Large-scale identification of recurrent RNA edits in human embryos suggests their role in enhancing maternal mRNA clearance

Yang Ding
Beijing Institute of Radiation Medicine

Yang Zheng
Beijing Institute of Radiation Medicine

Junting Wang
Beijing Institute of Radiation Medicine

Hao Li
Beijing Institute of Radiation Medicine

Chenghui Zhao
Medical Innovation Research Division of Chinese PLA General Hospital

Huan Tao
Beijing Institute of Radiation Medicine

Yaru Li
Beijing Institute of Radiation Medicine

Kang Xu
Beijing Institute of Radiation Medicine

Xin Huang
Beijing Institute of Radiation Medicine

Ge Gao
Biomedical Pioneering Innovation Center (BIOPIC), Beijing Advanced Innovation Center for Genomics (ICG), Peking University

Xiaochen Bo
Beijing Institute of Radiation Medicine

Hebing Chen (chb-1012@163.com)
Beijing Institute of Radiation Medicine

Genetics Analysis

Keywords: A-to-I RNA editing, early embryonic development, human embryo, maternal mRNA clearance, microRNA
Abstract

Posttranscriptional modification plays an important role in key embryonic processes. Adenosine-to-inosine RNA editing, a common example of such modifications, is widespread in human adult tissues and has various functional impacts and clinical consequences. However, whether it persists in a consistent pattern in most human embryos, and whether it supports embryonic development, are poorly understood. To address this problem, we compiled the largest human embryonic editome from 2,071 transcriptomes and identified for the first time thousands of Recurrent Edits (REs; >=50% chances of occurring in a given stage) for each early developmental stages. We found that these REs prefer exons consistently across stages, tend to target genes related to DNA replication, and undergo organized loss in abnormal embryos and embryos from elder mothers. In particular, REs are likely to enhance maternal mRNA clearance by introducing more microRNA binding sites to the 3'-untranslated regions of clearance targets. This study suggests a potentially important, if not indispensable, role of RNA editing in key human embryonic processes such as maternal mRNA clearance; the identified editome can aid further investigations.

Introduction

The successful development of human embryos is based on a well-regulated network that spans multiple omic layers [1], among which several types of posttranscriptional modifications have been confirmed to contribute to maternal mRNA clearance. The dysregulation of such clearance could lead to severe developmental defects in non-human model organisms [2–4], and has been observed frequently in arrested embryos from patients [5]. Few of these discoveries, however, have examined the famous adenosine-to-inosine (A-to-I) RNA editing (referred to simply as "RNA editing" thereafter) [6].

As one of the well-known posttranscriptional modifications, RNA editing converts the adenosines into inosines [7]. Because inosines are more like guanosines than the original adenosines, such editing can have various functional consequences, including the generation of non-synonymous substitutions during translation ("recoding") [8] or novel protein isoforms due to altered splicing [9], the alteration of microRNA-target binding affinity [10,11], and the disruption of long stem loops in endogenous mRNA that might aid the self-tolerance of innate immunity [12]. An estimated ~20% of human 3'-untranslated region (3'-UTR) edits may affect microRNA binding sites (MBSs), which possibly affects the targeting of many microRNAs [13]. In addition, previous studies have identified several disease-informative edits [14], suggesting their potential role in key developmental processes.

Previous studies, however, have been conducted with adult human tissues; whether and how RNA editing could consistently contribute to human embryonic development remains largely unclear. Several recent studies have been conducted to investigate edits in human embryos using pilot embryo RNA sequencing (RNA-Seq) datasets [15–17], but the sample sizes have been limited and whether their conclusions drawn apply to most embryos remains unclear. In addition, the rapid primate-specific expansion of Alu elements in mRNAs [18,19], which are hotspots of RNA editing [20], hinders the determination of the functional role
of RNA editing in human embryos by simple examination of their non-primate model organism counterparts [21].

In this study, we first compiled the first systematic A-to-I editome for human early embryonic development based on 2,071 early embryonic RNA-Seq samples. We then confirmed the existence of per-stage Recurrent Edits (REs; edits observed in >=50% of samples) along with several lines of evidences suggesting their potential functions in human early embryonic development. In particular, we discovered a likely supportive role of REs in enhancing maternal mRNA clearance through the regulation of microRNA-based mRNA decay.

Results

Construction of an adapted identification pipeline for 2,071 human embryonic RNA-Seq datasets

Screening for systematically published datasets in the National Center for Biotechnology Information's (NCBI's) Gene Expression Omnibus (GEO) database [22] yielded a catalog of 2,071 samples in 29 groups defined by developmental stages and cell types related to human embryonic development (Fig. 1A and Supplementary Table 1). Because none of these samples have genotype available, we chose a stringent approach with the use of RNA-Seq-data alone [23] for the identification of edits. As an adaptation for RNA-Seq datasets containing data on several-cell (e.g., 4-cell) and single-cell (e.g., oocytes) samples, we further minimized false positives (possibly the result of genomic contamination) by excluding all detected variant sites that overlapped with known genomic variants from worldwide genotyping studies [24–27]. When tested on an independent dataset with paired DNA and RNA sequenced for each single cell [28] (Fig. 1C and Supplementary Notes 2), this pipeline generated very low false-positive rates after filtering (Fig. 1D), supporting its application to the collected embryonic RNA-Seq datasets.

Identification of systematic A-to-I editome profile for human embryonic development

The application of the stringent pipeline to all 2,071 curated samples resulted in the identification of a total of 989,191 editing sites in normal and other samples (Fig. 2A), with hundreds to tens of thousands of sites identified in each stage (Fig. 2B). Consistent with previous large-scale identifications of RNA editing [23], we detected a high proportion of A-to-G mismatches (Fig. 2C), a high proportion of Alu edits among all edits similar to those in adult human tissues (Fig. 2D), and a signature RNA-specific adenosine deaminase (ADAR)-binding motif across all of these sites (Fig. 2E). These results supported the reliability of this human embryonic editome in revealing the dynamics of editing sites throughout embryonic development (see Fig. 2F for the example of the well-studied BLCAP Y2C recoding site [29]).

Detection of thousands of organized REs throughout early embryonic development

A per-stage search revealed that thousands of REs were present in all early embryonic stages (Fig. 3A and B).
Compared with all observed edits, REs were more likely to be exonic (<50% vs. >75%; Fig. 3C), and most were located in 3'-UTR regions (Fig. 3D). In addition, rather than being dispersed randomly like biological noises, >50% of REs persisted through stage transitions until the 2-cell stage, and ~30% of REs persisted through the 2-to-4-cell transition (Fig. 3E). It is also worth noting that most REs did not disappear completely upon stage transition, although they were no longer REs (as indicated by the scarcity of "not detected" edits in Fig. 3E). These results suggest a consistent, stable pattern (and thus a possibly functional role) of 3'-UTR REs in early human embryonic development.

**REs target similar genes enriched with DNA replication-related functions across early embryonic stages**

To gain insight into the functions that REs might affect, we selected genes that are frequently targeted by REs for each stage separately (Methods). We discovered hundreds of frequently targeted genes, >50% of which were targeted primarily in 3'-UTR REs in early embryonic stages (Fig. 4A and B). Similar to the REs, these RE-targeted genes also displayed a large degree of overlap from the oocytes (GV) to the 2-cell stages, and most such genes observed in 4-cell embryos were also observed in the 2-cell stage (Fig. 4C). Given this consistent pattern, we investigated the specific functions that these genes share, and found that functions enriched across >=3 stages were mostly related to DNA replication, a phenomenon observed only on genes targeted in exonic (primarily 3'-UTR) regions (Fig. 4D). These observations suggest a consistent functional impact of REs in early human embryogenesis.

**RE-matching edits undergo organized loss in embryos with uniparental disomy and those from elder mothers**

To further investigate the functional importance of REs, we examined for each early developmental stage whether RE-matching edits underwent an organized loss in embryos with particular phenotypes indicative of low embryo quality. An initial scan revealed 107 edits on 76 genes (Supplementary Table 3) that were REs in normal embryos, but completely lost in the same stage in pathological embryos (GSE133854 [30]) and embryos from elder mothers (GSE95477 [31]; Fig. 5A). Notably, one of these edits, chr9:132375956, induces missense recoding during the translation of transcription termination factor 1 for ribosomal gene transcription (TTF1; ENSG00000125482 [32]), and its complete lost in parthenogenetic zygotes suggests the potential clinical value of the detection of this edit in cases of uniparental disomy. In addition, gene ontology (GO) analysis of the genes with AG-lost REs revealed enrichment in various functions shared by four or more genes, and many of these functions were related to RNA metabolism, (Fig. 5B and Supplementary Table 4), suggesting a potential link between these REs and RNA metabolism in these pathological embryos.

The complete lost of certain RE-matching edits in these embryos suggests the possibility that most other REs also undergo a systematic loss; indeed, we detected fewer RE-matching edits in protein-coding genes of abnormal embryos and embryos from elder mothers (GSE133854, GSE95477, and GSE101571 [33]) than in stage-matched control embryos included in the same datasets, although most differences were not statistically significant, possibly due to limited sample size (Fig. 5C). To examine whether a subset of protein-coding genes actively regulated by REs underwent statistically significant loss in abnormal
embryos or embryos from elder mothers, we examined the average change in REs per gene in targets of maternal mRNA clearance (using other maternal genes as controls) with the exclusion of samples that were potentially outliers (Supplementary Notes 3). In GSE95477 samples (Fig. 5D), we observed statistically significantly fewer RE-matching edits per gene in targets of maternal mRNA clearance on average in oocytes (GV) from elder mothers than in those from young mothers (Fig. 5E), suggesting that REs are favored in these targets at least at this developmental stage (see Supplementary Fig. 3 for results from other datasets). These results raise the possibility that the loss of certain REs is indicative of certain phenotypes in a manner that may be related to key embryonic processes such as maternal mRNA clearance.

Targets of maternal clearance had more RE-induced microRNA binding sites than did nontargets

Having gained a preliminary understanding what genes and functions RE might affect, we then asked how RE would affect these genes. Because most exonic REs are located in 3'-UTRs (Fig. 3D), the gene element containing most MBSs, many 3'-UTR REs may affect genes by interfering with MBSs and thereby the microRNA-based regulatory program (see Fig. 6A for an example), a mechanism that has been studied extensively for RNA editing [10, 11, 13]. To confirm this, we annotated all MBSs on all editing-targeted transcripts before and after editing by TargetScan (with edited inosine treated as guanosine), and analyzed their associations with 3'-UTR edits. The 3'-UTR REs were much more likely to alter MBSs than were other edits (~50% vs. ~25% or less; Fig. 6B); in particular, they were more likely to result in MBS gain (mean, ~1 vs. ~0; Fig. 6C), suggesting their potential role in the enhancement of the microRNA-mediated degradation of targeted transcripts.

Based on this observation, we speculated that REs help to degrade mRNAs targeted by maternal mRNA clearance (referred to as "clearance targets" hereafter) [34] by introducing more MBSs (Fig. 6A). This hypothesis was immediately validated by the observation that REs result in bringing more MBSs on clearance targets than on other maternal genes (Fig. 6D), even with consideration of the net MBS change (i.e., accounting for the loss of preexisting MBSs by RE; Supplementary Fig. 4). Specifically, SUV39H2, a known target of maternal mRNA clearance during human embryrogenesis [5], received eight new MBSs and lost one preexisting MBS, suggesting a potential RE-MBS-based mechanism of clearance regulation (Fig. 6E). These results suggest a novel role of REs (and possibly other RNA edits) in the enhancement of maternal mRNA clearance through the introduction of more MBSs.

Discussion

By curating and analyzing the largest human embryonic editome to date, we showed that the early embryonic stages harbour thousands of REs that are preferably exonic and highly shared between stages at the editing site and target gene levels. We also showed that these REs could potentially enhance maternal mRNA clearance, a process that has been found to be associated with RNA editing in mouse embryos [6], by introducing more MBSs to clearance targets than to other maternal genes.
The role of A-to-I RNA editing in human has been ambiguous for a long time. Although several studies have demonstrated the importance of certain editing events [35–37] and documented the adverse consequences of the disruption of the core editing enzyme ADAR1 [38–41], the possible functional roles of RNA editing in key embryonic developmental processes remains largely unclear. Based on our observation of associations among REs, MBSs, and maternal mRNA clearance, we propose a working model of how human embryos could take advantage of the RNA editing machine for better development: embryonic A-to-I RNA edits, including the REs discovered in this study and possibly other accompanying edits, occur and result in the introduction of MBSs to (at least some) clearance targets more often than to other maternal genes; these targets are then more efficiently targeted and degraded by the microRNA machinery than they were in unedited form, thereby enhancing the maternal mRNA clearance (and thus the embryonic development [42–44]) (Fig. 7, left). Recent research has revealed the impairment of RNA editing in mouse oocytes upon knockout of Cnot6l, a deadenylase in the carbon catabolite repression 4-negative on TATA-less complex that is required for deadenylation-based maternal mRNA clearance [6]; although the roles of RNA editing in human and mice may not be directly comparable, this finding suggests that the microRNA-based effect of RNA editing on maternal mRNA clearance discovered in the present study might cooperate with other posttranscriptional modifications [2–4], possibly in an additive way [2], to advance maternal mRNA clearance.

Apart from altering the MBS count in clearance targets, REs (and other edits) can, in theory, affect embryonic development in other ways (Fig. 7, right). In fact, in addition to the completely lost RE-matching edits identified (Supplementary Table 3), we discovered a subset of RE-matching edits that are nearly lost in cases of uniparental disomy [30] (Supplementary Table 5); these edits may be of additional critical value for scientific understanding and clinical applications. Likewise, one could also further examine the recoding edits (Supplementary Table 6) in the editome to identify additional edits with critical functional impacts. Potentially useful insights could also be gained from the examination of RE-targeted genes (and their accompanying REs) in postimplantation stages (Fig. 4B). Although scarce, several RE-targeted genes (Supplementary Table 7) are frequently edited by certain REs; these REs could be of special research interest, provided that they are validated to be non-somatic mutations by, for example, the examination of additional postimplantation embryos from independent individuals.

We'd note that we may have missed a certain number of edits (or even REs) in the current editome due to the relatively low sequencing depth of early single-cell RNA-Seq techniques, although we sought to cover as many reliable edits as possible by screening a set of thousands of samples with application of a stringent pipeline for candidate RNA edits. More informative REs (and their additional functions) may be discovered with deeper sequencing. In addition, the functional relevance of this editome to embryonic development is far from being extensively studied; specifically, REs might be functionally important in key embryonic processes other than maternal mRNA clearance, such as those involving DNA replication and repair (as suggested by the results illustrated in Fig. 4D).

Conclusions
In this study, we have introduced the first large-scale A-to-I RNA editome for early human embryos, the analysis of which revealed a consistent early-stage editing pattern (of REs) with probable functional importance in microRNA-based maternal mRNA clearance. These discoveries, along with the editome itself, are valuable resources for further examination of the interplay between RNA editing and other mechanisms involved in maternal mRNA clearance, as well as the identification of additional roles of A-to-I RNA editing in early human embryonic development.

Methods

Compilation of human embryonic RNA-Seq datasets

In addition to including human embryonic RNA-Seq datasets whose A-to-I editomes have been studied previously [15,16], we used GEOmetadb [45] to search GEO [22] for all RNA-Seq samples submitted before October 1st, 2020, using the keyword "embryo" and the species restriction of Homo sapiens. We filtered the datasets identified by this search to identify paired-end RNA-Seq data with read length >= 75 × 2 bp, to increase the accuracy of A-to-I RNA editome identification [46]. For single-cell RNA-Seq datasets, we required that the sequencing technology not be based on cell barcoding. This process yielded a total of 2,071 samples (1,852 normal and 274 abnormal) from 18 datasets (Supplementary Table 1), which were sent to the A-to-I RNA editome identification pipeline.

Identification of the A-to-I editome and REs within it

We adapted a published pipeline [23] used in the Genotype-Tissue Expression A-to-I editome study [47] (Supplementary Notes 1). Briefly, we: 1) generated a new reference genome by concatenating the hg38 assembly and all sequence fragments spanning known junction sites from the version 32 annotation of GENCODE [48]; 2) aligned quality-controlled reads to this new reference; 3) mapped these alignments back onto hg38 coordinates; 4) called variants with GATK [49]; and 5) filtered for A-to-G variants that did not overlap with common genomic variants or regions prone to algorithmic errors, and with enough read and sample support. REs were then identified for each stage by filtering for those edits observed in >= 50% samples in that stage.

To reduce false positive results in this pipeline as much as possible, we expanded the set of genomic variants used in step 5 above. Specifically, in addition to data from dbSNP version 151 [50], the University of Washington Exome Sequencing Project (https://evs.gs.washington.edu/EVS/), and the 1000Genomes Project [26], we used data from the Genome Aggregation Databse [27] and the NCBI's Allele Frequency Aggregator project [25] which span more than hundreds of thousands of individuals to exclude variants that overlapped with population genomic variants found in these studies or projects. Variants passing through this filter are very unlikely to be false positives arising from genomic variation.

Annotation of the A-to-I editome
We obtained from the GATK variant call format output the chromosome and position for each A-to-I edit in each sample, as well as its read coverage (AN), the number of reads supporting the editing (AC), and the editing frequency AF = AC / AN. We then annotated these edits using SnpEff [51] with GENCODE version 32 annotation, and classified them according to their SnpEff 'Annotation' Field: coding sequence (CDS) regions, 5'-UTRs, 3'-UTRs, exonic regions of non-coding transcripts, introns, and intergenic regions (Supplementary Table 2). When a given edit was of different types on different transcripts of a given gene locus (e.g., in the CDS region of one transcript and the 3'-UTR of another), we assigned the edit type in the following order: CDS > 5'-UTR > 3'-UTR > non-coding exonic > intronic > intergenic.

**Motif visualization for editing sites**

We used Two Sample Logo (version 1.23) [52] to plot the ADAR-binding sequence motif. For the background sequence file (file for the -N option), we chose all 7-bp subsequences of GENCODE version 32 transcript sequences (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_32/gencode.v32.transcripts.fa.gz) whose fourth nucleotide was adenine.

**Gene-level enrichment analysis**

For GO term enrichment analysis, we used the "enrichGO" function in clusterProfiler [53] with the org.Hs.eg.db database [54] to analyze enriched terms for each type of genes in each stage. To correct for multiple hypothesis testing, we pooled all enrichment results and adjusted them using the Benjamini-Hochberg method.

**Annotation of MBSs and effects of REs on them**

We used TargetScan [55] to annotate MBSs in 3'-UTRs. The multi-species 3'-UTR input for each chromosome was generated by subsetting the UCSC 30-way alignment in MAF format (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/multiz30way/) with the "interval_maf_to_merged_fasta.py" script from Galaxy tools [56] (https://github.com/galaxyproject/tools-iuc) and https://github.com/galaxyproject/galaxy (release 21.01)) and the BED file describing the 3'-UTRs for that chromosome. To annotate the effect of each RE on MBSs, we modified the multi-species 3'-UTR input for TargetScan by replacing the adenine at the RE site in the human 3'-UTR sequence with guanine, and fed this modified multi-species 3'-UTR input to TargetScan again; in this way, one modified multi-species 3'-UTR input for TargetScan was generated for each RE. We then annotated an RE on a given combination of gene and microRNA family as follows:

1. The RE was annotated as **MBS neutral** if, for each of all transcripts of the gene locus, the number of new MBSs of the microRNA family that it introduced was equal to the number of preexisting MBSs...
of the microRNA family that it removed;

2. The RE was annotated as *MBS gaining* if both of the following two conditions were satisfied: 1) for each of all transcripts of the gene locus, the number of new MBSs of the microRNA family that it introduced was no less than the number of preexisting MBSs of the microRNA family that it removed; and 2) for at least one transcript of the gene locus, the number of new MBSs of the microRNA family that it introduced was strictly greater than the number of preexisting MBSs of the microRNA family that it removed;

3. The RE was annotated as *MBS losing* if both the following two conditions were satisfied: 1) for each of all transcripts of the gene locus, the number of new MBSs of the microRNA family that it introduced was no greater than the number of preexisting MBSs of the microRNA family that it removed; and 2) for at least one transcript of the gene locus, the number of new MBSs of the microRNA family that it introduced was strictly smaller than the number of preexisting MBSs of the microRNA family that it removed.

**Annotation of maternal genes and targets of maternal mRNA clearance**

We used STAR [57] to align the trimmed reads from the adapted RNA identification pipeline onto hg38 and then StringTie [58] to estimate the expression level of each gene. We then defined maternal genes as those with median FPKM > 2 in at least one of oocyte (GV) and oocyte [metaphase of second meiosis (MII)] stages. Finally, we annotated a maternal gene as a target of maternal mRNA clearance if the smaller median FPKM value between the oocyte (GV) and oocyte (MII) values was more than twice the median FPKM in the 8-cell stage. All normal samples from oocyte (GV), oocyte (MII), and 8-cell were considered.

**Declarations**

**Data availability statement**

The datasets supporting the conclusions of this article are available in the Zenodo repository, [https://zenodo.org/record/5804921](https://zenodo.org/record/5804921). All raw sequencing datasets are available from original publication (Supplementary Table 1), and the compiled editome (and some intermediate results) is available from Zenodo, [https://zenodo.org/record/5804921](https://zenodo.org/record/5804921).

**Code availability**

Codes for reproduction of the results reported in this article are available from the GitHub repository, [https://github.com/gao-lab/HERE](https://github.com/gao-lab/HERE).

**Competing Interests statement**
The authors have no conflict of interests.

Acknowledgments

The authors wish to thank Qiya Huang, Longteng Wang, Dr. Cheng Li and Zi-Yu Chen for their helpful feedback and discussions. The analysis was supported by the High-performance Computing Platform of Peking University. The authors would like to thank the developers of Snakemake for this excellent workflow management tool.

Funding

This work was supported by the Beijing Natural Science Foundation [http://kw.beijing.gov.cn/; 5204040 to H.L.], the Beijing Nova Program of Science and Technology [https://mis.kw.beijing.gov.cn; Z191100001119064 to H.C.], the National Key Research and Development Program [2016YFC0901603 to G.G.], the China 863 Program [2015AA020108 to G.G.], the National Natural Science Foundation of China [http://www.nsfc.gov.cn; 81973244 to H.L., 31801112 to H.C., 61873276 to X.B.], and the State Key Laboratory of Protein and Plant Gene Research and the Beijing Advanced Innovation Center for Genomics (ICG) at Peking University [to G.G.]. The research of G.G. was supported in part by the National Program for Support of Top-notch Young Professionals. Funding for open access charge: the Beijing Natural Science Foundation [http://kw.beijing.gov.cn/; 5204040 to H.L.].

References

1. Theunissen TW, Jaenisch R. Mechanisms of gene regulation in human embryos and pluripotent stem cells. Development. 2017;144:4496–509.
2. Kontur C, Jeong M, Cifuentes D, Giraldez AJ. Ythdf m6A Readers Function Redundantly during Zebrafish Development. Cell Rep. 2020;33:108598.
3. Morgan M, Much C, DiGiacomo M, Azzi C, Ivanova I, Vitsios DM, et al. mRNA 3’ uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. Nature. 2017;548:347–51.
4. Chang H, Yeo J, Kim JG, Kim H, Lim J, Lee M, et al. Terminal Uridylyltransferases Execute Programmed Clearance of Maternal Transcriptome in Vertebrate Embryos. Mol Cell. 2018;70:72–82.e7.
5. Sha QQ, Zheng W, Wu YW, Li S, Guo L, Zhang S, et al. Dynamics and clinical relevance of maternal mRNA clearance during the oocyte-to-embryo transition in humans. Nat Commun. 2020;11:4917.
6. Brachova P, Alvarez NS, Christenson LK. Loss of Cnot6l Impairs Inosine RNA Modifications in Mouse Oocytes. Int J Mol Sci. 2021;22.
7. Eisenberg E, Levanon EY. A-to-I RNA editing - immune protector and transcriptome diversifier. Nat Rev Genet. 2018;19:473–90.
8. Hoopengardner B, Bhalla T, Staber C, Reenan R. Nervous system targets of RNA editing identified by comparative genomics. Science. 2003;301:832–6.
9. Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G. RNA-editing-mediated exon evolution. Genome Biol. 2007;8:R29.

10. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science. 2007;315:1137–40.

11. Pinto Y, Buchumenski I, Levanon EY, Eisenberg E. Human cancer tissues exhibit reduced A-to-I editing of miRNAs coupled with elevated editing of their targets. Nucleic Acids Res. 2018;46:71–82.

12. Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science. 2015;349:1115–20.

13. Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, Zhu Y, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol. 2012;30:253–60.

14. Costa Cruz PH, Kawahara Y. RNA Editing in Neurological and Neurodegenerative Disorders. Methods Mol Biol. 2021;2181:309–30.

15. Qiu S, Li W, Xiong H, Liu D, Bai Y, Wu K, et al. Single-cell RNA sequencing reveals dynamic changes in A-to-I RNA editome during early human embryogenesis. BMC Genomics. 2016;17:766.

16. Li T, Li Q, Li H, Xiao X, Ahmad Warraich D, Zhang N, et al. Pig-specific RNA editing during early embryo development revealed by genome-wide comparisons. FEBS Open Bio. 2020;10:1389–402.

17. Qiu J, Ma X, Zeng F, Yan J. RNA editing regulates IncRNA splicing in human early embryo development. PLoS Comput Biol. 2021;17:e1009630.

18. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet. 2002;3:370–9.

19. Daniel C, Behm M, Öhman M. The role of Alu elements in the cis-regulation of RNA processing. Cell Mol Life Sci. 2015;72:4063–76.

20. Schaffer AA, Levanon EY. ALU A-to-I RNA Editing: Millions of Sites and Many Open Questions. Methods Mol Biol. 2021;2181:149–62.

21. Buchumenski I, Holler K, Appelbaum L, Eisenberg E, Junker JP, Levanon EY. Systematic identification of A-to-I RNA editing in zebrafish development and adult organs. Nucleic Acids Res. 2021;49:4325–37.

22. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashovsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41:D991–5.

23. Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O’Connell MA, et al. Identifying RNA editing sites using RNA sequencing data alone. Nat Methods. 2013;10:128–32.

24. Exome Variant Server [Internet]. NHLBI GO Exome Sequencing Project (ESP), Seattle, WA; 2020 [cited 2020 Nov 3]. Available from: http://evs.gs.washington.edu/EVS/

25. Phan L, Jin Y, Zhang H, Qiang W, Shekhtman D, Shao D, et al. ALFA: Allele Frequency Aggregator [Internet]. National Center for Biotechnology Information, U.S. National Library of Medicine; 2020. Available from: https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/

26. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
27. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
28. Ding J, Lin C, Bar-Joseph Z. Cell lineage inference from SNP and scRNA-Seq data. Nucleic Acids Res. 2019;47:e56.
29. Levanon EY, Hallegger M, Kinar Y, Shemesh R, Djinovic-Carugo K, Rechavi G, et al. Evolutionarily conserved human targets of adenosine to inosine RNA editing. Nucleic Acids Res. 2005;33:1162–8.
30. Leng L, Sun J, Huang J, Gong F, Yang L, Zhang S, et al. Single-Cell Transcriptome Analysis of Uniparental Embryos Reveals Parent-of-Origin Effects on Human Preimplantation Development. Cell Stem Cell. 2019;25:697–712.e6.
31. Reyes JM, Silva E, Chitwood JL, Schoolcraft WB, Krisher RL, Ross PJ. Differing molecular response of young and advanced maternal age human oocytes to IVM. Hum Reprod. 2017;32:2199–208.
32. Evers R, Grummt I. Molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-I. Proc Natl Acad Sci U S A. 1995;92:5827–31.
33. Wu J, Xu J, Liu B, Yao G, Wang P, Lin Z, et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature. 2018;557:256–60.
34. Marco A. Clearance of Maternal RNAs: Not a Mummy’s Embryo Anymore. Methods Mol Biol. 2017;1605:1–0.
35. Gonzalez C, Lopez-Rodriguez A, Srikumar D, Rosenthal JJ, Holmgren M. Editing of human K(V)1.1 channel mRNAs disrupts binding of the N-terminus tip at the intracellular cavity. Nat Commun. 2011;2:436.
36. Hu X, Wan S, Ou Y, Zhou B, Zhu J, Yi X, et al. RNA over-editing of BLCAP contributes to hepatocarcinogenesis identified by whole-genome and transcriptome sequencing. Cancer Lett. 2015;357:510–9.
37. Chen L, Li Y, Lin CH, Chan TH, Chow RK, Song Y, et al. Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. Nat Med. 2013;19:209–16.
38. Jiang Q, Crews LA, Barrett CL, Chun HJ, Court AC, Isquith JM, et al. ADAR1 promotes malignant progenitor reprogramming in chronic myeloid leukemia. Proc Natl Acad Sci U S A. 2013;110:1041–6.
39. Zhou S, Yang C, Zhao F, Huang Y, Lin Y, Huang C, et al. Double-stranded RNA deaminase ADAR1 promotes the Zika virus replication by inhibiting the activation of protein kinase PKR. J Biol Chem. 2019;294:18168–80.
40. Stellos K, Gatsiou A, Stamatelopoulos K, Perisic Matic L, John D, Lunella FF, et al. Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation. Nat Med. 2016;22:1140–50.
41. Lazzari E, Mondala PK, Santos ND, Miller AC, Pineda G, Jiang Q, et al. Alu-dependent RNA editing of GLI1 promotes malignant regeneration in multiple myeloma. Nat Commun. 2017;8:1922.
42. Zhao BS, Wang X, Beadell AV, Lu Z, Shi H, Kuuspalu A, et al. m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. Nature. 2017;542:475–8.
43. Yu C, Ji SY, Sha QQ, Dang Y, Zhou JJ, Zhang YL, et al. BTG4 is a meiotic cell cycle-coupled maternal-zygotic-transition licensing factor in oocytes. Nat Struct Mol Biol. 2016;23:387–94.

44. Ivanova I, Much C, Di Giacomo M, Azzi C, Morgan M, Moreira PN, et al. The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. Mol Cell. 2017;67:1059–1067.e4.

45. Zhu Y, Davis S, Stephens R, Meltzer PS, Chen Y. GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus. Bioinformatics. 2008;24:2798–800.

46. Lo Giudice C, Tangaro MA, Pesole G, Picardi E. Investigating RNA editing in deep transcriptome datasets with REDItools and REDIportal. Nat Protoc. 2020;15:1098–131.

47. Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, et al. Dynamic landscape and regulation of RNA editing in mammals. Nature. 2017;550:249–54.

48. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47:D766–73.

49. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.

50. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.

51. Cingolani P, Platts A, Wang leL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80–92.

52. Vacic V, Iakoucheva LM, Radivojac P. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics. 2006;22:1536–7.

53. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.

54. Carlson M. Org.Hs.eg.db: Genome wide annotation for Human. R package version 3.8.2; 2019.

55. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120:15–20.

56. Afgan E, Baker D, Batut B, Beek M van den, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018;46:W537–44.

57. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

58. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.

Figures
Figure 1

Overview of datasets and the adapted stringent pipeline used in this study. (A) Overview of RNA-Seq curation from the GEO database. (B) Overview of the adapted stringent pipeline. (C, D) The pipeline yielded a low false-positive rate (D) in a paired DNA-RNA sequencing dataset for single cells (C).

Figure 2

Identification and validation of the A-to-I editome for human embryos. (A) Total number of edits identified in all samples. (B) Counts of edits identified per stage. Groups labelled with * have only one sample. GV, germinal vesicle. MI, metaphase of first meiosis. MII, metaphase of second meiosis. PN, pronuclear. TE, trophectoderm. ICM, inner cell mass. hESC, human embryonic stem cell. CTB, cytotrophoblast. STB, syncytiotrophoblast. EVT, extravillous trophoblast. MTB, migratory trophoblast. (C) A-to-G ratios for all variants detected across all samples. (D) Alu ratios for all edits across all samples. (E) The signature ADAR-binding motif computed from all edits. (F) The profile of the BLCAP Y2C recoding edit across stages; the horizontal stripes represent edited samples with the color denoting the editing level. Arrows indicate the direction from the 5'- to the 3'-end (for DNA and RNA) or from the N-terminal to the C-terminal (for protein).

Figure 3

Thousands of organized REs were detected in early human embryos. (A) Definition of RE. (B) Count of REs per stage. (C) Percentage of exonic edits and REs in the early stages of embryogenesis. (D) Percentage of 3'UTR REs in early stages of embryogenesis. See also Supplementary Fig. 1 for percentages of general edits. (E) Sankey plot describing the numbers of REs passed to subsequent early stages. For clarity, only REs observe in at least one of the two stages appear in each subplot. For (C)-(E), we considered only REs on protein-coding genes.

Figure 4

REs target similar genes enriched with DNA replication-related functions across the early stages of embryogenesis. (A) Definition of RE-targeted genes. Note that for further filtering against possible false-positive results, we used a more stringent criterion (edited by at least one RE in >=80% of samples) than used for the definition of REs. (B) Counts of RE-targeted genes per stage. (C) Sankey plot describing the number of RE-targeted genes passed to subsequent early stages. For clarity, only RE-targeted genes
observed in at least one of the two stages appear in each subplot. (D) Cross-stage (>=3 stages) enriched functions of RE-targeted genes. Note that we performed enrichment analyses on 3’-UTR- and intronic-RE-targeted genes separately, and discovered cross-stage enriched functions only for the former (Methods).

**Figure 5**

Fewer REs on targets of maternal mRNA clearance were observed in oocytes from elder mothers than in those from young mothers. (A) Count of RE-matching edits that were completely lost in abnormal embryos and embryos from elder mothers (AMA embryos). (B) Biological processes enriched by four or more genes targeted by AG-lost RE-matching edits. Only those with BH-adjusted $p$-values less than 0.1 were shown. (C) Distribution of RE-matching edits on protein-coding genes in datasets covering abnormal embryos and embryos from elder mothers. All $p$-values were unadjusted and were derived from one-tailed Wilcoxon's Rank Sum tests, with the alternative hypothesis being that mean value for the left (abnormal/elder-mother) group would be less than that for the right (control) group. (D) Design for the study of dataset GSE95477. (E) Average numbers of RE-matching edits per gene in different stages and gene groups. All $p$-values were unadjusted and were derived from one-tailed Wilcoxon's Rank Sum tests, with the alternative hypothesis being that the average number of RE-matching edits per gene would decrease from young to old mothers. Note that some outlier samples are excluded from this plot (Supplementary Notes 3 and Supplementary Fig. 2; see also Supplementary Fig. 3 for the analysis of other datasets).

**Figure 6**

REs induced more microRNA binding sites (MBSs) on maternal clearance targets than on non-targets. (A) An example model of how RNA editing can result in the gaining of new MBSs. (B) REs are more likely to alter MBSs than general edits. (C) REs are more likely to result in new MBS gains than in preexisting MBS losses. The $p$-value was derived from one-tailed Wilcoxon's Rank Sum test, with the alternative hypothesis being that mean value for the "MBS gain" group would be greater than that for the "MBS lost" group. (D) Genes targeted by maternal mRNA clearance have more MBS-gaining REs than do other maternal genes (also see Supplementary Fig. 4). All $p$-values were unadjusted and were derived from one-tailed Wilcoxon's Rank Sum test, with the alternative hypothesis being that mean value for the left (decay at 8-cell) group would be greater than that for the right (others/baseline) group. The baseline group is just a value of 0; we plotted it here for the sake of visual clarity. (E) An example of RE-induced MBS gains on a known clearance target SUV39H2.

**Figure 7**
Proposed model of how human embryos take advantage of REs (and other edits).

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables12.xlsx
- SupplementaryTable3.csv.csv
- SupplementaryTable4.csv.csv
- SupplementaryTable5.csv.csv
- SupplementaryTable6.csv.csv
- SupplementaryTable7.csv.csv
- GUIDEDOA2200405nrreportingsummary.pdf
- GUIDEDOA2200405SupplementaryInformation.docx