene had a significant increase at 12 h after infection and then decreased to pre-infection level at 24 h, whereas ADAR1 remained constant throughout. On the other hand, the normalized editing levels remained stable in kidney tissues along the course of infection, whereas in tongue tissues the normalized editing level had a slight drop on day three post infection and backed to normal on day 5. So neither the expression level of ADARs, nor the normalized RNA editing levels displayed significant changes during the course of *C. albicans* infection in mouse tongue and kidney tissues, except a short pulse with ADAR2 at 12 h post infection.

**Discussion**

Being the most prevalent human pathogenic fungi, *C. albicans* accounts for 50–90% of all cases of candidiasis in humans. Host infection by *C. albicans* often initiates by transforming the unicellular yeast-like form of *C. albicans* into the multicellular filamentous form, thus becoming invasive to human epithelial and endothelial tissues. The complex interaction between *C. albicans* and the host immune system, including innate immunity, is a crucial, but little understood process. As RNA editing was previously shown to play significant roles in the innate immune response of hosts toward viral infections, it is important to understand whether RNA editing in human epithelial and endothelial cells plays a role in the innate immune response to *C. albicans* infection. To answer this question, the current study took a approach by integrating and re-analyzing two sets of data from previous experiments with human endothelial cells and oral epithelial cells during in vitro infection with 2 *C. albicans* stains, WO1 or SC5314, and with mouse model of hematogenously disseminated candidiasis (HDC) [28].

From our analysis, 89,648 and 60,872 A-to-I RNA editing events were identified for OKF6 and HUVEC cells, respectively. This represents the first comprehensive analysis of RNA editing edtome in human epithelial and endothelial cells. The validity of these events was confirmed by cross-checking against the annotated RNA editing databases, DARNED, RADAR, and REDIPortal. The significant fraction of newly identified A-to-I RNA editing events are interesting, as they represent editing sites specific to human epithelial and endothelial cells. Previously, functions of A-to-I RNA editing mediated by ADARs were discovered and documented mostly in neuronal receptors [42], and ion transporters [43]. Our results provide a useful reference for studying A-to-I RNA editing events in epithelial and endothelial cells, which points to likely significant functions in corresponding tissues.
To investigate the changes in A-to-I RNA editing events during the course of *C. albicans* infection, we measured their activities using the mean editing level of all events and the normalized editing levels of common editing sites across all time points. While the former displayed a slight upward trend of RNA editing activity during the course of infection, the latter showed a significant increase of RNA editing levels for sites located in the three different regions, Alu, Nonre, and Renonalu (Fig. 2d). Unexpectedly, the increased RNA editing levels were not associated with an increased expression of ADAR1 and ADAR2 genes (ADAR3 gene expression was not detectable) in human epithelial and endothelial cells. Previous studies [5, 39, 40] showed that increased A-to-I RNA editing activities were often associated with the up-regulated expression of ADARs. In the cases of (influenza) viral infected host cells, increased RNA editing activities were found to be accompanied by upregulation of ADARs through the innate immune pathway, newly revealed by our parallelly submitted paper. The observed difference between infection of influenza virus and that of *C. albicans* raises an interesting question about the mechanism of upregulated A-to-I editing.

![Fig. 4 Pattern of RNA editing sites in mouse tissues.](image)
activities in the case of C. albicans infection in human epithelial and endothelial cells, presenting a significant challenge for researchers in the field.

To further examine the effect of C. albicans infection on A-to-I RNA editing activities in epithelial and endothelial cells in vivo, we investigated a C. albicans infected mouse model to analyze RNA editing changes during infection. We identified a total of 1133 and 955 RNA-editing events in mouse tongue and kidney tissues, respectively. Due to the lack of Alu elements, mouse has a much lower number of A-to-I editing sites detected compared to human. Comparing to the sequence feature flanking editing sites in human cell lines, the 3’ G depletion in A-to-I RNA sites was much less in mouse tissues. In addition, these editing sites exhibited a significant tissue specificity. These data represents the first documented A-to-I RNA editome for mouse tongue and kidney tissues.

To illustrate the changes of RNA editing in mouse model infected with C. albicans in vivo, similarly we measured the RNA editing activities by the mean editing level of all events and by the normalized editing levels of common editing sites cross all time points for mouse tongue and kidney tissues. However, different from the results of C. albicans infected human epithelial and endothelial cells, the normalized editing levels remained stable in mouse kidney tissues along the course of infection, whereas the normalized editing level in tongue tissues dropped slightly on day three post infection before returning to normal. The different results from the in vivo mouse model were likely due to the more complex in vivo environments. Interactions between C. albicans and host animals are undoubtedly more complex because of the various factors existing in in vivo environments, like the presence of circulating immune system cells. It is likely the immune cells modulate the response of epithelial and endothelial cells in the in vivo mouse model. It is also possible that the mixed cell types from mouse tongue and kidney tissues inundated the signals generated from epithelial and endothelial cells infected with C. albicans. Nonetheless, our current study suggested a likely roles of ADAR mediated A-to-I RNA editing in microorganism infections, like C. albicans. Future study may be pursued with additional microorganisms in pathogenic settings to understand the roles of RNA editing in innate immune response to whole scope of infectious agents.

Conclusion
This work represents the first comprehensive analysis of A-to-I RNA editome in human epithelial and endothelial cells. It was proposed that C. albicans infection of human epithelial and endothelial cells up-regulated A-to-I editing activities through a mechanism different from that of viral infections of human hosts, such as influenza viral infections in our parallel study. However, a mouse in vivo study with C. albicans infection did not reveal significant changes in A-to-I RNA editing activities in tongue and kidney tissues. It is likely that in vivo infection with C. albicans experienced a different kinetics due to the presence of more complex environments, like circulation and immune cells.

Methods
Collection of RNA-Seq data
RNA-seq data for C. albicans infected cells or mouse tissues were collected from the NCBI Gene Expression Omnibus (GEO) under accession number GSE56093 and GSE67688 [28, 44]. Cell and animal treatment and infection with C. albicans were as described [28, 44]. Preparation of RNA-seq libraries and generation of RNA-seq data were performed according to standard Illumina protocols. The RNA-seq data and sample information were summarized in Supplementary Table S1.

Processing and mapping of RNA-seq data
For RNA-seq data processing and mapping, our pipeline was modified from previous work [29–31, 45], from which small adjustment was made in order for it to work with our collected RNA-seq datasets. In brief, the Burrows-Wheeler algorithm (BWA) [46] was used for RNA-seq reads mapping on reference genomes (the human hg19 reference genome and mouse mm9 reference genome, bowtie alignment, bwa aln -t 4, bwa samse –n4). The PCR duplicates reads were removed by MarkDuplicates tools from Picard (version: picard-1.127; https://broadinstitute.github.io/picard/). Unmapped reads and those with mapping quality score lower than 20 were removed by Samtools (version 0.1.19, samtools view –bS – F 4 –q 20) [47].

Calling and characterizing A-to-I RNA editing sites
The protocol we used to call A-to-I RNA editing sites was derived from that described previously. In brief, the RNA-seq data mapped to reference genomes were subject to variant calling by the GATK analysis tool [48]. The called variant sites were filtered by rigorous parameters as described [29, 30]. Briefly, we required variants identified both in human and mouse to be supported by at least three mismatched reads on editing sites to reduce false positives. Both A-to-G and T-to-C mismatches were combined and counted as A-to-I editing sites. To remove possibly false positive RNA editing events due to SNPs, human SNP (Build 141 by NCBI) and mouse SNP (Build 128 by NCBI) were downloaded using the UCSC table browser data retrieval tool [49] and were used to filter human and mouse RNA editing data, respectively.

Computation of editing level and normalized editing level
Editing level is defined for an editing site as the number of reads with the edited base divided by the number of total reads mapped on the site. Normalized editing level
is defined as the ratios of the editing levels at different time points to the maximum editing level for a specific editing site [50].

Annotation of A-to-I RNA editing sites
We used CAVA [51] (that provides additional clinical information, like disease association, for base variants in human genes) to annotate the A-to-I RNA editing sites from human cell lines, and ANNOVAR [52] to annotate the RNA editing sites from mouse tongue and kidney tissues (CAVA is not developed to work for mouse genes). Sequence pattern around A-to-I RNA editing sites in human and mouse was delineated in two steps: 1) extracting the profile of up- and down-stream sequences (15 bases on each side) flanking editing sites using bedtools getfasta [53]; 2) visualizing the sequence context around RNA editing sites using WebLogo 3 (weblogo –A dna –c classic –units probability –first-index –10) [54].

Gene differential expression analysis
Level of expressed genes in RPKM (Reads Per Kilobase per Million mapped reads) was estimated from RNA-seq mapping results as described [28]. Briefly, HISAT2 [55] was used to map reads on reference genomes, HTSeq [56] was used to count mapped reads for expressed genes, and edgeR [57] was used to perform gene differential expression analysis. Differentially expressed genes were defined as fold-change greater than 2 and false discovery rate (FDR) smaller 0.05. All RNA-seq reads were first trimmed by Trimmmomatic-0.32 [58] with parameters: HEADCROP = 10, SLIDINGWINDOW = 4:20 and MINLEN = 36). In addition, duplicated reads were removed by Picard.

Statistics
R package Venn Diagram [59] was used to calculate and draw the overlapping between our identified RNA editing sites and those included in the databases: DARNED, RADAR and REDIportal. R package ggplot2 [60] was used for plotting other figures. For statistics testing with distribution of A-to-I RNA editing data, the nonparametric test, Kruskal-Wallis rank sum test, was performed. For correlation analysis between ADAR gene expression and normalized RNA editing levels, the Spearman’s rank correlation coefficient was computed with R.

Additional files

Additional file 1: The data sets information of collected RNA seq sample. (XLS 36 kb)
Additional file 2: RNA editing sites of OKF6 and HUVEC cell lines. (XLS 29 kb)
Additional file 3: Mouse organ or tissues RNA-seq mapping results. (XLS 13269 kb)

Acknowledgements
We thank Drs. Gokul Ramaswami and Erez Levanon for help with analysis of RNA editing events.

Funding
This work is supported in part by grants from Beijing Municipal Science & Technology Commission (Z1611100000116049), the National Natural Science Foundation of China (31571310, 81590766, 31771412), and the National Major Project of Science and Technology (2017ZX09101005). Publication cost for this article is funded from the National Natural Science Foundation of China (81590766).

About this supplement
This article has been published as part of BMC Genomics Volume 19 Supplement 1, 2018: 16th International Conference on Bioinformatics (InCoB 2017): Genomics. The full contents of the supplement are available online at https://bmcgenomics.biomedcentral.com/articles/supplements/volume-19-supplement-1.

Availability of data and materials
All original data are available from NCBI Gene Expression Omnibus (GEO) under accession number GSE56093 and GSE67688 [28].

Authors’ contributions
PH conceived the study. CC and WZ directed analysis and revision of the manuscript. YH, YC and YL built performed data analysis and drafted the manuscript. JL and XL helped analysis, writing and revision of the manuscript. All authors approved the manuscript.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
48. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garinella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.

49. Donna K, Angela SH, Terence SF, Krishna MR, Charles WS, David H, WJ K. The UCSC table browser data retrieval tool. Nucleic Acids Res. 2004;32(Database issue):D493–6.

50. Chen J, Peng Z, Zhang R, Yang XZ, Tan BC, Fang H, Liu CJ, Shi M, Ye ZQ, Zhang Y, et al. RNA editome in rhesus macaque shaped by purifying selection. PLoS Genet. 2014;10(4):e1004274.

51. Munz M, Ruark E, Renwick A, Ramsay E, Clarke M, Mahamdallie S, Cloke V, Seal S, Strydom A, Lunter G, et al. CSN and CAVA: variant annotation tools for rapid, robust next-generation sequencing analysis in the clinical setting. Genome Med. 2015;7:76.

52. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

54. Gavin EC, Gary H, John-Marc C, Steven EB. WebLogo: A Sequence Logo Generator. Genome Res. 2004;14:1188–90.

55. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.

56. Anders S, Pyl PT, Huber W. HTSeq—a python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

57. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.

58. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

59. Chen H, Boutros PC. VennDiagram: a package for the generation of highly customizable venn and euler diagrams in R. BMC Bioinformatics. 2011;12(1):35.

60. Wickham H. ggplot2: elegant graphics for data analysis. USA: Springer; 2009.