

Precise genome editing of the Kozak sequence enables bidirectional and quantitative modulation of protein translation to anticipated levels without affecting transcription

Jingke Xie^{1,2,3,†}, Zhenpeng Zhuang^{1,4,†}, Shixue Gou^{1,2,†}, Quanjun Zhang^{1,2,5}, Xia Wang⁶, Ting Lan^{1,4}, Meng Lian^{1,5}, Nan Li^{1,2,3}, Yanhui Liang^{1,2}, Zhen Ouyang^{1,2,3,5}, Yinghua Ye^{1,2,5}, Han Wu^{1,2,5}, Liangxue Lai^{1,2,3,5,*} and Kepin Wang^{1,2,3,5,*}

¹China–New Zealand Joint Laboratory on Biomedicine and Health, CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China, ²Sanya Institute of Swine Resource, Hainan Provincial Research Centre of Laboratory Animals, Sanya 572000, China, ³Guangdong Provincial Key Laboratory of Large Animal models for Biomedicine, Wuyi University, Jiangmen 529020, China, ⁴University of Chinese Academy of Sciences, Beijing 100049, China, ⁵Research Unit of Generation of Large Animal Disease Models, Chinese Academy of Medical Sciences (2019RU015), Guangzhou 510530, China and ⁶Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou 510005, China

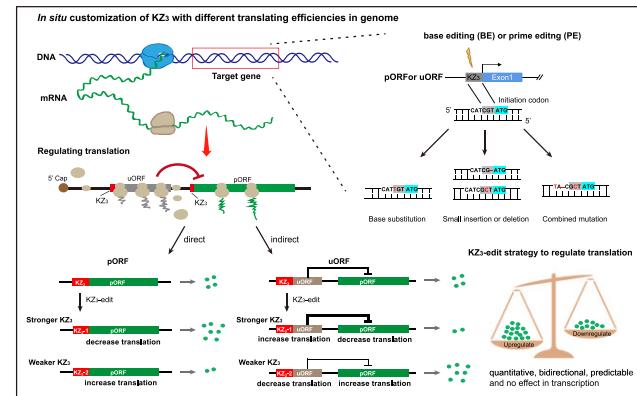
Received October 25, 2022; Revised July 31, 2023; Editorial Decision August 01, 2023; Accepted August 10, 2023

ABSTRACT

None of the existing approaches for regulating gene expression can bidirectionally and quantitatively fine-tune gene expression to desired levels. Here, on the basis of precise manipulations of the Kozak sequence, which has a remarkable influence on translation initiation, we proposed and validated a novel strategy to directly modify the upstream nucleotides of the translation initiation codon of a given gene to flexibly alter the gene translation level by using base editors and prime editors. When the three nucleotides upstream of the translation initiation codon (named KZ₃, part of the Kozak sequence), which exhibits the most significant base preference of the Kozak sequence, were selected as the editing region to alter the translation levels of proteins, we confirmed that each of the 64 KZ₃ variants had a different translation efficiency, but all had similar transcription levels. Using the ranked KZ₃ variants with different translation efficiencies as predictors, base editor- and prime editor-mediated mutations of KZ₃ in the local genome could bidirectionally and quantitatively fine-tune gene translation to the anticipated levels without affecting transcription *in vitro* and *in*

vivo. Notably, this strategy can be extended to the whole Kozak sequence and applied to all protein-coding genes in all eukaryotes.

GRAPHICAL ABSTRACT



INTRODUCTION

As a gene is a code of life and its expression is involved in all aspects of biological activities (1–4), the artificial regulation of gene expression is of great importance to many fields, such as investigation of gene function (5–7), disease

*To whom correspondence should be addressed. Tel: +86 20 3201 5304; Fax: +86 20 3201 5346; Email: wang_kepin@gibh.ac.cn
Correspondence may also be addressed to Liangxue Lai. Tel: +86 020 3201 5346; Fax: +86 020 3201 5346; Email: lai_liangxue@gibh.ac.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

treatment (8,9) and agricultural breeding (10–12). To date, several approaches have been developed to artificially up- or down-regulate gene expression. The most widely used approach to down-regulate gene expression is to use gene editing tools that induce frameshift mutations (13–16) or stop codons (17,18), or disrupt initiation codons (19,20) in the reading frame of the target gene and completely stop protein expression. RNA interference (RNAi) is another avenue that can be followed to obtain gene expression knockdown by reducing the mRNA level of a gene mediated by small RNA, including small interfering RNAs and microRNAs (21–24). Meanwhile, the main approach used to up-regulate gene expression involves transferring exogenous genes into the genomes of cells, which induces over-expression of the genes of interest (25–27). Many diseases occur due to abnormal protein production levels (28–31), and many traits of organisms, such as crops, are dependent on appropriate amounts of protein synthesis (32–34). Therefore, fine-tuning and customizing gene expression are necessary for the treatment of some diseases and agricultural breeding. However, gene expression cannot be quantitatively controlled to specific levels through gene knockout, RNAi or gene overexpression. Gene expression in plants has been fine-tuned using CRISPR/Cas9 [clustered regularly interspaced palindromic repeats (CRISPR)/CRISPR-associated protein 9], base editing or prime editing systems that edit the upstream open reading frames (uORFs) or generate *de novo* uORFs of genes (35–37). In addition, Rodríguez-Leal *et al.* developed a promoter editing approach to produce a wide range of gene expression levels and generate crops with quantitative trait variations (38). These strategies could bidirectionally regulate the expression of target genes, while they could not quantitatively control the protein expression to a specific level.

Translation is an important gene expression process and is responsible for synthesizing proteins from mRNAs. Controlling this process is a direct method that can be used to change the cellular contents of encoded proteins to maintain homeostasis. Growing evidence has shown that multiple *cis*-elements in the 5'-untranslated regions (UTRs) of genes, such as Kozak sequences and uORFs, have significant impacts on gene translation (39–41). The Kozak sequence is located near the initiation site of translation and significantly influences the performance of ribosomes in correctly identifying the initiation codon and starting protein synthesis (42). Different Kozak variants give rise to different amounts of protein synthesis, and thus modifications of Kozak sequences may have regulatory effects on translation efficiencies (43). In recent years, precise genome editing tools have undergone tremendous development, and base editors can achieve changes in single base accuracy (44–46). Newly established prime editors can achieve all types of base replacements, small DNA fragment deletions and insertions, and combinatorial mutations (47), thus facilitating the efficient manipulation of Kozak sequences.

Here, we proposed and validated a novel strategy that fine-tunes gene translation by specifically customizing the three nucleotides upstream of the translation initiation codon, part of the Kozak sequence with the most significant base preference (referred to as KZ₃, KZ as an abbreviation of the Kozak sequence and the number 3 as an abbreviation

of the three nucleotides upstream of the translation initiation codon), with base editors and prime editors. This strategy, referred to as KZ₃-edit, can be applied to a gene's primary ORF (pORF) and uORF to quantitatively down- or up-regulate protein expression to the desired levels at the translation stage.

MATERIALS AND METHODS

Animals

New Zealand White rabbits used in this study were purchased from Guangdong Medical Laboratory Animal Center (Foshan, China) and maintained at the rabbit facility (clean grade) of Guangzhou Institutes of Biomedicine and Health (GIBH), Chinese Academy of Sciences. The protocols for the use of rabbits were approved by the Institutional Animal Care and Use Committees of GIBH (Animal Welfare Assurance#A5748-01).

Plasmid construction

The plasmids pCMV-BE4max, pcDNA3.1-Target-AID and pCMV-PE2-T2A-PURO were obtained from our previous research in our laboratory. The plasmid pCMV-PE2-NG was obtained by partial replacement of the Cas9 sequence of pCMV-PE2 (addgene, Plasmid#132775) with a DNA fragment containing H840A, L111R, D1135V, G1218R, E1219F, A1322R, R1335V and T1337R mutations, which was amplified from the plasmid SpCas9-NG (addgene, Plasmid#138566). The plasmid pCMV-PE2-NG-T2A-PURO was constructed with a polymerase chain reaction (PCR)-amplified fragment T2A-PURO by the pEASY-UniSeamless Cloning and Assembly Kit (Trans-Gen, CU101). The pEF1 α -EGFP-CMV-mCherry dual-fluorescence expression plasmid was constructed by combining elongation factor 1 α (EF1 α)-enhanced green fluorescent protein (EGFP) and cytomegalovirus (CMV)-mCherry PCR-amplified DNA fragments and cloning into a mammalian expression vector with the ClonExpress MultiS One Step Cloning Kit (Vazyme, C113). A sequence containing two BpiI sites was set in front of the target regulatory gene in the reporter plasmid. To add a specific nucleotide sequence upstream of the initiation codon ATG of the EGFP ORF, two complementary oligo-primers containing 4 bp overhangs or PCR-amplified DNA fragments including the 5'-UTR of a specific gene were cloned into the Bpi I-digested pEF1 α -EGFP-CMV-mCherry cloning vector by ligating reaction with Solution I (Takara, 6022) or assembly reaction with the pEASY-UniSeamless Cloning and Assembly Kit (TransGen, CU101), respectively. The PCR was performed using Q5 High-Fidelity 2 \times Master Mix (New England Biolabs, M0492S). Plasmids expressing single guide RNAs (sgRNAs) or prime editing guide RNAs (pegRNAs) were constructed by ligating the annealed oligonucleotides into BpiI-digested acceptor vector under a U6 promoter with Solution I (Takara, 6022). All constructed plasmids were validated via Sanger sequencing by Guangzhou IGE Biotechnology. All primers were synthesized by Genewiz and are listed in Supplementary Table S1.

Cell culture and transfection

HEK293 cells or NIH3T3 cells were cultured and passaged in Dulbecco's modified Eagle's medium (DMEM; HyClone, SH30243.01) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco, 10270-106). Rabbit fetal fibroblasts (RFFs) or porcine fetal fibroblasts (PFFs) were cultured in DMEM supplemented with 15% (v/v) FBS (Gibco, 10099-141C) and 1× penicillin/streptomycin (Gibco, 15140122) with 1% non-essential amino acids (NEAs; Gibco, 11140050), 2 mM GlutaMAX (Gibco, 35050061) and 1 mM sodium pyruvate (Gibco, 11360070). HEK293 cells and NIH3T3 cells were maintained in a cell incubator at 37.5°C and 5% CO₂, and RFFs and PFFs were maintained at 38.5°C and 5% CO₂. For electroporation, the plasmids were delivered to HEK293 cells through the NeonTM transfection system (Life Technology) at 1110 V, 20 ms and two pulses, and to RFFs or PFFs at 1350 V, 30 ms and one pulse. For polyethylenimine (PEI)-based transfection, HEK293 or NIH3T3 cells were seeded on 48-well poly-D-lysine-coated plates (Corning). After 12–24 h, the cells were transfected at 60–80% confluence with 1.5 µg of PEI (Sigma-Aldrich, 408727) and suitable plasmids according to the manufacturer's protocols. Two days after transfection, the transfected cells were collected and used as samples for subsequent PCR for the detection of gene editing efficiency, or for flow cytometry for analysis of the expression of target genes.

Flow cytometry

For analyzing the expression of EGFP and mCherry for the dual-fluorescence system in HEK293 cells, NIH3T3 cells, RFFs and PFFs, cell samples were prepared by using 0.25% trypsin to dissociate adherent cells and then washed and resuspended with 1× phosphate-buffered saline (PBS; GENOM). High-quality samples were finally analyzed with a Fortessa cytometer (BD Biosciences). The intensity ratio of EGFP to mCherry was used to represent the translation efficiency of corresponding KZ₃ variants. The preparation and flow cytometry of the HEK293-EGFP cell line are similar to that for reporter detection. The average EGFP intensity is considered as the translation efficiency. The data are subsequently analyzed by FlowJo_V10.

Cell colony screening

HEK293-EGFP cells containing a single copy of an EGFP-expressing cassette in the human *ROSA26* locus were described in our previous study (48). For modifying the KZ₃ sequence of EGFP or endogenous genes, 2 µg of sgRNA-expressing or pegRNA-expressing plasmids and 6 µg of correlated gene-editing tool vectors were co-electroporated into HEK293-EGFP or HEK293 cells under the condition of 1120 V, 20 ms and two pulses by using the Neon Transfection System (Life Technology). After 24 h of recovery, the HEK293 cells were dissociated and seeded on 10 cm culture dishes. For cells co-electroporated with pCMV-PE2-T2A-PURO or pCMV-PE2-NG-T2A-PURO, 2 µg/ml puromycin were added to the culture medium for the selection of positive cells. After ~4–5 days, single-cell-derived colonies were picked and analyzed.

Genomic DNA extraction and genotyping

Whole genomic DNAs were extracted by using two approaches: (i) for HEK293 cells, a small fraction of cells was collected in 10 µl of lysis buffer (0.45% NP-40 plus 0.6% proteinase K) for 60 min at 56°C and 10 min at 96°C; and (ii) for newborn rabbits, the genomic DNAs of ear tissues were extracted using a TIANamp Genomic DNA kit (TIANGEN, DP304-03) in accordance with the manufacturer's instructions. Cell lysates or extracted DNAs were then harvested as PCR templates. The PCR products were directly sent for Sanger sequencing. Primers used for PCR are listed in Supplementary Table S1.

Editing efficiency detection and analysis by Sanger sequencing

To detect the primary efficiencies of gene editing in the editing region, we used specific primer pairs of the target region and 2 µl of cell lysates or 100 ng of extracted DNAs to amplify DNA fragments containing the edited site. PCR was performed by using 2× Rapid Taq Master Mix (Vazyme, P222-03) under the conditions of 95°C for 5 min; 35 cycles of 95°C for 15 s, 58°C for 15 s, 72°C for 15 s; 72°C for 5 min as a final extension; and 12°C for hold. Editing efficiency was calculated from Sanger sequencing results by using EdiTR (https://moriaritylab.shinyapps.io/editr_v10/) (49).

Amplicon deep sequencing and data analysis

For detailed genotyping of screened cell colonies, *PCSK9*-edited and *U2HR*-edited newborn rabbits, we performed amplicon deep sequencing to further define the frequencies of target mutations. The target sites were initially amplified through PCR with corresponding site-specific primer pairs from cell lysates of the transfected cells or extracted DNAs of the newborn rabbit tissues by using 2× Phanta Max Master Mix (Vazyme, P222-03) following the thermal cycler conditions: 95°C for 3 min; 20 cycles of 95°C for 10 s, 55°C for 10 s and 72°C for 10 s; 72°C for 5 min as a final extension; and 12°C for hold. Then, the PCR products were used as templates for subsequent amplification with different index primers of different samples under the condition of 95°C for 3 min; 30 cycles of 95°C for 10 s, 55°C for 10 s, and 72°C for 10 s; 72°C for 5 min as a final extension; and 12°C for hold by 2× Phanta Max Master Mix (Vazyme, P222-03). All the amplified PCR products were electrophoresed in 1.5% agarose gels with 1× TAE buffer, and the target bands were cut for extraction with the HiPure Gel Pure DNA Mini Kit (Magen, D1001-03). The concentration of purified products was quantified with an IPure Qubit dsDNA HS assay kit (IGE Biotech, IGE2019052901). Equivalent amounts of purified products were mixed for the preparation of a DNA library, and the library was sent to NOROAD (Beijing) for amplicon deep sequencing with a NovaSeq platform. The protospacer adjacent motifs (PAMs) in the reads were investigated to identify the expected genotypes and indels. The presented ratio was calculated by comparing individual reads with whole reads. The amplicons were sequenced in three of the repeated assays for each target site. Amplicon reads with a quality score of <30 were filtered.

CRISPResso2 (version 2.2.8) was used for analyzing and visualizing the amplicon results (50).

Real-time fluorescence quantitative PCR (q-PCR)

Total RNA was extracted from cell colonies or animal tissues by using the RaPure Total RNA Micro Kit (Magen, R4012-03) in accordance with the manufacturer's instructions. Reverse transcription was performed using ReverTra Ace qPCR RT Master Mix with gRNA Remover (TOYOBO, FSQ-301) or the PrimeScript RT Reagent Kit with gDNA Eraser (Takara, RR047A) according to the manufacturer's protocol. By using cDNAs diluted 10-fold as templates, q-PCR was conducted using the TB Green PCR Master Mix (Takara) in Bio-Rad CFX96 in a 10 μ l reaction volume according to the recommended protocols. Data were analyzed using the $2^{-\Delta Ct}$ method. The primers used in q-PCR are listed in Supplementary Table S1.

Bioinformatics analysis of the genome and transcript

To obtain the sequences around the initiation codon, we downloaded the human genome and transcriptome from the NCBI website (GCF000001405.39) as a reference for downstream analysis. To analyze the Kozak sequences in the genome, we first extracted the initiation codon position in the genome from the GTF file and extended 25 bp upstream and downstream of the initiation codon, combining each corresponding gene name and transcript to form a bed format file. Transcripts with identical initiation codons were then simplified using a custom R script to prevent multiple counting of Kozak sequences. The sequences around the initiation codon region (± 25 bp) were extracted from the genome using bedtools (version 2.28.0) and stored in fasta format for subsequent data statistics and visualization. We first extracted the initiation codon ATG and the following seven bases (a total of 10 bases), and corresponding protein ID and gene ID from the CDS file by custom shell command. The corresponding transcript IDs of the protein ID were then extracted from the GTF file, and the information including the transcript ID, gene name, protein ID and the sequences of ATG + 7 bp was saved to a local file. We used fasta2tab.pl to convert the transcriptome reference into a table format, and then read it into the R environment, using the transcript ID and the sequences of ATG + 7 bp as the query pattern to match and extract the sequences around ATG (± 20 bp) for subsequent data statistics and visualization. To analyze the Kozak sequences in the uORFs of the transcriptome, we extracted all the cDNA sequences before the initiation codon from the transcriptome reference, counted the number of ATGs in these sequences and extracted all the sequences around ATG (± 25 bp) in these sequences, just as for analyzing pORFs. Finally, all the statistical analyses and visualization of the genome and transcriptome were performed through custom R scripts.

Western blotting

The cells were collected and lysed in Pier RIPA Buffer (Thermo Fisher) supplemented with the inhibitor of proteinase K (100 mM phenylmethylsulfonyl fluoride, PMSF)

on ice for 30 min and then clarified by centrifugation at 14 000 rpm, 4°C for 10 min to generate a supernatant. Skin tissues collected from U2-1, U2-2 and wild-type (WT) rabbits, and liver tissues collected from PCSK9-1#, PCSK9-2#, PCSK9-3# and WT rabbits, were used to extract the total protein with the MinuteTM total protein extraction kit (for animal cultured cells and tissues) (Invent Biotechnologies, SD-001/SN-002) according to the manufacturer's protocol. The total protein concentrations of the above samples were determined with the Enhanced BCA Protein Assay Kit (Beyotime, P0010), and samples were boiled with sodium dodecylsulfate (SDS) loading buffer (62.6 mM Tris-HCl, 10% glycerol, 0.01% bromophenol blue and 2% SDS at pH 6.8) at 100°C for 15 min. Equal amounts of proteins were resolved by 10% SDS-polyacrylamide gel electrophoresis (PAGE) and then transferred to polyvinylidene fluoride membranes (Millipore). The membranes were blocked with 5% skim milk in TBST (Tris-buffered saline with Tween) for 2 h and then blotted with primary antibodies against TP53 (proteintech, 10442-1-AP, 1:2000 dilution), SMN1/2 (ZEN BIO, 383126, 1:2000 dilution), SOD1 (abcam, ab79390, 1:1000 dilution), RNASEH1 (ABclonal, A9116, 1:2000 dilution), SFXN3 (proteintech, 15156-1-AP, 1:2000 dilution), HR (OriGene, TA806882, 1:1000 dilution) and proprotein convertase subtilisin/kexin type 9 (PCSK9; Bioss Inc, bs-6060R, 1:1000 dilution) for 2 h at room temperature, followed by a horseradish peroxidase (HRP)-conjugated secondary antibody (Yeason, 33101ES60, 33201ES60; 1:2500 dilution) for 1 h at room temperature. Blots were developed with a Super ECL Detection Reagent (Yeasen, 36223ES60) and imaged on the Chemiluminescent Imaging System (SAGECREATION, minichemi601). In addition, the membranes were re-probed with a primary glyceraldehyde phosphate dehydrogenase (GAPDH) antibody (Affinity, AF7021; 1:3000 dilution) or a β -actin antibody (abcam, ab8227; 1:2000 dilution) and HRP-conjugated secondary antibody (Yeason, 33101ES60; 1:2500 dilution) as the loading control.

To evaluate the protein synthesis of screened cell colonies with KZ₃ manipulations of endogenous genes, the gray analysis of western blotting results was further performed to present the changes in translation. The gray levels of target regulated proteins and endogenous reference proteins were calculated by using the ImageJ tool. The gray ratio of the target protein compared with the reference protein was used as the amount of target protein synthesis.

Gene editing rabbit construction

Transcription of RNA in vitro. The pCMV-BE4max vectors containing an SP6 promoter in front of the coding sequence were linearized by the restriction endonuclease NotI and transcribed *in vitro* for the synthesis of cytosine base editor (CBE) mRNAs using the mMESSAGE mMACHINETM SP6 Transcription Kit (Thermo Fisher Scientific, AM1340). By using the specific forward primers T7-R-PCSK9-F or T7-R-U2HR-F with the reverse primer T7-template-R, sgRNA templates for *in vitro* transcription were amplified by PCR with the template of U6-sgRNA cloning vectors and then transcribed using the HiScribeTM T7 Quick High Yield RNA Synthesis Kit (New England

Biolabs, E2050S). The CBE mRNAs and sgRNAs were purified with the RNeasy MiniElute Cleanup kit (Qiagen, 74204) according to the manufacturer's instructions. The purified mRNAs of BE4max and sgRNA were diluted to 1000 and 300 ng/μl with RNase-free water, respectively, and stored at -80°C. The primers used for amplifying the templates of sgRNAs are listed in Supplementary Table S1.

Embryo microinjection and embryo transfer. Procedures of pronuclear-stage embryo microinjection and embryo transfer were performed as previously described (51). In brief, 6- to 8-month-old female New Zealand White rabbits were superovulated with 80 IU of pregnant mare serum gonadotropin (Ningbo Shusheng), mated with male rabbits and injected with 100 IU of human chorionic gonadotrophin (Ningbo Shusheng). After 18 h, the injected rabbits were euthanized, and the oviducts were flushed with 5 ml of warmed Dulbecco's phosphate-buffered saline-bovine serum albumin for the collection of pronuclear-stage embryos, which were then transferred to culture medium for microinjection. A mixture of BE4max mRNAs (150 ng/μl) and sgRNAs (100 ng/μl) was microinjected into the cytoplasm of one-cell stage zygotes. The injected embryos were then transferred to Earle's balanced salt solution medium at 38.5°C in 5% CO₂ for 30–60 min. Finally, the injected embryos were transferred into the oviduct of the surrogate recipient female rabbits.

Blood lipid testing

Beginning at 1 month of age, 1 ml of blood was collected once a week for 3 weeks from the ear vein of the rabbits with the PCSK9 Kozak mutation and the WT control rabbits of the same age. The collected blood samples were treated without anticoagulation and left at room temperature for 30 min. Centrifugation was performed at 3000 rpm for 15 min. Supernatant plasmas were collected for total cholesterol (CHO), triglyceride (TG), high-density lipoprotein-cholesterol (HDL-C) and low-density lipoprotein-cholesterol (LDL-C) detection. The testing experiments were conducted by Guangdong Laboratory Animal Monitoring Institute.

Histological analysis

Skin tissues on the back of 2-week-old MUHH (Marie Unna hereditary hypotrichosis) rabbits and age-matched WT rabbits were collected. The collected tissues were fixed with 4% paraformaldehyde for 48 h, embedded in paraffin and sectioned. The tissue sections were stained with hematoxylin and eosin (H&E), and slides were scanned with an Aperio ScanScope. For analyzing the number of hair follicles for different rabbit skin tissues, five regions of each represented tissue slide were selected, and the follicle numbers at 1 mm distance were individually counted. The relative numbers of follicles were applied to represent the hair distribution level.

Statistical analysis

Data were statistically analyzed using Excel and GraphPad Prism v.8.0. The editing frequencies of cell mixes or cell

colonies for the target sites were analyzed with an online tool (EditR 1.0.10) with the representative Sanger sequencing results. The expected editing frequencies and indels were calculated from the data of amplicon deep sequencing. The numerical values of data were presented as means ± standard deviation (SD) or means ± standard error of the mean (SEM). The significant difference analysis of data was performed with two tailed Student's *t*-test or one-way analysis of variance (ANOVA). *P* ≤ 0.05 was considered significant (**P* ≤ 0.05, ***P* ≤ 0.001, ****P* ≤ 0.0001).

RESULTS

Design and justification of the KZ₃-edit strategy for altering the translation levels of genes by customizing the upstream nucleotides of the translation initiation codon

The Kozak consensus sequence with optimal gene translation initiation is 5'-GCCGCC^A/GCCATGG-3' (the underlined bold sequence is the initiation codon) (43). However, in the intrinsic genes of an organism, the sequences of the Kozak region are actually diverse, giving rise to variations in the translation efficiency of corresponding mRNAs (52,53). Therefore, we hypothesized that customization of the Kozak sequences of specific genes can be implemented to achieve the pre-determined regulatory goal based on the preference of Kozak sequences. To justify this hypothesis, we first performed a statistical analysis of the sequences in regions near the translation initiation codon (ATG) of all human protein-coding transcripts. The ratios of each base of the 20 nucleotides before and behind the initiation codon in the Kozak region are illustrated in Figure 1A. The statistical results showed that the combinatorial sequence of the most preferred bases at the initiation site of translation (-9 to +4, the ATG was set at +1 to +3) was 5'-GCCGCCACCATGG-3', consistent with the previously defined optimal Kozak consensus sequence (43). Apart from the ATG, we found that the regions with the most significant base preferences were located from positions -3 to -1 and +4 (Figure 1A). These regions have been proven to play an important regulatory role in the initiation of gene translation (54), and thus these regions can have the greatest potential as target sites using genome editing to influence gene translation. Given that the mutation of the nucleotide at the +4 position would alter the corresponding amino acids and may thus interfere with the normal function of a target protein, we set the three nucleotides from -3 to -1 (referred to as KZ₃) as the editing region to alter the translation levels of proteins (Figure 1A). To verify the effect of mutating the KZ₃ sequence on gene translation, we constructed a dual fluorescence reporter system in which the expression of mCherry fluorescent protein was driven by the CMV promoter and used as an internal control, and EGFP driven by the EF1 α promoter was used as the target gene (Figure 1B). A total of 64 KZ₃ variants with different combinations of the triple nucleotides immediately in front of the initiation codon of the EGFP were designed, and 64 reporter plasmids were constructed. In addition, the 5'-UTR sequences of endogenous human *TP53* and *SMN* genes were inserted downstream of the EF1 α promoter and upstream of the initiation initiation codon of EGFP (Figure 1B). Similarly, 64 reporter plasmids with different KZ₃

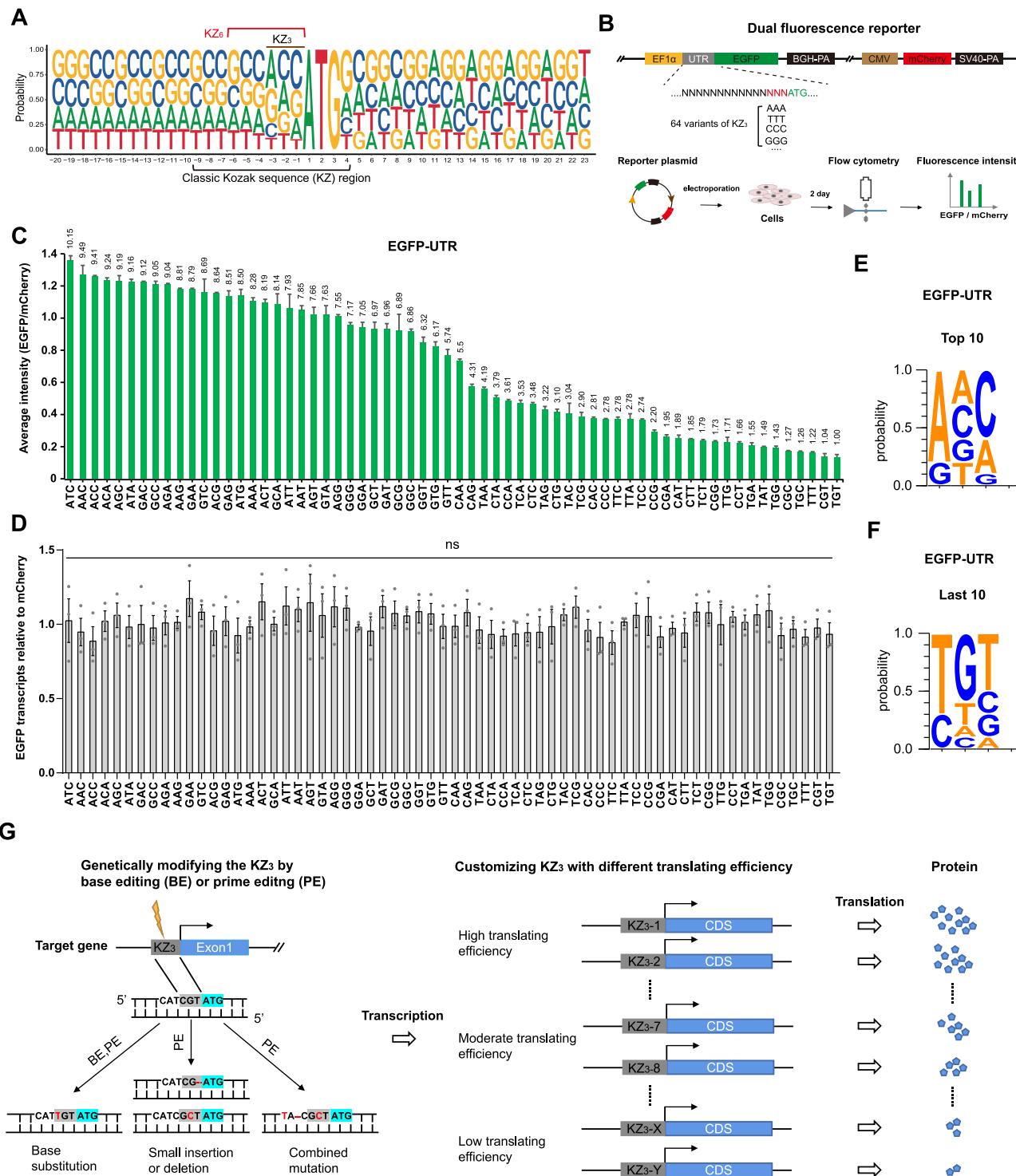


Figure 1. Establishment of the KZ₃-edit strategy to regulate gene translation based on Kozak manipulation. **(A)** The distribution of nucleotides near the initiation codon of all human protein-coding gene transcripts. Twenty nucleotides before and behind the initiation codon were analyzed. KZ₃ and KZ₆ represent the regions of three nucleotides and six nucleotides in front of the initiation codon, respectively. The classic Kozak sequence region includes the nucleotides from position -9 to +4, in which the ATG was set at the +1 to +3 positions. **(B)** Diagram of the dual-fluorescence reporter to test the effect of KZ₃ variants on translation. **(C)** Translating efficiency analysis of all 64 KZ₃ variants in EGFP-UTR with dual-fluorescence reporters. The average intensity ratios of EGFP to mCherry presented the translating efficiencies of the indicated KZ₃ variants ($n = 3$). **(D)** Transcription analysis of EGFP relative to mCherry in 64 reporters of KZ₃ by q-PCR ($n = 3$). **(E and F)** The probabilities of different nucleotides for the top 10 (**E**) and the last 10 (**F**) ranked KZ₃ variants of EGFP-UTR. **(G)** The diagram of strategy to regulate gene translation to high, moderate or low levels by modifying the KZ₃ sequence with base editors and/or prime editors, referred to as the KZ₃-edit strategy. Exon1, the first exon of the pORF. All values are represented as means \pm SEM with three experimental replicates. One-way ANOVA was used to analyze the statistical difference of transcription levels among the reporter KZ₃ variants. ns, no significant difference.

variants of the 5'-UTR of human *TP53* or *SMN* genes were also constructed. These constructed plasmids were individually transfected into HEK293 cells. Two days after transfection, the cells were collected for fluorescence intensity analysis through flow cytometry. The average intensity ratio of EGFP to mCherry was used as an indicator of the corresponding translation efficiency of each KZ₃ variant (Figure 1B). After conducting statistical analysis, we ranked the translation efficiencies of the 64 KZ₃ variants. As shown in Figure 1C and Supplementary Figure S1A and C, the KZ₃ variants differed in their translation efficiencies. Variants with A/G at the -3 nucleotides of KZ₃ had significantly higher translation efficiencies than those with C/T (Figure 1C, E, F; Supplementary Figure S1A, C, E–G). The average intensity ratios of EGFP to mCherry ranged from 0.13 to 1.36, 0.34 to 1.69 and 0.22 to 1.72, with maximum differences of 10.15-, 4.95- and 7.86-fold for EGFP-UTR, *TP53*-UTR and *SMN*-UTR, respectively (Figure 1C; Supplementary Figure S1A, C). The 64 ranked KZ₃ variants provided an indicator for predicting the mutation patterns in KZ₃ sequences and achieving the desired expression level of a specific protein. Furthermore, no significant differences in transcription levels were found among the KZ₃ variants, as confirmed by q-PCR (Figure 1D; Supplementary Figure S1B, D). Therefore, KZ₃ sequences are ideal targets for genome editing and obtaining the desired gene translation level. Although no obvious preferences for the bases at the -2 and -1 positions of KZ₃ were found, their combinations can induce fine-tuning effects (Figure 1C; Supplementary Figure S1A, C). Five KZ₃ variants (e.g. ATC, GAA, GTT, TAC and TGT of the EGFP-UTR) with different regulatory translational effects in human cells were further selected for confirmation in three other mammalian cell lines, namely a mouse embryo fibroblast cell line (NIH3T3), RFFs and PFFs. The results showed that KZ₃ variants with ATC and GAA sequences had robust regulatory translational effects, while KZ₃ variants with TAC and TGT sequences had weak regulatory translational effects, corresponding to the results in human cells (Supplementary Figure S2A). In addition, no significant differences in transcription levels were found for these five KZ₃ variants in these tested mammalian cells (Supplementary Figure S2B). Therefore, we proposed a novel strategy, KZ₃-edit, for regulating gene translation to desired levels in mammalian systems by customizing the KZ₃ sequence with base editors or primer editors (Figure 1G).

Validation of the KZ₃-edit strategy for altering the translation levels of genes by using base editors and prime editors

We then further verified whether *in situ* Kozak sequence modification at the local genome could effectively alter gene translation levels. An HEK293-EGFP cell line (48), in which a single copy of the EGFP-expressing cassette was knocked in downstream of the *ROSA26* promoter, was previously established (Figure 2A). A CBE, Target-AID, which has an editing function of C-to-T substitution as well as small amounts of C-to-A and C-to-G substitutions (55), was used to modify the EGFP KZ₃ sequence. The EGFP-PreATG-sgRNA targeting the -3C of the EGFP KZ₃ sequence and Target-AID were trans-

fected into HEK293-EGFP cells. Two days after transfection, the cells were collected for flow cytometry analysis, and non-transfected HEK293-EGFP cells were used as the control. HEK293-EGFP cells transfected with Target-AID + EGFP-PreATG-sgRNA presented distinct cell populations with different EGFP expression intensities, which were classified into EGFP-low, EGFP-medium and EGFP-high (Supplementary Figure S3A). The three cell populations with different EGFP expression intensities were separately collected by flow cytometry, and the gene editing status was evaluated. The Sanger sequencing results showed that ~48% of the cells in the EGFP-low population contained C-to-T substitutions at the target site, and 65% and 17% of cells in the EGFP-high population had C-to-G and C-to-A substitutions, respectively, whereas most of the cells in the EGFP-medium population, in which the green fluorescence intensity is similar to that of the non-transfected HEK293-EGFP cells, presented no mutations (Supplementary Figure S3B, C). Subsequently, we screened the EGFP-high and EGFP-low cell populations and obtained single-cell colonies with the KZ₃ sequences of CGT (WT) ($n = 6$), TGT ($n = 4$), GGT ($n = 7$) and AGT ($n = 4$) (Figure 2B; Supplementary Figure S3D). Via fluorescence microscopy, we found that the cell colonies with the TGT KZ₃ sequence had weaker fluorescence and that the cell colonies with GGT and AGT KZ₃ had significantly enhanced fluorescence compared with the WT cells with the CGT KZ₃ sequence (Figure 2B). These cell lines were further expanded to determine the relationship between the EGFP fluorescence intensity and transcription level. Flow cytometry analysis showed that when the -3C of the KZ₃ sequence became G and A, the average EGFP intensities of the cells were increased to 5.58- and 8.86-fold that of the original cell line, respectively. When the -3C was mutated to T, the EGFP intensity of the corresponding cell colony decreased to 0.57-fold that of the original intensity (Figure 2C, D). However, the q-PCR analysis showed that each single-nucleotide mutation in -3C above resulted in no changes in the EGFP at the transcriptional level (Figure 2E). These results suggested that KZ₃ *in situ* mutations mediated by the base editor can effectively regulate the gene translation efficiency, and even a single-base substitution can achieve significant changes in the synthesis of proteins of interest without affecting the transcription levels of genes.

Base editing can efficiently change single or multiple bases (56), but the editing patterns of KZ₃ are relatively limited. To further increase the diversity of mutation patterns, a prime editor was utilized to engineer the KZ₃ sequence. Five pegRNAs, namely EGFP-pegRNA1, 2, 3, 4 and 5, were designed and used to target the EGFP KZ₃ sequences, which achieved del T, G>T, GT>CC, del GT & ins AAC and del GT & ins CAAC mutations, respectively (Figure 2F; Supplementary Figure S4A). We transfected HEK293-EGFP cells with the PE2 or PE3 system of each pegRNA for cell colony screening and obtained six corresponding genotypes of interest (Supplementary Figure S4B, C), which would influence the sequences of KZ₃ or KZ₆. Compared with the WT cell line, the EGFP expression in all of the edited cell colonies with the various mutations of KZ₃ or KZ₆ sequences were up-regulated (Figure 2G, H). The average fluorescence intensities of the

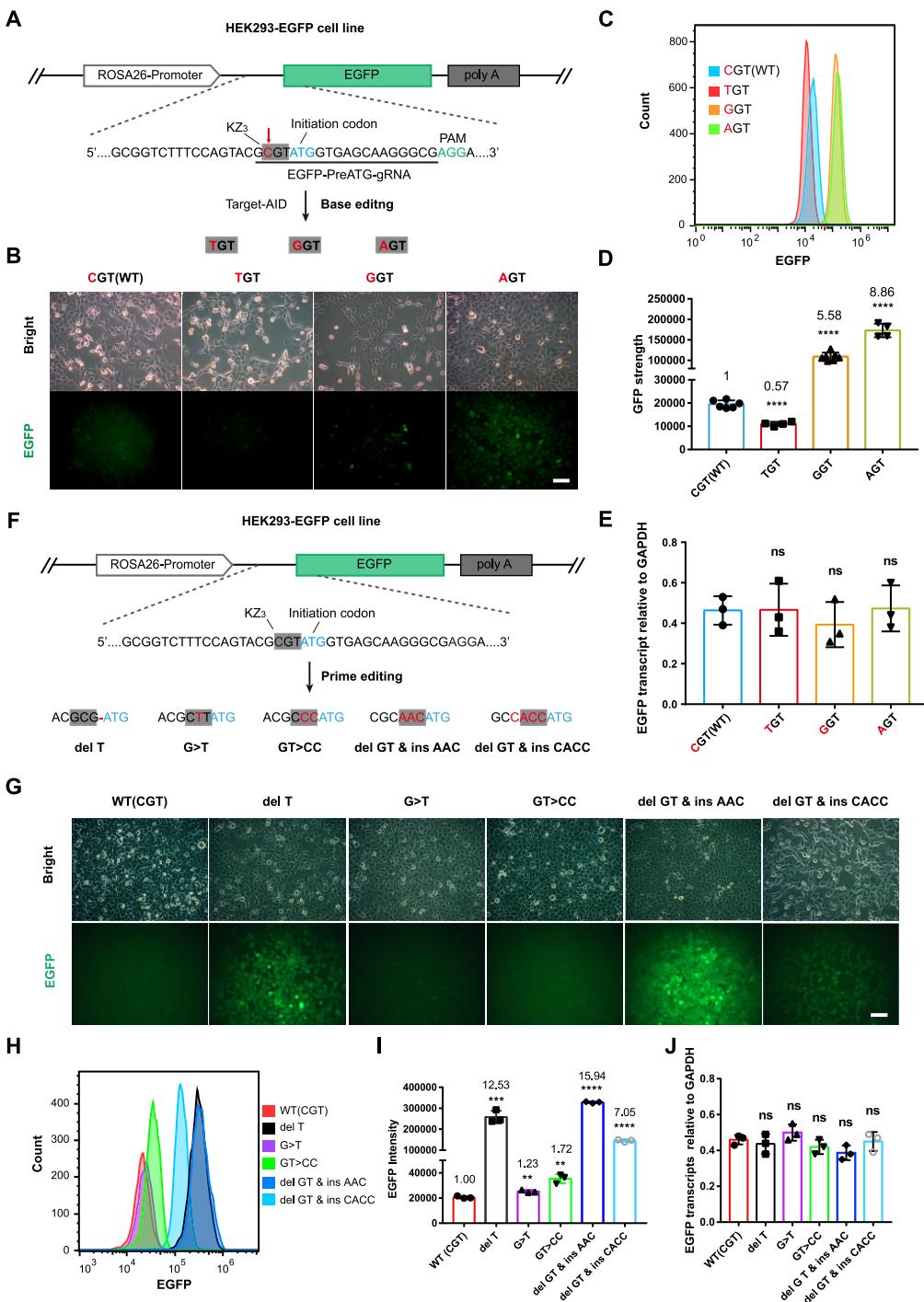


Figure 2. Efficient regulation of gene translation by modifying the KZ₃ sequence with a base editor or prime editor. (A) Diagram of the design of KZ₃ modification by base editing in the HEK293-EGFP cell line. The target nucleotide indicated by a red arrow was the -3C in front of the ATG of the EGFP, which can be substituted to T, G and A by a CBE, Target-AID. The initiation codon is blue and the PAM sequence is in green. The sequence underlined with a black line is the spacer of EGFP-PreATG-gRNA. (B) Representative fluorescence images of these four single-cell-derived colonies with different KZ₃ sequences. Stronger fluorescence indicates higher efficiency of EGFP translation. Scale bar, 50 μ m. (C) Fluorescence intensities of these four types of colonies were analyzed by flow cytometry. (D) Statistical analysis of average fluorescence intensities of four types of colonies with CGT (WT) ($n = 6$), TGT ($n = 4$), GGT ($n = 7$) and AGT ($n = 4$). The average EGFP intensity of the original cell line was normalized as 1.00. The numbers on the columns show the fold changes. (E) The transcript levels of different cell colonies with varied KZ₃ or KZ₆ modifications that included base substitutions, deletions and/or insertions. (F) Schematic of prime editing to flexibly create multiple types of KZ₃ or KZ₆ modification that included base substitutions, deletions and/or insertions. (G) Representative fluorescence images of six cell colonies with different KZ₃ or KZ₆ manipulations. Scale bar, 50 μ m. (H) The results of flow cytometry to show the fluorescence intensities of six representative colonies. (I) Statistical analysis of average fluorescence intensities for six types of cell colonies by flow cytometry (three colonies with each KZ₃ or KZ₆ variant were analyzed). (J) The transcript levels of different cell colonies with varied KZ₃ or KZ₆ modifications. All statistical data are shown as the mean \pm SD; statistical significance were determined in comparison with the WT as follows: ns, no significant difference; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ by two-tailed Student's *t*-test.

mutant cell colonies with del GT & ins AAC ($n = 3$), del T ($n = 3$), G>T ($n = 3$), GT>CC ($n = 3$) and del GT & ins CAAC ($n = 3$) were 15.94-, 12.53-, 1.23-, 1.72- and 7.05-fold that of the WT (CGT) cell colonies ($n = 3$), respectively (Figure 2I). These changes in the small fragments of the Kozak sequence did not significantly affect the transcription levels of target genes (Figure 2J). These results revealed that a prime editor can create more abundant mutation modes of Kozak sequences and achieve more flexible translation regulation effects than base editors with the KZ₃-edit strategy.

Anticipated alteration of protein expression levels at the translation stage via direct modification of the KZ₃ sequences of endogenous genes

We explored the feasibility of applying the KZ₃-edit strategy to human endogenous genes for translation regulation. We selected three endogenous genes, namely *TP53*, *SMN* (two copies, named *SMN1* and *SMN2*, respectively) and *SOD1*, as the target genes of interest. The KZ₃ sequence GCC of the *TP53* gene was mutated to TTT, GAA, TGT insertion (insTGT) and TGCC deletion (delTGCC) by the prime editor, which was predicted to decrease, increase, decrease and decrease the translation efficiency, respectively (Figure 3A; Supplementary Figure S5A), whereas the KZ₃ sequence GCT of the *SMN* gene was altered to AAC by the prime editor, which was predicted to slightly increase the translation efficiency (Supplementary Figures S5B and S6A). In addition, we induced mutations in the KZ₃ sequences of *SOD1* with the CBE, Target-AID. Due to the bystander effect of base editing, KZ₃ sequence conversion of GTT to ATT or CTT as well as the whole KZ₆ sequence conversion of CGAGTT to CAAATT or CAACTT would be generated, which were predicted to slightly increase, increase, decrease and decrease the translation rate of *SOD1*, respectively (Figure 3F). To verify whether these KZ₃ variants could cause the predicted regulatory translational effects, the corresponding dual fluorescence reporter vectors containing these KZ₃ variants in front of the initiation codon were constructed and electroporated into HEK293 cells (Supplementary Figure S5C, F). As predicted, the average intensity ratio of EGFP to mCherry suggested that KZ₃ variants with TTT, insTGT and delTGCC sequences of the human *TP53* gene had decreased translational effects, while the KZ₃ variants with GAA sequences had increased translational effects (Supplementary Figure S5D). For the human *SOD1* gene, KZ₃ variants with ATT sequences had increased translational effects, while KZ₃ variants with CTT sequences had decreased translational effects (Supplementary Figure S5G). No significant differences in transcription levels were observed in these groups with KZ₃ variants (Supplementary Figure S5E, H). Next, we verified whether KZ₃ mutations could lead to regulatory translation effects on endogenous *TP53*, *SMN* and *SOD1* genes. The constructed plasmids expressing the gene editing systems for *TP53*, *SMN* and *SOD1* were electrotransfected into WT HEK293 cells, and a small number of cells were collected after 2 days to detect the editing effects. The Sanger sequencing results of the PCR products covering the

target sites showed that the desired mutations occurred at all three editing target sites with efficiencies ranging from 5% to 53% (Supplementary Figure S5I). Then, we performed monoclonal screening of the remaining cells and obtained the corresponding edited colonies of interest. Twelve cell colonies with *TP53* KZ₃ editing (*TP53*-TTT-14, -23 and -35; *TP53*-GAA-2 and -22; *TP53*-insTGT-4, -16 and -18; and *TP53*-delTGCC-43, -44, -47 and -65), three cell colonies with *SMN* KZ₃ editing (*SMN1*/2-AAC-33, -40 and -54) and four cell colonies with *SOD1* KZ₃ editing (*SOD1*-ATT-17 and -25; and *SOD1*-CTT-13 and -27) were selected for further analysis. As shown in Figure 3B and G and Supplementary Figure S6B, the proportions of targeted cells with different mutations in each colony, as determined by amplicon deep sequencing, ranged from 24.6% to 99.8%. Western blotting and gray ratio tests showed that the protein expression of KZ₃-edited cells was altered to varying degrees compared with those of their corresponding WT counterparts. The protein expression of the *TP53*-TTT-14, -23 and -35 cell colonies, *TP53*-insTGT-4, -16 and -18 cell colonies and *TP53*-delTGCC-43, -44, -47 and -65 cell colonies decreased to 0.20-, 0.35-, 0.61-, 0.63-, 0.63-, 0.60-, 0.84-, 0.65-, 0.56- and 0.37-fold of the original colonies, respectively (Figure 3C, D), while for the *TP53*-GAA-2 and -22 cell colonies, the *TP53* protein expression increased to 1.37- and 1.13-fold that of the WT cells. The *SMN1*/2-AAC-33, -40 and -54 cell colonies with GCT>AAC point mutations increased by 1.58-, 1.95- and 2.25-fold, respectively (Supplementary Figure S6C, D). The *SOD1*-ATT-17 cell colony with CGAGTT>CAAATT and GTT>ATT mutations and the *SOD1*-ATT-25 cell colony with CGAGTT>CAAATT mutations showed only slightly increased translational levels (1.77- and 1.69-fold) compared with the WT, while the *SOD1*-CTT-13 and -27 cell colonies harboring GTT>ATT and/or CGAGTT>CAAATT mutations decreased the translational levels to 0.68- and 0.38-fold, respectively (Figure 3H, I). The results produced by these KZ₃ sequence mutations were consistent with the predictions made by using the expression ranking of KZ₃ variants as the indicator. Noticeably, no significant changes in transcriptional levels in all colonies with mutations were observed (Figure 3E, J; Supplementary Figure S6E). These results suggested that direct editing of the KZ₃ sequence at the local genome can quantitatively control gene translation to predictable levels without affecting transcription.

Generation of rabbit models with low cholesterol by direct editing of the KZ₃ sequence of the *PCSK9* gene

Following the *in vitro* characterization, we investigated whether animal models carrying mutated KZ₃ sequences could influence the translation effects of endogenous genes *in vivo* and lead to the desired phenotypes. *PCSK9*, a therapeutically relevant gene involved in cholesterol homeostasis, was selected as the target gene to generate KZ₃-edited rabbits. One sgRNA targeting the rabbit *PCSK9* gene was designed to modify the Kozak sequence (Figure 4A). *In vitro* transcribed CBE mRNA and R-*PCSK9*-sgRNA were mixed together and microinjected into rabbit zygotes (Supplementary Figure S7A). A total of 38 injected zygotes were

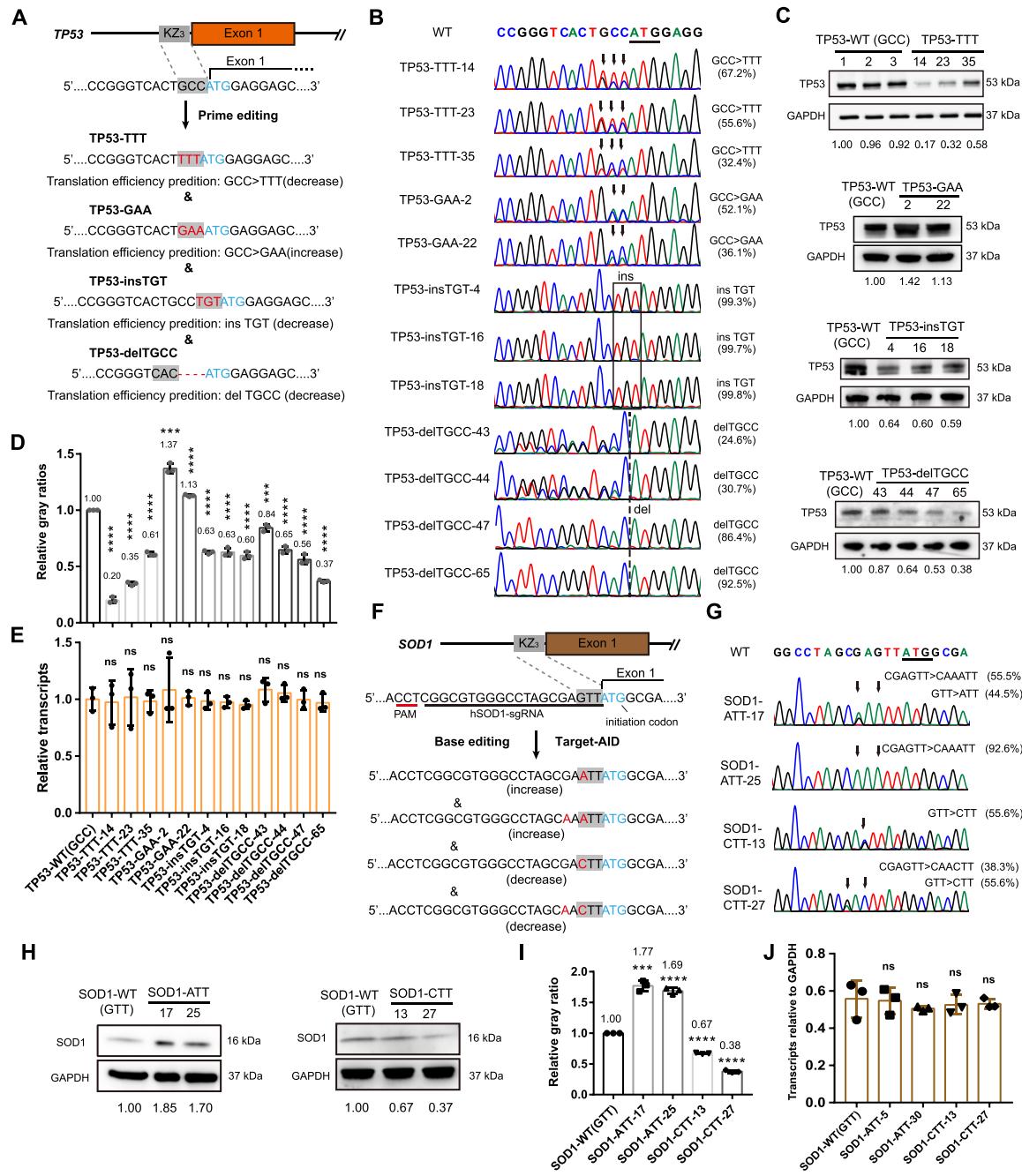


Figure 3. Translating regulation of endogenous genes via direct modification of the KZ₃ sequence. **(A)** The design of prime editing to mutate the KZ₃ sequence of the human TP53 gene. The initiation codon is labeled with blue color. The KZ₃ region is indicated by gray shading. The translation efficiency prediction was based on the dual-fluorescence reporter assay. **(B)** Sanger sequencing results of representative cell colonies with KZ₃ modifications of the human TP53 gene. The black line marks the initiation codon and the arrows indicate the editing nucleotides. The black box shows the inserted nucleotides and the dotted line indicates the deleted nucleotides. The editing efficiencies of different single-cell-derived colonies determined by amplicon deep sequencing are presented in parentheses. **(C)** Representative western blotting results of cell colonies with KZ₃ modifications of the TP53 gene from three experimental replicates. The gray ratio of TP53-WT was normalized as 1.00. **(D)** Statistical analysis of relative gray ratios of cell colonies with KZ₃ modifications of the TP53 gene. **(E)** The analysis of relative transcript levels of cell colonies with KZ₃ modifications of the TP53 gene by q-PCR. **(F)** The design of base editing to mutate the KZ₃ or KZ₆ sequence of the human SOD1 gene. The initiation codon is shown in blue and the KZ₃ region is indicated by gray shading. The underlined sequence is hSOD1-sgRNA and its PAM was labeled by a red line. **(G)** Sanger sequencing results of cell colonies with KZ₃ and/or KZ₆ modification of the SOD1 gene. The black line marks the initiation codon and the arrows indicate the editing nucleotides. The editing efficiencies determined by amplicon deep sequencing are presented in parentheses. **(H)** The representative western blotting results of cell colonies with KZ₃ and/or KZ₆ modification of SOD1 gene to indicate the expression levels of SOD1 proteins. The gray ratio of SOD1-WT was normalized as 1.00. **(I)** Statistical analysis of gray ratios for western blotting results. The data were calculated on the representative results from three independent replicates. **(J)** The relative transcript levels of cell colonies with KZ₃ and/or KZ₆ modification of the SOD1 gene. The data in (D–E) and (I–J) are shown as the mean \pm SD with three experimental replicates. Statistical significances were determined in comparison with the WT as follows: ns, no significant difference; *P < 0.05; **P < 0.01; ***P < 0.001 by two-tailed Student's *t*-test.

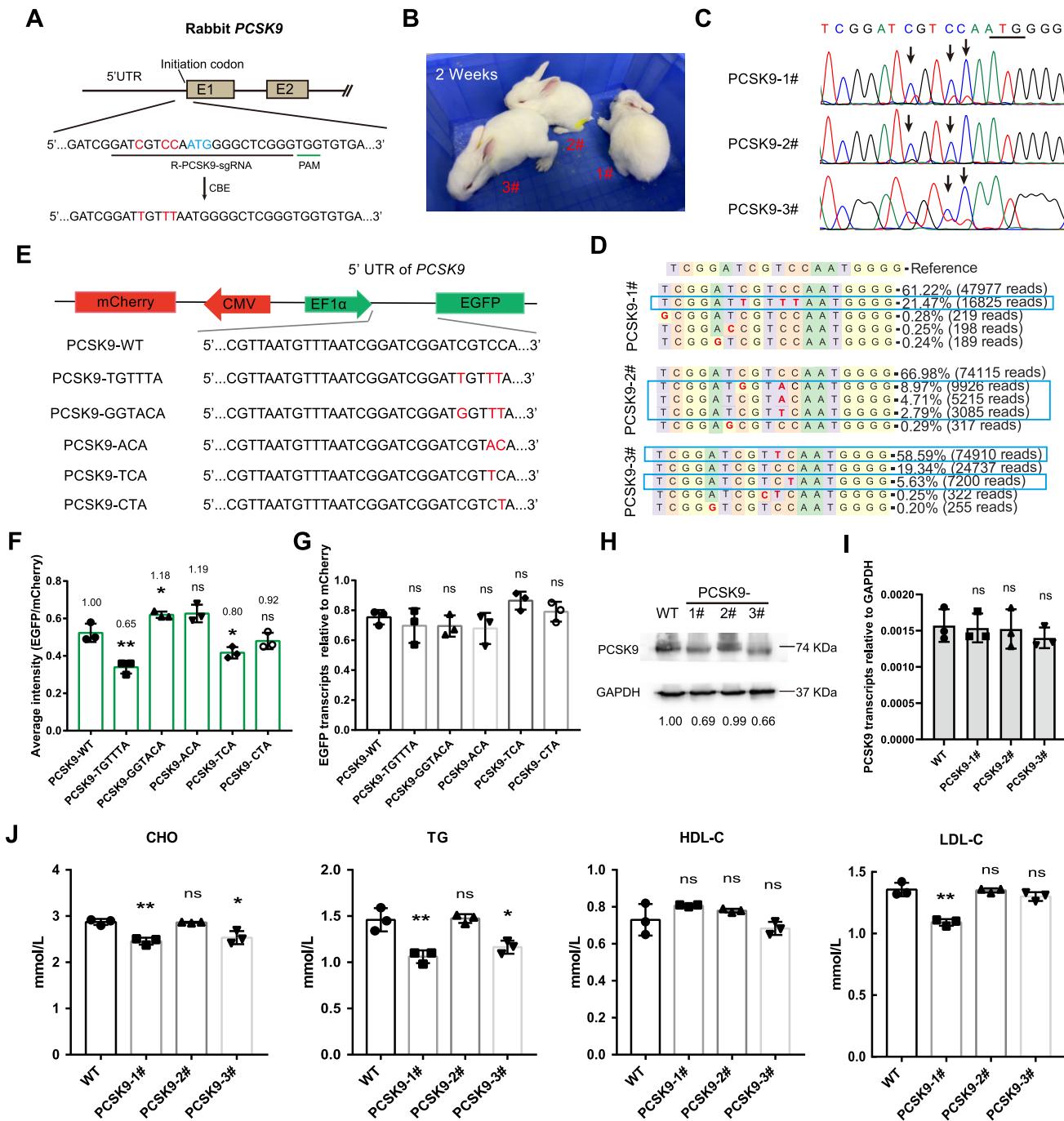


Figure 4. Down-regulation of the *PCSK9* expression in rabbits by using the KZ₃-edit strategy. (A) The base editing design of the rabbit *PCSK9* gene with the KZ₃-edit strategy. The initiation codon of the rabbit *PCSK9* gene was labeled with blue, while the target bases are indicated with red. The green underline shows the PAM sequence and the black underline indicates the spacer sequence of the guide RNA. With CBE, the target C would potentially be mutated to T. (B) Representative picture of the 2-week-old rabbits with base editing at the Kozak sequence of the *PCSK9* gene. (C) The genotyping of three rabbits with Sanger sequencing. The black arrows indicated the base mutations. (D) The amplicon deep sequencing results of the top five genotypes in each rabbit. The blue boxes show the main mutation genotypes. (E) Detection of the translational effects of rabbit *PCSK9* gene with KZ₃ and/or KZ₆ modifications by using the dual-fluorescence reporter system. The red color represents the base mutations. (F) Statistical analysis results of the EGFP intensity relative to mCherry. (G) The q-PCR result of the EGFP transcript levels relative to mCherry. (H) Detection of the PCSK9 protein expression in the livers of PCSK9-1#, PCSK9-2#, PCSK9-3# and age-matched WT rabbits by western blotting. The gray ratio indicated the PCSK9 protein expression relative to GAPDH and the gray ratio of WT rabbits was normalized as 1.00. (I) Detection of the transcription levels of the *PCSK9* gene in KZ₃-edited and WT rabbits by using q-PCR. The data are shown as mean \pm SD with three experimental replicates. (J) Blood lipid testing for the rabbits beginning at 1 month after birth, once a week ($n = 3$). The data in (F), (G), (I) and (J) are shown as the mean \pm SD; statistical significance was determined in comparison with the WT control as follows: ns, no significant difference; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ by two-tailed Student's *t*-test.

surgically transferred into two surrogate mothers (Supplementary Figure S7B). One surrogate was confirmed to be pregnant and gave birth to three rabbits (PCSK9-1#, PCSK9-2# and PCSK9-3#) after ~30 days of gestation (Figure 4B). The ear tissues of these three rabbits were collected and subjected to genomic DNA extraction. The PCR products surrounding the target site were first subjected to Sanger sequencing. The results showed that all three rabbits were mosaic animals harboring C-to-T/G/A mutations (Figure 4C). The efficiencies and patterns of base editing for these three KZ₃-edited rabbits were further analyzed by amplicon deep sequencing. The quantification of the high-throughput sequencing reads showed that PCSK9-1# rabbits harbored CGTCCAATG>TGTTTAATG (PCSK9-TGTTTA) with efficiencies of 21.47%; PCSK9-2# rabbits harbored CGTCCAATG>GGTACAATG (PCSK9-GGTACA), CCAATG>ACAATG (PCSK9-ACA) and CCAATG>TCAATG (PCSK9-TCA) with efficiencies of 8.97%, 4.71%, and 2.79%, respectively; and PCSK9-3# rabbits harbored CCAATG>TCAATG (PCSK9-TCA) and CCAATG>CTAATG (PCSK9-CTA) with efficiencies of 58.59% and 5.63%, respectively (Figure 4D). The dual-fluorescence reporter system further confirmed that these Kozak variants could regulate protein expression without affecting transcription (Figure 4E–G). PCSK9-TGTTTA, TCA and CTA could down-regulate the EGFP expression to 0.65-, 0.80- and 0.92-fold that of PCSK9-WT, respectively, while PCSK9-GGTACA and ACA could up-regulate the expression to 1.18- and 1.19-fold that of PCSK9-WT, respectively (Figure 4F). The liver tissues of PCSK9-1#, PCSK9-2#, PCSK9-3# and age-matched WT rabbits were further collected by needle biopsy and subjected to western blotting and q-PCR analysis. The gray ratio tests of western blotting results showed that the PCSK9 protein expression in the livers of PCSK9-1# and PCSK9-3# decreased to 0.69- and 0.66-fold that of WT rabbits, while no change of PCSK9-2# was observed (Figure 4H). These results were consistent with the predictions by the dual-fluorescence reporter system. Furthermore, no significant differences in transcription levels of the PCSK9 gene were found among PCSK9-1#, PCSK9-2#, PCSK9-3# and age-matched WT rabbits (Figure 4I). Next, we verified whether these mutations of the Kozak sequence of the PCSK9 gene could regulate cholesterol levels in rabbits. Serum from blood samples of 1-month-old PCSK9-1#, PCSK9-2#, PCSK9-3# and WT rabbits were extracted, and CHO, TG, LDL-C and HDL-C levels were analyzed. Compared with WT rabbits, rabbit PCSK9-1# had significantly reduced CHO, TG and LDL-C levels; rabbit PCSK9-3# had significantly reduced CHO and TG levels, but no change in LDL-C levels; and rabbit PCSK9-2# had no change in any of the detected indices (Figure 4J). In addition, the HDL-C levels of all tested rabbits were not significantly different because the expression changes in the PCSK9 gene did not affect HDL-C homeostasis. These *in vivo* results corresponded to the *in vitro* observations of the dual-fluorescence reporter system. These results suggested that direct modification of KZ₃ sequence of pORFs could regulate the translation of endogenous genes in animal models with predicted phenotypes.

Regulation of protein expression by delicate modifications of the KZ₃ sequences of uORFs

The uORFs in the 5'-UTRs of genes are able to influence the translation of downstream pORFs (31,57) mainly by competing for the utilization of ribosomes with downstream pORFs (58,59). The translation of uORF dominantly starts from the ATG initiation codon, sharing the same translation mechanisms with the main reading frame (60,61). Therefore, the KZ₃-edit strategy can be used not only for the translation initiation of a downstream pORF to control its expression, but also for the control of uORF translation, consequently indirectly regulating pORF translation (Figure 5A). Human *RNASEH1* and *SFXN3* genes, which have been previously shown to contain uORFs that could influence translation of the two genes, were selected as target genes to determine whether editing the Kozak sequences of uORFs by using the KZ₃-edit strategy could indirectly influence pORF translation. The KZ₃ sequence GAA of the uORF of the *RNASEH1* gene was substituted to ACC and TGT by the prime editor, which were predicted to increase and decrease the expression of the uORF and indirectly down- and up-regulate the translation of the pORF, respectively (Figure 5B; Supplementary Figure S8A). In addition, the KZ₃ sequence GTG of the uORF of the *SFXN3* gene was substituted with TGG by the prime editor, which was predicted to decrease the expression of the uORF and indirectly up-regulate the translation of the pORF (Figure 5G; Supplementary Figure S8B). The point mutations in the PAM sequences were designed to avoid repetitive targeting (Figure 5B, G). Similarly, the results of the dual fluorescence reporter system were consistent with the predictions (Supplementary Figure S8C–G). Prime editor systems targeting the Kozak sequences of the uORFs of human *RNASEH1* and *SFXN3* genes were electroporated into HEK293 cells. The Sanger sequencing results of the transfected cells showed that the predicted point mutations at the Kozak sequences occurred with all three pegRNAs with efficiencies ranging from 14% to 22% (Supplementary Figure S8H). Furthermore, single-cell-derived colonies harboring GAA>ACC or GAA>TGT at the KZ₃ sequence of the uORF of the *RNASEH1* gene or GTG>TGG at the KZ₃ sequence of the uORF of the *SFXN3* gene were screened and identified for further analysis (*uRNASEH1*-ACC-7 and -57; *uRNASEH1*-TGT-3 and -49; and *uSFXN3*-TGG-21, -35 and -51) (Figure 5C, H). Western blotting and gray ratio tests showed that the RNASEH1 protein expression of the *uRNASEH1*-ACC-7 and -57 cell colonies decreased to 0.52- and 0.51-fold that of the WT cells, respectively, while for the *uRNASEH1*-TGT-3 and -49 cell colonies, the RNASEH1 protein expression increased to 1.25- and 1.22-fold that of the WT cells, respectively (Figure 5D, E). In addition, the SFXN3 protein expression of the *uSFXN3*-TGG-21, -35, -51 and -60 cell colonies increased to 2.83-, 2.17-, 1.88- and 1.86-fold of the WT expression, respectively (Figure 5I, J). Similar to direct mutation of the KZ₃ sequences of pORFs, modifying the KZ₃ sequences of uORFs had no influence on the transcription of pORFs (Figure 5F, K). These results suggested that direct editing of the KZ₃ sequence of uORFs could also indirectly regulate the translation of pORFs without transcriptional changes.

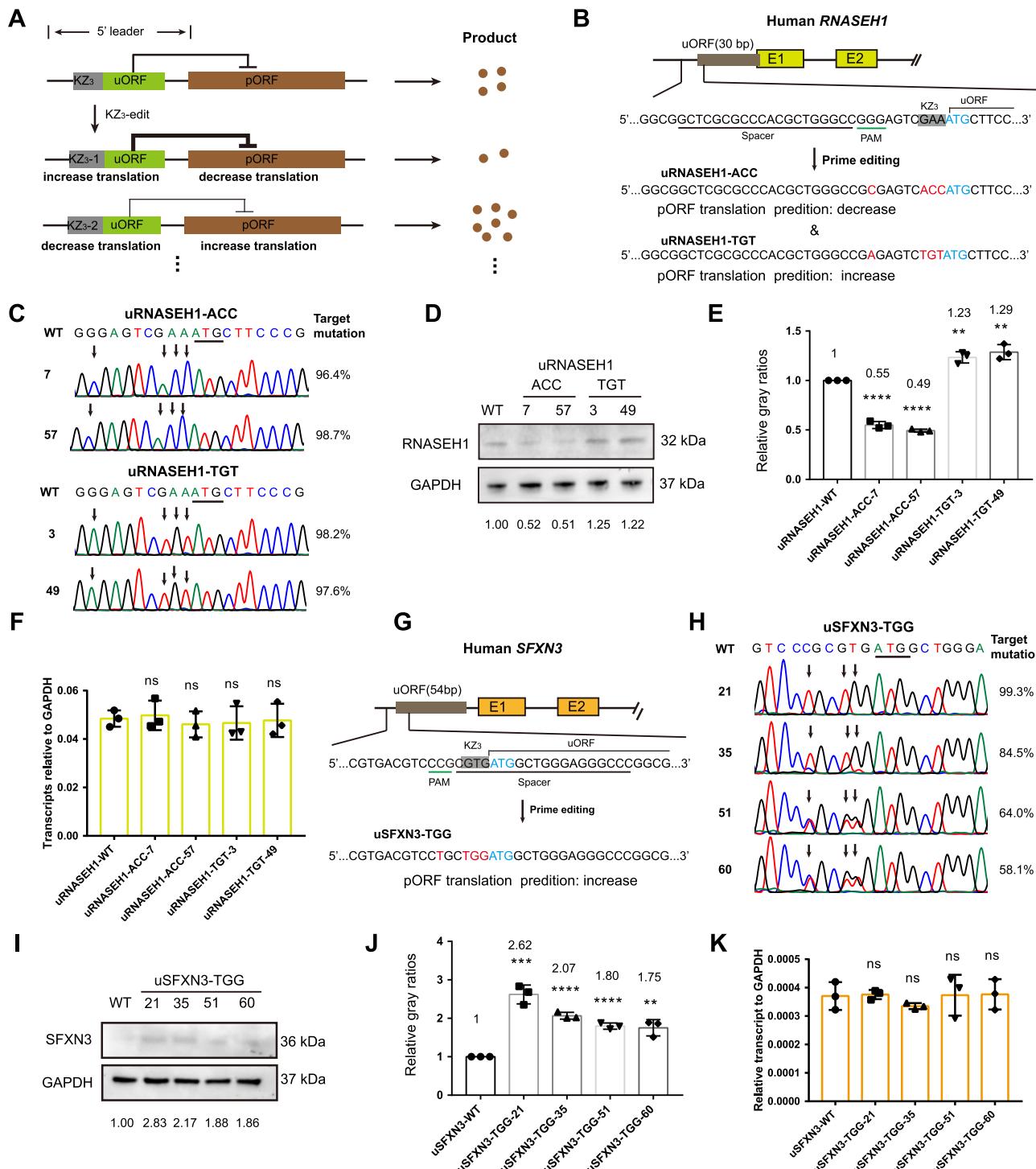


Figure 5. Application of the KZ₃-edit strategy to the uORF. **(A)** Schematic of the KZ₃-edit for uORFs to indirectly control the translation of pORFs. **(B and G)** The prime editing design for the uORF Kozak sequence of human RNASEH1 **(B)** and SFXN3 **(G)** genes. The PAMs are labeled by green color and the target mutations are indicated with red. **(C and H)** Sanger sequencing results of uRNASEH1-ACC and TGT **(C)**, and uSFXN3-TGG **(H)** cell colonies. The black arrows indicate the target mutations. The frequencies of target mutations were calculated from amplicon deep sequencing. **(D and I)** Representative western blotting results of the translation level of uRNASEH1-ACC and TGT **(D)**, and uSFXN3-TGG **(I)** cell colonies. The gray ratio of WT was normalized as 1.00. **(E and J)** Statistical analysis of gray ratios for uRNASEH1-ACC and TGT **(E)**, and uSFXN3-TGG **(J)** cell colonies. The data were calculated from three independent replicates. **(F and K)** The q-PCR results for the relative transcript levels of uRNASEH1-ACC and TGT **(F)**, and uSFXN3-TGG **(K)** cell colonies. The data are shown as mean \pm SD; all statistical significance was determined in comparison with the WT control as follows: ns, no significant difference; * P < 0.05; ** P < 0.01; *** P < 0.001 by two-tailed Student's *t*-test.

Validation of a practical application of the KZ₃-edit strategy to generate animal models with desired traits by delicately editing the KZ₃ sequence of uORFs

The hairless gene, *HR*, encodes a protein that is involved in hair growth. The pORF translation of the *HR* gene is affected by multiple uORFs, and *U2HR* appears to be the gene that is most closely related to the growth cycle of hair (62). A previous study showed that knocking out the upstream *U2HR* can significantly promote the synthesis of the *HR*-encoded protein, leading to MUHH, which is characterized by less hair or gradual shedding of hair with age (63). In theory, we can indirectly increase or decrease the translation of the *HR* pORF by using the KZ₃-edit strategy to regulate the translation of the upstream *U2HR*. To verify this speculation, we attempted to create a novel MUHH rabbit model by manipulating *U2HR*. We first compared the nucleotides and amino acids of the *HR* gene and its 5'-UTR between humans and rabbits by using the BLAST software on the NCBI website (Supplementary Figure S9A). When humans were compared with rabbits, the nucleotide similarity of the coding region (CDS) of the *HR* gene was 82.94%, the amino acid similarity was 79.39%, the nucleotide similarity of the 5'-UTR of the *HR* gene was 89.28% and the similarities of the nucleotides and amino acids for *U2HR* were as high as 98.10% and 100%, respectively, indicating that *U2HR* is highly conserved in mammals. To generate an MUHH rabbit model, we modified the KZ₃ sequence of rabbit *U2HR* with the CBE, BE4max, combined with the designed R-*HR*-sgRNA to create CCC to TTC or CCC to TTT conversion, which was expected to indirectly increase the translation of the *HR* pORF and trigger the phenotypes of MUHH (Figure 6A). The BE4max mRNA and corresponding sgRNA were transcribed and purified *in vitro*. An RNA mixture of BE4max (150 ng/μl) and sgRNA (100 ng/μl) was injected into rabbit zygotes. A total of 40 injected embryos were transferred to surrogate rabbits (Supplementary Figure S7A, B). Two rabbits with *U2HR* KZ₃ editing were successfully obtained and named U2-1 and U2-2. The hair of U2-1 and U2-2 grew slowly and sparsely compared with that of WT rabbits of the same age, displaying the typical phenotypes of MUHH (Figure 6B). The Sanger sequencing results showed that the desired mutation occurred in the KZ₃ sequence of *U2HR* in the U2-1 and U2-2 rabbits (Figure 6C). Amplicon deep sequencing further showed that the U2-1 rabbit had CCC>TTC (45%), CCC>TTT (32.64%) and 5 bp deletion (16.73%) mutations at the target site, whereas the U2-2 rabbit mainly had the mutation of CCC>TTC (93.83%) (Figure 6D). The average intensities of EGFP/mCherry varied among different groups, while there were no significant changes in the transcriptional ratios of EGFP relative to mCherry (Supplementary Figure S9C, D). The fluorescence intensity ratios of *U2HR*-TTC, -TTT and -DATG changed to 1.14, 0.96 and 1.96 times that of *U2HR*-WT, respectively (Supplementary Figure S9C). The skin tissues of U2-1, U2-2 and age-matched WT rabbits were collected and subjected to further analysis. The western blotting results showed that HR protein synthesis increased in the U2-1 (1.41-fold) and U2-2 (1.38-fold) rabbits compared with the WT rabbits (Figure 6E), whereas the transcript levels showed no changes

(Figure 6F). Additionally, the H&E staining results of skin tissues showed that the number of hair follicles in U2-1 and U2-2 decreased by 32% and 26%, respectively, relative to those of the WT rabbits (Figure 6G, H). These results demonstrated that manipulating the KZ₃ of the uORF in the genome can delicately regulate gene translation and facilitate the establishment of animal models with the desired phenotypes.

Universality of the KZ₃-edit strategy across the whole human genome and accessibility with existing commonly used base editors and prime editors

The above experiments have shown that precise genome editing tools can flexibly adjust the translation levels of target genes. Next, we further explored the universality of this strategy across the whole human genome by statistical analysis of the frequency and translation efficiency of each KZ₃ variant of pORFs and uORFs. The statistical results suggested that 71% of protein-encoding gene transcripts had varied numbers of ATG sequences in their 5'-UTR upstream of the actual initiation codon, indicating that the translation of 71% of gene transcripts might be regulated by uORFs (Figure 7E). As shown in Figure 7A and F, the 25 bp DNA sequences before and after the translation initiation codon sequences of all pORFs and uORFs presented significant preferences, which were consistent with the performance of the transcript sequence. Specifically, as an example, we listed the numbers of human pORFs and uORFs for 64 KZ₃ variants and ranked them according to the frequency and translation efficiency of each KZ₃ variant (Figure 7B, G). For the human pORFs, the gene transcripts containing the KZ₃ sequence GCC ranked number one, with as many as 3264, whereas those with the KZ₃ sequence ATG were least frequent, with as few as 19 (Figure 7B). For human uORFs, gene transcripts containing the KZ₃ sequence AAA ranked number one, with as many as 13 916, whereas those with the KZ₃ sequence CGT were least frequent, with as few as 887 (Figure 7G). According to the KZ₃ ranking of translation efficiency defined by the *in vitro* reporter system above, the variants in the top 50% of the list account for 72.2% and 55.1% of all human pORFs and uORFs, respectively, while those in the bottom 50% account for 27.8% and 44.9%, respectively (Figure 7B, G). These statistical results mean that 27.8% of pORFs and 44.9% of uORFs can be subjected to KZ₃-edit for the substantial up-regulation of translation, whereas 72.2% of pORFs and 55.1% of uORFs can be subjected to KZ₃-edit for the substantial down-regulation of translation.

We then analyzed the accessibility of existing commonly used base editors and prime editors at each pORF and uORF. We focused on the analysis of the accessibility of commonly used gene editing tools that identify four PAM sequences, NG, NGG, TTN and TTTN (46,56,64). The results showed that gene editing tools that recognize NG-PAM can find suitable target sites in all human pORFs and uORFs within the range of 25 bp around the translation initiation codon (Figure 7C, H). The numbers of human pORFs that can be targeted by gene editors with PAM recognition of NGG, TTN and TTTN were 34 922 (96.0%),

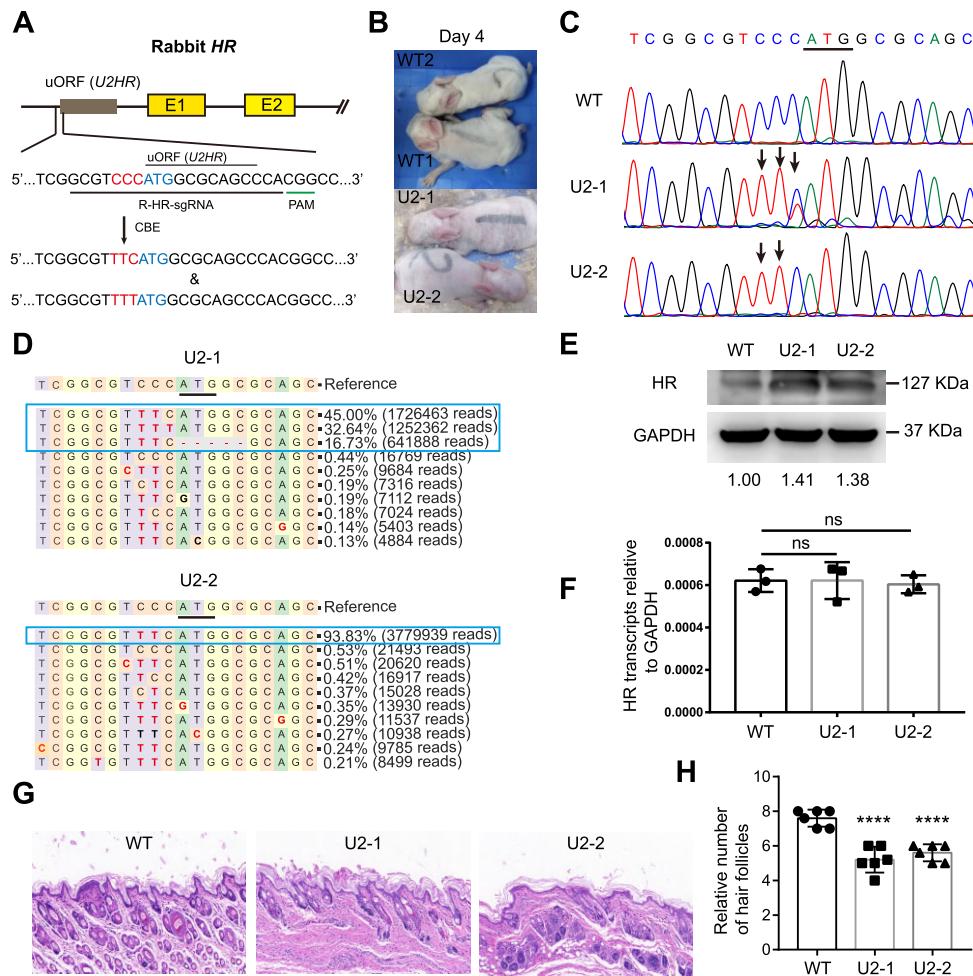


Figure 6. Generation of a novel MUHH rabbit model with the KZ₃-edit strategy. **(A)** The KZ₃-edit design of the uORF at the rabbit *HR* gene using a base editor. R-HR-sgRNA was indicated by black underline is the sgRNA to induce CCC>TTC or CCC>TTT substitutions; the PAM sequence is indicated by a green line and the target KZ₃ is labeled with red. **(B)** A representative picture of 4-day-old newborn rabbits. U2-1 and U2-2 are the two rabbits with a KZ₃ mutation at the uORF. **(C)** Sanger sequencing results of newborn rabbits. The black line marks the initiation codon of the uORF and the arrows indicate the edited bases. **(D)** Amplicon deep sequencing results for the two MUHH rabbits to analyze the detailed alleles. The blue boxes indicate the main mutations. **(E)** Detection of the *HR* translation level by western blotting. The gray ratio indicates the HR protein expression relative to GAPDH. **(F)** The transcription level analysis for *HR* of MUHH rabbits using q-PCR. The data are shown as the mean \pm SD with three experimental replicates. **(G)** Representative H&E staining pictures of the skin from WT, U2-1 and U2-2 rabbits. Scale bar, 100 μ m. **(H)** Counts of the hair follicles that were distributed at a 1 mm line. The data were calculated with five regions for each sample. Statistical significance in comparison with WT as follows: ns, no significant difference; * P < 0.05; ** P < 0.01; *** P < 0.001 by two-tailed Student's *t*-test.

27 822 (76.5%) and 14 057 (38.7%), respectively, while the numbers of human uORFs that can be targeted by gene editors with PAM recognition of NGG, TTN and TTTN were 53 615 (98.6%), 49 549 (91.1%) and 38 080 (70.0%), respectively (Figure 7C, H). These statistical results indicate that existing tools are sufficient to manipulate KZ₃ sequences on a vast majority of gene transcripts to regulate translation. Given that the translation of pORFs and uORFs in eukaryotes is often initiated by ATG, 100% of the pORFs and uORFs that encode proteins naturally contain a suitable 'TG' PAM. In addition, it was found that a guanine deoxynucleotide was located after the ATG sequence in >51.8% (18 829/36 367) of human pORFs and 69.2% (37 620/54 402) of uORFs, which provides the possibility to apply NGG-PAM, a commonly used and efficient PAM of gene editing tools (Figure 7D, I). Based on the

sequence characteristics of protein-coding gene transcripts and the KZ₃-edit strategy of this study, a universal edit design was proposed. For any target pORF or uORF with ATG as the start site of translation, the prime editing system that recognizes NG-PAM can be used for the targeted editing of the KZ₃-edit strategy, as the ATG sequence can work as the common TG-PAM (Figure 7J). For this design, the nick site is located between the -3 and -2 nucleotide positions, and the two nucleotides located upstream of the ATG can be replaced by any desired sequence using prime editing to achieve programmed translation regulation. The feasibility of this universal edit design was confirmed by generating precise point mutations, insertions and/or deletions with PE2-NG and corresponding pegRNAs at the KZ₃ sequences of human *TP53*, *RNASEH1* and *SOD1* genes (Supplementary Figure S10A–C).

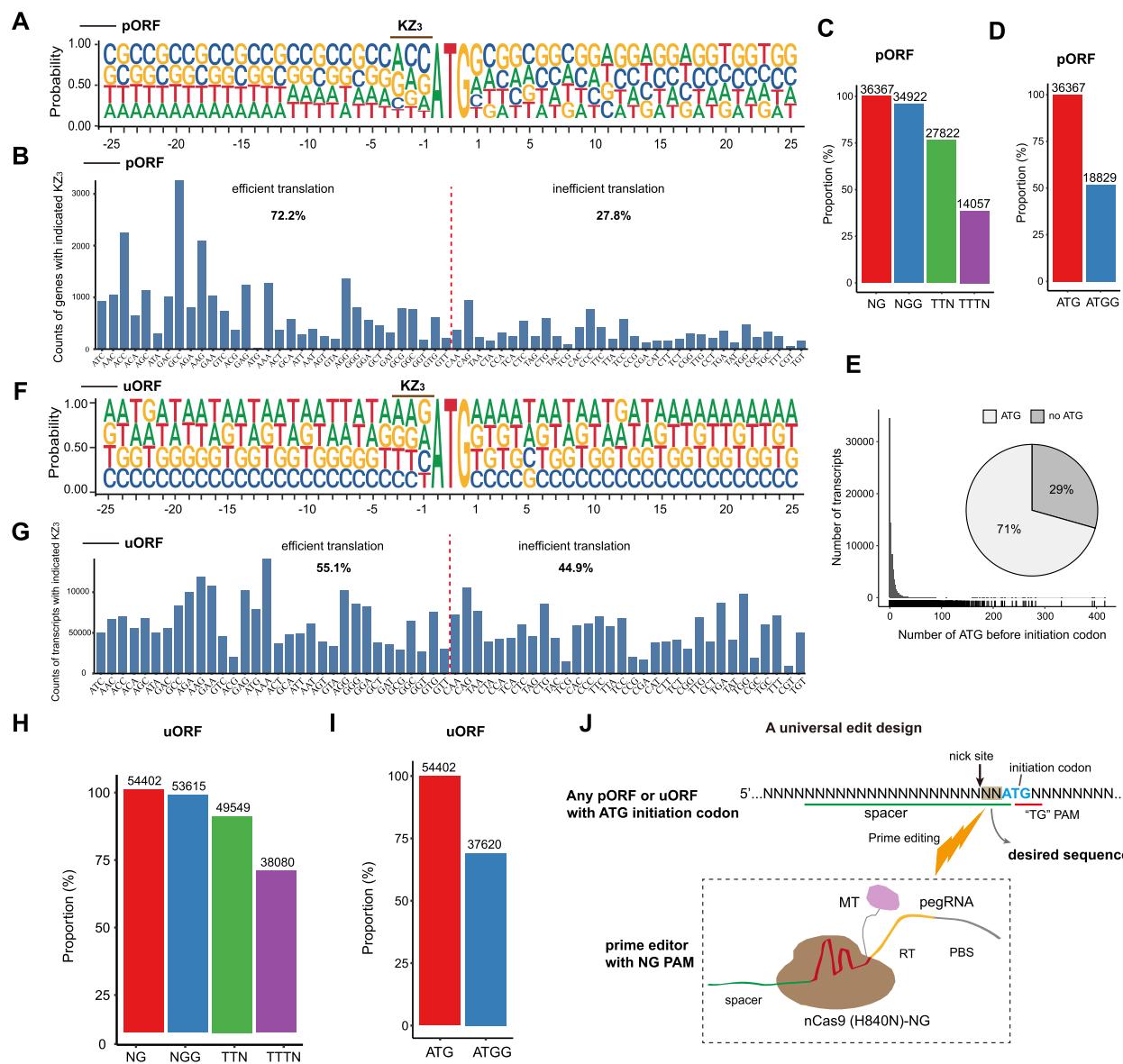


Figure 7. Analysis of the accessibility of the KZ₃-edit strategy on the genome. (A) Statistical analysis of 25 bp DNA sequence before and after the initiation codon at the local genome for all human protein-coding transcripts. The data show the probability of four kinds of nucleotides at each calculated position. (B) Gene counts of different KZ₃ sequences. The ranking of KZ₃ variants was the same as that in the KZ₃ reporter assay of the EGFP-UTR background. The KZ₃ sequences ranked in the first half were defined as efficient translation and those in the last half were considered as inefficient translation. (C) The proportion of genes that had the targetable PAM sequence within the 25 bp DNA around the initiation codon. The number of gene transcripts that contain the corresponding PAM sequence is shown on the column. (D) Statistical analysis of the gene transcripts with the 'ATG' or 'ATGG' initiation pattern. The number of gene transcripts is listed on the column. (E) Statistical analysis of the transcripts with or without ATG in the 5'-UTR. (F) Statistical analysis of 25 bp DNA sequence before and after the ATG of the uORF. (G) Transcript counts of uORFs with different KZ₃ sequences. The ranking of KZ₃ variants was the same as that in the KZ₃ reporter assay of the EGFP-UTR background. (H) The proportion of uORFs that had the targetable PAM sequence within the 25 bp DNA around the uORF initiation codon. (I) Statistical analysis of the uORF with the 'ATG' or 'ATGG' initiation pattern. (J) A universal KZ₃-edit design for any protein-coding genes or uORFs with the ATG initiation codon. MT, rat reverse transcriptase; RT, reverse transcription template; PBS, primer-binding sequence. The universal PAM is indicated by a red line and the ATG is labeled with blue colour.

DISCUSSION

In this study, we proposed a strategy to manipulate nucleotides located in front of a translation initiation codon in a genome with precise gene editing technology. This strategy can achieve effective control over the translation level of a target gene without affecting the transcription level. The effectiveness and flexibility of the strategy were verified at four levels: an *in vitro* dual-fluorescence reporting system, exogenous genes in cell lines expressing EGFP, endogenous genes

of local genomes *in vitro* and endogenous genes of local genomes *in vivo*. This strategy has several important advantages over existing methods for gene expression regulation, including long-lasting and bidirectional regulation and predictable and quantitative expression levels. Current gene expression regulation methods, such as gene knockout, RNAi and gene overexpression are ultimately unidirectional. The previously reported fusion system of dead Cas9 and a transcriptional activator (65,66) or suppressor (67)

can up- or down-regulate expression at the transcriptional level, but its effect is usually transient, given that the regulatory elements are not integrated into genomes. In contrast, for the KZ₃-edit strategy, the up- or down-regulation of protein expression could be long lasting since alteration of KZ₃ sequences occurs in the genome. In addition, the above methods to regulate protein expression by either DNA or RNA manipulation cannot quantitatively control gene expression to specific levels, and the resulting expression levels of proteins are unpredictable. However, for our KZ₃-edit strategy, the expression level ranking, as we performed for the 64 KZ₃ variants by utilizing an *in vitro* dual-fluorescence reporting system, can serve as an indicator to predict an ideal variant in KZ₃ sequences to achieve the desired expression level of a given protein. Although slight deviations of the real expression level of a given KZ₃ variant might exist in different UTR backgrounds or different cell types, with the aid of general trends, an appropriate KZ₃ variant for regulating the translation of endogenous genes to the desired levels can be pre-screened by using the same *in vitro* dual-fluorescence reporting system that is established in this study. For a given gene, translation will increase if it is mutated to a stronger KZ₃ variant and will decrease if it is mutated to a weaker variant. Since the regions with the most significant base preferences were the -3 to -1 nucleotides in front of ATG, which had previously been proven to play the most important regulatory role in the initiation of gene translation (54), we mainly focused on manipulating the three trinucleotides located before the initiation codon to validate the KZ₃-edit strategy. By using this strategy, 64 different KZ₃ variants can be customized, and the customized KZ₃ variants correspond to a wide range of protein translation efficiencies. We also confirmed that the KZ₃-edit strategy was a robust tool for controlling protein expression to specific levels under the guidance of the expression efficiency rank of 64 KZ₃ variants. For example, if the goal is to significantly change the translation level, the -3 bases (A>G>C>T) can be mutated preferentially, while if the goal is to finely tune the translation level, the -1 and -2 nucleotides can be selected for mutation. Each base of the classic Kozak sequence region was believed to have an effect on ribosome recognition of the initiation codon and translation initiation, so the KZ₃-edit strategy could also be extended to KZ₆ and even the whole Kozak sequence, allowing a richer variety of regulation. According to the sequence analysis of the whole genome, all human genes encoding proteins and these related uORFs could be the subject of the KZ₃-edit strategy. Furthermore, as eukaryotes share similar translation mechanisms, we speculated that this strategy should be applicable to gene expression control in all eukaryotes. During the preparation of our manuscript, an independent study also reported that base editing of the Kozak sequence could result in translational enhancement of endogenous genes and could be used for correcting dozens of haploinsufficient monogenic disorders independently of the causative mutation, which confirmed part of our findings (68).

Two precise gene editing tools, base editors and prime editors, were used to modify the KZ₃ sequences. The former can efficiently alter a single base (e.g. C>T, C>G or A>G), whereas the latter allows abundant and diverse target sequence customizations. When the KZ₃-edit strategy is used to regulate gene expression, base editors could be the

first choice to achieve high editing efficiency. Prime editors have relatively low efficiency but have been greatly improved by using dual pegRNAs and modified pegRNA structures (69–71), which could be the priority choice for sequence mutations requiring high editing accuracy. In fact, except for base editors and prime editors, other artificial nuclease-mediated gene editing techniques, such as CRISPR/Cas9, could be used to generate deletions and insertions in front of the initiation codon and increase the diversity of Kozak variants (72). If previously reported disruptions of initiation codons in pORFs (pATGs) (19,20) and uORFs (uATGs) (35,36), which could eradicate and up-regulate gene expression, respectively, are applied, more abundant variants for translation control could be achieved (Supplementary Figure S11A). Due to its versatile consequences on protein expression regulation, KZ₃-edit performed by using the AGBE editor, which could create multiple mutation patterns, including single nucleotide variants and indels (73), can be extended to the study of gene function through high-throughput processes. In particular, it will compensate for the lack of access to gain-of-function screening when using canonical transcript interference.

Finally, based on the sequence characteristics of protein-coding gene transcripts and the KZ₃-edit strategy, a common editing model was designed (Figure 7J). Then, a straightforward operating procedure was developed, as shown in Supplementary Figure S11B. These tools facilitate researchers to carry out the KZ₃-edit strategy to regulate translation and allow the widespread application of this strategy.

In conclusion, our study successfully validated a new strategy to regulate gene translation in a general and effective manner, which can directly edit the local KZ₃ sequence to achieve the goal of changing the translation levels of target endogenous genes. The effectiveness, simplicity and diversity of this strategy will greatly enrich existing methods used for gene expression regulation and greatly promote gene function investigations, disease treatments and the generation of animals and crops with desirable traits.

DATA AVAILABILITY

Amplicon deep sequencing data from this work have been deposited at the National Genomics Data Center, China, under accession code PRJCA010995 (<https://ngdc.cncb.ac.cn/ngdc-human/s/YZbEB5ko>). The authors state that all data necessary for confirming the conclusions of this article are presented fully within the article or are available from the authors upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Yahai Shu from Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Science for providing technical assistance with cell sorting. We are also grateful to the anonymous reviewers of NAR for their extensive and thoughtful suggestions and comments.

Author contributions: J.X., K.W. and L.L. conceived and designed the project. K.W. and L.L. supervised the project

and provided the funding support. J.X. and Z.Z. performed most of the experiments and analyzed the data. S.G. performed the informatics analyses and arranged some of data. Q.Z. performed embryo injection and generated gene-edited rabbits. Q.Z. and N.L. provided animal care and breeding. X.W., T.L., M.L., Y.L., Z.O., Y.Y. and H.W. provided technical assistance. L.L., K.W., J.X., Z.Z. and S.G. prepared the manuscript.

FUNDING

The National Key Research and Development Program of China [2022YFA1105403 and 2021YFA0805903]; the National Natural Science Foundation of China [82201736, 32170542 and 82101553]; the Key Research & Development Program of Hainan Province [ZDYF2021SHFZ052]; the Major Science and Technology Project of Hainan Province [ZDKJ2021030]; the 2020 Research Program of Sanya Yazhou Bay Science and Technology City [202002011]; the Postdoctoral Science Foundation of China [2021M703232 and 2022M713167]; the Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology City [2021JJLH0024]; the Youth Innovation Promotion Association of the Chinese Academy of Sciences [2019347]; the Young Elite Scientist Sponsorship Program by CAST [YESS20200024]; the Research Unit of Generation of Large Animal Disease Models, Chinese Academy of Medical Sciences [2019-I2M-5-025]; the Science and Technology Planning Project of Guangdong Province, China [2020B1212060052, 2021B1212040016 and 2020A1515110208]; and the Science and Technology Program of Guangzhou, China [202201010520].

Conflict of interest statement. None declared.

REFERENCES

1. Portin,P. and Wilkins,A. (2017) The evolving definition of the term 'gene'. *Genetics*, **205**, 1353–1364.
2. Huber,C.D., Durvasula,A., Hancock,A.M. and Lohmueller,K.E. (2018) Gene expression drives the evolution of dominance. *Nat. Commun.*, **9**, 2750.
3. Frye,M., Harada,B.T., Behm,M. and He,C. (2018) RNA modifications modulate gene expression during development. *Science*, **361**, 1346–1349.
4. Lee,T.I. and Young,R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
5. Wang,J.Q., Li,L.L., Hu,A., Deng,G., Wei,J., Li,Y.F., Liu,Y.B., Lu,X.Y., Qiu,Z.P., Shi,X.J. *et al.* (2022) Inhibition of ASGR1 decreases lipid levels by promoting cholesterol excretion. *Nature*, **608**, 413–420.
6. Bartonicek,N., Rouet,R., Warren,J., Loetsch,C., Rodriguez,G.S., Walters,S., Lin,F., Zahra,D., Blackburn,J., Hammond,J.M. *et al.* (2022) The retroelement Lx9 puts a brake on the immune response to virus infection. *Nature*, **608**, 757–765.
7. Dibble,C.C., Barratt,S.A., Perry,G.E., Lien,E.C., Geck,R.C., DuBois-Coyne,S.E., Bartee,D., Zenguya,T.T., Cohen,E.B., Yuan,M. *et al.* (2022) PI3K drives the de novo synthesis of coenzyme A from vitamin B5. *Nature*, **608**, 192–198.
8. Liu,X.J., Bao,X.H., Hu,M.J., Chang,H.M., Jiao,M., Cheng,J., Xie,L.Y., Huang,Q., Li,F. and Li,C.Y. (2020) Inhibition of PCSK9 potentiates immune checkpoint therapy for cancer. *Nature*, **588**, 693–698.
9. Verhaart,I.E.C. and Aartsma-Rus,A. (2019) Therapeutic developments for Duchenne muscular dystrophy. *Nat. Rev. Neurol.*, **15**, 373–386.
10. Wang,H.W., Sun,S.L., Ge,W.Y., Zhao,L.F., Hou,B.Q., Wang,K., Lyu,Z.F., Chen,L.Y., Xu,S.S., Guo,J. *et al.* (2020) Horizontal gene transfer of Fhb7 from fungus underlies fusarium head blight resistance in wheat. *Science*, **368**, eaba5435.
11. Wang,N., Fan,X., He,M.Y., Hu,Z.Y., Tang,C.L., Zhang,S., Lin,D.X., Gan,P.F., Wang,J.F., Huang,X.L. *et al.* (2022) Transcriptional repression of TaNOX10 by TaWRKY19 compromises ROS generation and enhances wheat susceptibility to stripe rust. *Plant Cell*, **34**, 1784–1803.
12. Zhu,X.X., Zhan,Q.M., Wei,Y.Y., Yan,A.F., Feng,J., Liu,L., Lu,S.S. and Tang,D.S. (2020) CRISPR/Cas9-mediated MSTN disruption accelerates the growth of Chinese Bama pigs. *Reprod. Domest. Anim.*, **55**, 1314–1327.
13. Anzalone,A.V., Koblan,L.W. and Liu,D.R. (2020) Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824–844.
14. Adli,M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat. Commun.*, **9**, 1911.
15. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
16. Cong,L., Ran,F.A., Cox,D., Lin,S.L., Barretto,R., Habib,N., Hsu,P.D., Wu,X.B., Jiang,W.Y., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
17. Billon,P., Bryant,E.E., Joseph,S.A., Nambiar,T.S., Hayward,S.B., Rothstein,R. and Ciccia,A. (2017) CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol. Cell*, **67**, 1068–1079.
18. Kuscu,C., Parlak,M., Tufan,T., Yang,J.K., Szlachta,K., Wei,X.L., Mammadov,R. and Adli,M. (2017) CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods*, **14**, 710–712.
19. Wang,X.J., Liu,Z.W., Li,G.L., Dang,L., Huang,S.S., He,L., Ma,Y.E., Li,C., Liu,M., Yang,G. *et al.* (2020) Efficient gene silencing by adenine base editor-mediated start codon mutation. *Mol. Ther.*, **28**, 431–440.
20. Chen,S.Y., Xie,W.H., Liu,Z.Q., Shan,H.H., Chen,M., Song,Y.N., Yu,H., Lai,L.X. and Li,Z.J. (2020) CRISPR start-loss: a novel and practical alternative for gene silencing through base-editing-induced start codon mutations. *Mol. Ther. Nucleic Acids*, **21**, 1062–1073.
21. Tang,W., Luo,X.Y. and Sanmuels,V. (2001) Gene silencing: double-stranded RNA mediated mRNA degradation and gene inactivation. *Cell Res.*, **11**, 181–186.
22. Hannon,G.J. (2002) RNA interference. *Nature*, **418**, 244–251.
23. Neumeier,J. and Meister,G. (2021) siRNA specificity: RNAi mechanisms and strategies to reduce off-target effects. *Front. Plant Sci.*, **11**.
24. Alagia,A. and Eritja,R. (2016) siRNA and RNAi optimization. *Wiley Interdiscip. Rev. RNA*, **7**, 316–329.
25. Higuchi,M., Kondou,Y., Ichikawa,T. and Matsui,M. (2011) Full-length cDNA overexpressor gene hunting system (FOX hunting system). *Methods Mol. Biol.*, **678**, 77–89.
26. Abe,K. and Ichikawa,H. (2016) Gene overexpression resources in cereals for functional genomics and discovery of useful genes. *Front. Plant Sci.*, **7**, 1359.
27. Saunders,T.L. (2020) The history of transgenesis. *Methods Mol. Biol.*, **2066**, 1–26.
28. Baker,R.I., Eikelboom,J., Lofthouse,E., Staples,N., Afshar-Kharghan,V., Lopez,J.A., Shen,Y., Berndt,M.C. and Hankey,G. (2001) Platelet glycoprotein Ib alpha Kozak polymorphism is associated with an increased risk of ischemic stroke. *Blood*, **98**, 36–40.
29. Jacobson,E.M., Concepcion,E., Oashi,T. and Tomer,Y. (2005) A Graves' disease-associated Kozak sequence single-nucleotide polymorphism enhances the efficiency of CD40 gene translation: a case for translational pathophysiology. *Endocrinology*, **146**, 2684–2691.
30. Kondo,S., Schutte,B.C., Richardson,R.J., Bjork,B.C., Knight,A.S., Watanabe,Y., Howard,E., de Lima,R., Daack-Hirsch,S., Sander,A. *et al.* (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat. Genet.*, **32**, 285–289.
31. Silva,J., Fernandes,R. and Romao,L. (2019) Translational regulation by upstream open reading frames and human diseases. In: Romao,L. (ed.) *The mRNA Metabolism in Human Disease. Advances in Experimental Medicine and Biology*. Springer, Cham. Vol. **1157**, pp. 99–116.
32. Li,S.N., Lin,D.X., Zhang,Y.W., Deng,M., Chen,Y.X., Lv,B., Li,B.S., Lei,Y., Wang,Y.P., Zhao,L. *et al.* (2022) Genome-edited powdery mildew resistance in wheat without growth penalties. *Nature*, **602**, 455–460.

33. Liu,X.X., Liu,H.L., Zhang,Y.Y., He,M.L., Li,R.T., Meng,W., Wang,Z.Y., Li,X.F. and Bu,Q.Y. (2021) Fine-tuning flowering time via genome editing of upstream open reading frames of heading date 2 in rice. *Rice*, **14**, 59.
34. Reis,R.S., Deforges,J., Sokoloff,T. and Poirier,Y. (2020) Modulation of shoot phosphate level and growth by PHOSPHATE1 upstream open reading frame. *Plant Physiol.*, **183**, 1145–1156.
35. Zhang,H.W., Si,X.M., Ji,X., Fan,R., Liu,J.X., Chen,K.L., Wang,D.W. and Gao,C.X. (2018) Genome editing of upstream open reading frames enables translational control in plants. *Nat. Biotechnol.*, **36**, 894–898.
36. Xing,S.N., Chen,K.L., Zhu,H.C., Zhang,R., Zhang,H.W., Li,B.B. and Gao,C.X. (2020) Fine-tuning sugar content in strawberry. *Genome Biol.*, **21**, 230.
37. Xue,C., Qiu,F., Wang,Y., Li,B., Zhao,K.T., Chen,K. and Gao,C. (2023) Tuning plant phenotypes by precise, graded downregulation of gene expression. *Nat. Biotechnol.*, <https://doi.org/10.1038/s41587-023-01707-w>.
38. Rodríguez-Leal,D., Lemmon,Z.H., Man,J., Bartlett,M.E. and Lippman,Z.B. (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell*, **171**, 470–480.
39. Kozak,M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13–37.
40. Garcia,V.E., Dial,R. and DeRisi,J.L. (2022) Functional characterization of 5' UTR cis-acting sequence elements that modulate translational efficiency in *Plasmodium falciparum* and humans. *Malar. J.*, **21**, 15.
41. Ambrosini,C., Garilli,F. and Quattrone,A. (2021) Reprogramming translation for gene therapy. In: Petris,G. (ed.) *Curing Genetic Diseases Through Genome Reprogramming*. Elsevier, Amsterdam. pp. 439–476.
42. Kozak,M. (1984) Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature*, **308**, 241–246.
43. Kozak,M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**, 8125–8148.
44. Komor,A.C., Kim,Y.B., Packer,M.S., Zuris,J.A. and Liu,D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.
45. Gaudelli,N.M., Komor,A.C., Rees,H.A., Packer,M.S., Badran,A.H., Bryson,D.I. and Liu,D.R. (2017) Programmable base editing of A-T to G-C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.
46. Rees,H.A. and Liu,D.R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.*, **19**, 770–788.
47. Anzalone,A.V., Randolph,P.B., Davis,J.R., Sousa,A.A., Koblan,L.W., Levy,J.M., Chen,P.J., Wilson,C., Newby,G.A., Raguram,A. et al. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.
48. Wu,H., Liu,Q.S., Shi,H., Xie,J.K., Zhang,Q.J., Ouyang,Z., Li,N., Yang,Y., Liu,Z.M., Zhao,Y. et al. (2018) Engineering CRISPR/Cpf1 with tRNA promotes genome editing capability in mammalian systems. *Cell. Mol. Life Sci.*, **75**, 3593–3607.
49. Kluesner,M.G., Nedveck,D.A., Lahr,W.S., Garbe,J.R., Abrahante,J.E., Webbor,B.R. and Moriarity,B.S. (2018) EditR: a method to quantify base editing from Sanger sequencing. *CRISPR J.*, **1**, 239–250.
50. Clement,K., Rees,H., Canver,M.C., Gehrke,J.M., Farouni,R., Hsu,J.Y., Cole,M.A., Liu,D.R., Joung,J.K., Bauer,D.E. et al. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.*, **37**, 224–226.
51. Song,Y.N., Yuan,L., Wang,Y., Chen,M., Deng,J.C., Lv,Q.Y., Sui,T.T., Li,Z.J. and Lai,L.X. (2016) Efficient dual sgRNA-directed large gene deletion in rabbit with CRISPR/Cas9 system. *Cell. Mol. Life Sci.*, **73**, 2959–2968.
52. Nakagawa,S., Niimura,Y., Gojobori,T., Tanaka,H. and Miura,K. (2008) Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Res.*, **36**, 861–871.
53. McClements,M.E., Butt,A., Piotter,E., Peddle,C.F. and MacLaren,R.E. (2021) An analysis of the Kozak consensus in retinal genes and its relevance to gene therapy. *Mol. Vis.*, **27**, 233–242.
54. Kozak,M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.
55. Nishida,K., Arazoe,T., Yachie,N., Banno,S., Kakimoto,M., Tabata,M., Mochizuki,M., Miyabe,A., Araki,M., Hara,K.Y. et al. (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, **53**, aaf8729.
56. Yang,L., Tang,J., Ma,X.L., Lin,Y., Ma,G.R., Shan,M.H., Wang,L.B. and Yang,Y.H. (2021) Progression and application of CRISPR-Cas genomic editors. *Methods*, **194**, 65–74.
57. Somers,J., Poiry,T. and Willis,A.E. (2013) A perspective on mammalian upstream open reading frame function. *Int. J. Biochem. Cell Biol.*, **45**, 1690–1700.
58. Wethmar,K. (2014) The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA*, **5**, 765–778.
59. Chen,H.H. and Tarn,W.Y. (2019) uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. *RNA Biol.*, **16**, 1327–1338.
60. Chew,G.L., Pauli,A. and Schier,A.F. (2016) Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat. Commun.*, **7**, 11663.
61. Johnstone,T.G., Bazzini,A.A. and Giraldez,A.J. (2016) Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.*, **35**, 706–723.
62. Potter,G.B., Beaudoin,G.M.J., DeRenzo,C.L., Zarach,J.M., Chen,S.H. and Thompson,C.C. (2001) The hairless gene mutated in congenital hair loss disorders encodes a novel nuclear receptor corepressor. *Genes Dev.*, **15**, 2687–2701.
63. Wen,Y.R., Liu,Y., Xu,Y.M., Zhao,Y.W., Hua,R., Wang,K.B., Sun,M., Li,Y.H., Yang,S., Zhang,X.J. et al. (2009) Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. *Nat. Genet.*, **41**, 228–233.
64. Collias,D. and Beisel,C.L. (2021) CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun.*, **12**, 555.
65. Perez-Pinera,P., Kocak,D.D., Vockley,C.M., Adler,A.F., Kabadi,A.M., Polstein,L.R., Thakore,P.I., Glass,K.A., Ousterout,D.G., Leong,K.W. et al. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods*, **10**, 973–976.
66. Maeder,M.L., Linder,S.J., Cascio,V.M., Fu,Y.F., Ho,Q.H. and Joung,J.K. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods*, **10**, 977–979.
67. Gilbert,L.A., Larson,M.H., Morsut,L., Liu,Z.R., Brar,G.A., Torres,S.E., Stern-Ginossar,N., Brandman,O., Whitehead,E.H., Doudna,J.A. et al. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442–451.
68. Ambrosini,C., Destefanis,E., Heir,E., Broso,F., Alessandrini,F., Longhi,S., Battisti,N., Pescce,I., Dassi,E., Petris,G. et al. (2022) Translational enhancement by base editing of the Kozak sequence rescues haploinsufficiency. *Nucleic Acids Res.*, **50**, 10756–10771.
69. Chen,P.J., Hussmann,J.A., Yan,J., Knipping,F., Ravisankar,P., Chen,P.F., Chen,C.D., Nelson,J.W., Newby,G.A., Sahin,M. et al. (2021) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*, **184**, 5635–5652.
70. Choi,J.H., Chen,W., Suiter,C.C., Lee,C., Chardon,F.M., Yang,W., Leith,A., Daza,R.M., Martin,B. and Shendure,J. (2022) Precise genomic deletions using paired prime editing. *Nat. Biotechnol.*, **40**, 218–226.
71. Nelson,J.W., Randolph,P.B., Shen,S.P., Everette,K.A., Chen,P.J., Anzalone,A.V., An,M., Newby,G.A., Chen,J.C., Hsu,A. et al. (2022) Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.*, **40**, 402–410.
72. Hsu,P.D., Lander,E.S. and Zhang,F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.
73. Liang,Y.H., Xie,J.K., Zhang,Q.J., Wang,X.M., Gou,S.X., Lin,L.H., Chen,T., Ge,W.K., Zhuang,Z.P., Lian,M. et al. (2022) AGBE: a dual deaminase-mediated base editor by fusing CGBE with ABE for creating a saturated mutant population with multiple editing patterns. *Nucleic Acids Res.*, **50**, 5384–5399.