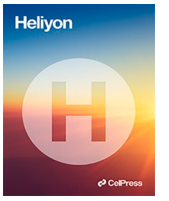
Heliyon 9 (2023) e18133

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Profiling A-to-I RNA editing during mouse somatic reprogramming at the single-cell level

Tianhang Lv [a,b,1], Siyuan Jiang [a,b,1], Xiaoshan Wang [b], Yong Hou [a,b,*]

[a] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China
[b] BGI-Shenzhen, Shenzhen, 518083, China

ARTICLE INFO

Keywords:
Single-cell RNA editing
Mouse somatic reprogramming
ADAR1

ABSTRACT

Mouse somatic cells can be reprogrammed into induced pluripotent stem cells through a highly heterogeneous process regulated by numerous biological factors, including adenosine-to-inosine (A-to-I) RNA editing. In this study, we analyzed A-to-I RNA editing sites using a single-cell RNA sequencing (scRNA-seq) dataset with high-depth and full-length coverage. Our method revealed that A-to-I RNA editing frequency varied widely at the single-cell level and underwent dynamic changes. We also found that A-to-I RNA editing level was correlated with the expression of the RNA editing enzyme ADAR1. The analysis combined with gene ontology (GO) enrichment revealed that ADAR1-dependent A-to-I editing may downregulate the expression levels of *Igtp*, *Irgm2*, *Mndal*, *Ifi202b,* and *Tapbp* in the early stage, to inhibit the pathways of cellular response to interferon-beta and regulation of protein complex stability to promote mesenchymal-epithelial transition (MET). Notably, we identified a negative correlation between A-to-I RNA editing frequency and the expression of certain genes, such as *Nras*, *Ube2l6*, *Zfp987*, and *Adsl*.

1. Introduction

Somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) by ectopically expressing four transcription factors (Oct4, Sox2, Klf4, and c-Myc) [1]. iPSCs are highly attractive for research and regenerative medicine due to their ability to differentiate into all adult cell types. There are various modulations to the reprogramming system at the transcriptional and epigenetic levels, such as key transcription factors [2], the transforming growth factor-β signaling pathway [3], and miRNAs [4]. Recent studies have also shown that RNA modification play a role in this system [5–9].

RNA modification is a process where a specific nucleotide in the RNA sequence is chemically modified. To date, over 100 types of cellular RNA modifications have been identified in both coding and non-coding RNA [10]. One of the most prevalent post-transcriptional RNA modifications on eukaryotic mRNA is m^6A modification, which is the methylation of adenosine bases at the nitrogen-6 position [11]. The presence of m^6A on transcripts contributes to diverse cellular functions, including pre-mRNA splicing, nuclear stability, and microRNA biogenesis [12–14]. miRNA can regulate the binding of METTL3, a methyltransferase that catalyzes the formation of m^6A, to target RNA through sequence pairing with mRNA that contain miRNA targeting sites, resulting in an increase in m^6A modification [15]. Increased m^6A modification promotes the reprogramming of mouse embryonic fibroblasts (MEFs) to iPSCs [15].

* Corresponding author. College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China.
E-mail address: houyong@mgi-tech.com (Y. Hou).
[1] Tianhang Lv and Siyuan Jiang contributed equally to this paper.

https://doi.org/10.1016/j.heliyon.2023.e18133
Received 8 May 2022; Received in revised form 5 July 2023; Accepted 7 July 2023
Available online 13 July 2023

Apart from m^6A modification, A-to-I RNA editing (also known as "A-to-I editing") is one of the most abundant RNA modifications in mammalian cells, which is related to multiple biological processes, including embryonic development [16], neuron generation [17], innate immunity [18], and somatic cell reprogramming [9]. A-to-I RNA editing refers to the deamination of adenine (A) to inosine (I), which is catalyzed by members of the adenosine deaminase acting on RNA (ADAR) family [19]. Inosine can be recognized as guanosine (G) when the edited base is transcribed [19]. A-to-I editing can change the sequence, coding potential, and secondary structure of RNA [19]. Knockdown of ADAR1 (a member of the ADAR family) in human embryonic stem cells results in an elevation of the expression of differentiation-related genes [16]. ADAR1 level changes in human foreskin fibroblast cells before induction of reprogramming result in varied reprogramming efficiencies [20]. In mouse reprogramming system, the absence of ADAR1-dependent RNA editing induces aberrant innate immune responses, unleashes endoplasmic reticulum (ER) stress, and impedes epithelial fate acquisition [9].

Although RNA modification affecting the acquisition of pluripotency has been illustrated, these studies have been conducted at the bulk level. Initially, cell reprogramming was thought to be a sequential intracellular process in an orderly manner. Studies found that the pluripotent gene expression was gradually increased, while the MEF-related gene expression was gradually decreased with the pursuit of reprogramming procedures [21,22]. In the mouse reprogramming system, the specific events that occur are as follows: (a) activation of alkaline phosphatase, (b) silencing of somatic-specific expression, (c) expression of SSEA1, and (d) gradual silencing of exogenous genes with concurrent upregulation of endogenous Nanog and Oct4 [23]. However, it was found that, in such a continuous process, only a small proportion of MEFs could be reprogrammed into iPSCs [24]. This low reprogramming efficiency implies that there are underlying mechanisms of reprogramming process in individual cells, which leads to different fates chosen by cells [25]. Therefore, the analysis results based on averaging the whole cell population tend to mask the unusual biological events of individual cells and the rare cell fate transformation during the reprogramming process [26]. Thus, studies at single-cell resolution can better reveal the true mechanism of inefficiency, randomness, and heterogeneity of the reprogramming system.

Through the use of scRNA-seq, recent studies have shown that the regulation of somatic and pluripotent genes can be independently triggered in individual cells [24–28], leading to significant variation in cell fates and reprogramming efficiency within single cells. Guo et al. [26] used single-cell transcriptome maps to reconstruct the reprogramming developmental trajectory of MEFs being reprogrammed into iPSCs and found that cells bifurcate into two categories: cells with reprogramming potential (RP) or non-reprogramming (NR). Francesconi et al. [24] observed that Myc activity affected transdifferentiation and reprogramming at the single-cell level, and cells with low Myc activity could efficiently transdifferentiate into macrophages but could not be reprogrammed into iPSCs. These studies revealed a wider range of developmental programs than previously characterized and confirmed key transcriptional factors that enhance reprogramming efficiency [25,27,28]. However, the role of A-to-I editing, an important RNA modification, in the somatic reprogramming system in single cells has not yet been reported. Does A-to-I editing modulate the expression of pluripotent transcripts in single cells? In this study, we aimed to investigate the role of A-to-I editing in the somatic reprogramming system at the single-cell level.

Accurate identification of A-to-I editing sites requires appropriate and effective analysis methods. Currently, most RNA editing detection methods are developed based on bulk RNA-seq data, such as AEI [29], REDITools [30], and GIREM [31]. However, compared to bulk RNA-seq data, scRNA-seq data has lower sequencing depth, lower sequence coverage, and more sequencing errors, which can introduce false positives and reduce the accuracy of A-to-I editing detection [32]. Therefore, methods and models developing based on bulk RNA-seq data could not be applied to scRNA-seq data straightly [33]. Key factors that affect the accuracy of A-to-I editing detection include sequencing quality, sequencing depth, single nucleotide polymorphism (SNP), and base position on reads [34]. Taking these data features and limitations of existing methods, our group developed a method for A-to-I RNA editing detection in single cells, using a strict filtering strategy and several statistical tests to remove potential false positives [35] (see Materials and Methods). We successfully applied this method to early human embryogenesis research and demonstrated dynamic changes in genome-wide RNA editing in a stage-specific fashion [35]. Additionally, Gabay et al. [36] developed a method for RNA editing site detection and applied this method to a single-cell dataset of mouse organs. However, this method only detects A-to-I editing sites located in the coding sequence (CDS) region which could not obtain editing sites coming from gene regulatory regions. In this study, we used both our developed method and Gabay et al.'s method to detect RNA editing sites. The application of two effective methods will help us identify authentic editing sites while avoiding potential false positive sites.

As previously mentioned, the relatively lower quality and coverage of scRNA-seq data make high-depth and full-length datasets crucial for detecting RNA editing sites at the single-cell level. Full-length scRNA-seq methods have incomparable advantages in RNA editing identification compared to 3′-end and 5′-end counting protocols, owing to their superior transcript coverage. In this study, we selected the GSE103221 [26], a dataset which was generated using the Smart-seq2 protocol and includes 912 isolated single cells sampled on day 0 (D0), 1 (D1), 2 (D2), 3 (D3), 5 (D5), 7 (D7), and 8 (D8), as well as MEFs, iPSCs, and embryonic stem cells (ESCs) from a continuous period generated by the iCD1-OSK (OCT4, SOX2, KLF4) system. To avoid potential false positives in the single-cell dataset detection, we also selected two bulk RNA-seq datasets (GSE93029 [37] and GSE106334 [38]) of the mouse somatic reprogramming system to perform a comparative analysis with the single-cell dataset, helping us identify authentic editing sites.

Through our comprehensive analysis, we investigated the temporal dynamics of genome-wide A-to-I RNA editing during mouse somatic cell reprogramming at a single-cell level. Our study sheds light on the intricate mechanisms of somatic reprogramming and provides valuable insights into this complex process.

2. Materials and Methods

2.1. Read alignment

We downloaded GSE103221 [26], a single-cell RNA-seq dataset of which cells sampled from MEFs to ESCs. The libraries with higher depth and coverage were constructed using the Smart-seq2 method [39]. To ensure greater accuracy in detecting editing sites, we also downloaded bulk datasets (GSE93029 and GSE106334) of this reprogramming system as a reference for our single-cell research. The mm10 (GRCh37) genome sequences were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu) and we used Tophat2 (version 2.1.0) [40] with default parameters to align the reads of bulk and single-cell RNA-seq to the genome.

2.2. RNA editing detection by our published method

We utilized our published method [35] to detect candidate RNA editing sites. Specifically, we collated the base calls of preprocessed and aligned RNA reads to the mouse reference in a pileup format. Bases with a base quality score of at least 20 were retained for analysis. The specific process is as follows:

1) We performed statistical tests based on binomial distribution B (n, p) to distinguish true variants from sequencing errors on every mismatch site, where p denotes the background mismatch rate of each transcriptome, and n denotes sequencing depth on this site. On a given specific site with k reads supporting variant in all n mapped reads, we used B (k, n, p) to calculate the probability that the k mismatches are all due to sequencing errors. This probability is adjusted using the Benjamini-Hochberg method. We only retained candidate sites with adjusted P-value < 0.01. In addition, the candidate variant sites should be with mapped reads ≥ 4, variant-supporting reads ≥ 3, and mismatch frequencies (variant-supporting-reads/mapped-reads) ≥ 0.1.
2) We filtered out variants with strand bias which may introduce false positives, referring to abnormal distribution of sense strand variant supporting reads and antisense strand variant supporting reads in sequencing data. We estimated strand bias and filtered out variants with strand bias as follows. (a) We performed a two-tailed Fisher's exact test (FET) using the following two-by-two table:

Sense strand variant supporting reads	Sense strand reference supporting reads
Antisense strand variant supporting reads	Antisense strand reference supporting reads

(b) We estimated variant strand frequency (sense-strand variant-supporting reads divided by total variant-supporting reads), variant strand preference [absolute (variant strand frequency minus 0.5)], reference strand frequency (sense-strand reference minus supporting read number divided by total reference-supporting read number), and reference strand preference [absolute (reference strand frequency minus 0.5)]. (c) We filtered out variant sites displaying significant strand bias, defined as either FET P-value < 0.005 plus variant strand preference $>$ reference strand preference, or variant strand frequency > 0.9, or variant strand frequency < 0.1.

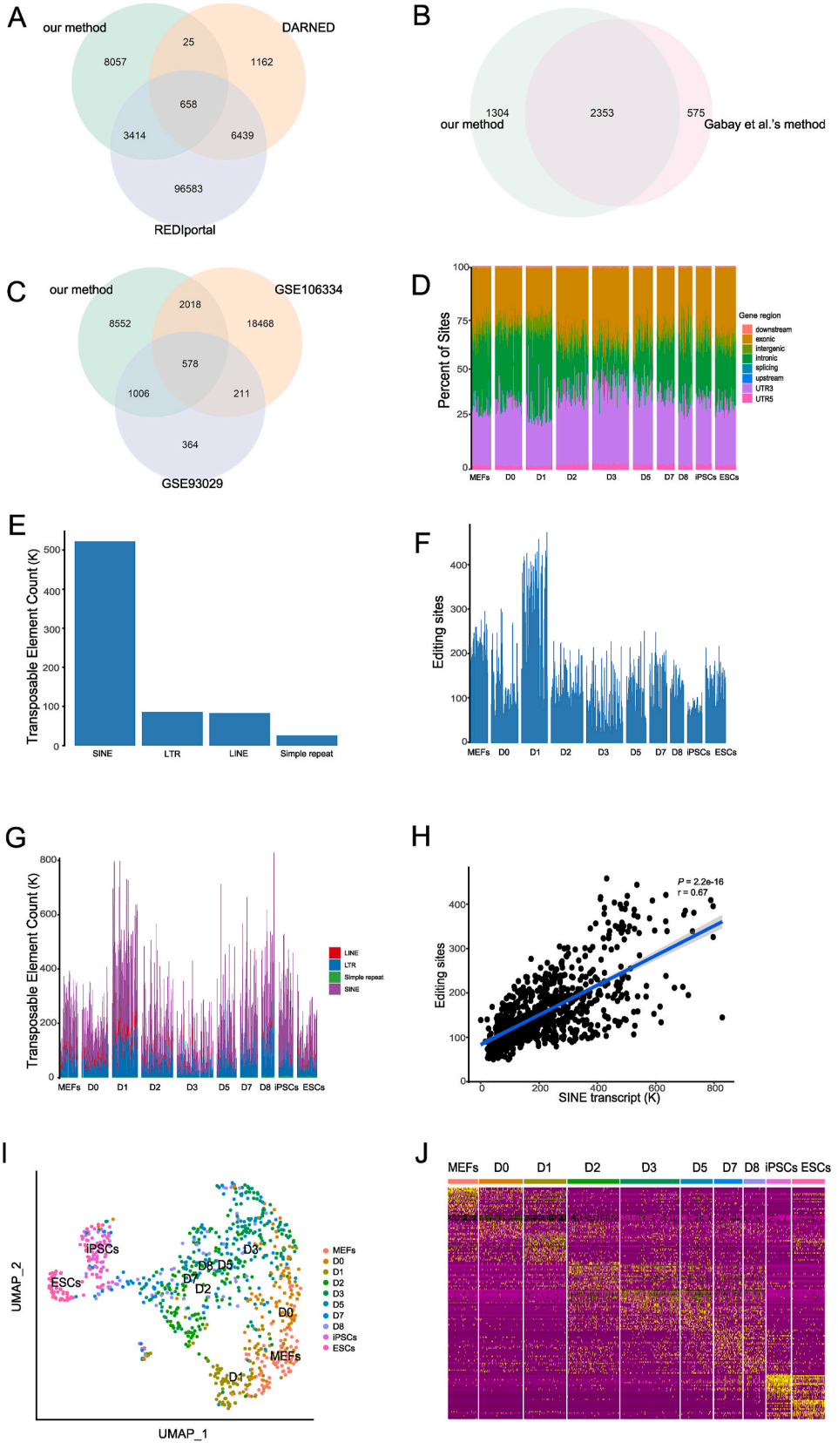
3) Because the mismatches at the ends of sequencing reads are less credible than the ones in the middle of reads, we filtered out variants with position bias. Read end is defined as 10 bp at 3′-end or 5 bp at 5′-end. Position biases are defined as either FET P-value <0.05 plus read end frequency $>$ read middle frequency, or read end frequency > 0.9.
4) We removed variant sites that are either in simple repeat regions (http://hgdownload.cse.ucsc.edu/goldenpath/mm10/database/) or in homopolymer regions (runs of ≥ 5 bp).
5) We filtered out candidate variants if they are found in mouse DNA dbsnp (V142).
6) We used ANNOVAR for functional annotation of the RNA editing sites.

2.3. RNA editing detection by Gabay et al.'s method

Due to the current lack of research on identifying A-to-I editing from single-cell RNA-seq data, we employed Gabay et al.'s method [36] with default parameters to detect RNA editing sites in CDS regions.

2.4. Single-cell clustering and visualization

After obtaining the candidate A-to-I editing sites, we generated a matrix in which each column represents a cell, each row represents an A-to-I RNA editing site, and each value is the A-to-I editing frequency. Then, we filtered out cells with less than 200 editing sites and editing sites with less than 5 cells. We then used the Seurat (version 4.1.1) [41] to process data by normalizing, scaling, and dimension reduction with the parameters default. The data points were visualized in UMAP (uniform manifold approximation and projection) two-dimensional (2D) space.



(caption on next page)

Fig. 1. Characteristics of A-to-I RNA editing sites during mouse somatic cell reprogramming. (A) A Venn diagram showing the A-to-I editing sites detected by our method and databases. (B) A Venn diagram showing the A-to-I editing sites located in CDS regions detected by our method and Gabay et al.'s method. (C) A Venn diagram showing the A-to-I editing sites detected in the single-cell dataset and two bulk datasets. (D) The proportion of gene regions in which A-to-I editing sites were located, with each column representing one cell. (E) The distribution of the number of transposable elements. (F) The distribution of RNA editing sites, with each column representing one cell. (G) The distribution of transposon elements, with each column representing one cell and colors representing transposon element types. (H) The correlation between the number of SINE transcripts and editing sites. (I) The cell cluster display in UMAP 2D space colored by time point using the A-to-I RNA editing frequency matrix. (J) A heatmap of marker A-to-I RNA editing sites for each stage.

2.5. Detection of transposable elements

We used TEtranscripts [42] to detect the expression level of transposable elements for each cell with the parameters defaulted.

2.6. GO analysis

We identified genes that were significantly edited in each stage and then conducted the GO analysis for these genes using Metascape [43].

3. Results

3.1. Characteristics of A-to-I RNA editing during the mouse somatic reprogramming process

Using our published method [35], we identified 1,145,189 candidate A-to-I editing sites across the 912 cells. After filtering out cells

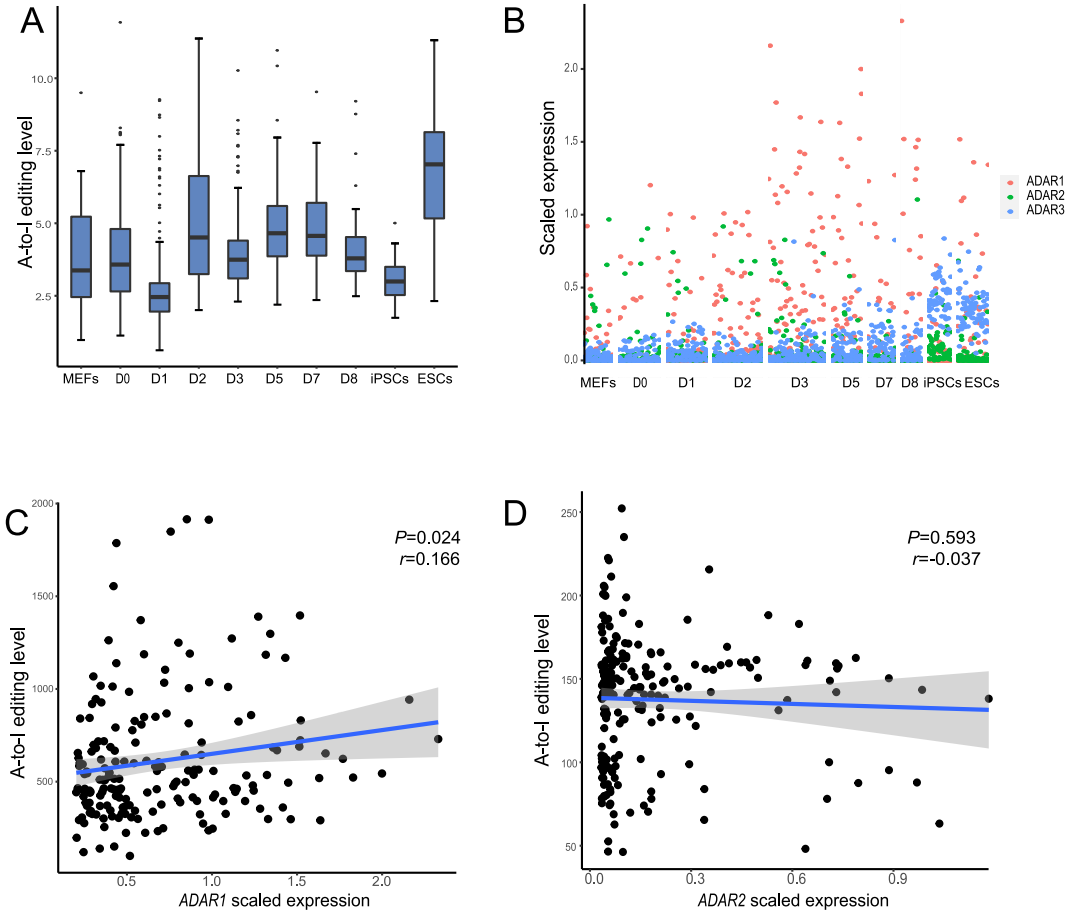
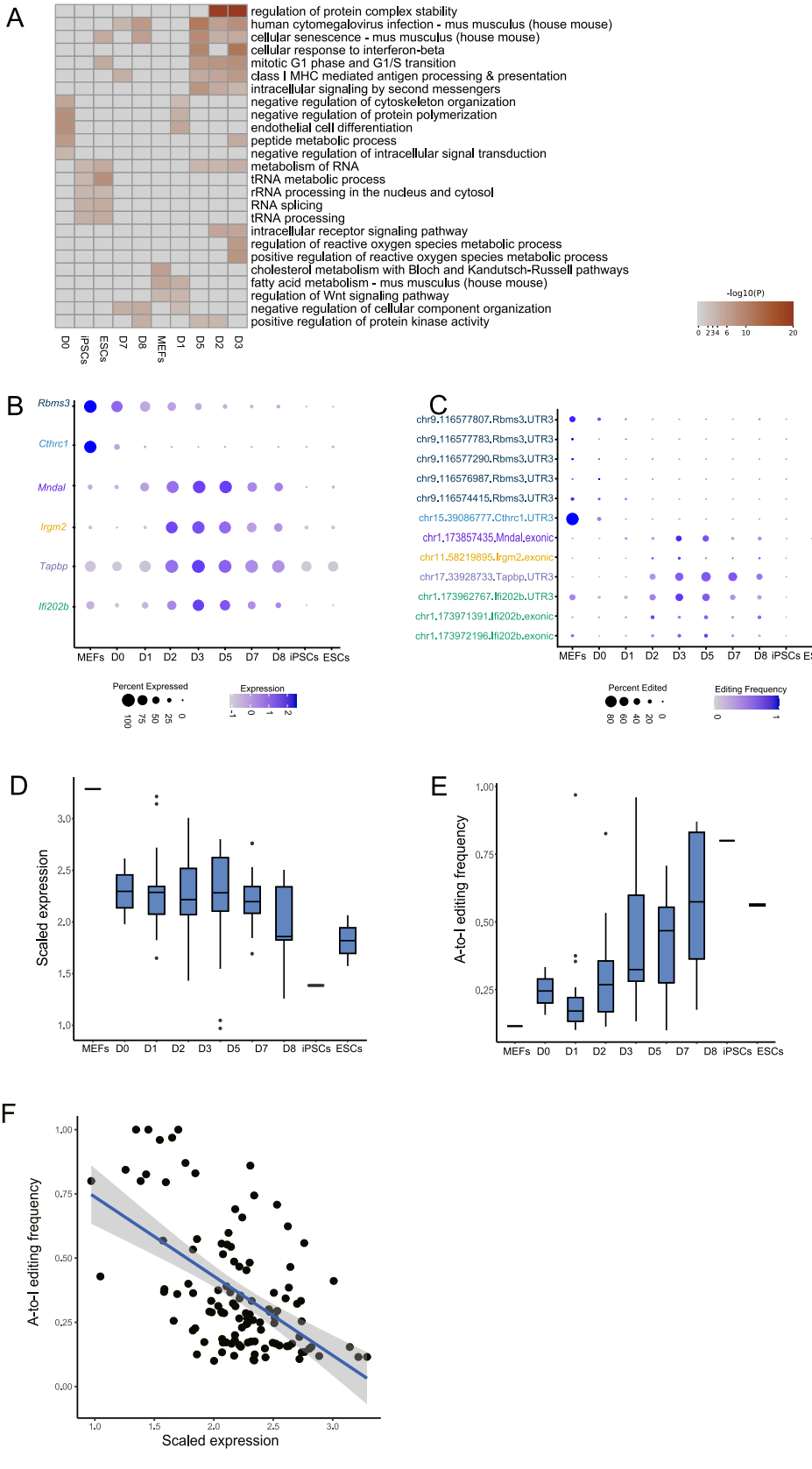


Fig. 2. ADARs govern changes in A-to-I RNA editing levels during mouse somatic cell reprogramming. (A) Changes in RNA editing levels during mouse somatic cell reprogramming. (B) Changes in gene expression of *ADAR1*, *ADAR2*, and *ADAR3* during mouse somatic cell reprogramming. (C) Correlation between editing levels and expression levels of *ADAR1*. (D) Correlation between editing levels and expression levels of *ADAR2*.



(caption on next page)

Fig. 3. GO enrichment of edited genes and correlation of gene expression with A-to-I editing frequency. (A) A heatmap of GO enrichment generated using marker-edited genes in each stage. (B) A dot plot of A-to-I editing sites, with color representing the editing frequency and the size of circle representing the ratio of cells that were positive for these sites. (C) Gene expression corresponding to editing sites in (B), with color representing the expression level and size of circle representing the ratio of cells which were positive for genes. (D) Changes in *Nras* expression in each stage. (E) Changes in the frequency of chr3.103066726.Nras.UTR3. (F) Correlation between *Nras* expression and chr3.103066726.Nras.UTR3 editing frequency.

with less than 200 editing sites and editing sites with less than 5 cells, we identified 12,153 A-to-I editing sites (Tables S1–2). Of these, 658 sites were present in the DARNED [44] database, and 4,072 sites were present in the REDIportal [45] database (Fig. 1A, Table S3). The mean number of RNA editing sites of single cells was 209 (Fig. S1). To validate the editing sites identified by our published method, we used Gabay et al.'s method [36] to detect A-to-I editing sites located in the CDS region. We then applied the same filtering parameters to the result detected by Gabay et al.'s method. We found that 2,353 sites (64.3% of our method detecting) were detected by the two methods, indicating the robustness of these sites (Fig. 1B).

However, we also acknowledged that A-to-I editing analysis in scRNA-seq data may yield false positives due to artifacts in bioinformatic analysis. To address this issue, we compared our used single-cell data to two previously published bulk RNA-seq datasets (GSE93029 and GSE106334), which can provide more depth and coverage to reduce false positives of candidate editing sites (Fig. 1C). There are 3,602 sites (29.7%) both existed in single-cell data and the bulk data. Considering the bias of sequencing and batch effect of samples, this shared ratio was comparatively high, which also showed that the A-to-I editing sites detected by our method are relatively reliable and robust.

We conducted an annotation of all A-to-I editing sites and found that the majority of them were located in the intronic region (31.7%), followed by the 3′ untranslated region (UTR3) (30.3%), exonic region (28.38%), and 5′ untranslated region (UTR5) (1.98%) (Fig. 1D). This distribution is consistent with the findings of a study by Gu et al. [46] in mice. As dsRNA produced from transposable elements are important ADAR substrates [47], we investigated whether A-to-I editing sites were enriched in these regions. Among the 12,153 sites, 6,909 of them were in the repeat region. Among the sites in the repeat region, 26% of them were in the short interspersed nuclear element (SINE) region, 4.14% were in the long interspersed nuclear element (LTR) region, 4.27% were in the long terminal repeat (LINE) region, and 1.28% were in the simple repeat region (Fig. 1E). This relative proportion of diverse transposable elements is also characterized by other studies in mice [48,49]. Given that transposable elements are more likely to form double-stranded RNA as substrates for ADARs [47], we quantified the expression level of transposable elements (Fig. 1E). It was obvious that the number of A-to-I editing sites sharply increased at D1 (Fig. 1F), reaching up to 299 ± 80 per cell (Table S4). Interestingly, the number of A-to-I editing sites was strongly correlated with the SINE expression level (Fig. 1G–H). The high expression level of SINE on D1 explained the spike of A-to-I editing sites on D1.

To assess changes in A-to-I RNA editing during the reprogramming process, we used the filtered 891-cell dataset to create a matrix in which each column represented a cell, each row represented an A-to-I editing site, and each value represented the A-to-I editing frequency. We then employed Seurat to cluster the cells in this matrix. Following data normalization, scaling, principal component analysis reduction, and neighbor finding, we clustered all data points and visualized them with time points labeled in the UMAP 2D space (Fig. 1I). The RNA editing cluster pattern closely resembled the single-cell gene expression cluster pattern (Fig. S3), suggesting that MEFs, D0, D1, iPSCs, and ESCs were distinctly separated from one another, while D2, D3, D5, D7, and D8 showed mixed patterns. Thus, RNA editing might be linked to gene expression to some extent. Additionally, we plotted a heatmap of the marker RNA editing sites at each time point (Fig. 1J) to identify specific marker RNA editing sites for each stage, revealing that A-to-I RNA editing was extremely heterogeneous and dynamic during reprogramming.

3.2. The A-to-I RNA editing level is positively related to ADAR1 expression

In order to profile dynamic A-to-I RNA editing patterns at a genome-wide level, we defined the A-to-I RNA editing level in each cell as the number of edited bases per million mapped bases, divided by the product of the base depth and normalized gene expression. This approach was employed to prevent bias due to the number of mapped bases (Fig. S2) and gene expression. We profiled the dynamic changes of A-to-I RNA editing level in each stage with a box plot (Fig. 2A). The RNA editing level in D1 was the lowest compared to other time points. However, the editing level significantly increased in the D2 stage, slightly decreased in the D3 stage, and increased again in the D5 stage, remaining relatively stable until D7. Subsequently, in the D8 and iPSCs stages, the editing level continuously decreased, while it reached the highest value during the ESCs stage (Fig. 2A). To investigate the cause of these changes, we profiled the expression of the ADAR family using the single-cell expression matrix (Fig. 2B). The pattern of *ADAR1* expression matched the change in the A-to-I RNA editing level (Fig. 2A–B), with expression at a low level on D1 and a higher level on D2, D3, D5, and D7. Pearson correlation analysis showed that the *ADAR1* expression level was significantly correlated with the A-to-I RNA editing level ($r = 0.166$, $P = 0.0024$) (Fig. 2C), which agrees with previous results obtained in other mouse tissues [8]. Pearson correlation analysis also showed that the *ADAR2* expression level was not correlated with the editing level (Fig. 2D). In addition, *ADAR3* was expressed at a lower level compared to *ADAR1* and *ADAR2* (Fig. 2B), which was also previously observed during mouse forebrain development [8].

3.3. GO enrichment analysis of marker-edited genes

We identified 749 marker RNA editing sites (log fold-change > 0.2) from MEFs to ESCs (Table S5). Among these sites, we identified

corresponding marker-edited genes in MEFs, D0, D1, D2, D3, D5, D7, D8, iPSCs, and ESCs, with 89, 37, 132, 60, 44, 36, 57, 48, 83, and 110 genes, respectively. We used Metascape, an online GO enrichment tool, to enrich GO terms (Fig. 3A, Table S6). It is essential to note that these GO terms were affected by A-to-I editing, which can have either a positive or negative effect on these pathways.

ADAR1-dependent A-to-I editing promotes MET by inhibiting MDA5 (*Ifih1* gene)-mediated innate immune response and ER stress [9]. ER stress, which is positively correlated with the accumulation of unfolded proteins, can impede cell fate decisions [48,49]. Guallar et al. [9] found that genes in the cellular response to the interferon-beta pathway were upregulated in ADAR1-absent mutants in the early stage of the reprogramming system. Similarly, Guo et al. [50] found that ADAR1 with the D1113H mutation in the catalytic domain in mice dramatically increased the expression level of *Igtp* (a gene involved in response to the interferon-beta pathway). Furthermore, Mannion et al. [51] found that *Igtp* and *Irgm2* are upregulated at least 3-fold in mouse embryos whose ADAR1 mutations cause Aicardi-Goutières syndrome. It is apparent that higher ADAR1 activity plays a key role in maintaining the expression level of genes involved in the pathways of innate immune response and ER stress in the early stage of the reprogramming system. *Igtp, Irgm2, Tapbp, Mndal,* and *Ifi202b* involved in the pathways of cellular response to interferon-beta and regulation of protein complex stability (Table S6), were highly expressed from D2 to D5 and then expressed less in the later stage (Fig. 3B). We speculate that ADAR1 may affect their expression by editing the bases in UTR3, UTR5, and intron regions, resulting in the inhibition of pathways of cellular response to interferon-beta and regulation of protein complex stability, promoting MET from D2 to D5 (Fig. 3C).

Meanwhile, we also observed that *Cthrc1* and *Rbms3*, the negative regulatory genes of the Wnt signaling pathway, were significantly upregulated in D0. And the bases in UTR3 of *Cthrc1* and *Rbms3* were highly edited in the early reprogramming stage (Fig. 3B–C). Notably, Wnt signaling was closed in the early reprogramming stage, in contrast to its high activity in the later stages [52]. Our findings suggested that A-to-I RNA editing may play a role in modulating the Wnt signaling by targeting *Cthrc1* and *Rbms3* during reprogramming.

3.4. A-to-I editing frequency is negatively correlated with corresponding gene expression level

We observed that marker-edited genes in D2, D3, D5, D8, and ESCs were enriched in the cellular senescence pathway. *Nras*, a classical gene involved in the cellular senescence pathway, was highly expressed in the early stage and lowly expressed in the later stage (Fig. 3D). Nras interacts preferentially with RAF (a Ras effect factor), to activate the MAPK pathway in the presence of bFGF (basic fibroblast growth factor), which makes human iPSCs (hiPSCs) in a pluripotent state [53]. The activity of the MAPK pathway is necessary for maintaining the pluripotency of hiPSCs, and Nras links bFGF signaling to the MAPK pathway. Additionally, Nras can promote the expression of Oct4 and Sox2 by activating ERK signaling pathway [53]. We found that the expression level of *Nras* (Fig. 3D), was contrasted with the editing frequency of chr3.103066726.Nras.UTR3 (indicating a site at position 103,066,726 of chromosome 3 located in the UTR3 of *Nras*) (Spearman's correlation test, $r = -0.485$, $P = 3.85e\text{-}08$; Fig. 3E–F). Combined with previous study [53], we guess this negative correlation showed that higher A-to-I editing may lead to lower expression level of *Nras*, and then inhibit cellular senescence pathway to promote iPSC generation.

Meanwhile, this negative correlation was also observed for *Adsl* (Spearman's correlation test, $r = -0.472$, $P = 1.208e\text{-}06$; Fig. S4A), *Cyp4x1* (Spearman's correlation test, $r = -0.367$, $P = 2.2e\text{-}04$; Fig. S4B), and *Ube2l6* (Spearman's correlation test, $r = -0.7407$, $P = 1.7e\text{-}12$; Fig. S4C). However, what we had to emphasize was this negative correlation was only found in a small part of all captured genes.

4. Discussion

Until now, A-to-I RNA editing profiling under single-cell resolution in the mouse reprogramming system is still unsolved. Given that cell fates are triggered independently in single cells, we used a single-cell dataset with high depth and high coverage to uncover the A-to-I editing sites in individual cells in a stage fashion. The intersection of the sites located in CDS regions detected by our published method and Gabay et al.'s method showed 2,353 shared sites indicating that these sites had high robustness. Of these sites, 4,730 sites also existed in databases (DARNED and REDIPortal). Moreover, considering the detection of editing sites using single-cell data could yield artificial irrelevant edited transcripts due to artifacts upon bioinformatic analysis, we also selected two bulk data to compare the editing sites generated by single-cell and bulk data. All of these intersection characteristics confirmed a higher reliability of the sites detected by our published method. However, we have to recognize that until now there is a lack of methods for detecting RNA editing sites at single-cell level. Thus, we tried to utilize diverse methods and datasets to make the detecting sites more reliable and avoid potential negative false in this study.

Given that repetitive elements, such as SINE, LINE, LTR, and simple repeat, are more likely to be substrates of the ADAR family, we profiled the expression of these elements and found that SINE expression was strongly correlated with the number of editing sites. Moreover, the expression level of *ADAR1* was highly correlated with the A-to-I editing level. These findings were also proved by other studies in mice, and the same principle was followed in the mouse somatic reprogramming system.

We enriched the GO terms by using the marker-edited genes, showing that multiple edited genes were enriched on classical pathways which were suggested to affect the reprogramming system, such as regulation of Wnt signaling pathway, cellular response to interferon-beta and regulation of protein complex stability from D2 to D5. We guess that A-to-I editing may participate reprogramming system by affecting these pathways in the early stage of the reprogramming system.

Interestingly, *Nras* (a gene in the cellular senescence pathway) were edited, and *Nras* expression was negatively related to the editing frequency of chr3.103066726.Nras.UTR3. microRNAs can regulate mRNA expression by translational inhibition and mRNA destabilization. ADAR1 edits microRNA to redirect the actions of microRNAs to modulate gene (microRNA targets) expression [54].

Additionally, HuR and ADAR could form a HuR-ADAR complex to bind to transcripts and regulate their expression cooperatively [55]. We found that the expression levels of *Nras*, *Adsl*, *Cyp4x1,* and *Ube2l6* were negatively correlated with the editing frequencies of the editing sites located in these genes. However, we only observed a small fraction of detected genes with expression levels that were negatively related to the editing frequencies. The exact mechanism underlying this phenomenon is not known and requires rigorous experimental verification.

Declarations

Author contribution statement

Tianhang Lv: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Analyzed and interpreted the data; Wrote the paper.

Siyuan Jiang: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Xiaoshan Wang: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Yong Hou: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Dr Yong Hou was supported by the Natural Science Foundation of Guangdong Province (2018A030313379).

Data availability statement

Data associated with this study has been deposited at GEO database under the accession number GSE103221, GSE93029 and GSE106334.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18133.

References

[1] M. Stadtfeld, K. Hochedlinger, Induced pluripotency: history, mechanisms, and applications, Genes Dev. 24 (2010) 2239–2263, https://doi.org/10.1101/gad.1963910.

[2] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, Cell 126 (2006) 663–676, https://doi.org/10.1016/j.cell.2006.07.024.

[3] R. Li, J. Liang, S. Ni, T. Zhou, X. Qing, H. Li, W. He, J. Chen, F. Li, Q. Zhuang, B. Qin, J. Xu, W. Li, J. Yang, Y. Gan, D. Qin, S. Feng, H. Song, D. Yang, B. Zhang, L. Zeng, L. Lai, M.A. Esteban, D. Pei, A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts, Cell Stem Cell 7 (2010) 51–63, https://doi.org/10.1016/j.stem.2010.04.014.

[4] F. Anokye-Danso, C.M. Trivedi, D. Juhr, M. Gupta, Z. Cui, Y. Tian, Y. Zhang, W. Yang, P.J. Gruber, J.A. Epstein, E.E. Morrisey, Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency, Cell Stem Cell 8 (2011) 376–388, https://doi.org/10.1016/j.stem.2011.03.001.

[5] P.J. Batista, B. Molinie, J. Wang, K. Qu, J. Zhang, L. Li, D.M. Bouley, E. Lujan, B. Haddad, K. Daneshvar, A.C. Carter, R.A. Flynn, C. Zhou, K.-S. Lim, P. Dedon, M. Wernig, A.C. Mullen, Y. Xing, C.C. Giallourakis, H.Y. Chang, m6A RNA modification controls cell fate transition in mammalian embryonic stem cells, Cell Stem Cell 15 (2014) 707–719, https://doi.org/10.1016/j.stem.2014.09.019.

[6] J. Liu, M. Gao, S. Xu, Y. Chen, K. Wu, H. Liu, J. Wang, X. Yang, J. Wang, W. Liu, X. Bao, J. Chen, YTHDF2/3 are required for somatic reprogramming through different RNA deadenylation pathways, Cell Rep. 32 (2020), 108120, https://doi.org/10.1016/j.celrep.2020.108120.

[7] M.A. Zipeto, A.C. Court, A. Sadarangani, N.P. Delos Santos, L. Balaian, H.-J. Chun, G. Pineda, S.R. Morris, C.N. Mason, I. Geron, C. Barrett, D.J. Goff, R. Wall, M. Pellecchia, M. Minden, K.A. Frazer, M.A. Marra, L.A. Crews, Q. Jiang, C.H.M. Jamieson, ADAR1 activation drives Leukemia stem cell self-renewal by impairing let-7 biogenesis, Cell Stem Cell 19 (2016) 177–191, https://doi.org/10.1016/j.stem.2016.05.004.

[8] M.M. Jacobs, R.L. Fogg, R.B. Emeson, G.D. Stanwood, ADAR1 and ADAR2 expression and editing activity during forebrain development, Dev. Neurosci. 31 (2009) 223–237, https://doi.org/10.1159/000210185.

[9] D. Guallar, A. Fuentes-Iglesias, Y. Souto, C. Ameneiro, O. Freire-Agulleiro, J.A. Pardavila, A. Escudero, V. Garcia-Outeiral, T. Moreira, C. Saenz, H. Xiong, D. Liu, S. Xiao, Y. Hou, K. Wu, D. Torrecilla, J.C. Hartner, M.G. Blanco, L.J. Lee, M. López, C.R. Walkley, J. Wang, M. Fidalgo, ADAR1-Dependent RNA editing promotes MET and iPSC reprogramming by alleviating ER stress, Cell Stem Cell 27 (2020) 300–314.e11, https://doi.org/10.1016/j.stem.2020.04.016.

[10] M. Schaefer, U. Kapoor, M.F. Jantsch, Understanding RNA modifications: the promises and technological bottlenecks of the 'epitranscriptome', Open Biol. 7 (2017), 170077 https://doi.org/10.1098/rsob.170077.

[11] X. Jiang, B. Liu, Z. Nie, L. Duan, Q. Xiong, Z. Jin, C. Yang, Y. Chen, The role of m6A modification in the biological functions and diseases, Signal Transduct. Targeted Ther. 6 (2021) 74, https://doi.org/10.1038/s41392-020-00450-x.

[12] J.-Y. Roignant, M. Soller, m6A in mRNA: an ancient mechanism for fine-tuning gene expression, Trends Genet. 33 (2017) 380–390, https://doi.org/10.1016/j.tig.2017.04.003.

[13] A. Shafik, U. Schumann, M. Evers, T. Sibbritt, T. Preiss, The emerging epitranscriptomics of long noncoding RNAs, Biochim. Biophys. Acta (BBA) - Gene Regul. Mech. 1859 (2016) 59–70, https://doi.org/10.1016/j.bbagrm.2015.10.019.

[14] Y. Yang, X. Fan, M. Mao, X. Song, P. Wu, Y. Zhang, Y. Jin, Y. Yang, L.-L. Chen, Y. Wang, C.C. Wong, X. Xiao, Z. Wang, Extensive translation of circular RNAs driven by N6-methyladenosine, Cell Res. 27 (2017) 626–641, https://doi.org/10.1038/cr.2017.31.

[15] T. Chen, Y.-J. Hao, Y. Zhang, M.-M. Li, M. Wang, W. Han, Y. Wu, Y. Lv, J. Hao, L. Wang, A. Li, Y. Yang, K.-X. Jin, X. Zhao, Y. Li, X.-L. Ping, W.-Y. Lai, L.-G. Wu, G. Jiang, H.-L. Wang, L. Sang, X.-J. Wang, Y.-G. Yang, Q. Zhou, m6A RNA methylation is regulated by MicroRNAs and promotes reprogramming to pluripotency, Cell Stem Cell 16 (2015) 289–301, https://doi.org/10.1016/j.stem.2015.01.016.

[16] R. Shtrichman, I. Germanguz, R. Mandel, A. Ziskind, I. Nahor, M. Safran, S. Osenberg, O. Sherf, G. Rechavi, J. Itskovitz-Eldor, Altered A-to-I RNA editing in human embryogenesis, PLoS One 7 (2012), e41576, https://doi.org/10.1371/journal.pone.0041576.

[17] I.H. Greger, L. Khatri, X. Kong, E.B. Ziff, AMPA receptor tetramerization is mediated by Q/R editing, Neuron 40 (2003) 763–774, https://doi.org/10.1016/S0896-6273(03)00668-8.

[18] M.M. Lamers, B.G. van den Hoogen, B.L. Haagmans, ADAR1: "Editor-in-Chief" of cytoplasmic innate immunity, Front. Immunol. 10 (2019) 1763, https://doi.org/10.3389/fimmu.2019.01763.

[19] C.E. Samuel, Adenosine deaminase acting on RNA (ADAR1), a suppressor of double-stranded RNA–triggered innate immune responses, J. Biol. Chem. 294 (2019) 1710–1720, https://doi.org/10.1074/jbc.TM118.004166.

[20] I. Germanguz, R. Shtrichman, S. Osenberg, A. Ziskind, A. Novak, H. Domev, I. Laevsky, J. Jacob-Hirsch, Y. Feiler, G. Rechavi, J. Itskovitz-Eldor, ADAR1 is involved in the regulation of reprogramming human fibroblasts to induced pluripotent stem cells, Stem Cell. Dev. 23 (2014) 443–456, https://doi.org/10.1089/scd.2013.0206.

[21] B.A. Schwarz, M. Cetinbas, K. Clement, R.M. Walsh, S. Cheloufi, H. Gu, J. Langkabel, A. Kamiya, H. Schorle, A. Meissner, R.I. Sadreyev, K. Hochedlinger, Prospective isolation of poised iPSC intermediates reveals principles of cellular reprogramming, Cell Stem Cell 23 (2018) 289–305.e5, https://doi.org/10.1016/j.stem.2018.06.013.

[22] E.R. Zunder, E. Lujan, Y. Goltsev, M. Wernig, G.P. Nolan, A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry, Cell Stem Cell 16 (2015) 323–337, https://doi.org/10.1016/j.stem.2015.01.015.

[23] R. Teshigawara, J. Cho, M. Kameda, T. Tada, Mechanism of human somatic reprogramming to iPS cell, Lab. Invest. 97 (2017) 1152–1157, https://doi.org/10.1038/labinvest.2017.56.

[24] M. Francesconi, B. Di Stefano, C. Berenguer, L. de Andrés-Aguayo, M. Plana-Carmona, M. Mendez-Lago, A. Guillaumet-Adkins, G. Rodriguez-Esteban, M. Gut, I. G. Gut, H. Heyn, B. Lehner, T. Graf, Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming, Elife 8 (2019), e41627, https://doi.org/10.7554/eLife.41627.

[25] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, E.S. Lander, Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming, Cell 176 (2019) 928–943.e22, https://doi.org/10.1016/j.cell.2019.01.006.

[26] L. Guo, L. Lin, X. Wang, M. Gao, S. Cao, Y. Mai, F. Wu, J. Kuang, H. Liu, J. Yang, S. Chu, H. Song, D. Li, Y. Liu, K. Wu, J. Liu, J. Wang, G. Pan, A.P. Hutchins, J. Liu, D. Pei, J. Chen, Resolving cell fate decisions during somatic cell reprogramming by single-cell RNA-seq, Mol. Cell 73 (2019) 815–829.e7, https://doi.org/10.1016/j.molcel.2019.01.042.

[27] T. Zhao, Y. Fu, J. Zhu, Y. Liu, Q. Zhang, Z. Yi, S. Chen, Z. Jiao, X. Xu, J. Xu, S. Duo, Y. Bai, C. Tang, C. Li, H. Deng, Single-cell RNA-seq reveals dynamic early embryonic-like programs during chemical reprogramming, Cell Stem Cell 23 (2018) 31–45.e7, https://doi.org/10.1016/j.stem.2018.05.025.

[28] K.A. Tran, S.J. Pietrzak, N.Z. Zaidan, A.F. Siahpirani, S.G. McCalla, A.S. Zhou, G. Iyer, S. Roy, R. Sridharan, Defining reprogramming checkpoints from single-cell analyses of induced pluripotency, Cell Rep. 27 (2019) 1726–1741.e5, https://doi.org/10.1016/j.celrep.2019.04.056.

[29] S.H. Roth, E.Y. Levanon, E. Eisenberg, Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity, Nat. Methods 16 (2019) 1131–1138, https://doi.org/10.1038/s41592-019-0610-9.

[30] E. Picardi, G. Pesole, REDItools: high-throughput RNA editing detection made easy, Bioinformatics 29 (2013) 1813–1814, https://doi.org/10.1093/bioinformatics/btt287.

[31] Q. Zhang, X. Xiao, Genome sequence–independent identification of RNA editing sites, Nat. Methods 12 (2015) 347–350, https://doi.org/10.1038/nmeth.3314.

[32] Y. Hu, K. Wang, M. Li, Detecting differential alternative splicing events in scRNA-seq with or without Unique Molecular Identifiers, PLoS Comput. Biol. 16 (2020), e1007925, https://doi.org/10.1371/journal.pcbi.1007925.

[33] G. Chen, B. Ning, T. Shi, Single-cell RNA-seq technologies and related computational data analysis, Front. Genet. 10 (2019) 317, https://doi.org/10.3389/fgene.2019.00317.

[34] Y. Pinto, E.Y. Levanon, Computational approaches for detection and quantification of A-to-I RNA-editing, Methods 156 (2019) 25–31, https://doi.org/10.1016/j.ymeth.2018.11.011.

[35] S. Qiu, W. Li, H. Xiong, D. Liu, Y. Bai, K. Wu, X. Zhang, H. Yang, K. Ma, Y. Hou, B. Li, Single-cell RNA sequencing reveals dynamic changes in A-to-I RNA editome during early human embryogenesis, BMC Genom. 17 (2016) 766, https://doi.org/10.1186/s12864-016-3115-2.

[36] O. Gabay, Y. Shoshan, E. Kopel, U. Ben-Zvi, T.D. Mann, N. Bressler, R. Cohen-Fultheim, A.A. Schaffer, S.H. Roth, Z. Tzur, E.Y. Levanon, E. Eisenberg, Landscape of adenosine-to-inosine RNA recoding across human tissues, Nat. Commun. 13 (2022) 1184, https://doi.org/10.1038/s41467-022-28841-4.

[37] D. Li, J. Liu, X. Yang, C. Zhou, J. Guo, C. Wu, Y. Qin, L. Guo, J. He, S. Yu, H. Liu, X. Wang, F. Wu, J. Kuang, A.P. Hutchins, J. Chen, D. Pei, Chromatin accessibility dynamics during iPSC reprogramming, Cell Stem Cell 21 (2017) 819–833.e6, https://doi.org/10.1016/j.stem.2017.10.012.

[38] B. Amlani, Y. Liu, T. Chen, L.-S. Ee, P. Lopez, A. Heguy, E. Apostolou, S.Y. Kim, M. Stadtfeld, Nascent induced pluripotent stem cells efficiently generate entirely iPSC-derived mice while expressing differentiation-associated genes, Cell Rep. 22 (2018) 876–884, https://doi.org/10.1016/j.celrep.2017.12.098.

[39] S. Picelli, O.R. Faridani, Å.K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2, Nat. Protoc. 9 (2014) 171–181, https://doi.org/10.1038/nprot.2014.006.

[40] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol. 14 (2013) R36, https://doi.org/10.1186/gb-2013-14-4-r36.

[41] Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L.M. Fleming, B. Yeung, A.J. Rogers, J.M. McElrath, C.A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data, Cell 184 (2021) 3573–3587.e29, https://doi.org/10.1016/j.cell.2021.04.048.

[42] Y. Jin, O.H. Tam, E. Paniagua, M. Hammell, TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets, Bioinformatics 31 (2015) 3593–3599, https://doi.org/10.1093/bioinformatics/btv422.

[43] Y. Zhou, B. Zhou, L. Pache, M. Chang, A.H. Khodabakhshi, O. Tanaseichuk, C. Benner, S.K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 10 (2019) 1523, https://doi.org/10.1038/s41467-019-09234-6.

[44] A. Kiran, P.V. Baranov, DARNED: a database of RNA editing in humans, Bioinformatics 26 (2010) 1772–1776, https://doi.org/10.1093/bioinformatics/btq285.

[45] E. Picardi, A.M. D'Erchia, C. Lo Giudice, G. Pesole, REDIportal: a comprehensive database of A-to-I RNA editing events in humans, Nucleic Acids Res. 45 (2017) D750–D757, https://doi.org/10.1093/nar/gkw767.

[46] T. Gu, F.W. Buaas, A.K. Simons, C.L. Ackert-Bicknell, R.E. Braun, M.A. Hibbs, Canonical A-to-I and C-to-U RNA editing is enriched at 3′UTRs and microRNA target sites in multiple mouse tissues, PLoS One 7 (2012), e33720, https://doi.org/10.1371/journal.pone.0033720.

[47] K. Nishikura, A-to-I editing of coding and non-coding RNAs by ADARs, Nat. Rev. Mol. Cell Biol. 17 (2016) 83–96, https://doi.org/10.1038/nrm.2015.4.

[48] M. Corazzari, M. Gagliardi, G.M. Fimia, M. Piacentini, Endoplasmic reticulum stress, unfolded protein response, and cancer cell fate, Front. Oncol. 7 (2017) 78. https://doi.org/10.3389/fonc.2017.00078.

[49] C. Hetz, The unfolded protein response: controlling cell fate decisions under ER stress and beyond, Nat. Rev. Mol. Cell Biol. 13 (2012) 89–102. https://doi.org/10.1038/nrm3270.

[50] X. Guo, R.A. Steinman, Y. Sheng, G. Cao, C.A. Wiley, Q. Wang, An AGS-associated mutation in ADAR1 catalytic domain results in early-onset and MDA5-dependent encephalopathy with IFN pathway activation in the brain, J. Neuroinflammation 19 (2022) 285, https://doi.org/10.1186/s12974-022-02646-0.
[51] N.M. Mannion, S.M. Greenwood, R. Young, S. Cox, J. Brindle, D. Read, C. Nellåker, C. Vesely, C.P. Ponting, P.J. McLaughlin, M.F. Jantsch, J. Dorin, I.R. Adams, A.D.J. Scadden, M. Öhman, L.P. Keegan, M.A. O'Connell, The RNA-editing enzyme ADAR1 controls innate immune responses to RNA, Cell Rep. 9 (2014) 1482–1494, https://doi.org/10.1016/j.celrep.2014.10.041.
[52] F. Aulicino, I. Theka, L. Ombrato, F. Lluis, M.P. Cosma, Temporal perturbation of the Wnt signaling pathway in the control of cell reprogramming is modulated by TCF1, Stem Cell Rep. 2 (2014) 707–720, https://doi.org/10.1016/j.stemcr.2014.04.001.
[53] F. Haghighi, J. Dahlmann, S. Nakhaei-Rad, A. Lang, I. Kutschka, M. Zenker, G. Kensah, R.P. Piekorz, M.R. Ahmadian, bFGF-mediated pluripotency maintenance in human induced pluripotent stem cells is associated with NRAS-MAPK signaling, Cell Commun. Signal. 16 (2018) 96, https://doi.org/10.1186/s12964-018-0307-1.
[54] D.J. Luciano, H. Mirsky, N.J. Vendetti, S. Maas, RNA editing of a miRNA precursor, RNA 10 (2004) 1174–1177, https://doi.org/10.1261/rna.7350304.
[55] I.X. Wang, E. So, J.L. Devlin, Y. Zhao, M. Wu, V.G. Cheung, ADAR regulates RNA editing, transcript stability, and gene expression, Cell Rep. 5 (2013) 849–860, https://doi.org/10.1016/j.celrep.2013.10.002.