# Investigating the role of ADAR-mediated RNA-editing in Cardiomyopathy

532

## Statement of originality

I hereby declare that to the best of my knowledge, the content of this dissertation is my own work. I certify that the intellectual content of this dissertation is the product of my own work and has been done during the specified period of my research project. All the assistance received in preparing this dissertation and sources have been properly acknowledged. The content acquired outside the specified period of my research project has been properly indicated in the dissertation.

Student's name: 

Date: 2021.5.10

## Acknowledgments

I would like to acknowledge Prof Roger Foo's supervise, guidance, and direction suggestions on this project, Dr Willson Tan, and Dr LEE Chang Jie Mick's suggestions in bioinformatics pipeline establishment and improvement. Also, I would like to thank Prof Polly Chen and her postdoc Dr Omer An's support for providing CSI bioinformatic portal and their related pipeline documents, Prof Gokul Ramaswami for providing the original RNA-editing pipeline codes. Thanks to Professor Robert Young for being my link-supervisor, giving me frequent supervising on my project progress, and providing access and technical support for the use of the UoE Eddie Remote Server.

I also appreciate Zenia Tiang's work in preparing my Singapore TEP visa although I'm not able to go due to school policy in COVID-19, and thanks to my family and friends' tolerance for my strange sleep schedule. Thanks to all the staff and people that had helped me in my four years of undergraduate education.

## Abstract

RNA editing is an important part of post-transcriptional modification which involves the majority of human cellular diversity. The ADAR-mediated Adenosine to inosine deamination, as the majority type of RNA editing, has been found with various RNA-specific functions to modify cellular activity and were involved in various disease pathogenesis. However, although some study has indicated an essential regulatory role of ADAR in cardiovascular diseases, the specific role of RNA editing was not comprehensively studied yet. In this study, we aimed to re-analyze the RNAseq data of cardiomyopathy and healthy hearts to systematically compare the RNA editing level and identify potential functional RNA-editing sites. Our results indicated no overall RNA editing level difference but diverse site-specific differences between groups. We obtain a list of 410 differentially edited genes, among which 4 of the top 10 genes were of great interest and worth laboratory verification. Besides, most editing sites were located in the 3'UTR functional region, indicating a close relationship between miRNA binding and RNA editing. Our study is the first to systemically analyze the RNA editing events in human hearts, which could give novel insights into human CVD pathogenesis and promote the development in the field of CVD epitranscriptomatics.

(word count: 197)



**Keywords:** ADAR, RNA-editing, Cardiovascular Diseases, Cardiomyopathy, RNA-seq, Bioinformatics

# Table of Contents

## List of figures and tables

## List of abbreviations

| Abbreviation | Full Name |
|---|---|
| CVD | Cardiovascular disease |
| CM | Cardiomyopathy |
| DCM | Dilated Cardiomyopathy |
| HCM | Hypertrophic cardiomyopathy |
| PPCM | Peripartum cardiomyopathy |
| Diff-edited | Differentially edited |
| A-to-I Editing | Adenosine to Inosine Editing |
| C-to-U Editing | Cytidine to Uridine Editing |
| G | Guanine |
| SNPs | Single-nucleotide polymorphisms |
| dsRNA | Double strand RNA |
| ADAR | Adenosine deaminases acting on RNA |
| AID/APOBEC | Activation-induced cytidine deaminase/apolipoprotein B editing complex |
| HIF | hypoxia-inducible factor |
| ANKRD2 | Ankyrin repeat domain 2 |
| CFLAR | CASP8 and Fas-associated protein with death domain-like apoptosis regulator |
| GEO | Gene Expression Omnibus |
| STAR | Spliced Transcripts Alignment to a Reference |
| GO | Gene Ontology |
| BLAT | BLAST-like alignment tool |
| AEI | Alu Editing Index |
| SVM | Support Vector Machine |
| RF | Random Forest |
| CI | confidential intervals |

# 1. Introduction

Post-transcriptional modifications contribute a lot to human genetic and functional diversity. The RNA modification, a subtype of post-transcriptional modifications, has been indicated with rather complex regulatory patterns. After transcribed from the DNA template, RNAs will be spliced, folded, and edited by various enzymes in either nucleus, cytoplasm, mitochondria, or plastids (Dorn, 2019), which would modify the cellular activities by affecting the protein function or RNA degradation sensitivity and accessibility. Therefore, it is not surprising to find that RNA modification involves in multiple developmental progress and disease pathogenesis (Moore Joseph, 2020; Baysal, 2017).

## *1.1. RNA editing*

RNA-editing is a type of post-transcriptional RNA modification that alters a single nucleotide in the sequence of RNAs but not DNA templates (Benne, 1986). Mechanically, a specific type of enzymes, named deaminases, could recognize the adenosine (A) or cytidine (C) sites in the double-strand RNAs (such as mRNA, tRNA, miRNA, etc), and deaminate them into Inosine (I) or Uridine (U), which could bring functional changes in downstream protein-coding or RNA accessibilities (Christofi, 2019; figure 1). Recent studies found that RNA editing events occurred in all living organisms (Li, 2013) and were evolutionarily conserved among Drosophila, zebrafish, mice, and human (Slavov, 2000; Keegan, 2011; Grice, 2015), which attract high concern among researchers. However, the specific functions of RNA editing diverse, depending on the cell types, editing types, RNA types, and editing locations (Ramaswami, 2014; Uchida, 2018), making the functionality study rather complicated.

For example, the role of RNA-editing in different RNA types varies. Editing in non-coding RNA could change the RNA stability by stabilizing the RNA secondary structures and altering the accessibility of non-coding RNAs (Wang, 2013). Editing in exons would mainly result in protein recoding in the mRNA translation process due to the changes of amino acids (Powell, 1987). Besides, editing in the 3' UTR or miRNA could regulate gene expression through miRNA biogenesis and binding site recognition (Yang, 2006), and editing in introns would affect the pre-mRNA splicing and introduce new protein isoforms (Mazloomian, 2015) (figure 2). In humans, most RNA editing events occur in noncoding regions, especially the introns of protein-coding and non-coding RNAs (89%) (Ramaswami, 2014). And two major types of deamination exist, including the adenosine to inosine (A-to-I) editing and cytidine to uridine (C-to-U) editing, while the former is deaminated by adenosine

deaminases acting on RNA (ADARs) family, the latter by cytidine deaminases within the activation-induced cytidine deaminase/apolipoprotein B editing complex (AID/APOBEC) family (Christofi, 2019).

### *1.2. ADAR-mediated A-to-I Deamination*

The ADAR-mediated adenosine (A) to inosine (I) deamination contributes to 90% of overall RNA editing events (Uchida, 2018). The converted inosine would then be recognized as Guanine (G), which could no longer pair with uracil (U) in the dsRNA and induces changes in RNA secondary structure. The subsequent binding ability of ADARs is thus changed, resulting in the change of RNA stability and cellular activity (Dorn, 2019). In human, three subtypes of ADARs (ADAR1, ADAR2, ADAR3), together with several other isoforms (such as ADAT) of the ADAR family were found (Ng, 2013), while ADAR3 was found expressed mainly in neurons (Mladenova, 2018) and less involved in the RNA-editing events (Chen, 2000). Research has shown different but essential developmental functions between Adar1 and Adar2 in mice, since depletion of Adar1 would induce embryonic death (Wang, 2004), while depletion of Adar2 would induce postnatal death instead (Higuchi, 2000). Besides, mutations in ADAR1 in human were found associated with dyschromatopsia symmetrical hereditria and Aicardi–Goutières syndrome (Slotkin, 2013), and high ADAR expression were found pivotal in the vasculature (Stellos, 2016). These ADAR-mediated A-to-I RNA editings are also regarded as an impetus of evolution, which could introduce new protein functions to adapt to natural selection. A specific example is the NARF gene, where the A-to-I editing introduces novel exons and changes the stop codon to affect the nonsense-mediated decay (NMD) of NARF RNA. Together, a novel and stable function of NARF protein was introduced (Schaffer, 2020).

Human A-to-I editing mostly occurs in introns and 3’ UTRs of protein-coding genes (Dorn, 2019), indicating a strong regulative role of RNA-editing in RNA metabolism. Besides, miRNAs and 3’UTR regions are also popular targets of ADARs (ADAR1 and ADAR2) to control miRNA decay and their binding ability (Yang, 2006). In particular, the most frequent target of ADAR-mediated RNA editing events is the Alu repeat, a repetitive short-interspersed element (SINE) abundant in humans which always forms long and stable dsRNAs (Peng, 2012). In human, more than 10% of the overall genome consists of Alu elements (Schaffer, 2021), which means these special sites provide millions of sites for RNA-editing (Bazak, 2014). These RNA editing events in Alu elements were also found effective in multiple diseases, such as promotion of malignant regeneration in multiple myeloma due to Alu-dependent RNA editing of GLI1 (Lazzari, 2017), development of autoimmune disease due to

additional activation of protein kinase R by Alu editing (Nakahama, 2020), and inducing of ALS type 12 due to deficient of GRIA2 editing in spinal motor neurons (Krestel, 2018). A global screen of ADAR expression in all human tissues found an extremely high ADAR expression in neurons, together with an oppositely low ADAR expression in cardiomyocytes (Tan, 2017). However, compared to neurology, the specific role of RNA editing in cardiovascular diseases (CVDs) and cardiomyocytes remains largely unexplored.

### *1.3. RNA editing in CVDs*

In contrast to the high occurrence and top mortality contribution of cardiovascular diseases (CVD) (Lozano, 2012), only a few studies linked RNA editing to this important disease, let alone the corresponding therapeutic strategies. Among these studies, some research did have indicated an essential role of the ADAR enzyme in cardiomyocyte survival and proliferation. Moore's study (2020) showed increased apoptosis of cardiomyocytes in Adar1 Knockout mice embryos, while Witman's study (2013) showed that the Adar1 overexpression in newt hearts would lead to cardiac injury. However, no direct research of ADAR function has been done in human hearts. On the other side, several reports indicated disease-related edited sites that involved in the pathogenesis. For example, a pilot study reported a higher A-to-I RNA editing level in the MED13 RNA in cyanotic congenital heart disease than acyanotic, suggesting an influence of RNA editing in the cellular and metabolic pathways (Borik, 2011). Another study indicated that the expression of cathepsin S (CTSS), which encodes angiogenesis and atherosclerosis-associated cysteine protease, was increased in the endothelial cells in atherosclerotic inflammatory diseases (Stellos, 2016). More recently, van der Kwast's group (2018) identified an increased level of editing in miR487b of murine muscle tissue during postischemic neovascularization, which changes the targets of the proangiogenic RNA. What's more, Jain's study (2018) revealed that an ADAR2-triggered loss of editing in Filamin A pre-mRNA led to misregulation of PLC and ROCK signaling in smooth muscle and induced persistently elevated diastolic blood pressure, resulting in left ventricular hypertrophy in mice. These studies gave insight into the new field of cardiovascular epitranscriptomics (which means post-transcriptional RNA modification) which requires further investigations. Therefore, a more specific illustration of ADAR activity and a more comprehensive comparison of RNA-editing sites in the cardiovascular system is in urgent need.

### *1.4. Cardiomyopathy*

Cardiomyopathy is a group of cardiovascular diseases that affect heart muscles and induces heart failure. It could be further categorized into primary and secondary cardiomyopathy, and the primary cardiomyopathy could be further classified into genetic (eg. Hypertrophic), acquired (eg. Peripartum), and mixed (eg. Dilated and restrictive) in etiology (Brieler, 2017). Regardless of the pathogenesis, all these categories showed common pathologies in heart failure with reduced ejection fraction, which is always associated with hypoxia. The current treatment of cardiomyopathy is mostly symptom-relief targeted, while the RNA-related pathogenesis remains unclear. Therefore, identifying the relationship between RNA-editing events and cardiomyopathy could provide novel therapeutic targets to improve treatment efficiency.

### *1.5. Aim and Hypothesis*

With the fast-increasing amount of available high-throughput RNA sequencing data and the recent development of bioinformatics pipelines that identify RNA-editing sites from RNA sequencing alone, we are now available to perform a comprehensive comparison of RNA editing events in CVDs based on large sample size. Therefore, in this research, we aimed to re-analyze the RNA-seq data in cardiomyopathy to compare the ADAR-mediated A-to-I editing events in the Alu regions between different cardiomyopathy etiology and healthy hearts and identify potential functional alteration in ADARs and RNA-editing sites which involves in the pathogenesis. We hypothesize that an ADAR-induced RNA-editing alteration of specific functional groups of mRNAs or miRNAs with specific functional target regions would affect the pathogenesis of cardiomyopathy. Thus, by recovering the normal RNA structure by RNAi or CRISPR-Cas9 system, new drug targets could be identified to promote cardiomyopathy treatments.

So far as we know, this is the first study that analyzes the RNA-editing events in human hearts, which could give a more direct indication of the situation in human hearts. The results could also give clues to other CVDs and promote the development of the cardiovascular epitranscriptomics field.

## 2. Methods

### 2.1. *Datasets*

To support the analysis of RNA-editing sites in all types of RNAs in relatively large sample size, the overall RNA sequencing data (Illumina HiSeq 2500, Homo sapiens) of the left ventricle from non-failing donors and heart failure samples from the MAGNet consortium was downloaded from Gene

Expression Omnibus (GEO: GSE141910). 324 RNAseq samples were used, including 147 samples of Dilated cardiomyopathy (DCM), 27 samples of hypertrophic cardiomyopathy (HCM), and 6 samples of peripartum cardiomyopathy (PPCM), together with 144 samples of healthy controls. Among all the samples, 213 Caucasians and 11 African American were included, while 168 males and 156 females were included. Also, the samples cover patients from 15 to 83 years old.

### *2.2. Bioinformatic pipeline for RNA-editing site detection*

With the recent development of RNA editing detecting bioinformatic pipelines, the RNA editing sites could now be identified directly from RNAseq data without DNA sequence as reference. In the A-to-I editing, the inosine (I) will be converted to guanine (G) during reverse transcription to cDNA. Therefore, by comparing the cDNA sequence to the reference genome, the editing sites could be directly identified. However, the biggest challenge of RNA-editing detection is the way to separate RNA editing sites from genome-encoded single-nucleotide polymorphisms (SNPs) and technical artifacts generated during sequencing or read mapping. To reduce the false-positive rate, we adopted Pamaswami's pipeline (2013) with a series of filters to exclude the technical biases and multiple databases of known SNPs as references (Diroma, 2019; An, 2020; figure 3A).

#### 2.2.1. Mapping of RNAseq reads

In brief, we first used Spliced Transcripts Alignment to a Reference (STAR: Dobin, 2013) toolkit two times for gap-awaring RNA alignment and integrated four known gene models from UCSC, RefSeq, GENCODE, and Ensembl known gene databases into the first round of STAR alignment. All the gene models were downloaded from the UCSC table browser. Then, we removed duplicate reads via *Picard MarkDuplicates* function and format the STAR outputs into GATK-acceptable style with *GATK SplitNCigarReads* function. After that, we filtered out unmapped reads and reads with mapping quality less than 20 with *samtools*. Then, indel realignment was conducted by GATK RealignerTargetCreator and IndelRealigner, and base quality score recalibration was done by *GATK BaseRecalibrator*.

#### 2.2.2. Variant calling and filtering

We first apply *samtools mpileup* function to call the initial variants with a base quality score $\geqslant$ 25 and a mapping quality score $\geqslant$ 20. Then we removed all known SNPs recorded in dbSNP (hg19 v150, http://www.ncbi.nlm.nih.gov/SNP/), the 1000 Genomes Project (https://www.internationalgenome.org/) and the University of Washington Exome Sequencing

Project (http://evs.gs.washington.edu/EVS/). After that, additional filters were used to remove false positive mismatches, including discarding mismatches in the first 6 bps of each read to avoid artificial mismatches derived from random-hexamer priming, and for non-Alu regions, removing sites in simple repeats, discarding intronic candidates located within 4 bp of all known splicing junctions, removing sites in homopolymer with more than 5 bp, and removing variants in highly similar parts by BLAST-like alignment tool (BLAT) (Ramaswami, 2013). The non-Alu sites were then separated into repetitive and non-repetitive (unique) sites. Finally, editing candidate sites with overall coverage $\geqslant$ 20 and mutation frequency $\geqslant$ 0.1 (An, 2020) were kept for downstream analysis (figure 3B).

#### 2.2.3. ANNOVAR annotation

All editing candidates were then annotated to genes, predicted functional consequence, and specific mutation site information via ANNOVAR (Wang, 2010). After that, the A-to-G and C-to-T mutation would be selected independently for easier visualization.

### *2.3. Alu Editing Index (AEI) calculation*

The Alu Editing Index (AEI) is an index reflecting the overall editing level in the Alu region of a sample. It was defined as the overall edited sites divided by the overall coverage of the Adenosine sites by Roth in his 2019's study. His study also showed AEI as a great representative of the overall RNA-editing level in Alu sites, which is highly correlated with ADAR expression level, while the ADAR expression itself is sometimes not well-correlated with editing levels. Therefore, here we apply Roth's *RNAEditingIndexer* software to calculate the A-to-I editing value in Alu regions based on the STAR alignment.

### *2.4. Selection of differentially edited sites & genes*

To filter the significantly differential edited sites between diseases and control groups, we performed a pairwise t-test comparison of editing frequency in each candidate site. Then all sites were sorted from the smallest p-value to the largest, and a cutoff of p-value at 0.05 was applied to filter only significant sites. After that, we count the significant diff-site numbers in each gene or functional region annotated by ANNOVAR and obtain the overall list.

### *2.5. GO pathway enrichment*

For all the genes selected in the above section, we then undergo GO pathway enrichment by *enrichGO and plotGOgraph* function in the R package '*clusterProfiler*' with options pvalueCutoff=0.05, qbalueCutoff=0.1, and ontology of BP.

### *2.6. Statistics*

Pair-wise t-tests were performed to compare the AEI value between groups, and standard deviation (sd) was calculated to indicate the diverse level of AEI within a group. Linear regression of AEI with race, ae, etiology, and sex was done in *R lm* function. Pearson correlation was conducted to compare the sample similarity based on site editing frequency.

## 3. Results

### *3.1. Bioinformatic analysis of RNA-seq data of human left ventricular presented thousands of RNA-edited genes.*

Through performing the established bioinformatic pipeline to 324 RNA-seq data of human left ventricular tissue of both cardiomyopathy and healthy control groups, the RNA-editing candidate sites, together with their overall coverage, mutated coverage, and corresponding editing frequency were listed for each sample. 33005 different editing sites were found in all samples, and 3268 genes were annotated of these sites. The sites are located in all 10 functional regions classified in ANNOVAR, and the specific edited sites vary among samples.

### *3.2. Overall editing level comparison indicated no difference in AEI and AEI distribution among all groups.*

The AEI value is compared between different etiology, different groups, different sexes, different ages, and different races, while no significant difference was found in this process. The distribution range is also similar with a similar standard deviation. The result of the linear regression between AEI and etiology, sexes, ages, races also showed no significant contribution of either index, indicating no significant contribution variables were found (figure 4).

### *3.3. Intra- and Inter- group correlation suggested complex editing event changes.*

To compare the sample similarity and relationship within and among groups, both site edited frequency-based and gene's site number-based Pearson correlation among samples were conducted. Both results showed intra-group and inter-group variation, which indicated complex but disease-related editing event changes (figure 5). Besides, a separated group of samples with extremely low AEI was found clustered, but not in high AEI samples (figure 5B).

### *3.4. Identification of differentially edited sites and genes between disease and control groups*

After the pairwise t-test comparison of each candidate site, 1491 sites with significantly different levels of RNA editing frequency were filtered and sorted ($P<0.05$). All top 10 differential edited sites showed extremely small confidential intervals (95% CI), indicating a strong and convincing difference between the disease group and control group. Meanwhile, 4 of the top 10 sites showed downregulation of editing level in all three cardiomyopathy etiologies, while the other 6 sites showed up-regulation (figure 6A, 6B). Besides, 410 genes were annotated to the 1491 significant diff-edited sites, with the number of diff-edited sites counted for each gene. We ranked the genes from the most site numbers to the least and analyzed the top 10 genes with the most diff-edited sites, namely HIF3A, ABHD18, ANKRD2, MRI1, CFLAR, MAVS, CCDC84, CALCOCO2, TXNDC15, and MDM4 (figure 6C). Meanwhile, 10 functional regions were also annotated to each site, and more than half of the diff-edited sites were located in the 3'UTR (figure 6D).

We also apply another calculating method to find the top clustered genes. We first annotated the editing candidate sites with genes and then count the number of sites in each gene of each sample. Then, we perform a pairwise t-test comparison between disease and control groups based on the site counts. Therefore, if the number of editing sites was significantly changed in the disease group, we believed that there might be changes in RNA editing levels. Based on this method, we obtain a list of 290 differentially edited genes, and the top 10 genes were visualized with 95% CI. Although most of the CIs are small, the CI of HIF3A was relatively large, indicating a less confident result (figure 6E, 6F).

### *3.5. Gene Ontology (GO) functional enrichment with diff-edited clustered genes*

After annotating the 410 site-frequency-based diff-edited genes with Gene Ontology (GO) pathways, we perform GO functional enrichment of Biological Process (BP) to these genes and obtain one significantly clustered accession named "translational elongation", which is defined as the successive addition of amino acid residues to a nascent polypeptide chain during protein biosynthesis (GO: 0006414) (figure 7). In specific, a particular gene named EEF2K was identified in this GO term. However, since the GO enrichment did not consider the significant level of diff-editing in genes, a result with misleading could happen.

### *3.6. Site-specific editing frequency comparison in top diff-edited genes*

Based on the two methods of diff-editing genes ranking, we took a detailed analysis on the genes that appeared in both methods and then focus on the genes identified in the site-frequency-based

method. In specific, in the top 10 genes, HIF3A, ABHD18, ANKRD2 (top 3 in the editing-frequency-based method), and CFLAR appeared in both methods, while MRI1, MAVS, CCDC84, CALCOCO2, TXNDC15, and MDM4 were found in the first method. Results showed that the HIF3A and ANKRD2 have a reduced RNA editing level in the cardiomyopathy group, while increased RNA editing was found in ABHD18 and CFALR (figure 8). What's more, almost all of the editing sites were found in introns for HIF3A (75/75), ABHD18 (25/35), and ANKRD2 (31/31), while all editing sites of CFALR are in the 3' UTR region (28/28).

## 4. Discussion

### *4.1. Complex changes in RNA editing events in disease groups*

In figure 4, we identified no significant differences between cardiomyopathy groups and control groups, as well as other variables including sex, race, and ages. Meanwhile, no significant difference in distribution width was found as well, indicating a seemly irrelevant relationship between RNA editing and disease pathogenesis. However, in figure 5 of sample correlations, differences did exist both between groups and within groups, which suggests a change in the site-specific RNA editing events. Since previous research has indicated a target-specific and region-specific function in RNA editing, we inferred that region-specific and location-specific changes in RNA-editing events might happen in cardiomyopathy groups, and compensative RNA editing regulations might finally resulting in no difference in the overall Alu editing level.

Meanwhile, a specific subgroup with extremely low AEI values was clustered in the correlation heatmap (figure 5), indicating a special situation under low AEI situation. Since we are not accessible to the ADAR expression level, it's hard to verify if these low AEI values are induced by low ADAR expression. However, based on previous research (Roth, 2019), we hypothesized that if the overall ADAR expression is low, the RNA editing events tend to be more similar, which also verified the conclusion that RNA editing events were a major source of genetic variance in human (Dorn, 2019).

### *4.2. Differential edited genes and their relation to cardiomyopathy*

When we considering RNA editing level changes, a strong change would result in significantly reduced edited site numbers, while a relatively milder change might result in no change in site numbers, but a significant difference in site editing frequencies. Therefore, in the two methods to get diff-edited genes, the site-frequency-based method should include the result of the site-number

based method, which is in accordance with our results that the site-frequency based method presented twice more genes than the site-number-based method. Therefore, in the following analysis, we decided to first analyze the genes that appeared in both methods, and then focus on the genes identified in the site-frequency-based method. In specific, in the top 10 genes, HIF3A, ABHD18, ANKRD2 (top 3 in the first method), and CFLAR appeared in both methods, while MRI1, MAVS, CCDC84, CALCOCO2, TXNDC15, and MDM4 were found in the first method.

#### 4.2.1. HIF3A

HIF3A is the gene that encodes the alpha-3 subunit of one of several alpha/beta-subunit heterodimeric transcription factors of hypoxia-inducible factor (HIF) which is conserved between humans and mice (Ensembl, 2017). The alpha-3 subunit is considered as a negative regulator of hypoxia-inducible gene expression (Hara, 2001), since inhibition of HIF3A expression resulted in increased physical endurance in the rat (Drevytska, 2012). A previous study also indicated a significant increase of HIF3A expression in failing myocardium as compensation of HIF target gene overexpression (Zolk, 2008), while in our study, a downregulated RNA editing event was identified in the failing myocardium group (figure 8A). Meanwhile, frequent interactions between miRNAs, such as miRNA-1290 (Wu, 2017) and miR-29b (Sun, 2020), and HIF3A expression were found in previous research as well, indicating an active epitranscriptomic activity in this RNA. Therefore, we hypothesized that heart failure would induce hypoxia in cardiomyocytes via overexpression of hypoxia-inducing genes, and to compensate for the overexpression, a series of HIF regulatory mechanisms would be activated. In specific, the downregulated RNA-editing event in HIF3A's introns might modify the alternative splicing of introns and thus expose or cover the miRNA binding sites, which resulted in downregulated miRNA-induced RNA decay and an increased expression of HIF3A in the protein level.

Therefore, wet-lab experiments could be performed to verify the functional mechanism. In general, we could perform ADAR conditional knockout in mice and test the downstream expression changes in HIF3A and its target genes and regulator miRNAs. And in specific, we could try the CRISPR-Cas9 system to silence the intron sites of HIF3A to test the corresponding changes in miRNAs and target genes.

#### 4.2.2. ABHD18

ABHD18 belongs to a family of ABHD gene which encodes the alpha/beta hydrolase domain-

containing proteins, a family of lipid metabolizing enzymes which are highly evolutionarily conserved. Research has found ABHD18 with highly variable numbers of Alu repeats, while low Alu repeat numbers were found to contribute to hepatocellular carcinoma (HCC) pathogenesis (Clifford, 2010). In our study, the edited level of ABHD18 was upregulated in the disease group (figure 8B), while most editing sites (25/35) are in introns. This might be caused by changes in Alu repeat numbers and increased Alu repeats are predicted. However, studied are limited and the physiological and functional role of ABHD18 with cardiomyopathy requires further investigation.

#### 4.2.3. ANKRD2

The ANKRD2 gene is responsible for encoding the ankyrin repeat domain 2 (ANKRD2) protein, also named Ankyrin Repeat, PEST sequence and Proline-rich region (ARPP), which is highly expressed in cardiac muscle, especially the intercalated discs of ventricular regions (Jasnic-Savovic, 2015). The role of ANKRD2 in muscles has been intensively studied. It is a strong regulator and co-factor of several regulatory complexes in cardiomyocytes, which participates in both intercellular and intracellular communication pathways (Belgrano, 2011). Physically, the muscle-specific mechanosensory ANKRD2 protein interacts with several transcription regulators, such as p53, YB-1, and PML (Kojic, 2004), and structural and signaling proteins such as ZASP (Cenni, 2011), calpain-3 (Hayashi, 2008), and titin (Kojic, 2004), to modify cardiogenesis, myofibrillar assembly and other muscle-state-decisive activities in response of stretch and stress (Jasnic-Savovic, 2015). Recent studies showed an upregulation of ANKRD2 gene in human dilated cardiomyopathy (DCM) (Nagueh, 2004), while silencing of Ankrd2 in human muscle cells would induce a series of changes in DCM- and HCM-involved gene pathways (Belgrano, 2011), indicating an important role of ANKRD2 in DCM pathogenesis. However, the knockout of ANKRD2 in mice cardiac muscles showed a normal cardiac function even under stress-induced cardiac hypertrophy, suggesting some functional differences between mice and humans.

In our study, significant downregulation of Alu editing levels was found in the cardiomyopathy group (figure 8C), while all edited sites were in introns (31/31), suggesting a potential alteration in splicing sites which might result in improvement of RNA stability or suppression of RNA decay. Therefore, overexpression of ANKRD2 protein could be found in the disease. To investigate the specific effect of RNA editing in the ANKRD2 expression, lab interventional experiments such as ADAR conditional knockout or CRISPR-Cas9 site-specific ADAR-like editing to retrieve the RNA

editing level could be applied.

#### 4.2.4. CFLAR

The CASP8 and Fas-associated protein with death domain-like apoptosis regulator (CFLAR), encoded by CFLAR gene, is a negative regulator of cell apoptosis with similar structures to apoptosis initiating protein caspase-8 (Wang, 2017). Therefore, CFLAR acts as the inhibitor of caspase-8 to regulate the downstream apoptosis cascading pathways. Previous studies mainly focused on the effect of CFLAR in cancer cells, while one research also indicated a significant downregulation of CFLAR mRNA expression, together with activation of a group of apoptosis inhibitors, in failing hearts (Haider, 2009). Therefore, it is possible that apoptosis would be induced in cardiomyopathy by CFLAR downregulation, although the downstream apoptosis pathway is interrupted by other apoptosis inhibitors. With the disease progress, the apoptosis might be strengthened, thus regulating CFLAR could be a possible way to release the apoptosis.

In our study, upregulation of CFLAR editing levels was found in the disease group, while all 28 diff-edited sites were placed in 3' UTR, which is a frequent binding region to miRNA. Together with a previous study which indicated an interaction between miR-20a and CFLAR to regulate the hepatocellular carcinoma apoptosis, we hypothesized that the upregulation of RNA editing in CFLAR might affect the miRNA binding and thus resulted in decreased CFLAR expression. Thus the risk of cardiomyocyte apoptosis might be increased in failure hearts. However, since few researches focused on the CFLAR in cardiomyopathy, the interpretation and verification should be more careful.

In conclusion, we analyzed four genes with the most significantly modified RNA editing events in the Alu repeats, and all of them showed different functional potentials in disease pathogenesis. Compare to the HIF3A which might be compensation in disease pathogenesis, ANKRD2 and CFLAR presented a role directly related to pathogenesis. Meanwhile, all RNA editing changes showed an opposite regulation in mRNA expression, suggesting a potential of RNA editing in RNA decay. The ANKRD2 worth particular attention since it is expressed specifically in muscle, and the potential therapeutic function could be considered.

### *4.3. Most differential editing sites in the 3'UTR functional region*

Among all the differential editing sites, more than half were found located in the 3'UTR region of a gene (Figure 6D), indicating a frequent interaction between miRNA and RNA editing events. Besides, in the four analyzed top modified genes, two of them (HIF3A, CFLAR) were verified with

miRNA interaction in previous studies. It was thus possible that the ADAR-induced RNA-editing might have some interaction with miRNA binding, such as co-factors for miRNA recruiting and binding. Therefore, it worth further investigation into the potential role of ADAR in miRNA site recognition. To verify the miRNA binding activity in the 3'UTR edited genes, we could link the genes to miRNAs in miRNA-target databases (both verified and predicted) and calculate the number of genes that are bindable to miRNAs. Therefore, a comprehensive comparison between miRNA and RNA editing could be performed.

### *4.4. Current limitation and future improvement*

In this study, we identify no significant difference in AEI between groups while intra-group differences did exist. Since we lack the identification method for actual ADAR expression, it is relatively uncertain whether a similar editing level would be found in samples with different ADAR expression levels. It is possible to conduct traditional RNA seq differential expression analysis based on the overall RNAseq data, although it might be relatively difficult to filter out the noises of recognizing poly-A ended RNA sequence from the overall RNAseq.

Besides, currently, we did a simple comparison based on diff-edited level, which might include some unimportant editing sites. Since we have a relatively large dataset, we could also try to perform machine learning, such as Support Vector Machine (SVM) or Random Forest (RF) to select the featured editing sites and genes.

What's more, the current RNA-editing pipeline requires relatively large core numbers and storage of Linux servers. Therefore, a simplification and improvement in the current bioinformatics pipeline, such as evaluating the RNAseq quality before running and exclude some steps if the quality reaches the requirement. A more integrative pipeline could also be built to include both RNA editing detection and transcriptome expression evaluation functions, with which we could save time for repetitive calculation.

## 5. Conclusion

In this research, we established a bioinformatic pipeline to identify RNA editing sites from RNAseq data alone, and samples of human left ventricular between cardiomyopathy patients and healthy donors were compared for their A-to-I RNA editing sites and levels in the Alu repeats. As a result, we identify 410 diff-edited genes with 1491 diff-edited sites, among which HIF3A, ABHD18,

ANKRD2, and CFLAR lie in the top diff-edited genes. In specific, HIF3A, ANKRD2, and CFLAR have been shown correlated with cardiomyopathy before, and ANKRD2 is particularly closely related to cardiomyopathy due to its muscle-specific expression and function. So far, our study is the first comprehensive RNA editing analysis done in the human heart, which could provide new insight into the epitranscriptomic field of CVDs and give valuable clues to the new field development.

(Word count: 4986)

## Reference


1. Altaf F, Vesely C, Sheikh A M, Munir R, Shah S T A and Tariq A (2019) Modulation of ADAR mRNA expression in patients with congenital heart defects, PLoS ONE, 14(4). doi.org/10.1371/journal.pone.0200968.
2. An O, Tan K-T, Li Y, Li J, Wu C-S, Zhang B, Chen L and Yang H (2020) CSI NGS Portal: An Online Platform for Automated NGS Data Analysis and Sharing, International Journal of Molecular Sciences, 21(11). doi.org/10.3390/ijms21113828.
3. Baysal B E, Sharma S, Hashemikhabir S and Janga S C (2017) RNA Editing in Pathogenesis of Cancer, Cancer Research, 77(14), pp. 3733–3739.
4. Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs F J, Rechavi G, Li J B, Eisenberg E and Levanon E Y (2014) A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes, Genome Research, 24(3), pp. 365–376.
5. Belgrano A, Rakicevic L, Mittempergher L, Campanaro S, Martinelli V C, Mouly V, Valle G, Kojic S and Faulkner G (2011) Multi-tasking role of the mechanosensing protein Ankrd2 in the signaling network of striated muscle, PloS One, 6(10), p. e25519.
6. Benne R, Van den Burg J, Brakenhoff J P, Sloof P, Van Boom J H and Tromp M C (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA, Cell, 46(6), pp. 819–826.
7. Borik S, Simon A J, Nevo-Caspi Y, Mishali D, Amariglio N, Rechavi G and Paret G (2011) Increased RNA editing in children with cyanotic congenital heart disease, Intensive Care Medicine, 37(10), p. 1664.
8. Brieler J, Breeden M A and Tucker J (2017) Cardiomyopathy: An Overview, American Family Physician, 96(10), pp. 640–646.
9. Burkhard P, Stetefeld J and Strelkov S V (2001) Coiled coils: a highly versatile protein folding motif, Trends in Cell Biology, 11(2), pp. 82–88.
10. Cenni V, Bavelloni A, Beretti F, Tagliavini F, Manzoli L, Lattanzi G, Maraldi N M, Cocco L and Marmiroli S (2011) Ankrd2/ARPP is a novel Akt2 specific substrate and regulates myogenic differentiation upon cellular exposure to H(2)O(2), Molecular Biology of the Cell, 22(16), pp. 2946–2956.
11. Chen C X, Cho D S, Wang Q, Lai F, Carter K C and Nishikura K (2000) A third member of the

RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains, RNA (New York, N.Y.), 6(5), pp. 755–767.

12. Christofi T and Zaravinos A (2019) RNA editing in the forefront of epitranscriptomics and human health, Journal of Translational Medicine, 17. doi.org/10.1186/s12967-019-2071-4.
13. Clifford R, Zhang J, Meerzaman D, Lyu M, Hu Y, Cultraro C, Finney R, Kelley J M, Efroni S, Greenblum S I, Nguyen C, Rowe W L, Sharma S, Wu G, Yan C, Zhang H, Chung Y-H, Kim J, Park N, Song I and Buetow K (2010) Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma, Hepatology. doi.org/10.1002/hep.23943.
14. Diroma M A, Ciaccia L, Pesole G and Picardi E (2019) Elucidating the editome: bioinformatics approaches for RNA editing detection, Briefings in Bioinformatics, 20(2), pp. 436–447.
15. Dobin A, Davis C A, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras T R (2013) STAR: ultrafast universal RNA-seq aligner, Bioinformatics (Oxford, England), 29(1), pp. 15–21.
16. Dorn L, Tual-Chalot S, Stellos K and Accornero F (2019) RNA epigenetics and cardiovascular diseases, Journal of molecular and cellular cardiology, 129, pp. 272–280.
17. Drevytska T, Gavenauskas B, Drozdovska S, Nosar V, Dosenko V and Mankovska I (2012) HIF-3α mRNA expression changes in different tissues and their role in adaptation to intermittent hypoxia and physical exercise, Pathophysiology: The Official Journal of the International Society for Pathophysiology, 19(3), pp. 205–214.
18. Grice L F and Degnan B M (2015) The origin of the ADAR gene family and animal RNA editing, BMC evolutionary biology, 15, p. 4.
19. Haider N, Arbustini E, Gupta S, Liu H, Narula N, Hajjar R, Moorjani N, Westaby S, Semigran M J, Dec G W, Chandrashekhar Y and Narula J (2009) Concurrent upregulation of endogenous proapoptotic and antiapoptotic factors in failing human hearts, Nature Clinical Practice. Cardiovascular Medicine, 6(3), pp. 250–261.
20. Hara S, Hamada J, Kobayashi C, Kondo Y and Imura N (2001) Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in human kidney: suppression of HIF-mediated gene expression by HIF-3alpha, Biochemical and Biophysical Research Communications, 287(4), pp. 808–813.
21. Hayashi C, Ono Y, Doi N, Kitamura F, Tagami M, Mineki R, Arai T, Taguchi H, Yanagida M, Hirner

S, Labeit D, Labeit S and Sorimachi H (2008) Multiple molecular interactions implicate the connectin/titin N2A region as a modulating scaffold for p94/calpain 3 activity in skeletal muscle, The Journal of Biological Chemistry, 283(21), pp. 14801–14814.

22. Hentze M W, Castello A, Schwarzl T and Preiss T (2018) A brave new world of RNA-binding proteins, Nature Reviews. Molecular Cell Biology, 19(5), pp. 327–341.
23. Higuchi M, Maas S, Single F N, Hartner J, Rozov A, Burnashev N, Feldmeyer D, Sprengel R and Seeburg P H (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2, Nature, 406(6791), pp. 78–81.
24. Jain M, Mann T D, Stulić M, Rao S P, Kirsch A, Pullirsch D, Strobl X, Rath C, Reissig L, Moreth K, Klein-Rodewald T, Bekeredjian R, Gailus-Durner V, Fuchs H, Hrabě de Angelis M, Pablik E, Cimatti L, Martin D, Zinnanti J, Graier W F, Sibilia M, Frank S, Levanon E Y and Jantsch M F (2018) RNA editing of Filamin A pre-mRNA regulates vascular contraction and diastolic blood pressure, The EMBO journal, 37(19). doi.org/10.15252/embj.201694813.
25. Jasnic-Savovic J, Nestorovic A, Savic S, Karasek S, Vitulo N, Valle G, Faulkner G, Radojkovic D and Kojic S (2015) Profiling of skeletal muscle Ankrd2 protein in human cardiac tissue and neonatal rat cardiomyocytes, Histochemistry and Cell Biology, 143(6), pp. 583–597.
26. Keegan L P, McGurk L, Palavicini J P, Brindle J, Paro S, Li X, Rosenthal J J C and O'Connell M A (2011) Functional conservation in human and Drosophila of Metazoan ADAR2 involved in RNA editing: loss of ADAR1 in insects, Nucleic Acids Research, 39(16), pp. 7249–7262.
27. Kiran A M, O'Mahony J J, Sanjeev K and Baranov P V (2013) Darned in 2013: inclusion of model organisms and linking with Wikipedia, Nucleic Acids Research, 41(Database issue), pp. D258-261.
28. Kojic S, Medeot E, Guccione E, Krmac H, Zara I, Martinelli V, Valle G and Faulkner G (2004) The Ankrd2 protein, a link between the sarcomere and the nucleus in skeletal muscle, Journal of Molecular Biology, 339(2), pp. 313–325.
29. Krestel H and Meier J C (2018) RNA Editing and Retrotransposons in Neurology, Frontiers in Molecular Neuroscience, 11. doi.org/10.3389/fnmol.2018.00163.
30. van der Kwast R V C T, van Ingen E, Parma L, Peters H A B, Quax P H A and Nossent A Y (2018) Adenosine-to-Inosine Editing of MicroRNA-487b Alters Target Gene Selection After Ischemia and Promotes Neovascularization, Circulation Research, 122(3), pp. 444–456.

31. Lazzari E, Mondala P K, Santos N D, Miller A C, Pineda G, Jiang Q, Leu H, Ali S A, Ganesan A-P, Wu C N, Costello C, Minden M, Chiaramonte R, Stewart A K, Crews L A and Jamieson C H M (2017) Alu-dependent RNA editing of GLI1 promotes malignant regeneration in multiple myeloma, Nature Communications, 8(1), p. 1922.
32. Lev-Maor G, Sorek R, Levanon E Y, Paz N, Eisenberg E and Ast G (2007) RNA-editing-mediated exon evolution, Genome Biology, 8(2), p. R29.
33. Li J, Peng X, He Y, Huang X, You N, Gu H, Dong R, Yang X and Zheng L (2020) Bioinformatics Analysis and Insights for the Role of COMMD7 in Hepatocellular Carcinoma. preprint. In Review. doi.org/10.21203/rs.3.rs-113846/v1.
34. Li S and Mason C E (2014) The pivotal regulatory landscape of RNA modifications, Annual Review of Genomics and Human Genetics, 15, pp. 127–150.
35. Lozano R, et al (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010, Lancet (London, England), 380(9859), pp. 2095–2128.
36. Mazloomian A and Meyer I M (2015) Genome-wide identification and characterization of tissue-specific RNA editing events in D. melanogaster and their potential role in regulating alternative splicing, RNA biology, 12(12), pp. 1391–1401.
37. Mladenova D, Barry G, Konen L M, Pineda S S, Guennewig B, Avesson L, Zinn R, Schonrock N, Bitar M, Jonkhout N, Crumlish L, Kaczorowski D C, Gong A, Pinese M, Franco G R, Walkley C R, Vissel B and Mattick J S (2018) Adar3 Is Involved in Learning and Memory in Mice, Frontiers in Neuroscience, 12, p. 243.
38. Moore Joseph B., Sadri Ghazal, Fischer Annalara G., Weirick Tyler, Militello Giuseppe, Wysoczynski Marcin, Gumpert Anna M., Braun Thomas, and Uchida Shizuka (2020) The A-to-I RNA Editing Enzyme Adar1 Is Essential for Normal Embryonic Cardiac Growth and Development, Circulation Research, 127(4), pp. 550–552.
39. Nakahama T and Kawahara Y (2020) Adenosine-to-inosine RNA editing in the immune system: friend or foe?, Cellular and molecular life sciences: CMLS, 77(15), pp. 2931–2948.
40. Ng S K, Weissbach R, Ronson G E and Scadden A D J (2013) Proteins that contain a functional Z-DNA-binding domain localize to cytoplasmic stress granules, Nucleic Acids Research, 41(21), pp. 9786–9799.

41. Nishikura K (2010) Functions and Regulation of RNA Editing by ADAR Deaminases, Annual review of biochemistry, 79, pp. 321–349.
42. Peng Z, Cheng Y, Tan B C-M, Kang L, Tian Z, Zhu Y, Zhang W, Liang Yu, Hu X, Tan X, Guo J, Dong Z, Liang Yan, Bao L and Wang J (2012) Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome, Nature Biotechnology, 30(3), pp. 253–260.
43. Powell L M, Wallis S C, Pease R J, Edwards Y H, Knott T J and Scott J (1987) A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine, Cell, 50(6), pp. 831–840.
44. Rahman R, Xu W, Jin H and Rosbash M (2018) Identification of RNA-binding protein targets with HyperTRIBE, Nature Protocols, 13(8), pp. 1829–1849.
45. Ramaswami G and Li J B (2014) RADAR: a rigorously annotated database of A-to-I RNA editing, Nucleic Acids Research, 42(Database issue), pp. D109-113.
46. Ramaswami G, Zhang R, Piskol R, Keegan L P, Deng P, O'Connell M A and Li J B (2013) Identifying RNA editing sites using RNA sequencing data alone, Nature Methods, 10(2), pp. 128–132.
47. Raney B J, Dreszer T R, Barber G P, Clawson H, Fujita P A, Wang T, Nguyen N, Paten B, Zweig A S, Karolchik D and Kent W J (2014) Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser, Bioinformatics (Oxford, England), 30(7), pp. 1003–1005.
48. Roth S H, Levanon E Y and Eisenberg E (2019) Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity, Nature Methods, 16(11), pp. 1131–1138.
49. Schaffer A A and Levanon E Y (2021) 'ALU A-to-I RNA Editing: Millions of Sites and Many Open Questions', in Picardi E and Pesole G (eds.) RNA Editing: Methods and Protocols. New York, NY: Springer US, pp. 149–162.
50. Slavov D, Crnogorac-Jurcević T, Clark M and Gardiner K (2000) Comparative analysis of the DRADA A-to-I RNA editing gene from mammals, pufferfish and zebrafish, Gene, 250(1–2), pp. 53–60.
51. Slotkin W and Nishikura K (2013) Adenosine-to-inosine RNA editing and human disease, Genome Medicine, 5(11), p. 105.
52. Stellos K, Gatsiou A, Stamatelopoulos K, Perisic Matic L, John D, Lunella F F, Jaé N, Rossbach

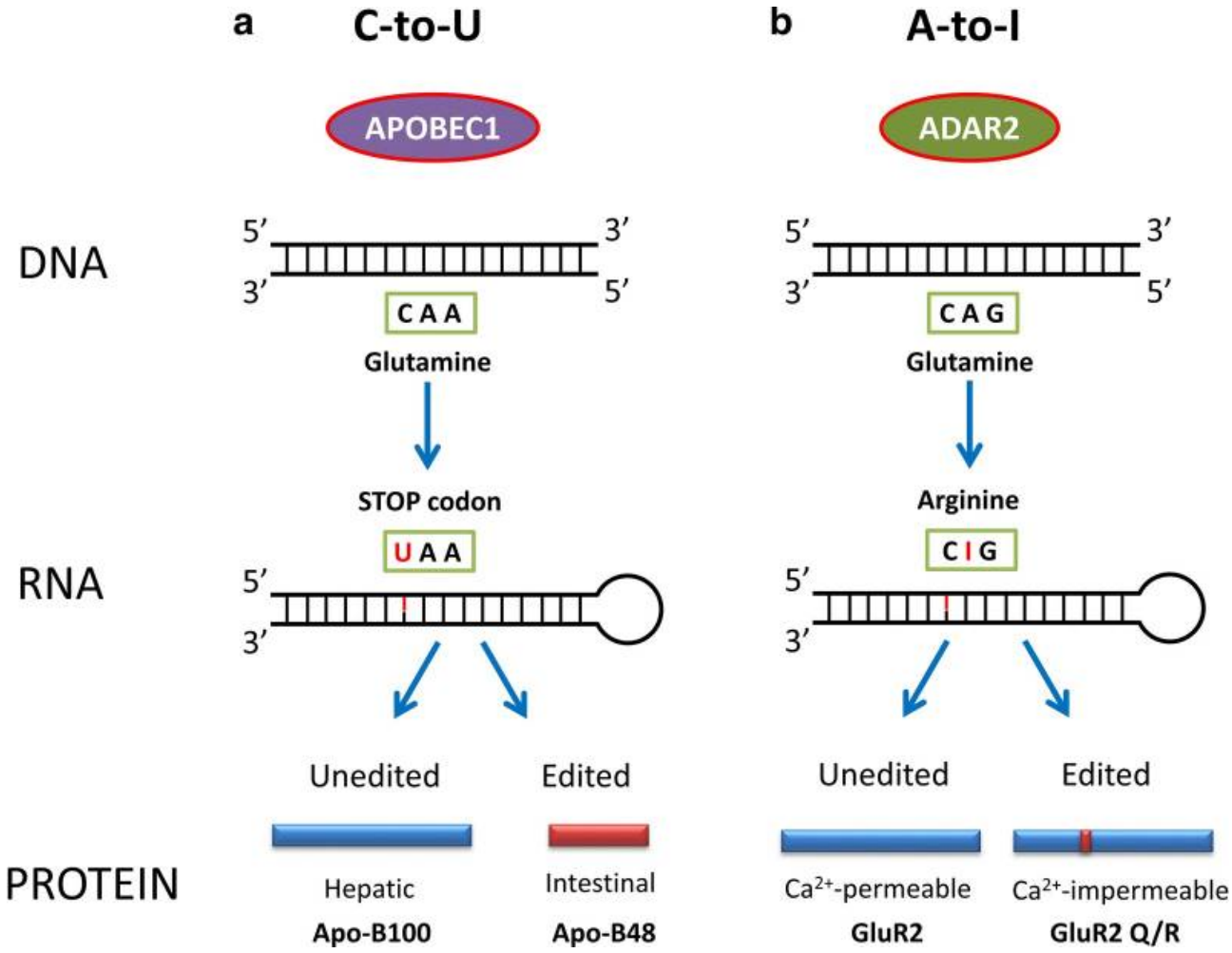
O, Amrhein C, Sigala F, Boon R A, Fürtig B, Manavski Y, You X, Uchida S, Keller T, Boeckel J-N, Franco-Cereceda A, Maegdefessel L, Chen W, Schwalbe H, Bindereif A, Eriksson P, Hedin U, Zeiher A M and Dimmeler S (2016) Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation, Nature Medicine, 22(10), pp. 1140–1150.

53. Sun D-G, Tian S, Zhang L, Hu Y, Guan C-Y, Ma X and Xia H-F (2020) The miRNA-29b Is Downregulated in Placenta During Gestational Diabetes Mellitus and May Alter Placenta Development by Regulating Trophoblast Migration and Invasion Through a HIF3A-Dependent Mechanism, Frontiers in Endocrinology, 11, p. 169.
54. Tan M H, Li Q, Shanmugam R, Piskol R, Kohler J, Young A N, Liu K I, Zhang R, Ramaswami G, Ariyoshi K, Gupte A, Keegan L P, George C X, Ramu A, Huang N, Pollina E A, Leeman D S, Rustighi A, Sharon Goh Y P, Chawla A, Del Sal G, Peltz G, Brunet A, Conrad D F, Samuel C E, O'Connell M A, Walkley C R, Nishikura K and Li J B (2017) Dynamic landscape and regulation of RNA editing in mammals, Nature, 550(7675), pp. 249–254.
55. Uchida S and Jones S P (2018) RNA Editing: Unexplored Opportunities in the Cardiovascular System, Circulation Research, 122(3), pp. 399–401.
56. Wang I X, So E, Devlin J L, Zhao Y, Wu M and Cheung V G (2013) ADAR regulates RNA editing, transcript stability, and gene expression, Cell Reports, 5(3), pp. 849–860.
57. Wang K, Li M and Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Research, 38(16), p. e164.
58. Wang Q, Miyakoda M, Yang W, Khillan J, Stachura D L, Weiss M J and Nishikura K (2004) Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene, The Journal of Biological Chemistry, 279(6), pp. 4952–4961.
59. Wang Y, Zhao Y-R, Zhang A-Y, Ma J, Wang Z-Z and Zhang X (2017) Targeting of miR-20a against CFLAR to potentiate TRAIL-induced apoptotic sensitivity in HepG2 cells, European Review for Medical and Pharmacological Sciences, 21(9), pp. 2087–2097.
60. Witman N M, Behm M, Ohman M and Morrison J I (2013) ADAR-related activation of adenosine-to-inosine RNA editing during regeneration, Stem Cells and Development, 22(16), pp. 2254–2267.
61. Wu K, Hu M, Chen Z, Xiang F, Chen G, Yan W, Peng Q and Chen X (2017) Asiatic acid enhances

survival of human AC16 cardiomyocytes under hypoxia by upregulating miR-1290, IUBMB life, 69(9), pp. 660–667.

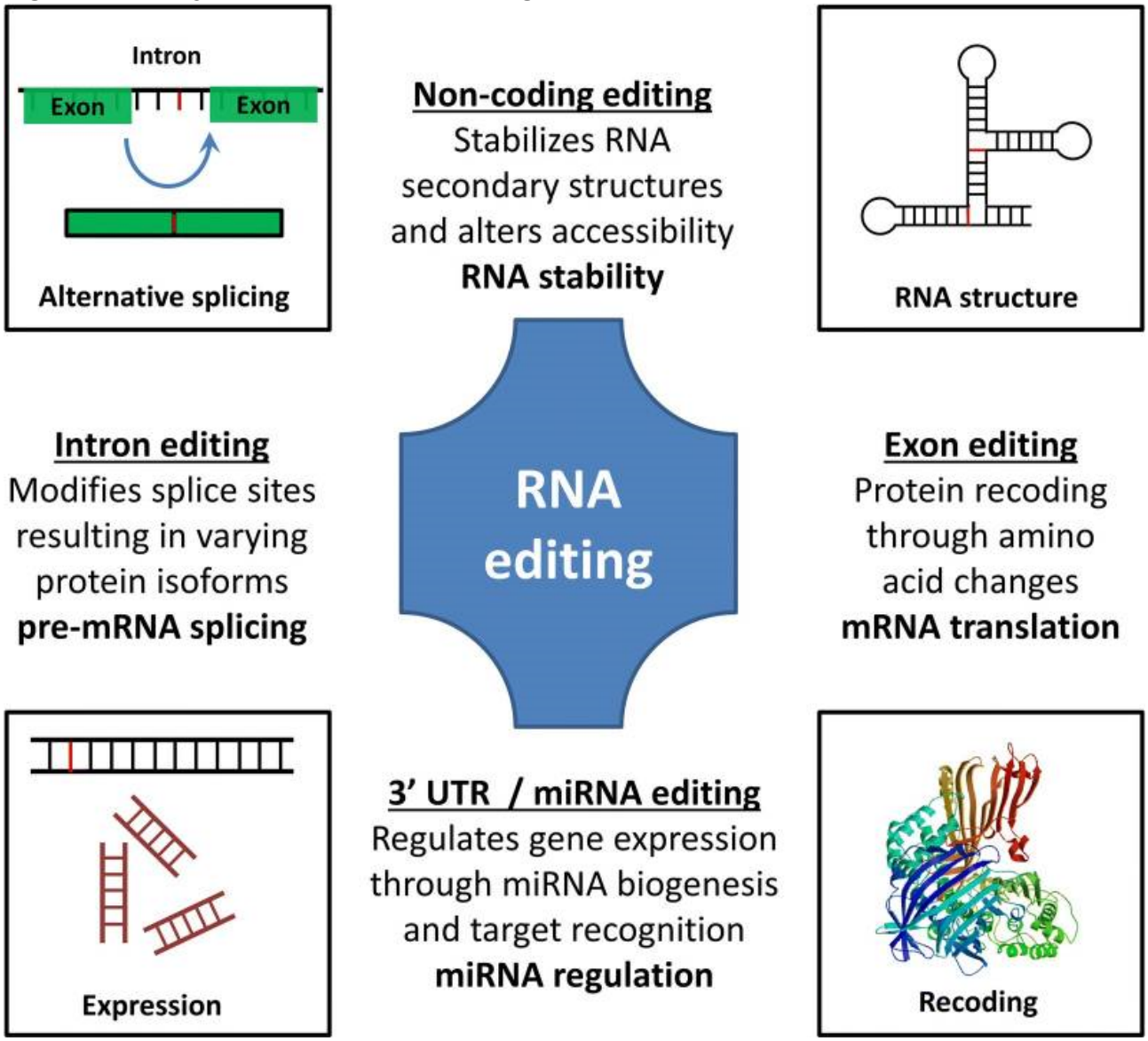
62. Yang W, Chendrimada T P, Wang Q, Higuchi M, Seeburg P H, Shiekhattar R and Nishikura K (2006a) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases, Nature Structural & Molecular Biology, 13(1), pp. 13–21.
63. Yang W, Chendrimada T P, Wang Q, Higuchi M, Seeburg P H, Shiekhattar R and Nishikura K (2006b) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases, Nature Structural & Molecular Biology, 13(1), pp. 13–21.
64. Zolk O, Solbach T F, Eschenhagen T, Weidemann A and Fromm M F (2008) Activation of negative regulators of the hypoxia-inducible factor (HIF) pathway in human end-stage heart failure, Biochemical and Biophysical Research Communications, 376(2), pp. 315–320.
65. (No date) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data - PubMed. Available at: https://pubmed.ncbi.nlm.nih.gov/20601685/ (Accessed 5 May 2021).
66. (No date) Gene: HIF3A (ENSG00000124440) - Summary - Homo sapiens - Ensembl genome browser 89. Available at: http://may2017.archive.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000124440;r=19:46297046-46343433 (Accessed 8 May 2021).
67. (No date) Gene: Hif3a (ENSMUSG00000004328) - Summary - Mus musculus - Ensembl genome browser 89. Available at: http://may2017.archive.ensembl.org/Mus_musculus/Gene/Summary?db=core;g=ENSMUSG00000004328;r=7:17031507-17062427 (Accessed 8 May 2021).

**Figure 1. Mechanism of RNA editing event in double-strand RNAs (Christofi, 2019)**



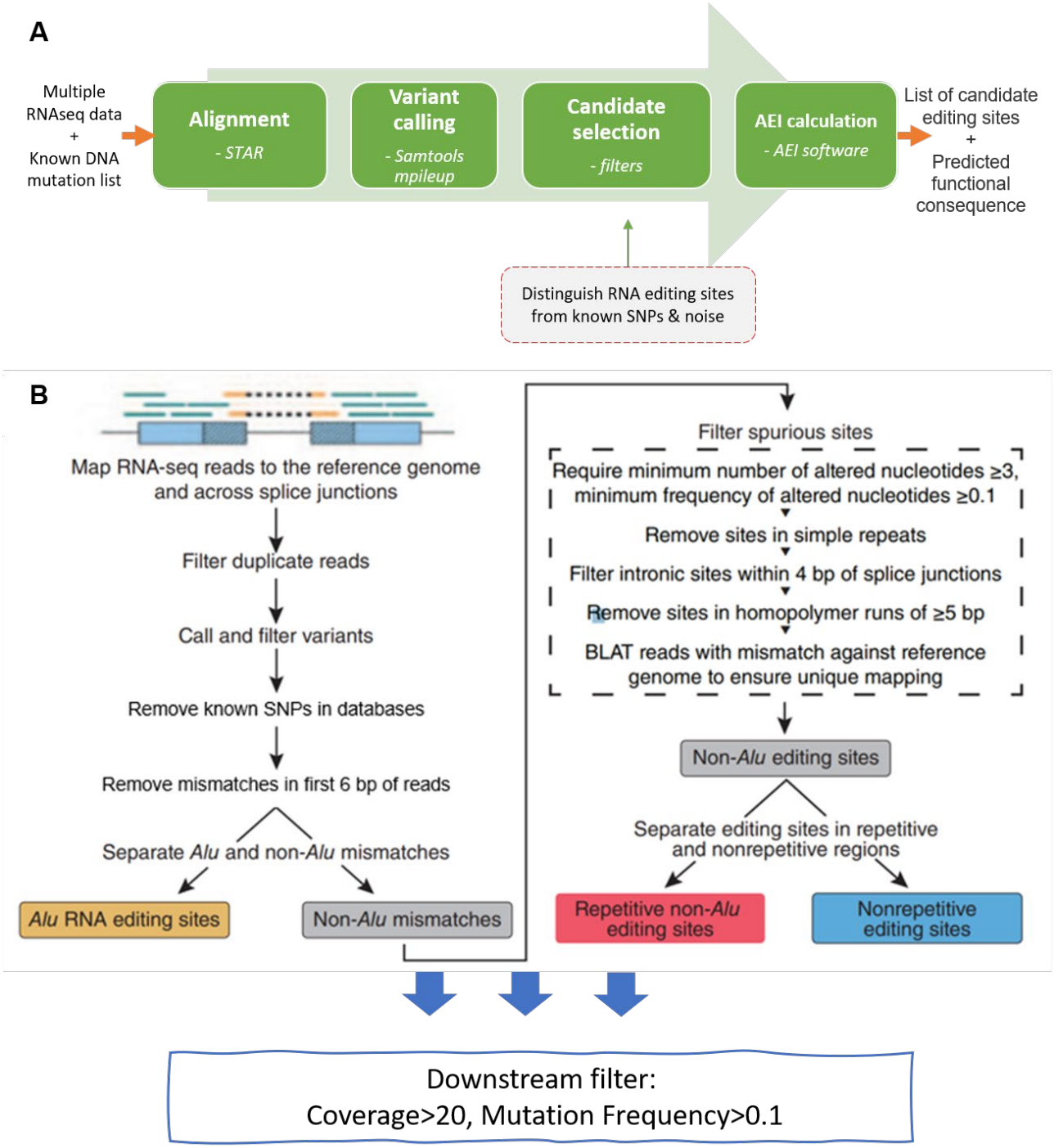

In the two major types of RNA editing, the specific family of deaminase (**a:** APOBEC, **b:** ADAR) could recognize the adenosine (A) or cytidine (C) site in the double-strand RNA (dsRNA) and then deaminate them into Inosine (I) or uridine (U), which change the translated amino acid type and thus results in protein structural and functional changes. The corresponding cellular activities are thus altered.

**Figure 2. RNA-type dependent RNA-editing functions (Christofi, 2019)**



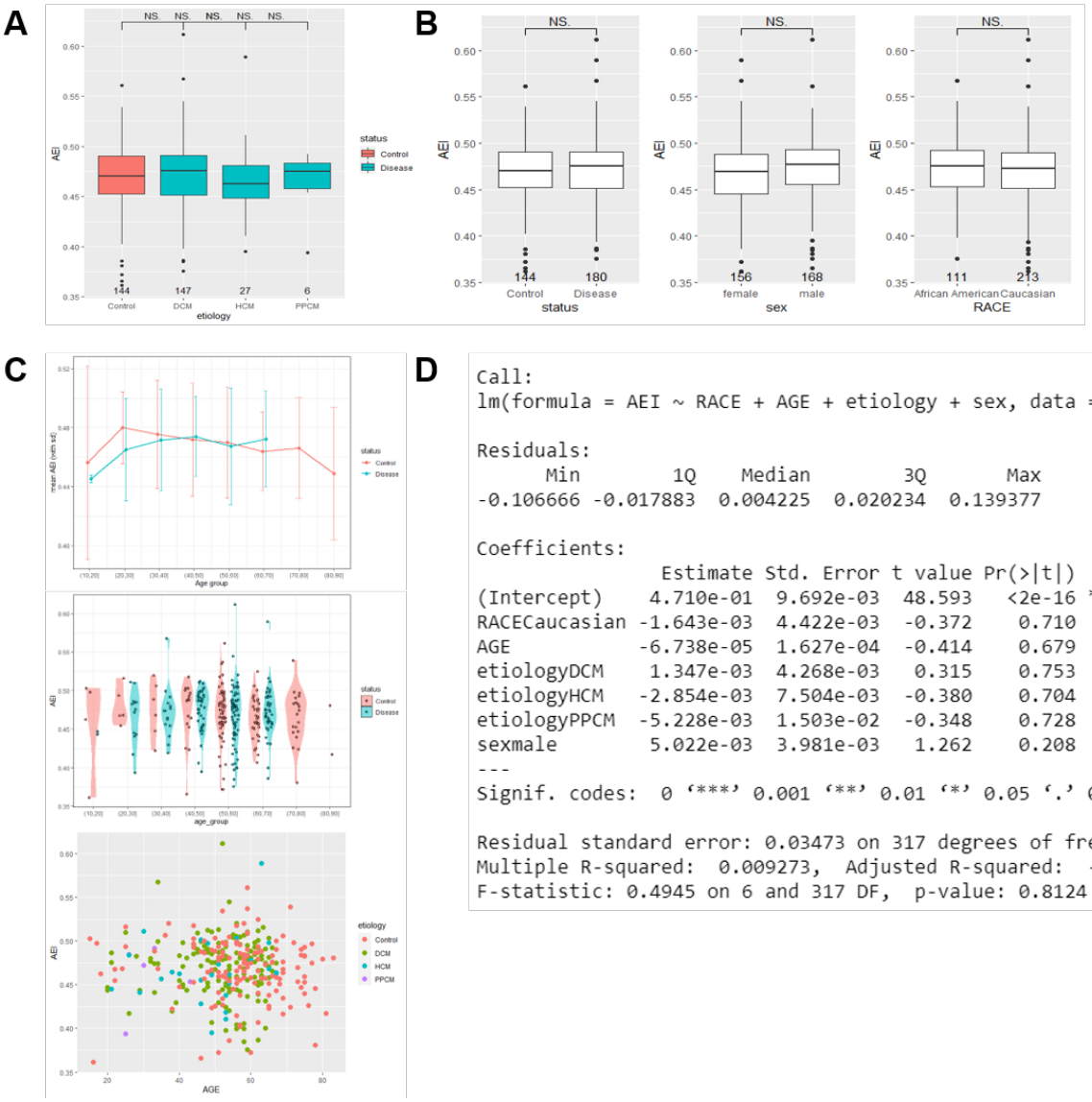

The functional roles of RNA editing vary in different RNA types. **Top:** change in RNA structure and modification in RNA stability due to RNA editing in non-coding sites. **Bottom:** Alteration in miRNA binding and decay due to RNA editing in 3'UTR or miRNA. **Left:** New protein isoforms introduced by RNA editing in introns. **Right:** changes in amino acids in mRNA translation due to exon editing.

**Figure 3. Bioinformatic pipeline for RNA editing site detection from RNAseq data alone (adopted from Ramaswami, 2012)**




Detailed workflow for RNA editing detection pipeline. A. The overall workflow for RNA editing identification from RNAseq data. B. The specific filtering options during candidate selection.

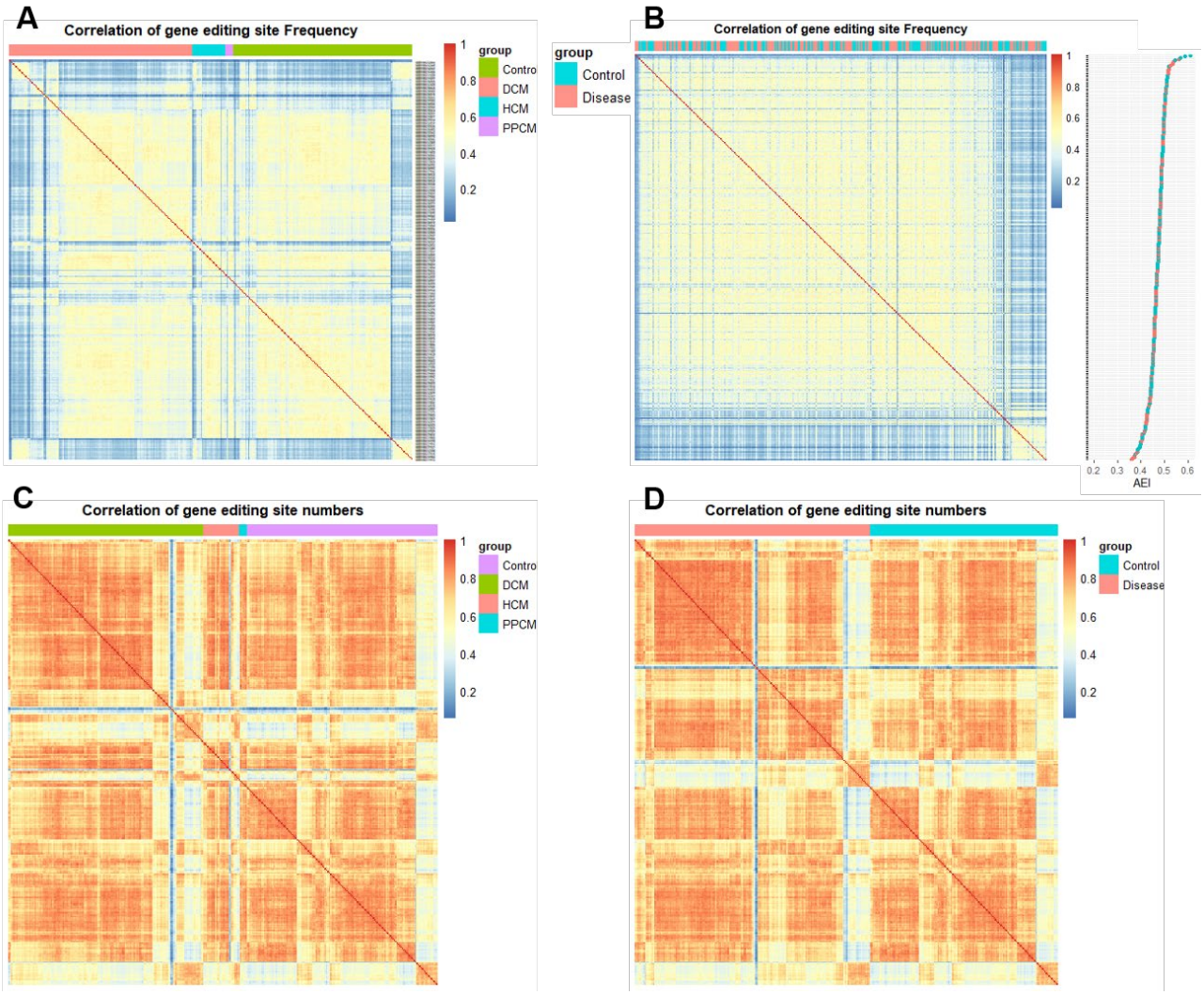
**Figure 4. Overall comparison of editing levels**




```
Call:
lm(formula = AEI ~ RACE + AGE + etiology + sex, data = aei)

Residuals:
      Min        1Q    Median        3Q       Max
-0.106666 -0.017883  0.004225  0.020234  0.139377

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.710e-01  9.692e-03  48.593   <2e-16 ***
RACECaucasian -1.643e-03  4.422e-03  -0.372    0.710
AGE           -6.738e-05  1.627e-04  -0.414    0.679
etiologyDCM    1.347e-03  4.268e-03   0.315    0.753
etiologyHCM   -2.854e-03  7.504e-03  -0.380    0.704
etiologyPPCM  -5.228e-03  1.503e-02  -0.348    0.728
sexmale        5.022e-03  3.981e-03   1.262    0.208
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03473 on 317 degrees of freedom
Multiple R-squared:  0.009273,  Adjusted R-squared:  -0.009479
F-statistic: 0.4945 on 6 and 317 DF,  p-value: 0.8124
```

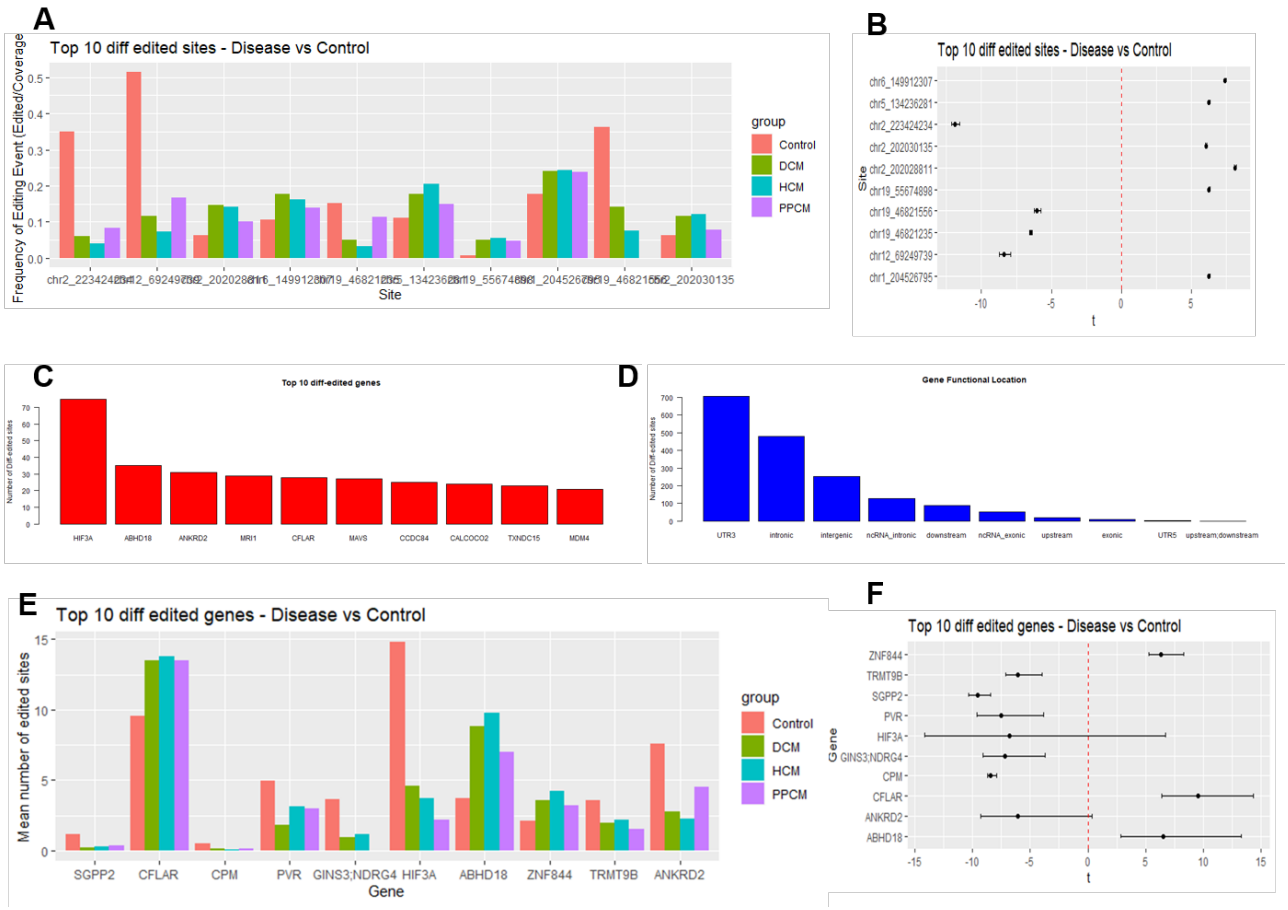
No significant difference in AEI between groups**. A.** Boxplot of AEI values between control and three cardiomyopathy etiologies (ECM, HCM, PPCM), no significant difference was identified by pair-wise t-test between either groups. **B.** Boxplot of AEI values between control and disease groups, males and females, and Caucasian and African American. **C.** Distributional difference among different age groups. The top is the line chart of mean AEI of each age interval, the error bars are standard deviation (sd) of each group. Violin plots and dot plots were also applied for distribution analysis. No significant difference or tendency was found. **D.** linear regression with calling of AEI ~ race + age + etiology + sex. No significantly involved variables were found.

**Figure 5. Intra-group and inter-group variation in sample correlations**



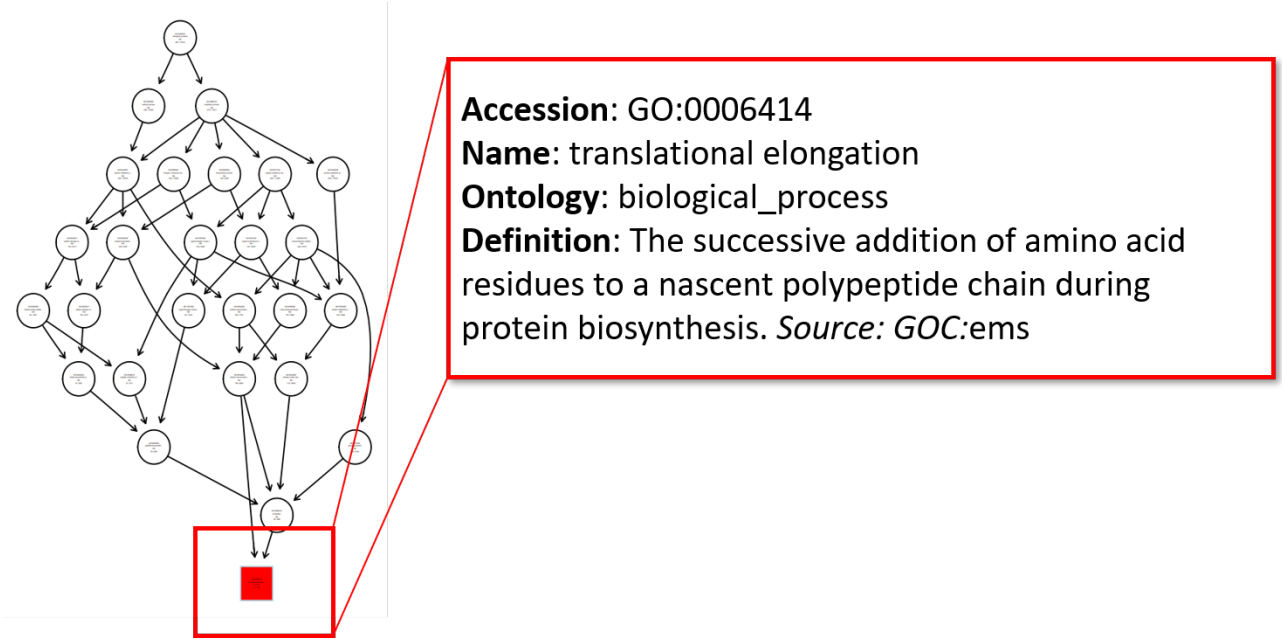

Sample variation within and between groups was found with different clusters. **A.** Correlation heatmap of four groups based on the editing site frequency (33005 sites in total for each sample). Sample clustered within each etiology group. **B.** Correlation heatmap of editing site frequency sorted by AEI value. The AEI value of each sample is listed on the right. A unique cluster of samples was found in the low AEI group. **C.** Correlation heatmap of four etiology groups based on the number of editing sites of each gene (3268 genes in total for each sample). Sample clustered within each etiology group **D.** Gene-based correlation heatmap of disease and control groups.

**Figure 6. Differential analysis of edited sites and genes between groups**




Top 10 sites and genes filtered through different methods. **A.** Editing frequency of top 10 differentially edited sites in four etiology groups. Pairwise t-tests of site editing frequency were performed as a ranking standard. **B.** 95% Confidential interval for top 10 diff-edited sites. **C.** Number of diff-edited sites of top 10 clustered genes with most diff-edited sites. **D.** Number of diff-edited sites in 10 functional regions. **E**. Top 10 diff-edited genes based on the difference in the number of editing candidate sites. The y axis is the mean number of edited sites of each etiology group. **F.** 95% Confidential Interval (CI) of top 10 count-based diff-edited genes.

**Figure 7. GO functional enrichment with significant diff-edited genes**




The Gene Ontology (GO) Biological Process (BP) enrichment of 410 genes with diff-edited sites showed a significantly clustered GO term of translational elongation (11/124). The pdf version of the GO enrichment graph is attached independently due to the image resolution limitation in word.

**Figure 8. Site-based editing frequency of top differentially edited genes**


The mean editing frequency of all diff-edited sites in HIF3A (A), ABHD18 (B), ANKRD2 (C), CFLAR (D) gene. Downregulation of RNA editing level in disease group was found in HIF3A and ANKRD2, while upregulation of RNA editing level in disease group was found in ABHD18 and CFLAR.