Supplementary information, Data S1 Experimental Procedures

Plasmids construction

Plasmids for different shRNAs were constructed into pLVTHM. A scramble shRNA plasmid was used as a negative control. Two efficient shRNAs were used to knock down ADAR1 in H9 cells, KD1 (sh4146): GAGGCAGAAACCTAAGAAG and KD2 (sh4974): GTTGACTAAGTCACATGTAAA(Toth et al., 2009).

To rescue ADAR1 expression, the full-length cDNA of ADAR1 p110 along with a N-flag tag was amplified from pcDNA3-ADAR1 plasmid (a gift from Dr. Yong Liu) with PhantaTM Super-Fidelity DNA Polymerase (vazyme) and inserted into plenti-EGFP vector by BamH I and Not I. Plasmids expressing ADAR1-E912A(Lai et al., 1995; Ota et al., 2013) and ADAR1 EAA(Valente and Nishikura, 2007), were then constructed using Quikchange site-directed mutagenesis kit (Stratagene) with mutated gene-specific primers.

For the 3'-UTR luciferase reporter assay, full-length or partial 3'-UTRs of various genes (Figure S6) were amplified from genomic DNAs of H9 cells with PrimeSTAR HS DNA Polymerase (Takara). PCR products were inserted into psiCHECK-2 vector (Promega) by Xho I and Not I. Mutagenesis of predicted miRNA-targeting sites was generated using the site-directed mutagenesis kit described above.

To study whether ADAR1 plays a role in the regulation of miR302s transcription, a DNA fragment of ~600bp containing the promoter of pri-miR302(Barroso-delJesus et al., 2008) was amplified and ligated into pGL3-basic vector (Promega) by MluI and BglII. All primers used were listed in Table S6 and all constructed plasmids were verified by Sanger sequencing.

hESCs maintenance and differentiation

H9 hESCs were maintained on irradiated-MEF feeder cells and passaged weekly as described previously(Yin et al., 2012). H9 cells were induced to differentiate into the region-specific neural cells according to the published protocols(Boisvert et al., 2013; Li et al., 2005; Li et al., 2009b; Wang et al., 2013; Zhang et al., 2010). Briefly, for forebrain neural cells induction, hESCs were detached from MEF cells and suspended in hESCs medium without FGF2 for 4 days in Petri dish. After the aggregates floated in neural induction medium (NIM) consisting of DMEM/F12, N2 supplement, 1xNEAA and heparin for additional 2-3 days, the aggregates were induced to adhere to laminin-coated substrate, and primitive neuroepithelial (NE) would be observed on days 8-10, followed by the definitive NE cells in the center of colonies occurring at days 14-17 of differentiation. The definitive NE were gently blown off (day 17) and suspended in NIM for one week and then replated on laminin-coated substrate for terminal differentiation. For motor neuron generation, primitive NE at day 10 was administrated with RA (0.1 μM) for caudalization. After isolation of definitive NE, RA, purmorpahmine and B27 supplement were added to NIM for ventralization. Samples from day 0 (d0), day 10 (d10), day 17 (d17) and day 35 (d35) were collected for subsequent analyses.

Lentivirus production and infection of hESCs

Lentiviral particles were produced by transient cotransfection of 20 μg transfer vector constructs (pLVTHM or plenti constructs), 15 μg of psPAX2 and 10 μg pMD2.G into 293FT cells (Invitrogen) at 1.25×10^7 cells in a 15-cm dish. The supernatant containing lentivirus particles was harvested at 48 and 72 hr after transfection, filtered through Fast PES filter unit (0.2 μm pore size, Thermo fisher Scientific). Subsequently viral particles were

concentrated 100-fold by sucrose gradient ultracentrifugation and resuspended in PBS containing 0.1% BSA, and stored at -80^{o}C ultra cold freezer until use.

To infect hESCs with shRNA lentivirus, H9 cells grown on Matrigel (BD Biosciences) were digested into single cells with Accutase (Millipore). Cells were incubated with MEF-condition medium (MEF-CM) containing 10 µl concentrated lentivirus, 5 µg/ml polybrene (sigma) and 10 µM Y-27632 (Tocris) at 37^{o}C for 1 hr in a 15 ml-conical tube and were then seeded onto Matrigel coated culture dishes. 24 hr later, cells were fed with pre-warmed MEF-CM containing 5 µg/ml polybrene for 15 min and then infected the second time on the plates. 4 days after 1st infection, cells were split and cultured for 5 additional days. GFP-positive colonies were mechanically detached and cultured individually. To obtain a pure and stable ADAR1 KD H9 cell population, the RNAi cells were further split and selected for more than 3 passages. ADAR1 KD H9 cells were infected with lentiviral particles that express ADAR1-wt, ADAR1-E912A, or ADAR1-EAA using the same method described above to rescue ADAR1 knockdown.

5-Bromo-2-deoxyuridune (BrdU) incorporation assay

50,000 H9 cells or H9 cells with different treatments were seeded onto 12-well plates coated with Matrigel in MEF-condition medium containing 8 ng/ml bFGF and were cultured for 4 days. On the 4th day, cells were incubated with fresh medium containing BrdU (final 10 µM) for 4 hr and then fixed with 4% paraformaldehyde (PFA) for 10 min. After the treatment with 1.5 M HCl for 30 min, BrdU incorporation was detected by immunocytochemistry with mouse anti-Brdu (1:1000; Millipore) and Cy3 conjugated goat anti-mouse IgG. Images were collected using a fluorescence microscope (BX41, Olympus). Cell nucleus was stained with DAPI (Sigma). Statistics of BrdU-incorporated cells were calculated in random image fields by comparing the number of Brdu positive cells to total 1000 counted cells.

Cloning formation efficiency

H9 cells or H9 cells with different treatments were incubated in MEF-condition medium containing 10 μM Y-27632 for 1 hr and then digested with Accutase into single cell suspention. Cells were seeded at the density of 3,000/well on 24-well Matrigel-coated plates and grew overnight in the same media with Y-27632. 24 hr later, the old medium was replaced by a standard conditioned medium without Y-27632. 5 days after the initial seeding, cells were fixed with 4% PFA and staining with the Alkaline Phosphatase Detection kit (Millipore). The cloning efficiency was calculated by comparing the number of colonies to the initially seeded cells.

Teratoma formation assay

After acclimatization for 7 days in the laboratory, the NOD/SCID mice were injected with 5 × 10^6 cells into the thigh muscle (3 mice per group). The animals were euthanized after 10 weeks, and the tumors were fixed in 4% PFA overnight, and then embedded in paraffin and sectioned at 6 μm thickness. For morphometric analysis of the toratoma, the slides were deparaffinized using xylene and ethanol (absolute, 95%, 90%, 80%, 70% diluted in water) and stained with hematoxylin and eosin (H&E) stain. Before scarification of the mice, diameter of the hind leg was measure using traditional caliper method. The care and treatment of animals were approved by the Institution Animal Care and Research Advisory Committee at Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.

RNA extraction, reverse transcription, and qPCR

Total RNAs were extracted using Trizol regent (Invitrogen) according to the

manufacturer's instructions. To discriminate pri-/pre-miRNAs, the large-sized RNA fractions (>200 bp) and small fraction RNAs (<200 bp) were isolated with *mirVana* miRNA Isolation Kit (Ambion). First-strand cDNA was synthesized according to the manufacturer's recommendations (Invitrogen) and PCR was done with 2X Taq Plus Master Mix (Vazyme biotech co.,ltd.). Quantitative PCR for cDNAs was performed using the SYBR Green PCR Master Mix (Applied Biosystems), and reactions were run on an ABI Prism@7500 Sequence Detection System. Expression values were normalized to *gapdh* mRNAs. For miRNA detection, mature miRNAs were reverse transcribed and then quantified using miScript PCR starter kit (Qiagen). All primers used for miRNA quantification were listed in Table S6. Expression values were normalized to *U6*.

RNA editing analysis by Sanger sequencing

First-strand cDNAs were synthesized using SuperScript™ III Reverse transcriptase (Invitrogen) from 2 μg total RNA with gene-specific (for IR*Alu*s containing mRNAs) or random primers. Individual PCR products that detect UTRs of mRNAs and pri-miRNAs were amplified from cDNAs and were subjected to Sanger sequencing. The Sanger sequencing results were further aligned with genomic DNAs to examine the edited Adenosines.

Luciferase reporter assay

For 3' UTR luciferase reporter assays, HeLa cells were seeded in 96-well plate and cotransfected with 50 nM miRNA mimics (Ribobio) together with 50 ng psiCHECK-2 reporter vectors (Promega) using Lipofectamine 2000 (Invitrogen). 48 hr after transfection, luciferase activity was measured using Dual-Glo luciferase assay system (Promega) using Synergy 2 Multi-Mode Microplate Reader (BioTek). The relative luciferase level was

calculated according to the manufacturer's instructions. To study the effect of ADAR1 on miR-302 promoter, 100 ng of pGL3basic empty vector or pGL3 inserted with the miR-302 promoter sequences and 50 ng of pRL-TK were transfected into scramble shRNA or ADAR1 shRNA treated cells. And luciferase activity was measured using the same method described above 24 hr after transfection.

Individual-nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP)

We followed the standard iCLIP protocol(Konig et al., 2010) with slight modifications. ADAR1 KD H9 cells infected with Flag-ADAR1 or Flag-ADAR1-E912A were washed with cold 1xPBS, then UV cross-linked at 400mJ/cm2 followed by immunoprecipitaiton with 8 μg anti-flag antibody at 4°C for 2 hours. Beads with protein-RNA complexes with 5 μg yeast tRNAs were treated with 5 Kunz U/ml of micrococcal nuclease for 5 min at 37°C, and RNA 3' ends were dephosphorylated by PNK at pH6.5. Then pre-adenylated linker L3 (3' modification with biotin) was linked to 3' end of RNA at 16°C overnight. 5' end of RNA was labeled with [γ-^{32}P] ATP and eluted with 1x Nupage loading buffer (Invitrogen) at 70°C for 10 min. Protein-RNA complexes were purified by monomeric avidin beads (Fischer Scientific) and eluted again with 1xNuPAGE loading buffer. Complexes were separated on 4-12% NuPAGE Bis-Tris gel (Invitrogen), transferred to a nitrocellulose membrane and exposure to a Fuji film at －80℃. Protein-RNA complexes were digested with proteinase K, and RNA were reverse transcribed by Superscript III reverse transcriptase (Invitrogen) with Rclip primers. cDNAs were separated on 6% TBE-urea gel, then three bands were cut at 120-200 nt (high), 85-120 nt (medium) and 70-85 nt (low). cDNAs were circularized by Circligase II (Epicentre). Single-stranded cDNAs were hybridized with cut_oligo, and then digested by BamHI. Linearized cDNAs were PCR amplified and purified on PAGEs to obtain iCLIP libraries, which were subjected to Illumina HiSeq 2000 at CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai, China.

Sequencing read mapping and peak calling for iCLIP

High-throughput sequencing reads were separated according to 4-nt experimental barcodes. 3-nt barcodes for identifying PCR duplicates were registered and removed, followed by trimming relevant sequencing adapters. Trimmed reads were aligned against the GRCh37/hg19 human reference genome with the UCSC Genes annotation (hg19 knownGene.txt updated at 2013/6/30) using TopHat2(Kim et al., 2013) 2.0.9 (parameters: -a 6 --microexon-search -m 2). Mapped reads with same genomic locations and the same 3-nt barcode, were combined. Peak calling and the false discovery rate (FDR) calculating were determined according previous work(Konig et al., 2010). In brief, the genomic location of 5' end of each combined read was considered as the cross-link site, and the binding counts for each cross-link site were calculated. To determine the FDR for each cross-link site, binding counts within -5nt to +5nt of each cross-link site were merged to represent the height of this site, and the background height distribution was computed through randomly placing the same number of binding counts within the gene for 10000 iterations. Modified FDR(Yeo et al., 2009) for each cross-link site was computed against the background height distribution, and cross-link sites with FDR < 0.05 were defined as iCLIP clusters.

To evaluate the distribution of ADAR1 binding sites on pri-miRNAs, 936 annotated pri-miRNA sequences were extracted from UCSC according to UCSC hg19 sno/miRNA track (pre-miRNA sequences from miRBase Release 15.0 plus +/- 50nt). iCLIP reads were aligned to these pri-miRNA sequences using Bowtie 0.12.9 with parameters -v 3 -k 1 --best. PCR duplicates removing and cluster identification were also employed, the same as the strategy used in iCLIP-seq genome-wide analysis.

Gene expression analysis

RNA-seq reads of each library were mapped to hg19 reference genome using TopHat2(Kim et al., 2013) with parameters “-g 1 --coverage-search --microexon-search -G knowGene.gtf -m 2”, respectively. The genome sequence and known gene annotations for hg19 were downloaded from UCSC genome browser database (http://genome.ucsc.edu/) (updated on 5/23/2012). Bedgraph files were generated using bedtools genomeCoverageBed(Quinlan and Hall, 2010) from BAM files, followed by performing UCSC tool bedGraphToBigWig (http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/) to generate bigwig files, which were uploaded to UCSC genome browser for visualization. Expression for each known gene from RefSeq was determined by covered reads and normalized with RPKM (reads per kilobase of exon model per million mapped reads).

miRNA expression analysis

We use miRDeep2(Friedlander et al., 2012) pipeline to quantify miRNA expression. Briefly, small RNA-seq reads were trimmed to remove 3' adapters, mapped to the genome (hg19), and piped up for miRNA expression. Besides 1,499 known miRNAs annotated in miRBase v.20(Kozomara and Griffiths-Jones, 2014), 1,281 new miRNA candidates were also identified in this analysis with miRDeep2. The homologies of expressed human miRNAs were determined by matching the 2-8th seed sequences to 5' end of known miRNAs from other 5 species (Gorilla, Rhesus macaque, Mouse, Rat, Gallus gallus) in miRBase v.20.

Function enrichment of selected gene sets

Enriched GO terms for each gene sets were determined by Fisher exact test, followed by Benjamini-Hochberg correction(Benjamini, 1995). GO annotations were downloaded from ftp://ftp.ncbi.nlm.nih.gov/gene/DATA on 4/14/2011, and assigned GO terms were filtered as described method(Xia et al., 2006).

Genome-wide A-to-I RNA editing analysis

Over 1.4 million reported editing sites were obtained from different datasets(Pinto et al., 2014; Zhu et al., 2013) after removing redundancy. To determine A-to-I editing level in these sites, RNA-seq reads were uniquely mapped to human genome hg19 and its exon junctions by bwa(Li and Durbin, 2009) with parameters “bwa aln -n 6 -t 40 -R 2; bwa samse -n 1”, and mismatches at all reported A-to-I sites were piped up with samtools(Li et al., 2009a). High confidence A-to-I sites were obtained by stringent cutoff criteria as described previously(Zhu et al., 2013), including i) at least two individual reads with A-to-I variants, ii) expression level is greater than 2 HPB (equivalent with RPKM for genes), and iii) A-to-I mismatch signal is greater than 95% at each reported A-to-I mismatch site. Mismatches in both ends of reads within 6 nt were removed to avoid mapping errors (Xue et al, unpublished).

For A-to-I editing in matured miRNAs, 19 known sites were examined from our small RNA-seq data sets with similar strategy as described previously(Alon et al., 2012). Briefly, raw reads were trimmed to remove adapters with cutadapt(Martin, 2011) and 2 more nucleotides at 3’ end, uniquely mapped to the reference genome hg19 with “bowtie -n 1 -m 1”, and then aligned to mature miRNAs with “bowtie -n 1 -m 1”. High quality bases (Phred score ≥ 30) at 19 known miRNA A-to-I sites were piled up with samtools (Li et al., 2009a),

and further determined by requiring i) expression level of host miRNA is greater than 1 RPM in at least one sample; ii) A-to-I editing level is greater than 0.1%.