

# 系统鉴定与分析新型冠状病毒感染心肌细胞内的 A-to-I RNA 编辑事件

赵珊珊, 丁素萍, 朱秀志, 袁丰华, 周江赢, 揭亚亮, 袁志栋\*

(赣南医学院 基础医学院, 中国江西 赣州 341000)

**摘要:** 严重急性呼吸综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) 感染增加了心肌损伤的风险, 其致病分子机制还不是很清楚。本文比较了 SARS-CoV-2 感染心肌细胞(感染组)与未感染病毒心肌细胞(对照组)中 RNA 编辑的变化, 以期从 RNA 编辑角度了解 SARS-CoV-2 感染对心肌细胞的影响。首先, 从 GEO 数据库中下载 SARS-CoV-2 感染的人诱导多能干细胞衍生的心肌细胞(human induced pluripotent stem cell-derived cardiomyocyte, hiPSC-CM)的原始测序数据(GSE150392); 然后, 用 SPRINT 软件识别 RNA 编辑位点(RNA editing site, RES), 比较感染组和对照组心肌细胞编辑水平的变化; 最后, 对编辑位点进行分析 and 注释, 并进行相关基因的功能分析。结果显示, 总共检测到约 92 899 个碱基替换位点, 其中 87 670 个被鉴定为 A-to-I RES, 这些 A-to-I 编辑位点发生在 Alu 区域上的有 78 978 个; A-to-I 编辑位点倾向于成簇分布, 主要分布在内含子、基因间区等区域; 与对照组相比, 感染组中编辑水平显著上调的 A-to-I RES 有 102 个, 显著下调的有 94 个, 这些含有显著差异 RES 的基因参与的病毒感染相关的 GO (Gene Ontology)生物学过程主要富集于病毒过程、病毒生命周期和响应病毒感染的防御等, 并且其 KEGG (kyoto Encyclopedia of Genes and Genomes)功能通路也富集在病毒感染等方面, 而筛选的含有高质量 A-to-I RES、组间编辑水平存在显著差异的 11 个基因的功能则富集于内吞、细胞因子-细胞因子受体相互作用、蛋白酶体、烟酸和烟酰胺代谢以及铁死亡等方面。本研究结果表明, SARS-CoV-2 感染影响了心肌细胞中 A-to-I 的 RNA 编辑, 这类 RNA 编辑事件的发生体现了宿主心肌细胞对病毒感染的一种响应机制。

**关键词:** 严重急性呼吸综合征冠状病毒 2 (SARS-CoV-2); 心肌细胞; RNA 编辑; A-to-I 编辑

中图分类号: Q81, R511, R542.2

文献标志码: A

文章编号: 1007-7847(2023)03-0196-14

## Systematic Identification and Analysis of A-to-I RNA Editing Events in SARS-CoV-2-infected Cardiomyocytes

ZHAO Shanshan, DING Suping, ZHU Xiuzhi, YUAN Fenghua, ZHOU Jiangying, JIE Yaliang, YUAN Zhidong\*

(College of Basic Medicine, Gannan Medical University, Ganzhou 341000, Jiangxi, China)

**Abstract:** Novel coronavirus (SARS-CoV-2) infection increases the risk of myocardial injury, and the molecular mechanisms of pathogenesis remain to be elucidated. Herein, the changes of RNA editing in SARS-CoV-2 infected cardiomyocytes (infection group) and uninfected cardiomyocytes (mock group) were compared to explore the effects of the viral infection on cardiomyocytes from the perspective of RNA editing. RNA-seq data (GSE150392) of SARS-CoV-2 infection of human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs) was downloaded from the GEO database, and RNA editing sites (RESs) were identified using SPRINT software. The changes in editing levels of cardiomyocytes in the infection and mock groups were compared, editing sites were annotated, and relevant functional analyses were performed. A total of about 92 899 base substitutions were detected, of which 87 670 were identified as A-to-I RESs, and 78 978

收稿日期: 2022-12-14; 修回日期: 2023-02-20; 网络首发日期: 2023-04-10

基金项目: 国家自然科学基金资助项目(82260288); 心脑血管疾病防治教育部重点实验室开放课题(XN202019)

作者简介: 赵珊珊(1997—), 女, 四川江油人, 硕士研究生; \*通信作者: 袁志栋(1979—), 男, 博士, 教授, 主要从事生物信息学、心脑血管疾病病理机制的研究, Tel: 0797-8169760, E-mail: yuanzhidong@gmu.edu.cn。

of these A-to-I editing sites were found to appear in the Alu regions. A-to-I editing sites tend to be distributed in clusters, mainly in regions such as introns and intergenic regions. By screening the differential RESs between the infection and mock groups in A-to-I editing, it was found that 102 of the significantly differential RESs are up-regulated and 94 are down-regulated in editing level. These genes with significant differences in RES are involved in the biological processes of viral infection related GO (Gene Ontology), which are mainly enriched in viral process, virus life cycle, and defense response to virus, *et al.*, and their Kyoto Encyclopedia of Genes and Genomes (KEGG) functional pathways are also mainly related to viral infection. In contrast, the 11 screened genes with high-quality A-to-I RESs and significant differences in editing levels between groups are enriched in endocytosis, cytokine-cytokine receptor interaction, proteasome, nicotine and nicotinamide metabolism, and ferroptosis, *et al.*. The results showed that SARS-CoV-2 infection affects A-to-I RNA editing in cardiac myocytes, and the occurrence of such RNA editing events is a response of host cardiac myocytes to the viral infection.

**Key words:** severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); cardiomyocyte; RNA editing; A-to-I editing

(*Life Science Research*, 2023, 27(3): 196-209)

严重急性呼吸综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) 是一种有包膜的单股正链 RNA 病毒, 属于  $\beta$  属冠状病毒, 使用其自身的 RNA 依赖性 RNA 聚合酶(nsp12)进行复制, 可以在短暂病毒血症期间或通过感染巨噬细胞渗入心肌到达心脏<sup>[1-2]</sup>。载脂蛋白 B mRNA 编辑酶催化多肽样蛋白(apolipoprotein B mRNA editing enzyme catalytic polypeptide-like, APOBEC)和双链 RNA 依赖的腺苷脱氨酶(adenosine deaminase acting on RNA, ADAR)都是机体先天免疫的一部分, 在病毒感染期间, ADAR 通过病毒 RNA 的超突变直接发挥作用, 或通过编辑宿主细胞转录本调节细胞反应从而间接发挥作用, 宿主先天免疫防御会驱动 SARS-CoV-2 的核苷酸替换和基因组进化<sup>[3-5]</sup>。有研究表明, SARS-CoV-2 能直接侵入心肌细胞, 引发多种心血管疾病, 如心肌损伤、心律失常、急性冠脉综合征和血栓等, 这类患者预后极差, 总体死亡率较高, 目前没有明确的策略来解决新冠病毒感染中出现的不良心脏疾病<sup>[6-7]</sup>。

RNA 编辑是指通过插入、删除或转换核苷酸来改变核苷酸序列, 从而增加转录本多样性的转录后事件; 其通过调节受影响基因的表达和功能来促进转录组多样性, 从而维持正常内环境稳定<sup>[8-9]</sup>。在动物中, 最常见的 RNA 编辑类型是 A-to-I 编辑, ADAR 通过与双链 RNA (double-stranded RNA, dsRNA)结合, 使腺苷(adenosine, A)脱氨基转化为肌苷(inosine, I), A-to-I 的 RNA 编辑可以

导致密码子改变, 进而改变蛋白质编码序列, 使蛋白质功能多样化<sup>[10]</sup>。ADAR 酶家族中的成员分别是 ADAR1 (dsRNA 特异性腺苷脱氨酶)、ADAR2 (dsRNA 特异性编辑酶 1)和 ADAR3。ADAR1 和 ADAR2 在大多数组织中表达, ADAR3 主要在大脑中表达<sup>[11-12]</sup>; 其中, ADAR1 和 ADAR2 催化了 RNA 序列中 A-to-I 编辑的发生<sup>[10]</sup>。有研究表明, SARS-CoV-2 感染后, 患者体内存在宿主细胞脱氨酶对 SARS-CoV-2 的 RNA 编辑事件, 其中 ADAR1 p110 亚型在 SARS-CoV-2 的 RNA 编辑活动中起作用<sup>[13]</sup>。目前, 关于宿主细胞脱氨酶对 SARS-CoV-2 进行 RNA 编辑的研究报道较多, 而 SARS-CoV-2 感染对宿主细胞自身 RNA 编辑影响的研究很少, 其中, 新冠病毒感染患者心肌或 SARS-CoV-2 体外感染心肌细胞的 RNA 编辑事件还没有相关报道。本文主要分析了 SARS-CoV-2 感染心肌细胞对宿主细胞转录本 RNA 编辑变化的影响, 有助于进一步了解 SARS-CoV-2 感染导致心肌细胞损伤及心肌细胞响应病毒感染的分子机制, 可为其诊断和治疗提供新的思路。

## 1 材料与方法

### 1.1 数据来源

SARS-CoV-2 感染的人诱导多能干细胞衍生的心肌细胞(human induced pluripotent stem cell-derived cardiomyocyte, hiPSC-CM)的转录组测序数据集 GSE150392 (下载于 2021-01-09), 来源于美国国家生物技术信息中心(National Center for Bio-

technology Information, NCBI)的 GEO (Gene Expression Omnibus)数据库。该数据集为单端测序数据,包括 3 个 SARS-CoV-2 感染的 hiPSC-CM 样本(感染组,即 SARS-CoV-2 组)、3 个正常的 hiPSC-CM 样本(对照组,即 mock 组)。

## 1.2 数据预处理

从 GEO 数据库下载 hiPSC-CM 转录组测序数据,使用 fastp (版本 0.19.5)对数据进行质量控制。

## 1.3 RNA 编辑位点的鉴定和注释

首先,从 UCSC 基因组浏览器(<https://genome.ucsc.edu>)下载人类 hg38 参考基因组数据和重复序列的注释文件,从 GENCODE 数据库(<https://www.gencodegenes.org>)下载基因组注释文件 gencode.v36.chr\_patch\_hapl\_scaff.annotation.gtf。然后,使用 SP-RINT 软件识别 RNA 编辑位点(RNA editing site, RES),构建映射索引,进而识别常规 RES 和超编辑 RES,并且从输出文件中提取 A-to-I 编辑位点。最后,利用 REDiportal 数据库<sup>[14]</sup>识别已知的 A-to-I 编辑位点,采用 ANNOVAR 软件<sup>[15]</sup>对所有得到的 A-to-I 编辑位点进行注释。将距离小于 500 bp 的相邻两个 A-to-I 编辑位点定义为成簇位点,使用 R 软件(版本 4.3.0)的 ggplot2 包(版本 3.4.2)绘制相邻成簇位点之间距离的密度图。

## 1.4 差异编辑位点的识别及分析

采用非参数 Wilcoxon 秩和检验比较 A-to-I 的 RNA 编辑中感染组和对照组的编辑水平。对于 A-to-I 中的 RES, RNA 编辑水平为支持编辑碱基的读段(read)与该位点上检测到的总读段数的比率。用 limma 包(版本 3.52.1)<sup>[16]</sup>对 A-to-I 编辑中感染组和对照组的编辑水平进行差异分析,组间差异显著的编辑位点的筛选标准为  $|\log_2(\text{FC})| > 0.5$  (FC: fold change)且  $P < 0.05$ 。采用 ANNOVAR 软件进行注释,利用 pheatmap (版本 1.0.12)、ggplot2 包对相关结果进行可视化处理。使用 clusterProfiler 包(版本 4.8.1)<sup>[17]</sup>对组间差异显著的编辑位点所在的基因进行京都基因和基因组数据库(Kyoto Encyclopedia of Genes and Genomes, KEGG)及基因本体论(Gene Ontology, GO)富集分析。在感染组、对照组编辑位点的上下游均取 5 个碱基,使用 R 软件的 ggseqlogo 包(版本 0.1)绘制序列标识图。使用每百万读段中来自于某基因每千碱基长度的读段数(reads per kilobase of exon model per million mapped reads, RPKM)标准化表达数据,利用 R 软件对 ADAR 家族成员(ADAR1、ADAR2 和 ADAR3)在感染组和对

照组中各自标准化的表达数据进行可视化分析。

## 1.5 不同编辑水平的 A-to-I 编辑位点的分析

对 A-to-I 编辑位点的编辑水平进行分析,用 R 软件对 A-to-I 编辑位点进行处理,筛选出高质量编辑位点。高质量 A-to-I 编辑位点的筛选标准为: RNA 编辑水平至少为 0.3,支持该位点发生 A-to-I 变化的读段至少有 3 条,且该位点上检测到的总读段数至少有 10 条覆盖该位点,同时组间变化的对数绝对值大于 0.5 且  $P$  值小于 0.05。使用 R 软件对高质量 A-to-I 编辑位点所在基因进行注释,并对注释的基因集进行富集分析。另外,对 GSE150392 转录组测序数据集进行基因表达差异分析,使用 R 软件处理并整合组间存在显著差异表达的基因数据 [ $|\log_2(\text{FC})| > 0.5$  且  $P < 0.05$ ] 和 ANNOVAR 软件注释后的含有组间 A-to-I 编辑水平存在显著差异编辑位点的基因数据 [ $|\log_2(\text{FC})| > 0.5$  且  $P < 0.05$ ], 并对筛选出的基因进行分析。

## 2 结果

### 2.1 心肌细胞 RNA 编辑的总体概况

由于 RNA 可转录自 DNA 双链的任一条,所以当转录组测序读段比对到参考基因组序列上时,被鉴定为发生 A-to-I RNA 编辑的位点呈现 A-to-G 或 T-to-C 的碱基替换形式<sup>[18]</sup>。本研究检测了 SARS-CoV-2 感染心肌细胞后的 RNA 编辑事件,通过对原始测序数据进行系统性的分析,总共识别到 92 899 个碱基替换位点;确定了比对结果中有 12 种碱基替换类型: AC、GT、AG、CA、CG、GC、AT、GA、CT、TG、TC 和 TA,其中,在数量上占绝大多数的碱基替换类型是 A-to-I RNA 编辑事件中的 AG、TC (图 1A)。在鉴定的 92 899 个碱基替换位点中,50 378 个被发现是 REDiportal 数据库已记录的 RES,约占总数的 54.23%;此外,感染组存在 44 869 个 A-to-I RES,对照组存在 46 878 个 A-to-I RES,感染组和对照组有 4 077 个共同的 A-to-I RES (图 1B)。本研究识别的替换位点绝大多数是 A-to-I 类型, A-to-I 编辑位点有 87 670 个,其中已知的 A-to-I 编辑位点有 50 465 个,约占 57.56%。这些 RES 倾向于在 500 bp 内成簇分布,其中相邻的位点大多数相距 100 bp 以内 (图 1C)。

有文献表明,几乎所有 A-to-I 编辑位点都位于 Alu 序列<sup>[19]</sup>。对心肌细胞中的 A-to-I 编辑位点进行注释,发现位于 Alu 元件的编辑位点有 78 978

个,约占总 A-to-I 编辑位点的 90.09%;位于重复序列非 Alu 元件的编辑位点有 3 553 个(约占 4.05%),而位于非重复序列区域的位点有 5 139 个(约占 5.86%)。由此可见,与前人研究结果<sup>[19]</sup>类似,本研究中 A-to-I 编辑位点也主要位于 Alu 元件。而且, A-to-I 编辑位点主要位于内含子(约占 45.59%)、基因间区(约占 30.90%)和 3'非翻译区(3'-untranslated region, 3'-UTR)(约占 12.37%)等区域(图1D);位于外显子区域的 A-to-I 编辑位点大多数为非

同义变化(图 1E)。研究表明,蛋白质编码区间的 RNA 编辑事件产生的非同义变化将可能导致氨基酸序列的改变,从而影响蛋白质的功能。例如, SARS-CoV-2 转录组中的非同义 RNA 编辑对 SARS-CoV-2 的传播、致病机制和免疫干预有影响<sup>[20-22]</sup>。

## 2.2 ADAR 对 A-to-I 编辑位点局部序列的识别偏好模式

为了探索 ADAR 对 A-to-I RES 序列的识别

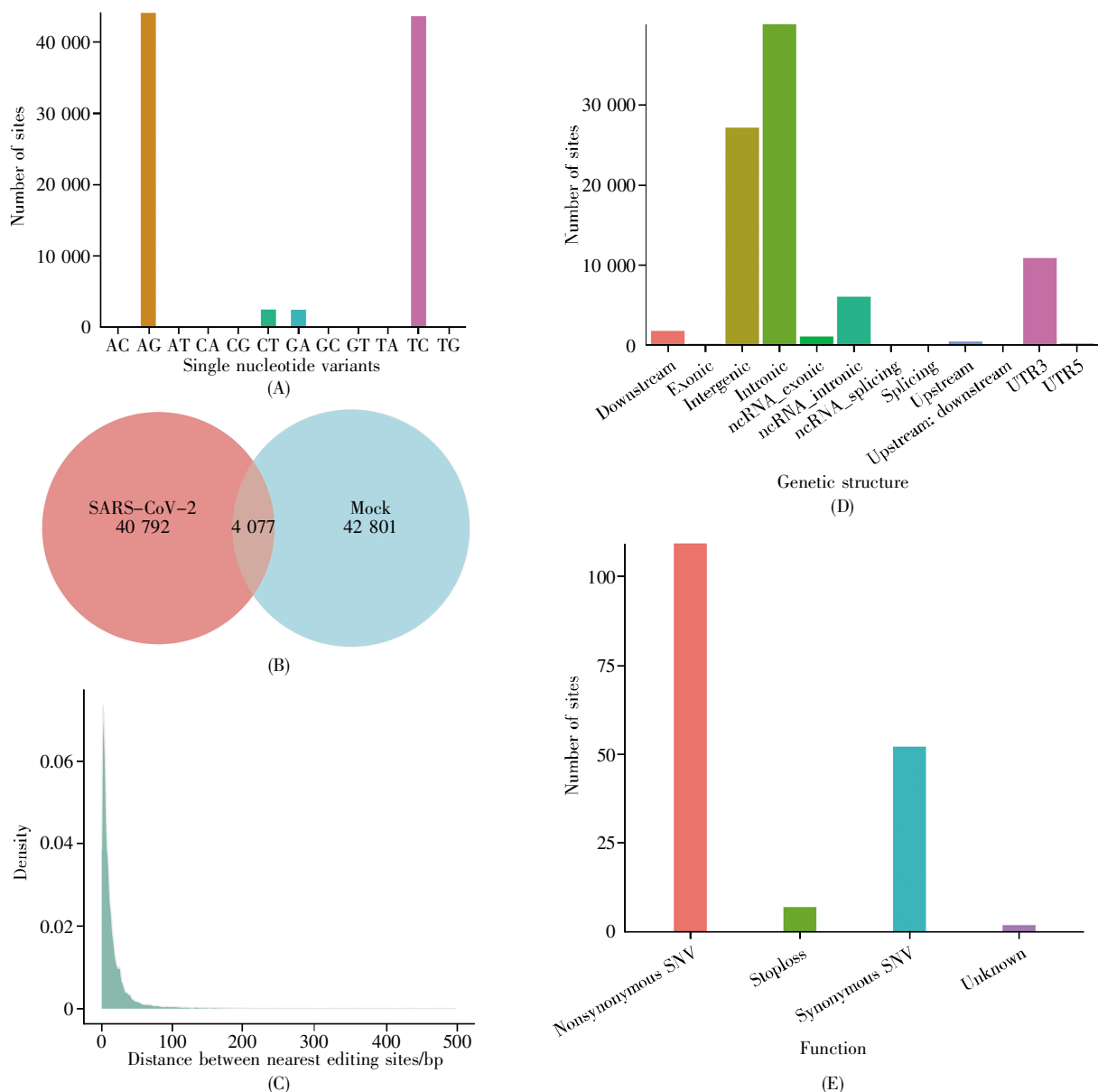


图1 RNA 碱基编辑位点的鉴定和分析

(A) 全部碱基替换类型分析; (B) 感染组和对照组之间的共同 A-to-I 编辑位点数量; (C) A-to-I RNA 编辑相邻成簇编辑位点的距离密度图; (D) A-to-I 编辑位点在整个基因组中的分布情况; (E) A-to-I RNA 编辑外显子功能分类图。

### Fig.1 RNA base editing site identification and analysis

(A) Distribution of all base substitution types; (B) Number of A-to-I co-editing sites between infection and mock groups; (C) Distance density plot of A-to-I RNA editing adjacent to clustered editing sites; (D) Distribution of the A-to-I editing sites through the whole genome; (E) Distribution of A-to-I RNA editing sites among exon function classification.

是否存在偏好性,对感染组和对照组编辑位点的序列模式进行了分析。结果显示,在感染组和对照组的编辑位点中,较高比例的序列组成模式分别为 ATATCAGCCGA 和 ATACCAGCCCA,它们之间仅有两个碱基的差异;除位点 4、9 和 10 的碱基出现比例差异较大外,其他位点的碱基呈现模式非常类似(图 2A~B)。在组间差异显著的编辑位点选取上下游各 5 个碱基绘制序列标识图,结果显示,虽然在编辑位点前后各 5 个碱基的区间均未发现非常保守的较长的序列模式,但是编辑

位点前后 1 个碱基存在非常保守的模式,即编辑位点 A 的前一个碱基 G 倾向于缺失,而后一个碱基 G 出现的概率较高,并且其后紧挨着的两个 C 出现的概率也比较高(图 2A~C)。这说明 ADAR 酶作用的 A-to-I RES 前后的局部序列有一定的保守性。

### 2.3 组间 A-to-I 编辑位点数量与编辑水平的比较

从图 3 可知,感染组和对照组发生碱基替换的位点在各碱基替换类型中的分布与数量上总体差别不大,只是对照组发生 RNA 编辑事件的位点稍多一些(图 3A);然而,两组 A-to-I 的编辑水平却呈现出极其显著的统计学差异(图 3B),总体上感染组的编辑水平显著高于对照组(Wilcoxon 检验,  $P < 2.2E-16$ );并且,与对照组相比,感染组有 102 个编辑位点的编辑水平显著升高,94 个编辑位点的编辑水平显著降低(图 3C)。对含有上述 196 个编辑水平在组间有显著差异的 A-to-I 编辑位点的基因进行 GO 富集分析,显著富集前 10 的结果显示:在生物过程(biological process, BP)方面这些基因主要富集在肌动蛋白丝过程的调控、肌动蛋白结构组织等;富集的细胞成分(cellular component, CC)为黏着连接、染色体与中心区、肌节等;富集的分子功能(molecular function, MF)包括磷脂酰肌醇-3, 4, 5-三磷酸酶结合、蛋白丝氨酸/苏氨酸磷酸酶活性和蛋白磷酸酶调节剂的活性等(图 3D)。而直接与病毒感染相关的生物学过程则主要富集于病毒生命周期、响应病毒感染的防御、病毒过程的调控等(图 3E)。KEGG 通路中显著富集前 10 的结果显示,这些 RNA 编辑水平在组间存在显著差异的基因主要富集在神经系统疾病与病毒感染相关的通路,即肌萎缩侧索硬化、阿尔茨海默病、丙型肝炎和人乳头瘤病毒感染等(图 3F)。

### 2.4 不同编辑水平的 A-to-I 编辑位点的分析

图 4A 显示了 A-to-I 编辑位点在 0 到 1 编辑水平的分布,其中以编辑水平在 0.9~1.0 的 A-to-I 编辑位点最多。进一步按照更严谨的高质量编辑位点筛选标准(RNA 编辑水平至少为 0.3,支持编辑碱基的测序读段至少有 3 条,该位点上检测到的总读段数至少有 10 条,组间编辑水平差异变化的对数绝对值大于 0.5 且  $P$  值小于 0.05),对 A-to-I 的 RES 进行筛选,发现位于 0.3~0.4 编辑水平的 A-to-I 编辑位点最多(图 4B),而其他编辑水平分布的位点较少,那是因为不少 A-to-I RES

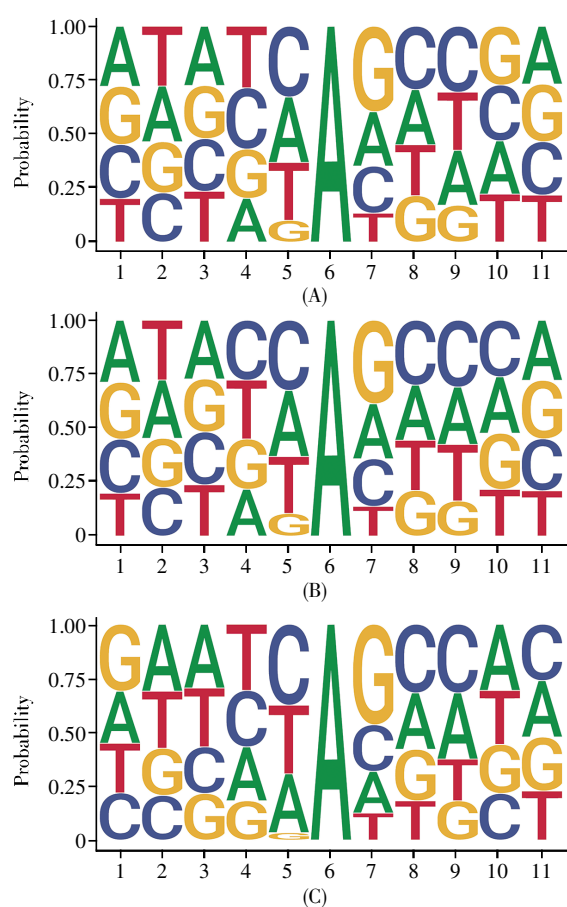


图 2 A-to-I 编辑的序列偏好

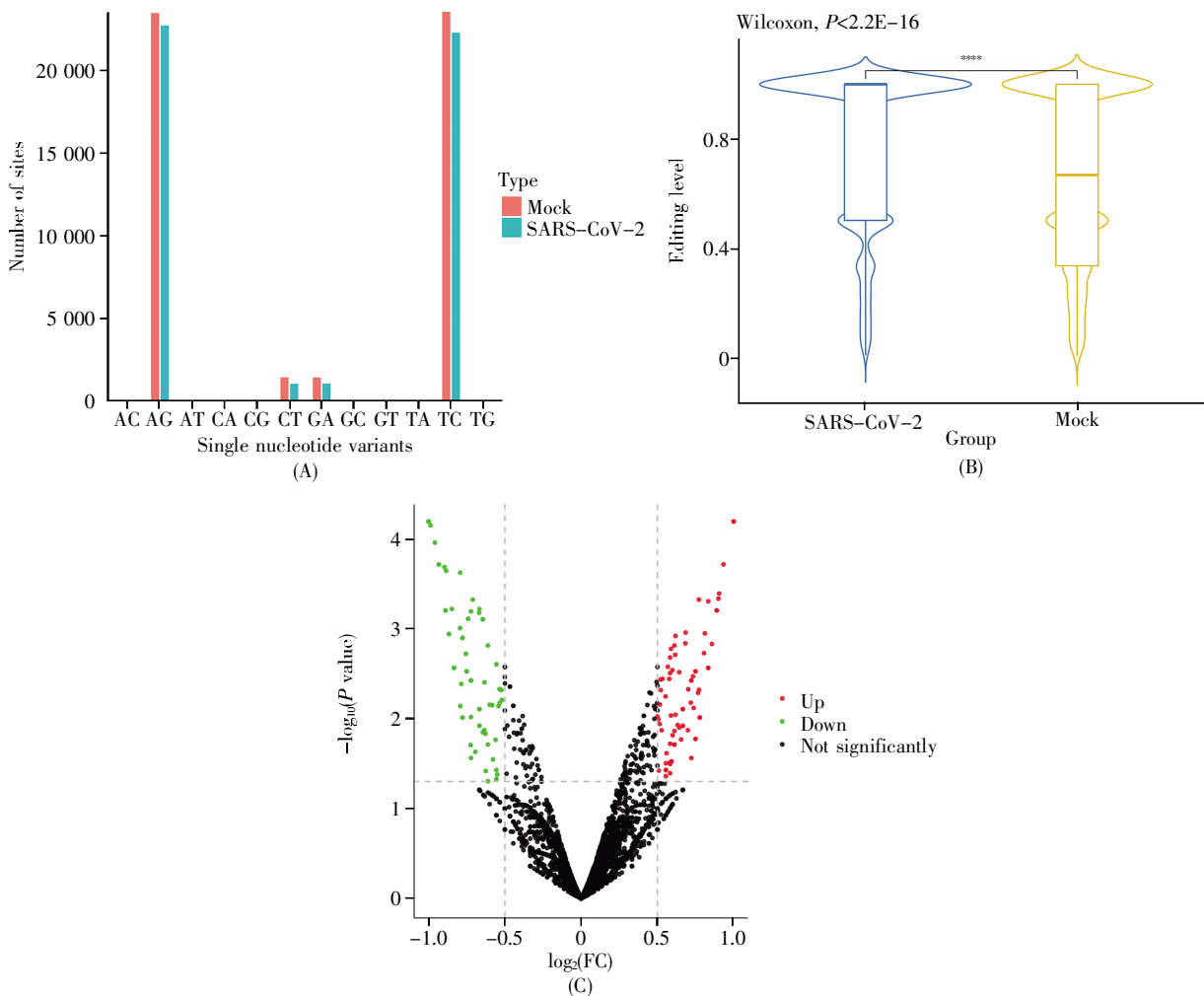
(A) 感染组 A-to-I 编辑位点局部序列的偏好模式; (B) 对照组 A-to-I 编辑位点局部序列的偏好模式; (C) 显著差异编辑位点局部序列的偏好模式。横坐标代表以 A-to-I 编辑位点为中心(坐标 6 的位置)的上下游各 5 个碱基,纵坐标表示碱基出现的百分比。

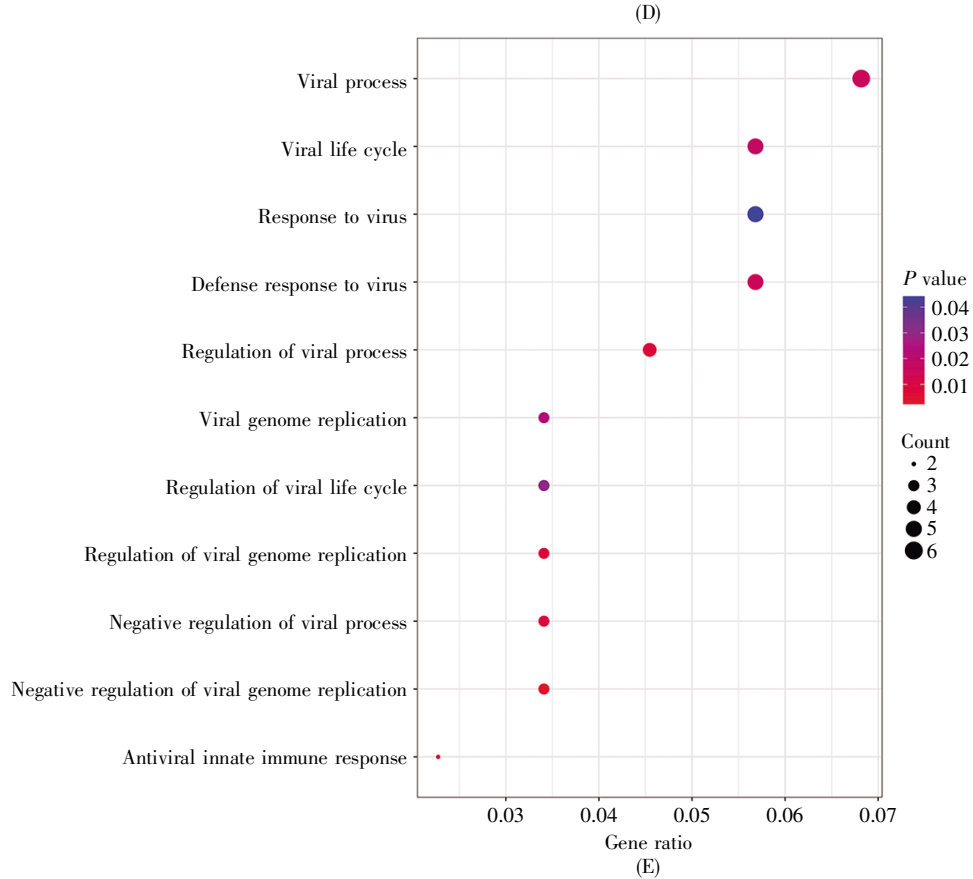
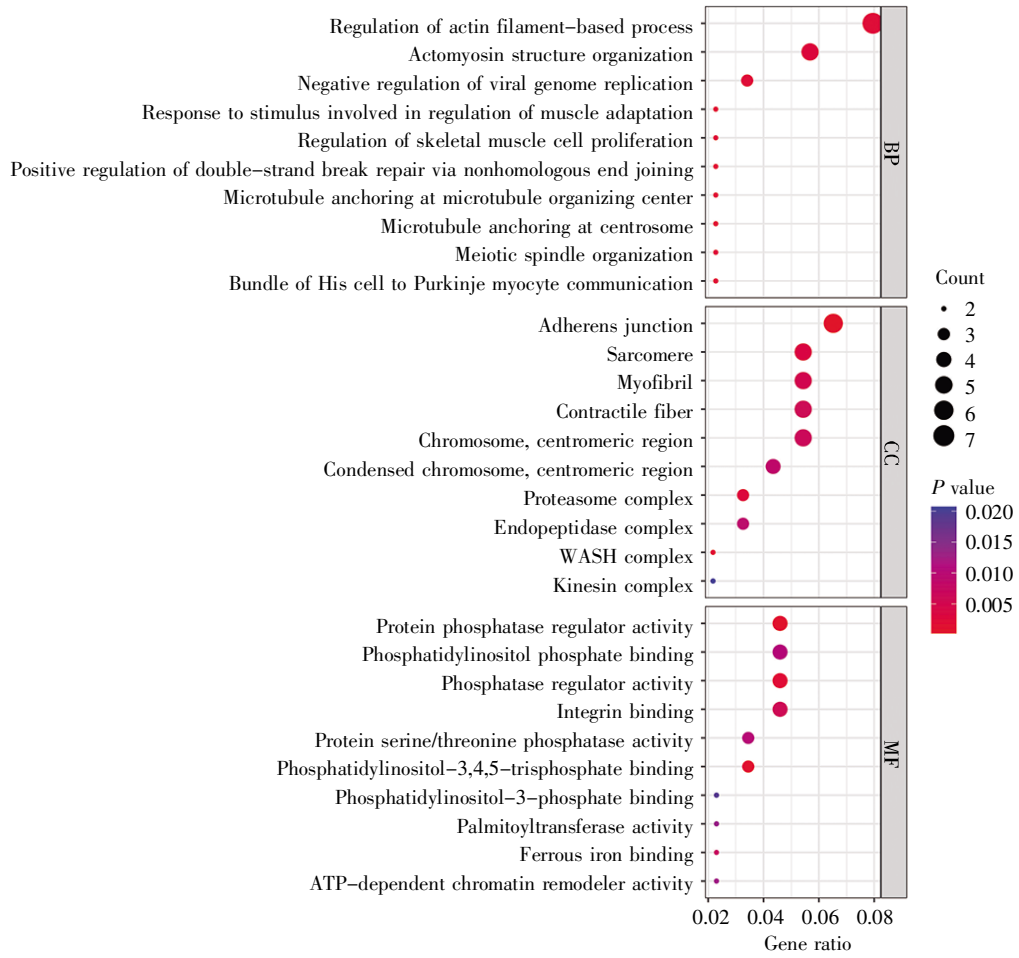
#### Fig.2 Sequence preference of A-to-I editing sites

(A) The local sequence preference surrounding the A-to-I sites in infection group; (B) The local sequence preference surrounding the A-to-I sites in mock group; (C) Preference patterns for local sequences at significantly different editing sites. The numbers on the horizontal coordinate represent five bases each upstream and downstream of the A-to-I editing site (centered at the position 6), and the vertical coordinate indicates the percentage of base occurrence.

的支持读段数少于 3 条, 或者总读段数小于 10 条。在所有鉴定的 A-to-I RES 中, 经 ANNOVAR 注释, 有 109 个含有在感染组和对照组间有显著差异编辑水平的 RES 的相关基因, 其中 43 个基因的转录水平在感染组和对照组间也存在显著差异(图 4C)。由于这些编辑位点主要位于编码基因的内含子、3'-UTR, 以及非编码基因内等, 所以可能对基因表达水平具有一定的影响, 但需要进一步的实验证据来支持。按照更严谨的高质量编辑位点的筛选标准过滤, 发现含有高质量 A-to-I RES、A-to-I 编辑水平在感染组和对照组间存在显著差异[ $|\log_2(\text{FC})| > 0.5$  且  $P < 0.05$ ]并经 ANNOVAR 注释的基因总共有 11 个, 分别是 *ADAM19*、*FTH1*、*MRI1*、*NMNAT1*、*PSMB2*、*SPON1*、*TNFRSF9*、*WASHC2A*、*WASHC2C*、*ZFP69B* 和 *ZNF815P*。对上述基因中表达上调[ $\log_2(\text{FC}) > 0$ ]的 *ADAM19*、*TNFRSF9*、*WASHC2A*、*MRI1*、*WASHC2C* 和 *ZFP69B* 进行富集分析, GO 分析结果显示: 这些基因富集的生物过程主要包括内体向高尔基体的逆向运输和

Arp2/3 复合物介导肌动蛋白成核的调控等, 富集的细胞成分为 WASH (Wiskott-Aldrich syndrome protein and SCAR homologue) 复合体、初级内体等, 富集的分子功能包括磷脂结合和磷脂酰肌醇磷酸结合等(图 4D); KEGG 通路的分析结果显示, 这些基因主要富集在内吞作用、细胞因子-细胞因子受体相互作用及半胱氨酸和蛋氨酸代谢等通路(图 4E)。对表达下调[ $\log_2(\text{FC}) < 0$ ]的基因 *NMNAT1*、*SPON1*、*PSMB2* 和 *FTH1* 进行富集分析, GO 分析结果显示: 这些下调基因富集的生物过程主要涉及金属离子的整合、蛋白质加工及成熟的正调控等, 富集的细胞成分为三级颗粒腔、次级溶酶体和蛋白酶体核心复合体等, 富集的分子功能包括作用于金属离子的氧化还原酶活性、核苷酸转移酶活性和铁离子结合等(图 4F); KEGG 通路的富集分析显示, 这 4 个基因主要富集在蛋白酶体、烟酸和烟酰胺代谢、铁死亡以及矿物质吸收等通路(图 4G)。需要指出的是, *ZNF815P* 基因不在差异表达基因列表里, 故未纳入相关注释分析。进一步筛选发





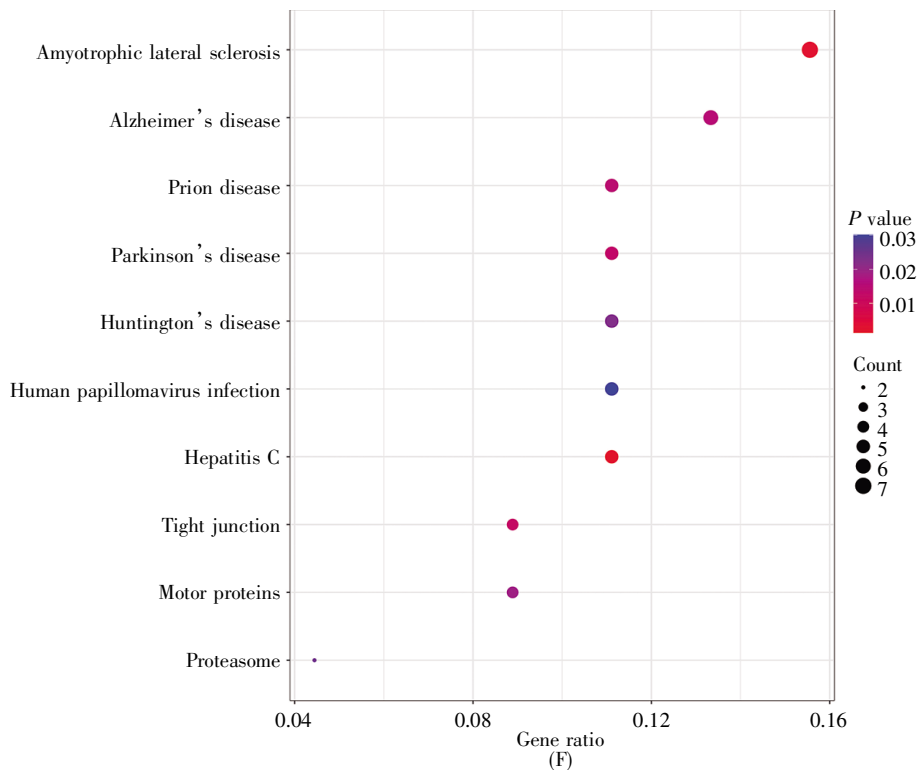


图 3 感染组和对照组 A-to-I 编辑位点的鉴定与分析

(A) 感染组和对照组 12 种碱基替换类型的分布; (B) 感染组和对照组 A-to-I 编辑水平的比较; (C) 两组间 A-to-I 编辑水平差异的火山图; (D) A-to-I 编辑水平差异显著的编辑位点所在基因的 GO 富集分析(前 10 个); (E) 显著富集的病毒感染相关的生物学过程; (F) A-to-I 编辑水平差异显著位点所在基因的 KEGG 富集分析(前 10 个)。

### Fig.3 Identification and analysis of A-to-I editing sites in infection and mock groups

(A) Distribution of the 12 base substitution types in infection and mock groups; (B) Comparison of A-to-I editing level between infection and mock groups; (C) Volcano plot of A-to-I editing level difference; (D) GO enrichment analysis of genes with significant differences in A-to-I editing level at editing sites (top 10); (E) Significantly enriched biological processes related to viral infection; (F) KEGG enrichment analysis of genes with significant differences in A-to-I editing level (top 10).

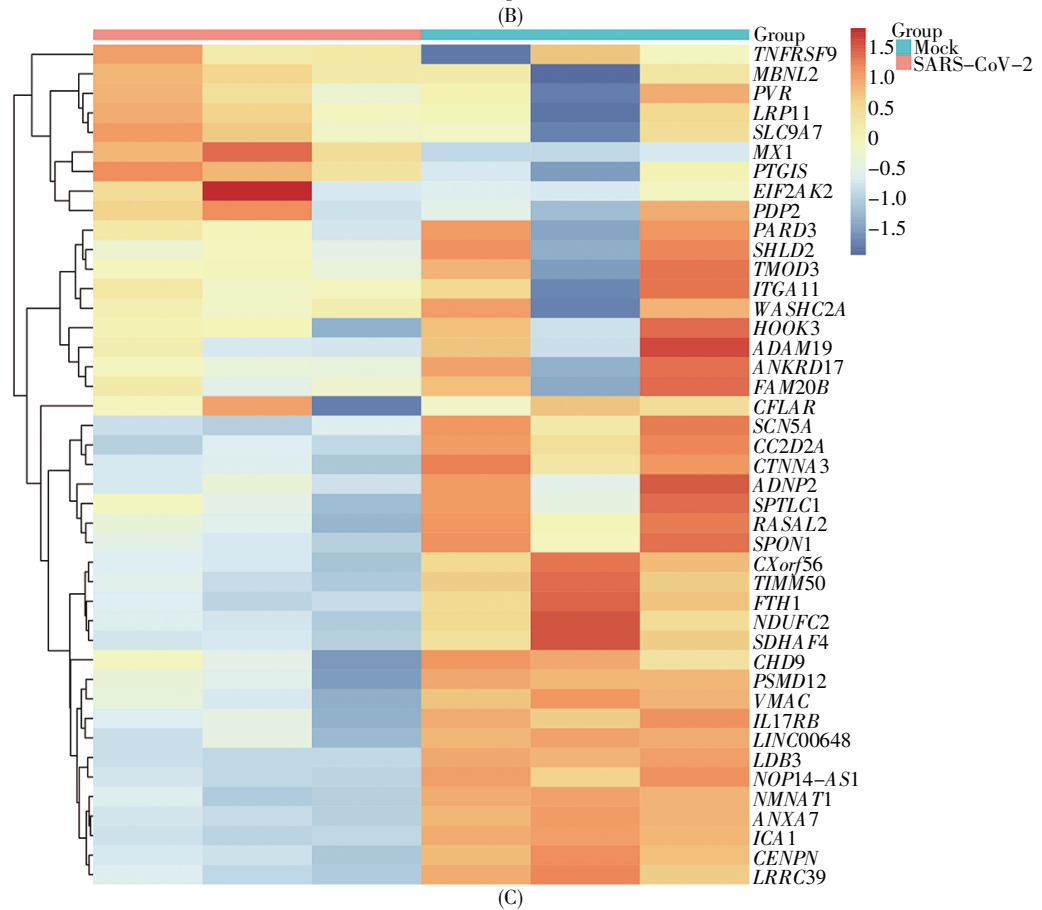
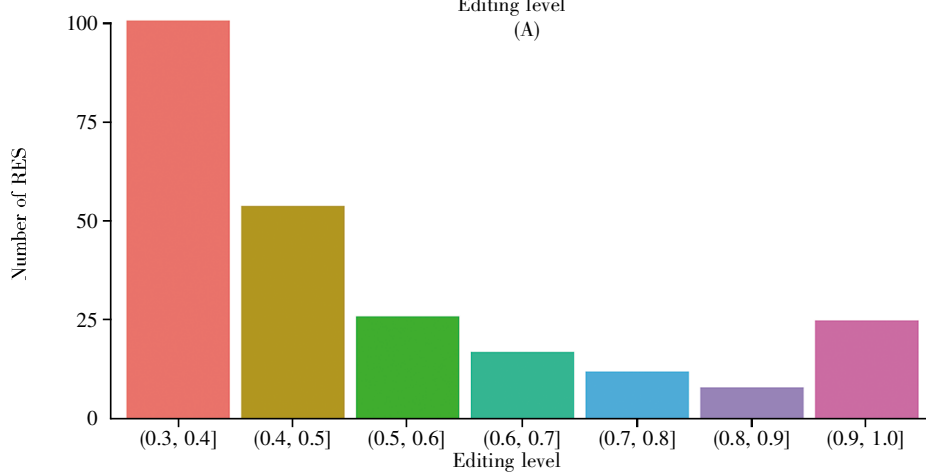
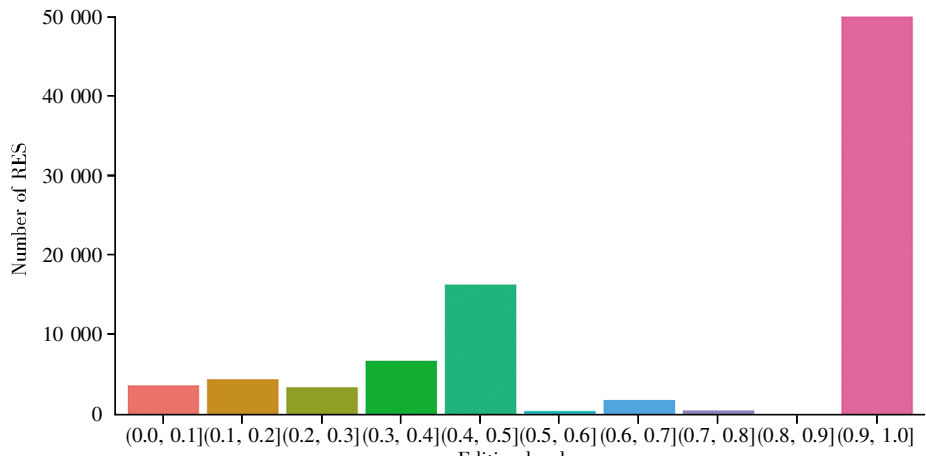
现, 这 11 个基因中的 6 个, 即 *ADAM19*、*FTH1*、*NMNAT1*、*SPON1*、*TNFRSF9* 和 *WASHC2A* (图 4C), 在感染组和对照组间的编辑水平及表达水平均存在显著性差异, 同时都是含有高质量 RES 的基因。在这 6 个基因中, *ADAM19*、*TNFRSF9*、*WASHC2A* 的表达水平在感染组显著上调, 且这三个基因的转录本仅在感染组发生了 A-to-I RNA 编辑事件; *FTH1*、*NMNAT1* 和 *SPON1* 的表达水平在感染组显著下调, 其中, *FTH1* 和 *SPON1* 基因的转录本仅在对照组发生了 A-to-I RNA 编辑事件, *NMNAT1* 基因的转录本则在两种情况下均发生了 A-to-I 编辑, 而且 *NMNAT1* 基因在感染组的编辑水平显著高于对照组(表 1)。

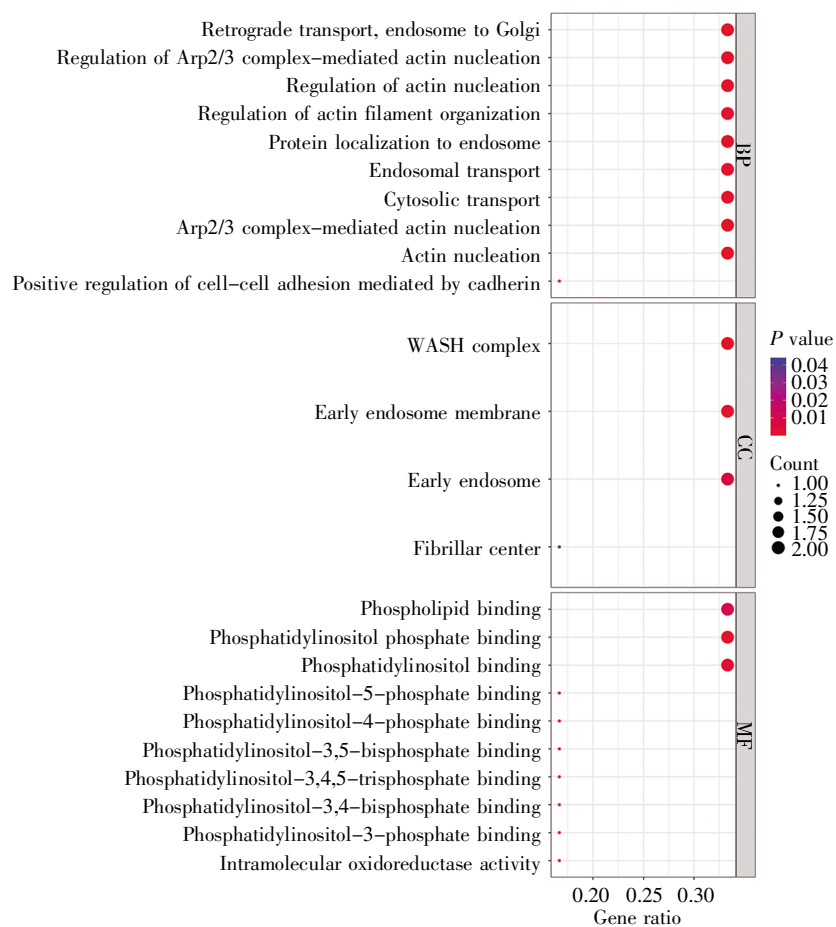
### 3 讨论

新冠疫情流行期间, SARS-CoV-2 病毒感染造成了有些病人除肺以外, 还包括心脏等多处重要器官受损, 而心肌的损伤往往是终身性的。目

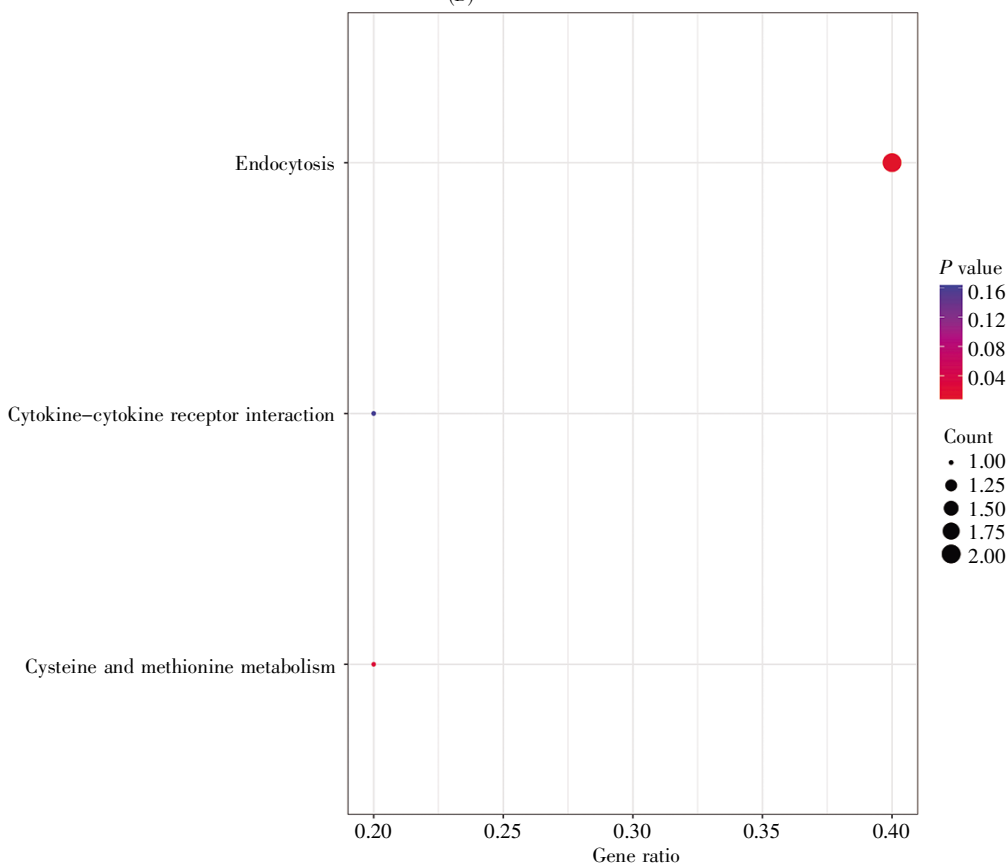
前, 不少研究报道, 宿主细胞的 ADAR 和 APOBEC 脱氨酶对 SARS-CoV-2 病毒基因组 RNA 进行编辑可能造成了病毒核苷酸的多样性<sup>[13, 23-24]</sup>, 并且影响了 SARS-CoV-2 病毒基因组的演化<sup>[25]</sup>。而关于宿主细胞在 SARS-CoV-2 病毒感染后改变自身 RNA 编辑事件以应对病毒感染的研究, 除了新冠病毒感染(COVID-19)患者肺活检样本的 RNA 编辑分析<sup>[26]</sup>外, 还未见其他的报道。为探究 A-to-I RNA 编辑事件在 SARS-CoV-2 病毒感染后的心肌细胞中的变化, 以及其在心肌细胞对抗病毒感染中发挥的潜在作用, 本研究检测了 SARS-CoV-2 感染心肌细胞的 RES, 对 RES 的分布、编辑水平及组间差异等进行了分析。

A-to-I RNA 编辑是人类转录组中的一种普遍现象, 有意思的是本研究发现, 绝大多数 A-to-I 编辑位点之间相距在 100 bp 以内, 成簇分布, 并且它们主要位于内源性 Alu 逆转录元件。究其原因, 主要是 Alu 元件局部形成的 dsRNA 区域是





(D)



(E)

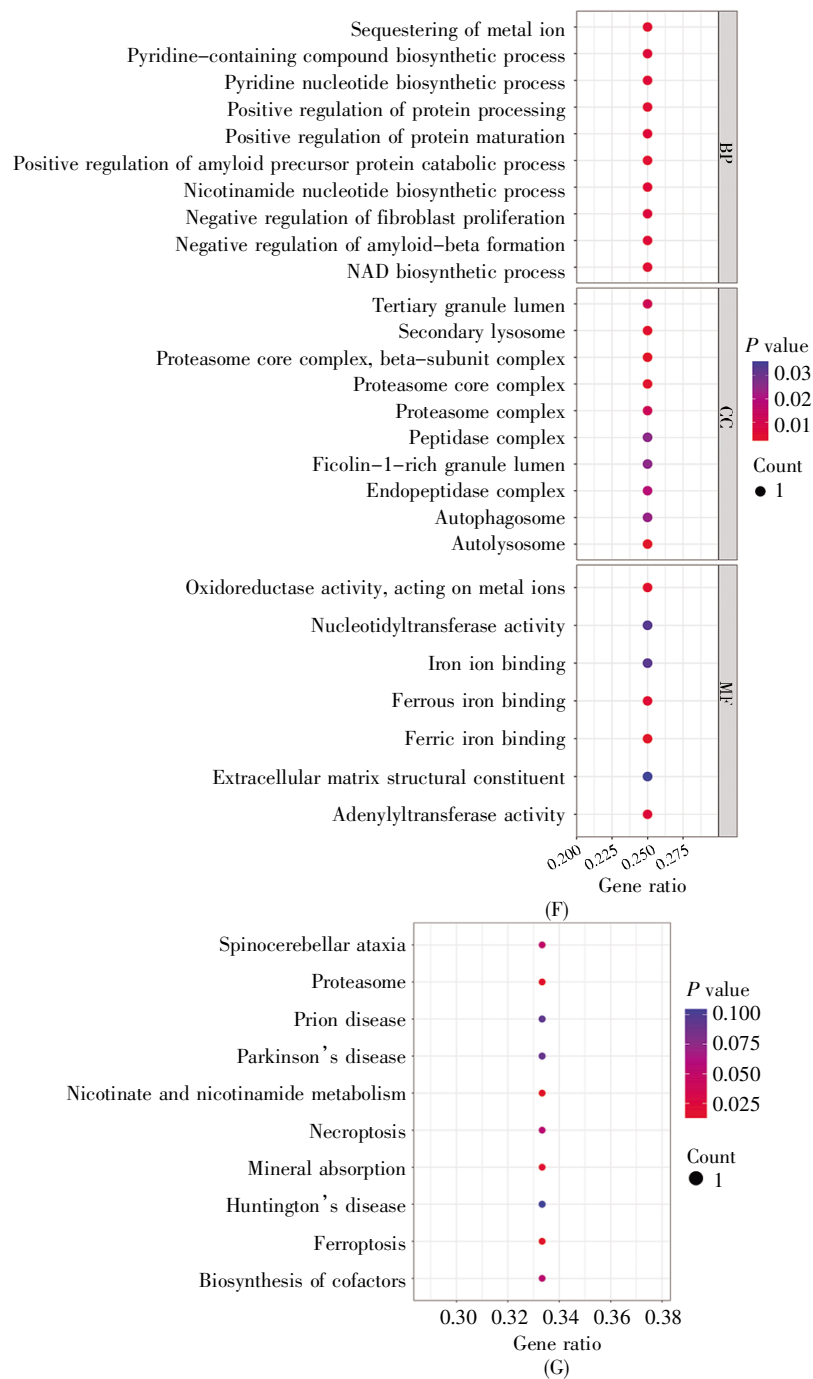


图 4 不同编辑水平的 A-to-I 编辑位点的分析

(A) A-to-I 编辑位点在不同编辑水平的数量分布直方图; (B) 经高质量筛选后 A-to-I 编辑位点在不同编辑水平的数量分布直方图; (C) 非严谨筛选条件下 A-to-I 编辑水平和转录水平均存在显著差异的 43 个基因的表达热图; (D) 经高质量筛选的表达上调基因的 GO 富集分析; (E) 经高质量筛选的表达上调基因的 KEGG 富集分析; (F) 经高质量筛选的表达下调基因的 GO 富集分析; (G) 经高质量筛选的表达下调基因的 KEGG 富集分析。图中富集分析仅显示最显著的前 10 个通路, 显著富集通路未达到 10 个的, 则显示所有显著富集的通路。

**Fig.4 Analysis of A-to-I editing sites with different editing levels**

(A) Histograms of editing levels between different A-to-I editing sites; (B) Histogram of editing levels between A-to-I editing sites after screening; (C) Heatmap of 43 genes with significant difference in transcriptional level and A-to-I editing level under non-rigorous screening conditions; (D) GO enrichment analysis of the screened genes with upregulated expression; (E) KEGG enrichment analysis of the screened genes with upregulated expression; (F) GO enrichment analysis of the screened genes with downregulated expression; (G) KEGG enrichment analysis of the screened genes with downregulated expression. The enrichment analysis in the figure only shows the most significant top 10 pathways, if there are less than 10 significantly enriched pathways, all significantly enriched pathways will be displayed.

表 1 具有高质量编辑位点并且在两组间的编辑水平和表达水平均有显著性差异的基因

Table 1 Genes with high-quality editing sites and significant differences in both editing and expression levels between the two groups

Gene symbol	Editing site	Expression level					
		SARS-CoV-2			Mock		
		1	2	3	1	2	3
<i>ADAM19</i>	chr5: 157 478 552	4 646.41	3 906.10	4 588.69	2 307.44	1 695.41	2 685.21
<i>TNFRSF9</i>	chr1: 7 920 478	654.86	431.86	535.40	57.17	266.19	133.00
<i>WASHC2A</i>	chr10: 50 069 315	6 838.98	6 215.23	8 881.55	5 079.27	1 099.22	4 223.91
<i>FTH1</i>	chr11: 62 009 756	53 892.98	44 027.36	55 617.09	68 097.94	182 839.98	73 844.67
<i>SPON1</i>	chr11: 14 081 004	4 867.15	4 394.67	3 869.36	13 722.77	3 960.85	15 589.28
<i>NMNA1</i>	chr1: 9 982 930	397.59	289.03	359.90	745.42	909.86	651.41

Gene symbol	Editing site	Editing level					
		SARS-CoV-2			Mock		
		1	2	3	1	2	3
<i>ADAM19</i>	chr5: 157 478 552	0.33	0.75	0.46	0	0	0
<i>TNFRSF9</i>	chr1: 7 920 478	0.56	0.60	0.70	0	0	0
<i>WASHC2A</i>	chr10: 50 069 315	0.80	0.70	0.56	0	0	0
<i>FTH1</i>	chr11: 62 009 756	0	0	0	1.00	0.96	1.00
<i>SPON1</i>	chr11: 14 081 004	0	0	0	0.67	0.75	0.72
<i>NMNA1</i>	chr1: 9 982 930	1.00	1.00	0.67	0.40	0.40	0.35

注: 表达水平值为标准化的基因表达值。

Note: Expression level values are normalized gene expression values.

ADAR 结合和催化脱氨基的先决条件<sup>[27]</sup>。从心肌细胞内 RNA 碱基变异的分析结果可知, A-to-I 和 C-to-U 的碱基替换占总体碱基替换数的绝大多数, 即 SARS-CoV-2 病毒感染后心肌细胞内发生的 RNA 编辑事件主要是 A-to-I 类型的 RNA 编辑(图 1A)。Alu 元件是灵长类基因组中普遍存在的短散在重复序列, 可以形成稳定的 dsRNA 结构, 是 ADAR 的主要底物, 蛋白质编码基因由于 Alu 元件的插入以及 RNA 编辑, 可能会产生数百种新的蛋白质<sup>[19, 26, 28-29]</sup>。与这些研究结果类似, 本研究发现 A-to-I RNA 编辑位点也主要位于 Alu 区域, 而且位于 Alu 元件的编辑位点约占总 A-to-I 编辑位点的 90.09%。此外, 与前人的研究结果<sup>[8]</sup>类似, 本研究鉴定的 A-to-I RES 也大多发生在内含子(约占 45.59%)、基因间区(约占 30.90%)和 3'-UTR(约占 12.37%)等区域(图 1D)。这说明本研究鉴定的 A-to-I 编辑位点主要位于基因组的非编码区的 Alu 元件。

RNA 编辑是转录后水平的重要修饰之一, 在不改变基因组序列的情况下, 改变了 RNA 的核苷酸序列。对蛋白质编码序列进行 RNA 编辑很可能导致其功能的重新编码, 从而促进转录组和蛋白质组的多样性<sup>[10, 30]</sup>。mRNA 的非同义替换会改变目标蛋白质的序列, 并且可能影响其结构的稳定性与功能, 这为病毒感染和疾病传播提供了线索<sup>[31]</sup>。本研究对 A-to-I 编辑位点的外显子区域进

行注释后发现, 编辑位点的碱基替换也主要是非同义替换, 但发现的这些非同义替换仅仅位于低编辑水平或含支持测序读段很少的编辑位点, 对细胞的功能影响可能很小, 其重要性需要进一步的实验去验证。

感染组和对照组存在大量的共同编辑事件, 但其中一些编辑位点的组间编辑水平有显著差异。本研究发现, 编辑水平在感染组显著更高的差异编辑位点有 102 个, 显著更低的差异编辑位点为 94 个。这些编辑水平具有显著差异的编辑位点所在的基因主要富集的通路包括与病毒感染相关的通路, 如丙型肝炎病毒和人乳头瘤病毒感染等, 反映心肌细胞通过 RNA 编辑作用改变了自身功能状态, 以应对新冠病毒感染。另外, 本研究发现, 含有高质量 A-to-I RES 且 A-to-I 编辑水平在组间存在显著差异的基因仅有 11 个: *ADAM19*、*FTH1*、*MRI1*、*NMNA1*、*PSMB2*、*SPON1*、*TNFRSF9*、*WASHC2A*、*WASHC2C*、*ZFP69B* 和 *ZNF815P*, 总体分析显示, 它们主要富集在内吞作用、细胞因子-细胞因子受体相互作用、蛋白酶体、烟酸和烟酰胺代谢以及铁死亡等通路(图 4E、图 4G)。有研究表明, RNA 编辑能对基因表达和功能产生影响, 从而直接或间接影响细胞过程和稳态<sup>[32-33]</sup>。在上述 11 个基因中, 有 6 个基因在感染组和对照组间的 A-to-I 编辑水平及表达水平均存在显著差异(表 1)。这说明 SARS-CoV-2 的感染导致了

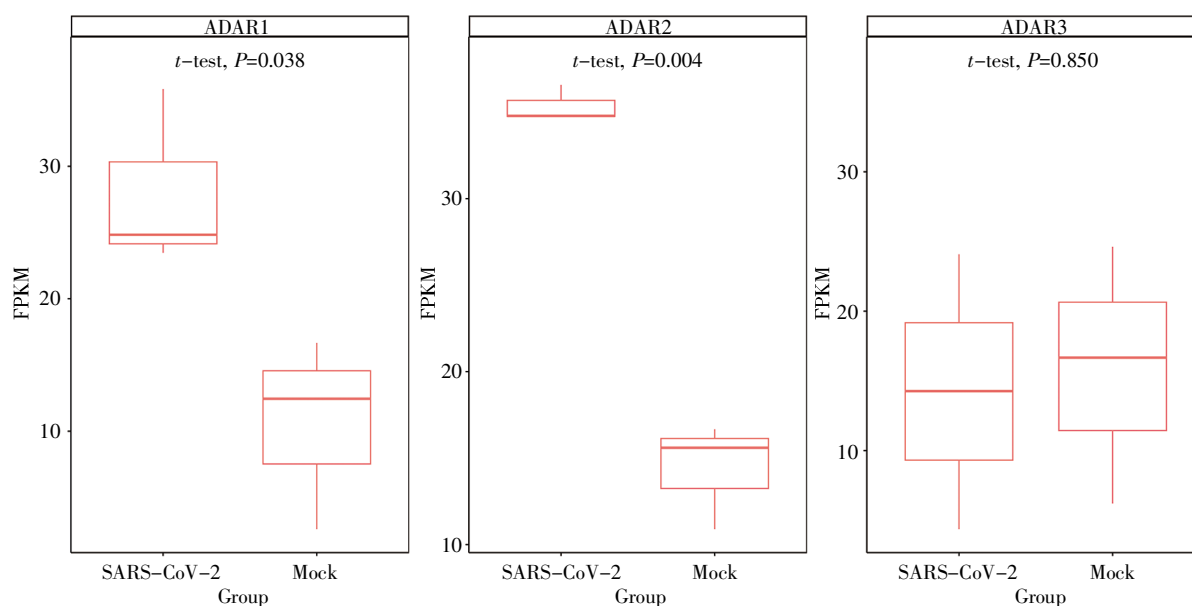


图5 ADAR 家族成员在感染组和对照组的表达水平比较

Fig.5 Comparison of the expression levels of ADAR family members between the infection and mock groups

心肌细胞 RNA 编辑事件的改变,因而改变了细胞内一些基因的转录本,进而可能影响相关基因的表达及其功能。

A-to-I RNA 编辑事件与腺苷脱氨酶(adenosine deaminase, ADA)相关。人体存在 3 种 ADAR 酶,其中, A-to-I 的 RNA 编辑由普遍表达的 ADAR (ADAR1)和 ADAR2 (ADAR2)两种腺苷脱氨酶催化,ADAR1 主要负责重复序列区域的 RNA 编辑,ADAR2 则负责非重复序列编码区的 RNA 编辑,而 ADAR3 (ADAR3)被认为是无催化活性的,并对 RNA 编辑起负调控的作用<sup>[34]</sup>。与对照组相比,感染组心肌细胞中 ADAR1 和 ADAR2 的表达水平显著升高,而 ADAR3 的表达水平没有显著变化(图 5)。因此,ADAR1 和 ADAR2 在感染组心肌细胞中更高的表达水平可能导致了总体 RNA 编辑位点的 A-to-I RNA 编辑水平在感染组偏高(图 3B)。

SARS-CoV-2 感染后的干扰素(interferon, IFN)反应主要受 MDA5 (melanoma differentiation-associated gene 5)调节,ADAR1 的主要生理功能是编辑内源性 dsRNA,以防 MDA5 将内源性 dsRNA 识别为非己 RNA<sup>[35]</sup>,MDA5 则识别病毒等外源的 RNA,从而引发 IFN 反应。有研究表明,在 SARS-CoV-2 感染者中,AG 突变随炎症的增加而增加,SARS-CoV-2 RNA 中的 AG 突变与死亡率呈负相关<sup>[30]</sup>。本文数据显示,ADAR1 和 ADAR2 在 SARS-CoV-2 感染的心肌细胞中表达增加(图 5),并且感染组心肌细胞的 A-to-I 编辑水平高于对照组(图 3B),由

此我们推测,感染组心肌细胞内更高表达水平的 ADAR1 和 ADAR2 介导的 A-to-I 编辑也可能会对 SARS-CoV-2 RNA 基因组的编辑发挥作用,因为已有多篇研究<sup>[13, 25, 30]</sup>报道,宿主细胞 ADAR1 对 SARS-CoV-2 基因组 RNA 进行编辑影响了病毒的适应性和感染性。同时,本文发现 ADAR 酶作用的 A-to-I RES 所处的局部序列模式有一定的保守性(图 2)。

综上所述,本研究显示 SARS-CoV-2 感染改变了心肌细胞内的 RNA 编辑事件。虽然该计算研究存在一些局限性,需要进一步的实验研究来评估 SARS-CoV-2 感染心肌细胞中 RNA 编辑的生物学作用,但对了解心肌细胞通过 RNA 编辑响应 SARS-CoV-2 感染的机制有一定的意义。

#### 参考文献(References):

- [1] GIOVANETTI M, BENEDETTI F, CAMPISI G, *et al.* Evolution patterns of SARS-CoV-2: snapshot on its genome variants[J]. *Biochemical and Biophysical Research Communications*, 2021, 538: 88-91.
- [2] TSCHÖPE C, AMMIRATI E, BOZKURT B, *et al.* Myocarditis and inflammatory cardiomyopathy: current evidence and future directions[J]. *Nature Reviews Cardiology*, 2021, 18(3): 169-193.
- [3] DI GIORGIO S, MARTIGNANO F, TORCIA M G, *et al.* Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2[J]. *Science Advances*, 2020, 6(25): eabb5813.
- [4] MOURIER T, SADYKOV M, CARR M J, *et al.* Host-directed editing of the SARS-CoV-2 genome[J]. *Biochemical and Biophysical Research Communications*, 2021, 538: 35-39.
- [5] KOCKLER Z W, GORDENIN D A. From RNA world to SARS-CoV-2: the edited story of RNA viral evolution[J]. *Cells*, 2021, 10(6): 1557.

- [6] TAJBAKSH A, GHEIBI HAYAT S M, TAGHIZADEH H, *et al.* COVID-19 and cardiac injury: clinical manifestations, biomarkers, mechanisms, diagnosis, treatment, and follow up[J]. *Expert Review of Anti-Infective Therapy*, 2021, 19(3): 345-357.
- [7] BOJKOVA D, WAGNER J U G, SHUMLIAKIVSKA M, *et al.* SARS-CoV-2 infects and induces cytotoxic effects in human cardiomyocytes[J]. *Cardiovascular Research*, 2020, 116(14): 2207-2215.
- [8] HARIHARAN A, SUN S N, WIPPLINGER M, *et al.* RNA editing in mesothelioma: a look forward[J]. *Open Biology*, 2020, 10(10): 200112.
- [9] TEOH P J, KOH M Y, CHNG W J. ADARs, RNA editing and more in hematological malignancies[J]. *Leukemia*, 2021, 35(2): 346-359.
- [10] NISHIKURA K. Functions and regulation of RNA editing by ADAR deaminases[J]. *Annual Review of Biochemistry*, 2010, 79: 321-349.
- [11] EISENBERG E, LEVANON E Y. A-to-I RNA editing-immune protector and transcriptome diversifier[J]. *Nature Reviews Genetics*, 2018, 19(8): 473-490.
- [12] ZINSHTEYN B, NISHIKURA K. Adenosine-to-inosine RNA editing[J]. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, 2009, 1(2): 202-209.
- [13] PENG X X, LUO Y K, LI H Y, *et al.* RNA editing increases the nucleotide diversity of SARS-CoV-2 in human host cells[J]. *PLoS Genetics*, 2022, 18(3): e1010130.
- [14] PICARDI E, D'ERCHIA A M, LO GIUDICE C, *et al.* REDIPortal: a comprehensive database of A-to-I RNA editing events in humans[J]. *Nucleic Acids Research*, 2017, 45(D1): D750-D757.
- [15] WANG K, LI M Y, HAKONARSON H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data[J]. *Nucleic Acids Research*, 2010, 38(16): e164.
- [16] RITCHIE M E, PHIPSON B, WU D, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Research*, 2015, 43(7): e47.
- [17] YU G C, WANG L G, HAN Y Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS*, 2012, 16(5): 284-287.
- [18] GABAY O, SHOSHAN Y, KOPEL E, *et al.* Landscape of adenosine-to-inosine RNA recoding across human tissues[J]. *Nature Communications*, 2022, 13: 1184.
- [19] LEVANON E Y, EISENBERG E. Does RNA editing compensate for Alu invasion of the primate genome?[J]. *BioEssays*, 2015, 37(2): 175-181.
- [20] VAN DORP L, ACMAN M, RICHARD D, *et al.* Emergence of genomic diversity and recurrent mutations in SARS-CoV-2[J]. *Infection, Genetics and Evolution*, 2020, 83: 104351.
- [21] HASSAN S S, CHOUDHURY P P, ROY B. SARS-CoV2 envelope protein: non-synonymous mutations and its consequences[J]. *Genomics*, 2020, 112(6): 3890-3892.
- [22] GUPTA A M, CHAKRABARTI J, MANDAL S. Non-synonymous mutations of SARS-CoV-2 leads epitope loss and segregates its variants[J]. *Microbes and Infection*, 2020, 22(10): 598-607.
- [23] SIMMONDS P. Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories[J]. *mSphere*, 2020, 5(3): e00408-20.
- [24] LIU X Y, LIU X, ZHOU J, *et al.* Rampant C-to-U deamination accounts for the intrinsically high mutation rate in SARS-CoV-2 spike gene[J]. *RNA*, 2022, 28(7): 917-926.
- [25] SONG Y L, HE X J, YANG W B, *et al.* Virus-specific editing identification approach reveals the landscape of A-to-I editing and its impacts on SARS-CoV-2 characteristics and evolution[J]. *Nucleic Acids Research*, 2022, 50(5): 2509-2521.
- [26] CROOKE P S, TOSSBERG J T, PORTER K P, *et al.* Reduced A-to-I editing of endogenous Alu RNAs in lung after SARS-CoV-2 infection[J]. *Current Research in Immunology*, 2021, 2: 52-59.
- [27] VLACHOGIANNIS N I, VERROU K M, STELLOS K, *et al.* The role of A-to-I RNA editing in infections by RNA viruses: possible implications for SARS-CoV-2 infection[J]. *Clinical Immunology*, 2021, 226: 108699.
- [28] AUNE T M, TOSSBERG J T, HEINRICH R M, *et al.* Alu RNA structural features modulate immune cell activation and A-to-I editing of Alu RNAs is diminished in human inflammatory bowel disease[J]. *Frontiers in Immunology*, 2022, 13: 818023.
- [29] LI C H, SHI X R, YANG J Y, *et al.* Genome-wide characterization of RNA editing highlights roles of high editing events of glutamatergic synapse during mouse retinal development[J]. *Computational and Structural Biotechnology Journal*, 2022, 20: 2648-2656.
- [30] RINGLANDER J, FINGAL J, KANN H, *et al.* Impact of ADAR-induced editing of minor viral RNA populations on replication and transmission of SARS-CoV-2[J]. *Proceedings of the National Academy of Sciences of USA*, 2022, 119(6): e2112663119.
- [31] ALJINDAN R Y, AL-SUBAIE A M, AL-OHALI A I, *et al.* Investigation of nonsynonymous mutations in the spike protein of SARS-CoV-2 and its interaction with the ACE2 receptor by molecular docking and MM/GBSA approach[J]. *Computers in Biology and Medicine*, 2021, 135: 104654.
- [32] NISHIKURA K. A-to-I editing of coding and non-coding RNAs by ADARs[J]. *Nature Reviews. Molecular Cell Biology*, 2016, 17(2): 83-96.
- [33] CORREIA DE SOUSA M, GJORGJIEVA M, DOLICKA D, *et al.* Deciphering miRNAs' action through miRNA editing[J]. *International Journal of Molecular Sciences*, 2019, 20(24): 6249.
- [34] TAN M H, LI Q, SHANMUGAM R, *et al.* Dynamic landscape and regulation of RNA editing in mammals[J]. *Nature*, 2017, 550(7675): 249-254.
- [35] LIDDICOAT B J, PISKOL R, CHALK A M, *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself[J]. *Science*, 2015, 349(6252): 1115-1120.