Overexpression-based detection of translatable circular RNAs is vulnerable to coexistent linear byproducts

Yanyi Jiang[1], Xiaofan Chen[1], Wei Zhang[1, 2, 3]

1. Biomedical Research Institute, Shenzhen Peking University - The Hong Kong University of Science and Technology Medical Center, Shenzhen, Guangdong Province, China.
2. Greater Bay Biomedical Innocenter, Shenzhen Bay Laboratory, Shenzhen, Guangdong Province, China.
3. Corresponding author. Email address: zhangweispace@yeah.net

Abstract

In RNA field, the demarcation between coding and non-coding has been negotiated by the recent discovery of occasionally translated circular RNAs (circRNAs). Although absent of 5' cap structure, circRNAs can be translated cap-independently. Complementary intron-mediated overexpression is one of the most utilized methodologies for circRNA research but not without bearing echoing skepticism for its poorly defined mechanism and latent coexistent side products. In this study, leveraging such circRNA overexpression system, we have interrogated the protein-coding potential of 30 human circRNAs containing infinite open reading frames in HEK293T cells. Surprisingly, pervasive translation signals are detected by immunoblotting. However, intensive mutagenesis reveals that numerous translation signals are generated independently of circRNA synthesis. We have developed a dual tag strategy to isolate translation noise and directly demonstrate that the fallacious translation signals originate from cryptically spliced linear transcripts. The concomitant linear RNA byproducts, presumably concatemers, can be translated to allow pseudo rolling circle translation signals, and can involve backsplicing junction (BSJ) to disqualify the BSJ-based evidence for circRNA translation. We also find non-AUG start codons may engage in the translation initiation of circRNAs. Taken together, our systematic evaluation sheds light on heterogeneous translational outputs from circRNA overexpression vector and comes with a caveat that ectopic overexpression technique necessitates extremely rigorous control setup in circRNA translation and functional investigation.

Key words

Circular RNA, circular RNA overexpression, rolling circle translation, non-AUG translation

1. Introduction

Circular RNAs (circRNAs), generated from backsplicing in which the splice donor of an exon covalently joints the splice acceptor site of an upstream exon, have been thrusted into the biomedicine limelight for their high abundance and versatile biological function (*1-4*). CircRNAs have recently attracted accumulating attention for their protein-coding potential, even in the absence of 5' cap and 3' poly(A) tail structures that both sustain the canonical cap-dependent translation machinery. It has been reported that circRNAs can serve as templates for cap-independent translation via virus or cellular internal ribosome entry site (IRES) (*5-7*), m6A modification (*8*), short IRES-like sequence (*9*), and other elusive mechanisms (*10*). While the translation of circRNAs and their cellular impact cannot reach consensus (*11-15*), tentative and massive efforts have been made to identify endogenous translatable circRNAs (*5*), investigate the function of circRNA-derived peptides (*16, 17*), and explore the bioengineering application for protein production (*18-20*). The circular structure renders circRNA not only tolerant to RNA exonuclease resulting in high stability, but also the possibility to enable rolling circle translation from infinite open reading frames (ORFs), ORFs without stop codons (*6, 9, 10, 18*).

Among current approaches for circRNA study, the complementary intron-mediated circularization stands out as a preferred wildly used methodology for circRNA overexpression, as flanking complementary intronic sequences efficiently facilitate backsplicing to spur circRNA biogenesis (*21, 22*). Unfortunately, coexistent linear overlapping sequence seems inevitable and usually gets entangled with other side products including concatenated RNAs (*5, 19, 23, 24*). These byproducts are prone to undermining experimental design and inducing misleading conclusions. For example, circ-ZNF609 was reported to be translationally active (*25*), but a following study denied this possibility, as the putative overexpressed circRNA-derived peptides were even generated when the synthesis of circ-ZNF609 was compromised (*24*). Thus, the "dark side" of circRNA overexpression strategy deserves a closer inspection, whereas the methodical evaluation remains limited.

Here, using FLAG tagged ZKSCAN1 minigene circRNA overexpression vector (*22*), we have conducted a coding potential assessment of 30 human circRNAs carrying infinite ORFs and annotated AUG start codons. While pervasive translation events are detected by immunoblotting, following mutagenesis study demonstrates numerous translation signals are not from circRNAs. Such aberrant translation is initiated independently of the AUG start codons within the circRNA sequence but can be masked upon epitope tags spanning backsplicing junction (BSJ). A FLAG-Myc dual tag construct is developed to discriminate translation noise of linear RNAs. Using this system, we ascertain that the fallacious translation events are derived from concomitant cryptically spliced linear RNAs. Unexpectedly, we further determine the translation of BSJ-involved linear RNA byproducts and identify linear RNA derived pseudo rolling circle translation signals. Additionally, we find non-AUG start codons may participate in the translation initiation of circRNAs.

2. Materials and methods

2.1. Plasmids

All the circular RNA overexpression constructs used pcDNA3.1(+) CircRNA Mini Vector-ZKSCAN1 (Addgene, Plasmid #60649) (*22*) as backbone. EcoRV (GATATC) and SacII

(CCGCGGAG), the two multiple cloning sites flanking ZKSCAN1, were removed. The splicing signals of all the constructs were predetermined and adjusted according to Human Splicing Finder (*26*) or "Splice Site Prediction by Neural Network" website NNSPLICE 0.9 version (January 1997). An extra CAG was added immediately upstream of splicing donor in a subset of circRNAs to enhance the splicing strength. For the circRNAs which UTR length cannot be divided by three, the ORF was shifted to make sure the FLAG located in a right ORF (Table. 1). For SF constructs, the FLAG was split behind the position 9 (…GACGACGACGATAAG-gene-GATTACAAG…). For FM constructs, Myc tag was inserted immediately upstream of the stop codon in the putative ORF (…GAAAAGAAT-Myc-TAGGCTCGG…). All constructs and mutants were created through standard PCR-based mutagenesis method and confirmed by DNA sequencing.

2.2. Cell Culture and transfection

HEK293T (ATCC) were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (Gibco), and 1% of penicillin-streptomycin at 37°C with 5% CO_2. Plasmids were transfected into HEK293T with ViaFect (#E4982, Promega). 2.5ng of total DNA was added per well in a 6-well plate at Reagent:DNA ratio of 4:1. Cells were harvested after 42-50h without medium change.

2.3. Western blotting

Total proteins from HEK293T cells were extracted using lysis buffer (150mM NaCl, 50 mM Tris-HCl, pH7.4, 0.5% Triton-X) adding protease inhibitor cocktail (#11836153001, Roche). Prepared protein samples were separated by 12% or 15% SDS-PAGE and subsequently transferred onto a 0.2nm PVDF membrane (Pall). The immunoblotting was performed with the following primary antibodies (overnight at 4°C) and secondary antibodies (1h at RT): mouse-anti-Flag (M2 monoclonal, #F3165, 1:3000, Sigma-Aldrich), rabbit-anti-GAPDH (monoclonal, #2118, 1:1000, Cell Signaling), rabbit-anti-GFP (polyclonal, 1:2000, #ab290, Abcam), rabbit-anti-copGFP (polyclonal, #PA5-22688, 1:5000, Invitrogen), rabbit-anti-Myc (monoclonal, #71D10, 1:1000, Cell Signaling), anti-mouse and anti-rabbit HRP secondary antibody (1:10000, #A4416 and #A6154, Sigma). The membranes were visualized by ECL kits (#NEL105001EA, PerkinElmer; #34580, Thermo Scientific).

2.4. RNA extraction, reverse transcription (RT), q-RT-PCR and RNase R treatment

Total RNA was extracted by RNA-easy Isolation Reagent (#R701-01, Vazyme) according to manufacturer's protocol. cDNA was generated by GoScript kit (# TM316, Promega) using 4 μg total RNA with random primers. q-RT-PCR was performed with Bio-Rad CFX96 using SYBR Master Mix (#Q311-02, Vazyme). Relative expression was calculated by the ΔΔCT method, GAPDH and copGFP as normalization control. Primer sequences are listed in Table S2. 1μg total RNA was incubated with 4U/μl RNase R (#RNR07250, Lucigen) at 37°C for 20min, then placed on ice and reverse transcribed directly. Water replaced RNase R in mock reactions.

2.5. DNA extraction and Sanger sequencing

The PCR products were purified from DNA gels by extraction kit (#20051, Qiagen), and were directly sent for Sanger sequencing with gene specific primers (the same ones used for q-RT-

PCR) by Sangon Biotech.

2.6. Immunoprecipitation assay

Anti-FLAG M2 beads (#M8823, Sigma-Aldrich) were used according to the technical bulletin. Specifically, 20μl precleared M2 beads were added into 80μl cell lysate, and the final volume was brought to 500μl by adding lysis buffer. The sample was incubated at 4°C for 3h, washed with 500μl TBS for 3 times, and eluted proteins by 10μl 2X SDS-PAGE buffer for western blot analysis.

3. Results

3.1. Circular GFP RNA can be translated into functional protein

Previous researches reported that overexpressed circularized GFP RNA alone can be translated without other modifications (*9*). To test whether the ZKSCAN1 minigene circRNA overexpression vector is feasible for translatable circRNA study, split GFP was engineered into the vector and transfected into HEK293T cells. A robust GFP protein synthesis was detected by anti-GFP western blot, but no fluorescence was observed probably due to the misfolding of tandem GFP protein (data not shown). Then we inserted a P2A self-cleavage peptide sequence between two split GFP fragments, which enables ribosome skipping and cleavage of the nascent polypeptide without translation termination (Fig. S1A, S1C) (*18, 27*). Functional GFP protein translated from circ-GFP was detected by both fluorescence microscopy and anti-GFP western blot (Fig. S1B, S1D). Moreover, compared with the linear counterpart, circ-GFP showed less translation efficiency but more resistance to co-expressed 4E-BP1, a cap-dependent translation inhibitor (*28*). Consistent with reported studies (*5, 9*), these results indicate overexpressed circular GFP RNA can be translated cap-independently into functional protein.

3.2. Pervasive translation signals are detected from overexpressed circRNAs

We next evaluated the coding potential of 30 human circRNAs (Table 1): (1) These circRNAs are derived from exon circularization including the annotated AUG start codons, coined AUG circRNAs. We expect they have more tendency to initiate translation, although this still attracts controversy (*12*). (2) There are no stop codons in their circular annotated ORF, creating infinite ORF. We cloned circRNA sequence into ZKSCAN1 minigene vector, and fused FLAG tag immediately downstream of the splicing acceptor so that only when circRNAs are formed and translated can FLAG-fusion proteins be visualized by anti-FLAG western blot analysis (Fig. 1A). No AUG start codon is located before FLAG in the expected ORF. The NeoR/KanR site in pCDNA3.1 vector was replaced by copGFP for estimating transfection efficiency and q-RT-PCR normalization. To validate the synthesis and the circularization of overexpressed circRNA, FLAG and gene specific divergent primers are designed to perform RT–PCR and BSJ sequencing. As expected, ZKSCAN1 minigene vector accurately facilitates the production of most circRNAs except for circ-HSPB11 and circ-TNRC6B (Fig. 1B, Table S1). Surprisingly, pervasive translation signals were obtained in more than 10 circRNAs by anti-FLAG immunoblotting (Fig. 1C). Circ-BBX (414nt), abbreviated as circ-BBX, shares two exons with circ-BBX (678nt), and both BBX circRNAs generated translation signals as a single protein band, indicating similar initiation mechanism underlies their translation. Circ-SYMPK and

circ-ZFAND6 were reported to associate with ribosomes in ribosome profiling (*29*), and we indeed observed flagged protein produced from them. More translation signals will be captured if transfection efficiency is pushed highly enough (data not shown). Our further study focused on circ-BBX, circ-MAD1L1, circ-PPP1R13L, circ-SYMPK and circ-ZFAND6, for their particularly strong signals. Their circularity was evidenced by RNase R treatment as linear RNAs were digested while circRNAs not (Fig. 1D).

3.3. Mutagenesis identifies aberrant translation from overexpressed circRNAs

We designed 3 subsets of mutants to examine whether the flagged proteins are translated from circRNAs (Fig. 2A). Firstly, the deletion of the upstream *AluSq2* sequence impaired, but not abolished, the formation of circRNAs, validated by RT-PCR and quantitative PCR (Fig. 2D, 2G). This observation is consistent to former study (*7*). Secondly, we mutated the splicing acceptor (AG to AC) to create SA-mut construct. For SA-mut constructs, the genesis of circ-PPP1R13L and circ-MAD1L1 was abrogated, and other 3 circRNAs formed new BSJ using other splicing sites on vector, which were investigated by RT-PCR coupled BSJ sequencing (Fig. 2D, S2A). Thirdly, we inserted a stop codon at the end of circularized exon to assess whether latent cryptic translation initiation site is formed from unexpectedly spliced linear RNA and promotes flagged protein translation. The circRNA-derived proteins will be masked in anti-FLAG western blot since their translation terminates just before FLAG tag, but linear RNA derived proteins can be visualized (Fig. 2B). As expected, the flagged protein synthesis was abolished in all SA-mut constructs (Fig. 2C). Most of ladder-like bands disappeared in circ-MAD1L1-ΔALU and circ-PPP1R13L-ΔALU constructs. However, for circ-BBX-ΔALU and circ-ZFAND6-ΔALU mutants, the flagged proteins were yielded when the genesis of circRNAs was curbed. The FLAG signals were still observed in circ-BBX-stop and circ-ZFAND6-stop constructs, and their molecular weights were close to the wild-type (WT). These results demonstrate the flagged proteins in circ-BBX and circ-ZFAND6 groups are not translated from circRNAs. In a more comprehensive mutagenesis on circ-PPP1R13L (Fig. 2E, 2F), the 15kD protein band remained upon splicing donor (GT) removed or downstream *AluSZ* sequence deleted, which rather corroborates the conclusion that the peptides of 15kD are not derived from circ-PPP1R13L. We also infer from these data that the expected splicing acceptor, but not the assumed splicing donor, is required for the genesis of byproducts. There is a discrepancy between this observation and previous findings (*7, 24*). With perturbation of inserting stop codon immediately behind FLAG (F+stop) and frame shifting (UTR+1, CDS+1) (Fig. 2E), rolling circle translation was disrupted, but the translation noise from byproducts differs (Fig. 2F). So, on the plus side, substantial *Alu*-dependent flagged rolling circle translation may originate from circRNAs; on the minus side, concurrent *Alu*-independent noise obfuscates *bona fide* translation signals. Remarkably, we mutated all 5 annotated AUG codons to ACG on circ-PPP1R13L but failed to perturb the translation pattern (AUG-mut construct, Fig. 2F), indicating the translation from circ-PPP1R13L may be non-AUG mediated.

3.4. Non-AUG start codons may initiate circRNAs translation

To further investigate the translation initiation of circRNAs, we mutated all the annotated AUG into GGG codons which are acknowledged not to initiate translation (Fig. 3A, 3B) (*30*). The rolling circle translation of circ-MAD1L1 was significantly impeded by AUG mutation, but the

aberrant translation remained. AUG mutation did not affect the production of circ-MAD1L1 (Fig. S2B). The translation noise of circ-BBX and circ-ZFAND6 did not change much, suggesting the aberrant translation is initiated by the elements beyond circRNA sequence, a scenario distinct from circ-ZNF609 (*24, 25*). The AUG mutation were also unable to abrogate the rolling circle translation of circ-SYMPK (Fig. 3B). With the same method, we tested circ-PAPOLA, a circRNA associates with ribosome (*29*), carries infinite ORF, but naturally lacks annotated AUG codons (Table. 1). The rolling circle translation of circ-PAPOLA were indeed observed (Fig. 3B). These results indicate the non-AUG start codons may function in the translation initiation of circRNAs.

3.5. The aberrant translation is arisen from coexistent linear RNA.

To disentangle the authentic translation signals from byproduct noise, we flanked circRNA sequence with split FLAG to create SF constructs (Fig. 3A). Since the previously verified fallacious signals disappeared, it seems to be an efficacious noise masking strategy (Fig. 3C). Strangely, we detected new protein bands (~100kD) in circ-BBX-SF which were never seen in the WT construct. More flagged proteins (~100kD) were occasionally visualized depending on transfection efficiency, but the signals were too weak to be studied (Fig. 3C). Inspired by the observation of circ-PPP1R13L SD-mut (Fig. 2E), we hypothesized that the putative intronic splicing donor is involved in the aberrant translation. Dual tag constructs (FM) were designed to test it: Myc tag was inserted into the end of downstream ORF between splicing donor and *AluSZ* sequence, which is not supposed to be backspliced (Fig. 3D). Therefore, the anti-Myc immunoblotting signals label the translation noise from linear RNAs. Such engineering did not inhibit the production of overexpressed circRNAs (Fig. 3G), a finding aligns with recent study (*19*). For FM constructs, anti-Myc signals were obtained, but their molecular weight was larger than the theoretical value of putative ORF (Fig. 3D, 3E). Moreover, AUG mutation on circRNA sequence was also unable to disturb the synthesis of Myc tagged proteins (Fig. 3F).

Considering flagged and Myc-tagged noise shows similar SDS-PAGE gel shift and resists AUG mutation, we asked whether they are derived from the same proteins. We further performed immunoprecipitation (IP) to purify FLAG fusion proteins from cell lysate and visualized them by anti-Myc western blot (Fig. 4A). Strong Myc signals representing previously defined aberrant translation were detected in FM IP groups, but not in SFM IP groups in which FLAG is separated (purple arrows in Fig. 4B). By virtue of IP enrichment, robust anti-Myc rolling circle translation signals unexpectedly came out. Moreover, translation noise was even observed in SFM IP groups of circ-PPP1R13L and circ-SYMPK (blue arrows in Fig. 4B), indicating linear RNAs involving intact FLAG BSJ region were created and translated. We excluded the possibility of novel circRNAs containing both FLAG and Myc sequence by RT-PCR using corresponding divergent primers (Fig. 4A, 4C). Eventually, we concluded that the heterogeneous aberrant translation originates from concomitant linear RNAs, and directly demonstrated that the circRNA-iconic rolling circle translation signals can be stimulated by linear transcripts, most likely from RNA concatemers (Fig. 4D).

4. Discussion

Translate or not? Our endeavors, technically, have not completely confirmed that the rolling

circle translation derives from circRNAs in this study, because BSJ-bridging epitope tag and mutagenesis done here cannot circumvent the latent artefacts arisen from RNA concatemers and other unforeseen transcripts (Fig. 4D). These byproducts, which share the same BSJ with circRNAs, escape from most PCR-based discrimination and can be only visualized via northern blot. Even worse, the rationale evidencing translatable circRNAs usually rests on three pillars: immunoblotting to detect backsplicing specific proteins, mass spectrometry to determine BSJ specific peptides, ribosome profiling to identify polysome-protected BSJ sequence; coexistent linear concatemers including BSJ region will take a sledgehammer to all of them if conclusions heavily rely on overexpression without careful control experiments. A discernable connection between translation signal and flanking *Alu* sequence is a prerequisite but not sufficient condition to prove circRNA translation because the complementary *Alu* sequence could also contribute to the formation of undesired linear transcripts and exon concatemerization. Exemplified by our study, the practical limitations for detecting and avoiding unwanted transcripts would ambush the overexpression-based circRNA investigation. Fortunately, the caveat echoes, the remedy follows (*2, 3, 14, 15, 31*). Given most of RNA concatemers are observed in complementary intron-mediated backsplicing, other rarely used circRNA overexpression methodologies may overcome this dilemma (such as the strategy of inserting quaking binding sites (*17*) or using twister-optimized RNA expression (*32*)) and deserve a further evaluation.

Overexpression of circRNA carrying infinite ORF, resulting in a plethora of unfolded repeating polypeptides, can overwhelm the cellular protein quality control (QC) system to artificially amplify the signals of naturally degraded proteins, supported by the recent finding that rolling circle translation products undergo proteasome-mediated degradation(*9*). However, too much attention has been attracted on justifying the existence of circRNA-derived protein to realize translation outcome is a total package enveloping entire translation process, which could shed the enigma of ribosome-associated but peptide-undetected circRNAs. It is very tempting to speculate translatable circular RNAs function not by the intact translated proteins, but by the translation process itself, to regulate cellular activities. Because once a circRNA is translating, it must be strictly and precisely governed by translation machinery and QC system, of which dysregulation is sufficient to raise functional impact. Impressive gains in the understanding of ribosome QC (*33*) invite our further hypothesis: unstopped circular ORF renders ribosome collision an imminent possibility, because translating circRNAs threading through ribosomes without stop codon-mediated termination are prone to becoming fully loaded and susceptible to ribosome stalling, whereas the relevant research stays rather scarce.

Finally, non-AUG start codons have emerged as important translation initiators with profound cellular function (*34, 35*). Although our results regarding non-AUG translation are vulnerable to coexistent byproducts, we speculated that non-AUG may function more efficiently in triggering rolling circle translation of circRNAs, compared with the scenario in linear mRNA. Non-AUG circRNA translation has been also observed by another group with more solid mutagenesis evidence (*9*).

The optimism about endogenous translationally active circRNAs would be tempered if flawed experiments piggyback leapfrogging functional studies. More interrogation and development of circRNA research toolset merits spotlight in the future.

Gene	circBase ID	circRNA length(nt)	5'UTR length(nt)	ORF Modification	Sequence (5'-3')
ANKIB1	hsa_circ_0135061	576	90	–	[FLAG] AATGTGAAA ··· ··· TGTGTAGAG
ARF3	hsa_circ_0003770	477	93	–	[FLAG] GATCAAGTC ··· ··· AACAAACAG
BBX	**hsa_circ_0001324**	414	9	–	[FLAG] GTCACAGTA ··· ··· GCCAAGGAG
BBX	hsa_circ_0001325	678	9	–	[FLAG] GTCACAGTA ··· ··· ACACCAGAG
CDK17	hsa_circ_0002762	312	29	ORF+1, +CAG	[FLAG] GGATTTTTCA ··· ··· GCAGATTA [CAG]
CDKAL1	hsa_circ_0001582	291	5	ORF+1, +CAG	[FLAG] GAGAATATGC ··· ··· AAATTACA [CAG]
CDNF	hsa_circ_0093170	270	191	ORF+4, +CAG	[FLAG] TATGTATGTAAAG ··· ··· TGAAA [CAG]
CTPS2	hsa_circ_0089952	678	39	–	[FLAG] ATTATATCT ··· ··· CCAGATCTG
DTNA	hsa_circ_0108308	363	1	ORF-1, +CAG	[FLAG] ATGATTGA ··· ··· GTTTGATCCA [CAG]
EPHX1	hsa_circ_0003740	369	5	ORF+1, +CAG	[FLAG] GGAGCCATGT ··· ··· AGATTGAA [CAG]
FAM169A	hsa_circ_0073027	321	3	–	[FLAG] AGGATGGCA ··· ··· CTTAAGCAG
HSPB11	hsa_circ_0113645	333	15	–	[FLAG] GTTAAACAC ··· ··· GAAATTGTG
HUWE1	hsa_circ_0001922	528	24	–	[FLAG] ATCGTAAAG ··· ··· TTGGCAGAG
INVS	hsa_circ_0087754	297	24	–	[FLAG] GTTGCTCCG ··· ··· GCCCAGAAG
LARP4B	hsa_circ_0004202	180	39	–	[FLAG] CATCCTGAG ··· ··· TTGAGTCAG
MAD1L1	**hsa_circ_0004961**	339	189	–	[FLAG] CAGCACTCC ··· ··· AGCATGCAG
NMNAT1	hsa_circ_0110116	495	56	ORF+1, +CAG	[FLAG] GACAACAAGG ··· ··· AAACAAAA [CAG]
PASK	hsa_circ_0058988	471	42	–	[FLAG] TTGTTAGAG ··· ··· ACCACAGAG
PMS1	hsa_circ_0004776	438	20	ORF+1, +CAG	[FLAG] GCTGCTCTGT ··· ··· TTGGTCAA [CAG]
POLR3GL	hsa_circ_0110693	297	41	ORF+1, +CAG	[FLAG] GGTTTATTCA ··· ··· CCAAGAGA [CAG]
PPP1R13L	**hsa_circ_0008615**	219	21	–	[FLAG] GCGCCCGCT ··· ··· CCTTCTAGG
RIMKLB	hsa_circ_0025372	231	56	ORF+1, +CAG	[FLAG] GGTAGACGTT ··· ··· GAAACCTG [CAG]
RIPK1	hsa_circ_0001571	381	60	–	[FLAG] GGTACAGCT ··· ··· AAAGCCGAG
SYMPK	**hsa_circ_0002105**	183	12	–	[FLAG] CTGACAGCA ··· ··· CTCAAACAG
TAF12	hsa_circ_0011162	330	84	–	[FLAG] TTAATTGGC ··· ··· GTGGAGGAG
TMEM50A	hsa_circ_0000032	219	13	ORF-1, +CAG	[FLAG] GACCTGAA ··· ··· CTTCCTAATT [CAG]
TNRC6B	hsa_circ_0116578	165	130	ORF-1, +CAG	[FLAG] GTTGTGGT ··· ··· AGTTCCAAGG [CAG]
UROS	hsa_circ_0008369	345	26	ORF+4, +CAG	[FLAG] CTAGTGCCCTATA ··· ··· CTTCT [CAG]
ZFAND6	**hsa_circ_0000643**	381	17	ORF+1	[FLAG] GGTGTGCAAC ··· ··· ATGTGCAG
ZNF782	hsa_circ_0139446	288	44	ORF+4	[FLAG] CCAGCTGAGGCTT ··· ··· ACTCC
PAPOLA	hsa_circ_0033126	138	0	–	[FLAG] GATGAAACA ··· ··· GCCTCTCAG

Table 1. The circRNAs for coding potential investigation

Several circRNA ORFs are adjusted as mentioned in methods, for their sequence details are described here. Five shortlisted circRNA candidates are highlighted in red. The annotated 3' terminal of circRNA sequence is marked in blue.

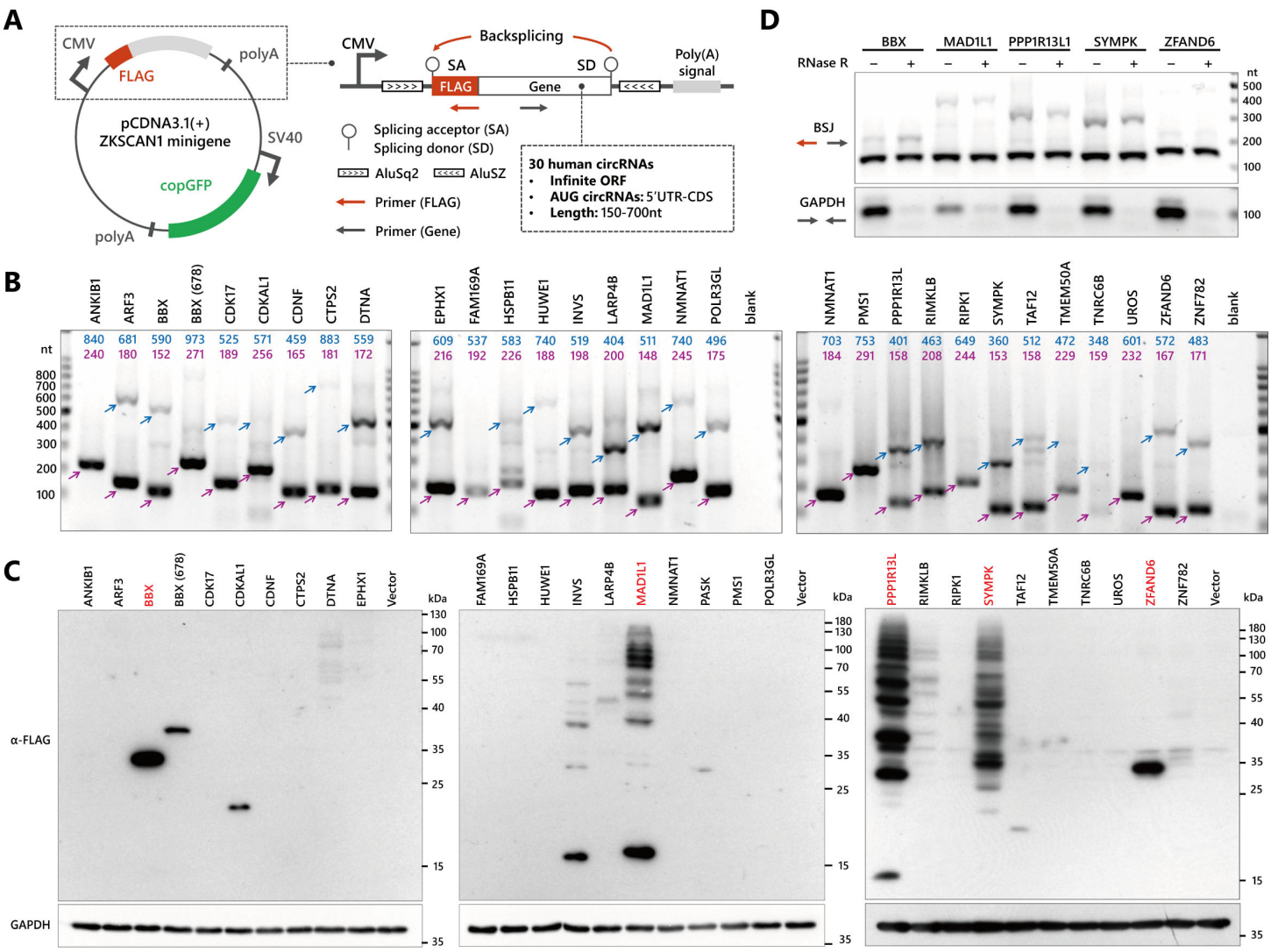


Figure 1. Pervasive translation events detected from overexpressed circRNAs.

A. Schematic representation of circRNA overexpression vector for translation activity assessment.

B. The rolling circle RT-PCR products from overexpressed circRNAs with divergent FLAG and gene specific primers. The purple/blue arrows and numbers denote the 0-cycle/1-cycle RT-PCR products and their theoretical length, respectively.

C. Western blot analysis with anti-FLAG antibody on proteins from HEK293T cells transfected with FLAG tagged circRNA overexpression vector. GAPDH as loading control. Five circRNAs for further investigation are marked in red.

D. RNase R treatment were implemented to verify the circularity of overexpressed circRNAs. GAPDH as linear RNA control.

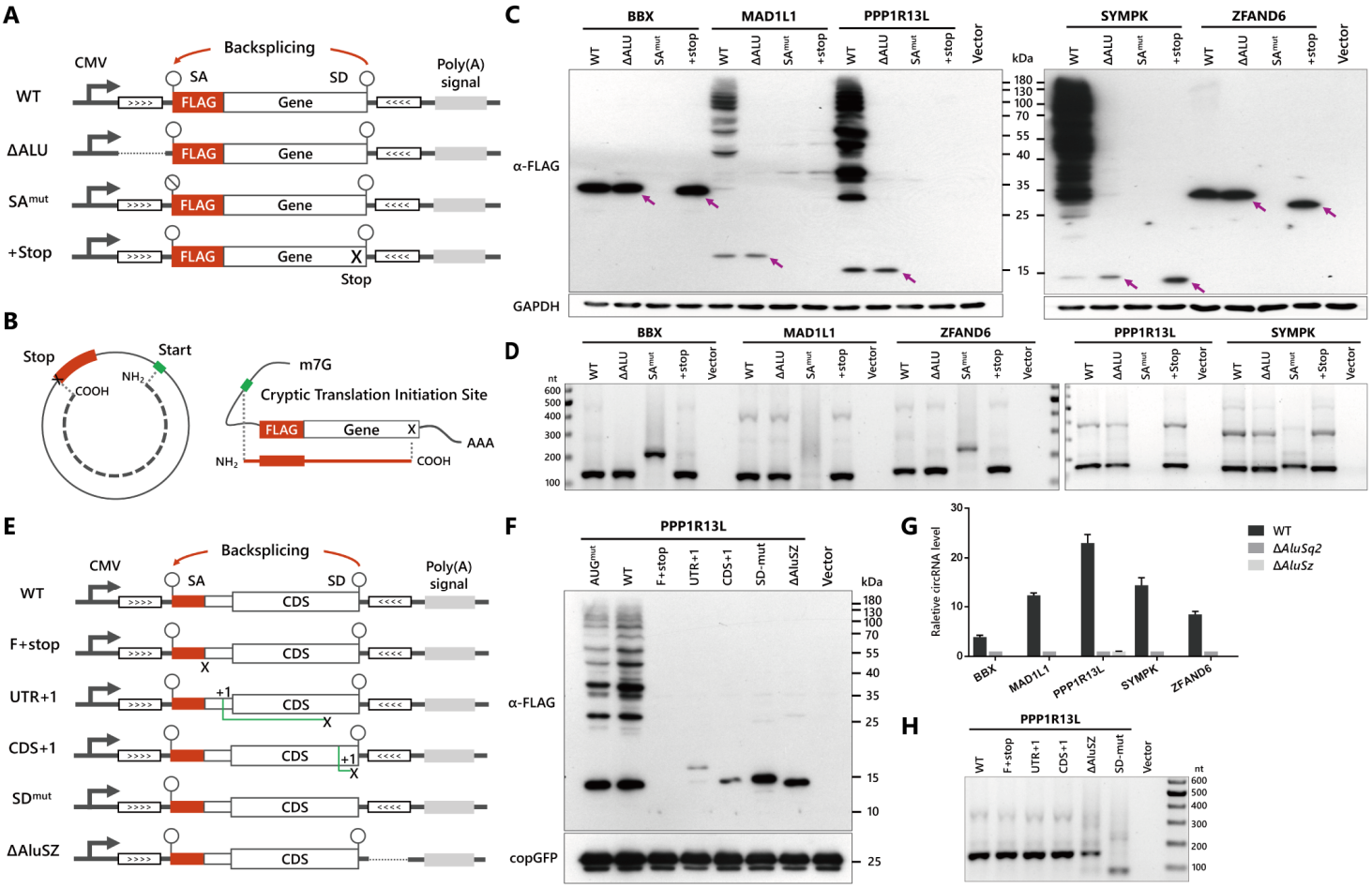


Figure 2. Mutagenesis-based validation of translation from overexpressed circRNAs

A. Scheme overview of mutants for translation validation. Specifically, ΔALU: the upstream *AluSq2* sequence (40bp) was deleted; SA-mut: the splicing acceptor AG was mutated into AC; +Stop: one stop codon was added at the end of circRNA sequence.

B. Schematic showing "+stop" mutants enable identification of the circRNA translation from lineate RNA output.

C. Western blot of HEK293T transfected with the mutants illustrated in Fig. 2A. The protein bands regarded as byproducts are marked by arrows. GAPDH as loading control.

D and H. RT-PCR assay for the overexpressed circRNAs and their mutants with gen and FLAG-specific divergent primers.

E. Scheme overview of circ-PPP1R13L mutants. Briefly, F+stop: a UAG stop codon was inserted immediately downstream of FLAG; UTR+1: a cytosine was inserted after position 7 of untranslated region (UTR); CDS+1: a cytosine was added before 6 nucleic acids of the coding sequence (CDS) end; SD-mut: the splicing donor AG was removed; ΔAluSZ: the downstream *AluSZ* sequence (36bp) was deleted. New ORFs created by frame shifting are shown as green lines.

F. Western blot of HEK293T transfected with the mutants of circ-PPP1R13L illustrated in Fig. 2D. copGFP as loading control.

G. q-RT-PCR for the measurement of relative circRNA level using divergent primer sets of FLAG and genes. CircRNA expression was normalized to the ΔAluSq2 groups, and copGFP was used to calibrate other RNA levels. Data is presented as mean ±SD (n=3).

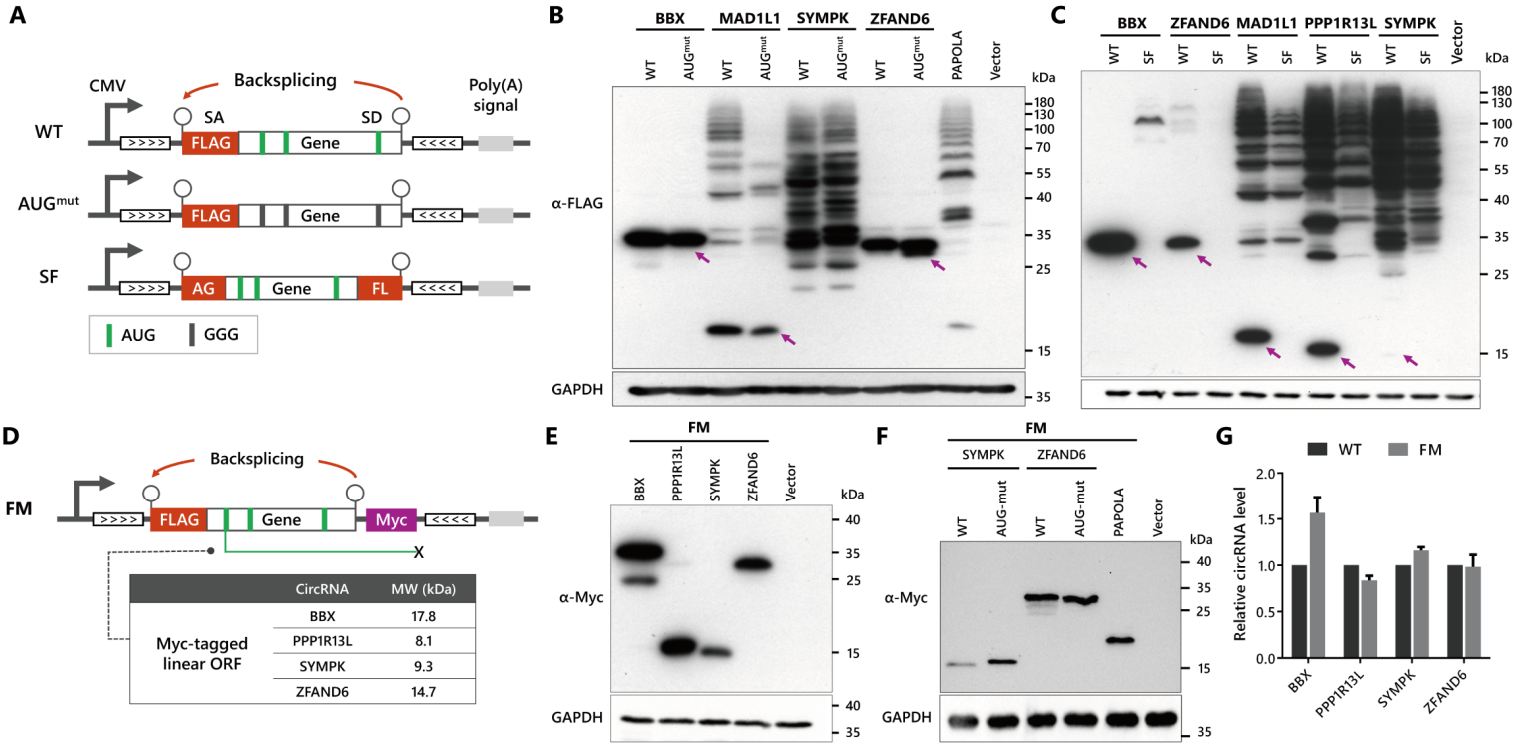


Figure 3. Analysis of the aberrant translation from overexpressed circRNAs

A. Scheme overview of constructs. AUG-mut: all the annotated AUG codons were mutated into GGG; SF: FLAG tag was split into two fragments flanking circRNA sequence.

B and C. Anti-FLAG western blot of HEK293T transfected with AUG-to-GGG mutant and split FLAG constructs. GAPDH as loading control. The fallacious translation signals are marked by arrows.

D. Schematic representation of FLAG-Myc construct design. Myc tag was embedded at the end of ORF beyond backsplicing cassette. The putative Myc tagged ORF is presented as green line, and the molecular weight of largest proteins translated from such ORF is listed.

E and F. Anti-Myc western blot of HEK293T transfected with FM constructs and the derivative mutants. All the annotated AUG start codons were mutated into GGG codons in AUG-mut constructs. GAPDH as loading control.

G. CircRNA levels of WT and FM were measured by q-RT-PCR with FLAG and gene specific divergent primers. CircRNA expression was normalized to the WT group. GAPDH was used for calibration. Data is presented as mean ±SD (n=3)

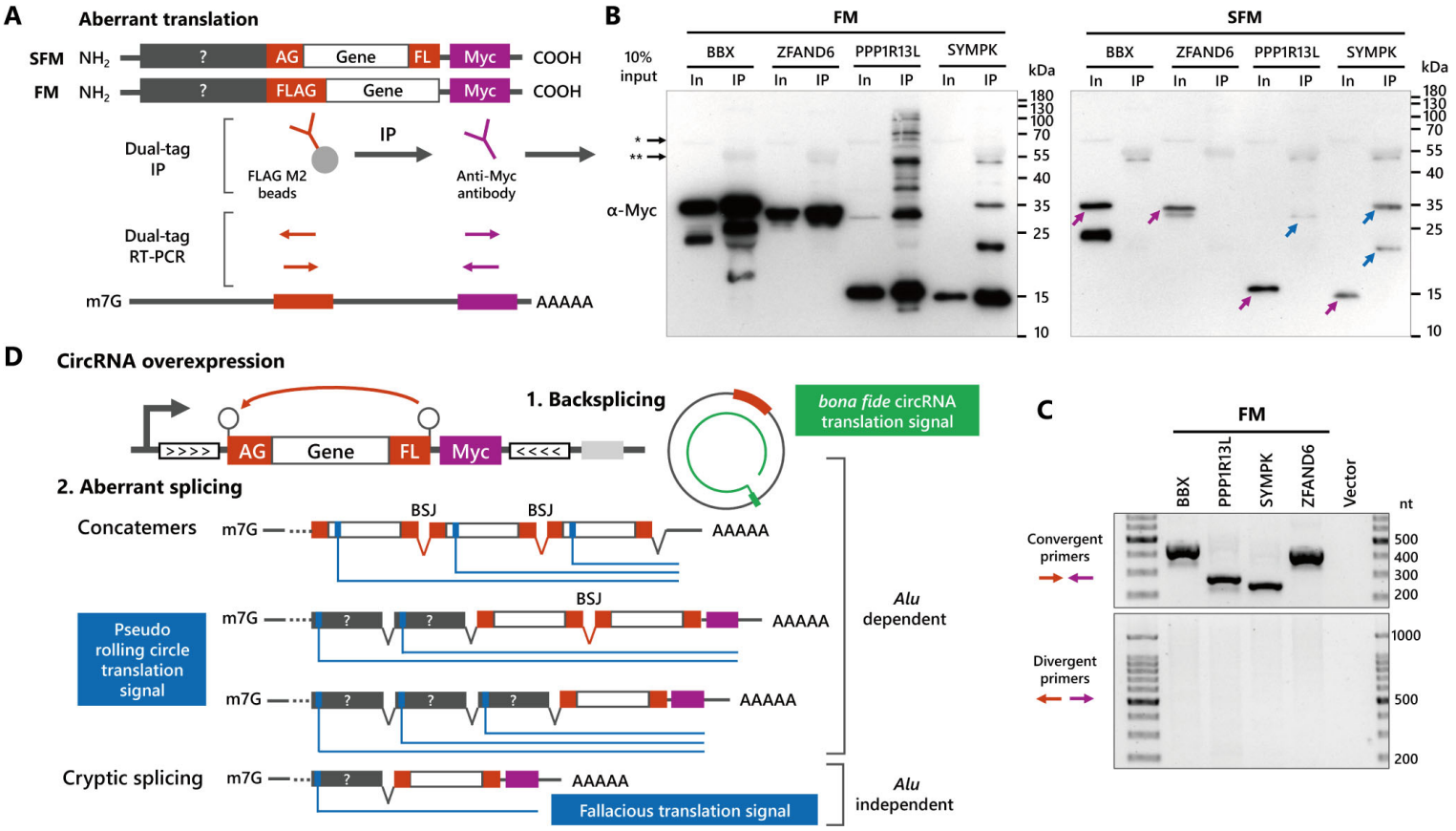


Figure 4. Heterogeneous outputs from complementary intron-mediated circRNA overexpression

A. Workflow for the examination of aberrant translation by dual tag IP and RT-PCR.

B. FLAG IP coupled anti-Myc western blot analysis of HEK293T transfected with FM and SFM constructs. Blue/purple arrows represent fallacious translation signals from linear coexistent transcripts with/without BSJ, respectively. The single and double asterisks denote the non-specific binding by endogenous protein and FLAG antibody fragment, respectively.

C. RT-PCR analysis of HEK293T transfected with FM constructs with divergent primer set of FLAG and Myc to detect circRNAs involving both FLAG and Myc sequence. RT-PCR using convergent primers as control.

D. Proposed various transcript outcomes of circRNA overexpression. The authentic translation from circRNAs is described as green curve, and the linear RNA derived translation is illustrated as blue lines.

References

1. J. E. Wilusz, A 360 degrees view of circular RNAs: From biogenesis to functions. *Wiley Interdiscip Rev RNA* **9**, e1478 (2018).
2. L. S. Kristensen *et al.*, The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.*, (2019).
3. L. L. Chen, The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol*, (2020).
4. X. Li, L. Yang, L.-L. Chen, The Biogenesis, Functions, and Challenges of Circular RNAs. *Molecular Cell* **71**, 428-442 (2018).
5. N. R. Pamudurti *et al.*, Translation of CircRNAs. *Mol Cell* **66**, 9-21 e27 (2017).
6. C. Chen, P. Sarnow, Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* **268**, 415-417 (1995).
7. Y. Wang, Z. Wang, Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**, 172-179 (2015).
8. Y. Yang *et al.*, Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res* **27**, 626-641 (2017).
9. X. Fan, Y. Yang, C. Chen, Z. Wang, Pervasive translation of circular RNAs driven by short IRES-like elements. *bioRxiv*, 473207 (2020).
10. N. Abe *et al.*, Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci Rep* **5**, 16435 (2015).
11. X. You *et al.*, Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* **18**, 603-610 (2015).
12. L. V. Stagsted, K. M. Nielsen, I. Daugaard, T. B. Hansen, Noncoding AUG circRNAs constitute an abundant and conserved subclass of circles. *Life Sci Alliance* **2**, (2019).
13. M. Guttman, P. Russell, N. T. Ingolia, J. S. Weissman, E. S. Lander, Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240-251 (2013).
14. T. B. Hansen, Signal and noise in circRNA translation. *Methods*, (2021).
15. M. Chekulaeva, N. Rajewsky, Roles of Long Noncoding RNAs and Circular RNAs in Translation. *Cold Spring Harb Perspect Biol*, (2018).
16. C. M. Weigelt *et al.*, An Insulin-Sensitive Circular RNA that Regulates Lifespan in Drosophila. *Mol Cell*, (2020).
17. J. Zhao *et al.*, Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus. *Nat Commun* **10**, 2300 (2019).
18. A. Costello, N. T. Lao, N. Barron, M. Clynes, Continuous translation of circularized mRNA improves recombinant protein titer. *Metabolic Engineering* **52**, 284-292 (2019).
19. R. M. Meganck *et al.*, Engineering highly efficient backsplicing and translation of synthetic circRNAs. *Mol Ther Nucleic Acids* **23**, 821-834 (2021).
20. R. M. Meganck *et al.*, Tissue-Dependent Expression and Translation of Circular RNAs with Recombinant AAV Vectors In Vivo. *Mol Ther Nucleic Acids* **13**, 89-98 (2018).
21. T. B. Hansen *et al.*, Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-388 (2013).
22. D. Liang, J. E. Wilusz, Short intronic repeat sequences facilitate circular RNA production. *Genes Dev* **28**, 2233-2247 (2014).
23. T. B. Hansen, in *Circular RNAs: Methods and Protocols,* C. Dieterich, A. Papantonis, Eds.

(Springer New York, New York, NY, 2018), pp. 143-157.
24. H. Ho-Xuan *et al.*, Comprehensive analysis of translation from overexpressed circular RNAs reveals pervasive translation from linear transcripts. *Nucleic Acids Res* **48**, 10368-10382 (2020).
25. I. Legnini *et al.*, Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol Cell* **66**, 22-37 e29 (2017).
26. F. O. Desmet *et al.*, Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**, e67 (2009).
27. A. L. Szymczak-Workman, K. M. Vignali, D. A. Vignali, Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harb Protoc* **2012**, 199-204 (2012).
28. C. M. Olson, M. R. Donovan, M. J. Spellberg, M. T. Marr, II, The insulin receptor cellular IRES confers resistance to eIF4A inhibition. *eLife* **2**, e00542 (2013).
29. S. van Heesch *et al.*, The Translational Landscape of the Human Heart. *Cell*, (2019).
30. Michael G. Kearse *et al.*, CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Molecular Cell* **62**, 314-322 (2016).
31. S. Dodbele, N. Mutlu, J. E. Wilusz, Best practices to ensure robust investigation of circular RNAs: pitfalls and tips. *EMBO Rep*, e52072 (2021).
32. J. L. Litke, S. R. Jaffrey, Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nature Biotechnology*, (2019).
33. C. C. Wu, A. Peterson, B. Zinshteyn, S. Regot, R. Green, Ribosome Collisions Trigger General Stress Responses to Regulate Cell Fate. *Cell*, (2020).
34. T. Zu *et al.*, Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* **108**, 260-265 (2011).
35. M. G. Kearse, J. E. Wilusz, Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes & Development* **31**, 1717-1731 (2017).

Acknowledgments

We thank Miaowei Mao and Tao Zu for the advice about *in silico* splicing signal prediction and non-AUG translation. This work was supported by National Natural Science Foundation of China (81874249), Guangdong Basic and Applied Basic Research Foundation (2020A1515011125), and Shenzhen Basic Research Grants (JCYJ20180223181224405, JCYJ20180507182657867).

Author contributions

Conceptualization: Y.J. and W. Z.; Methodology and investigation: Y.J.; Data analysis: Y.J.; Writing and visualization: Y.J.; Review and editing: Y.J. and X.C.; Project administration: Y.J.; Supervision and funding acquisition: W.Z.

Competing interests

The authors declare no competing interests.

Supplementary materials

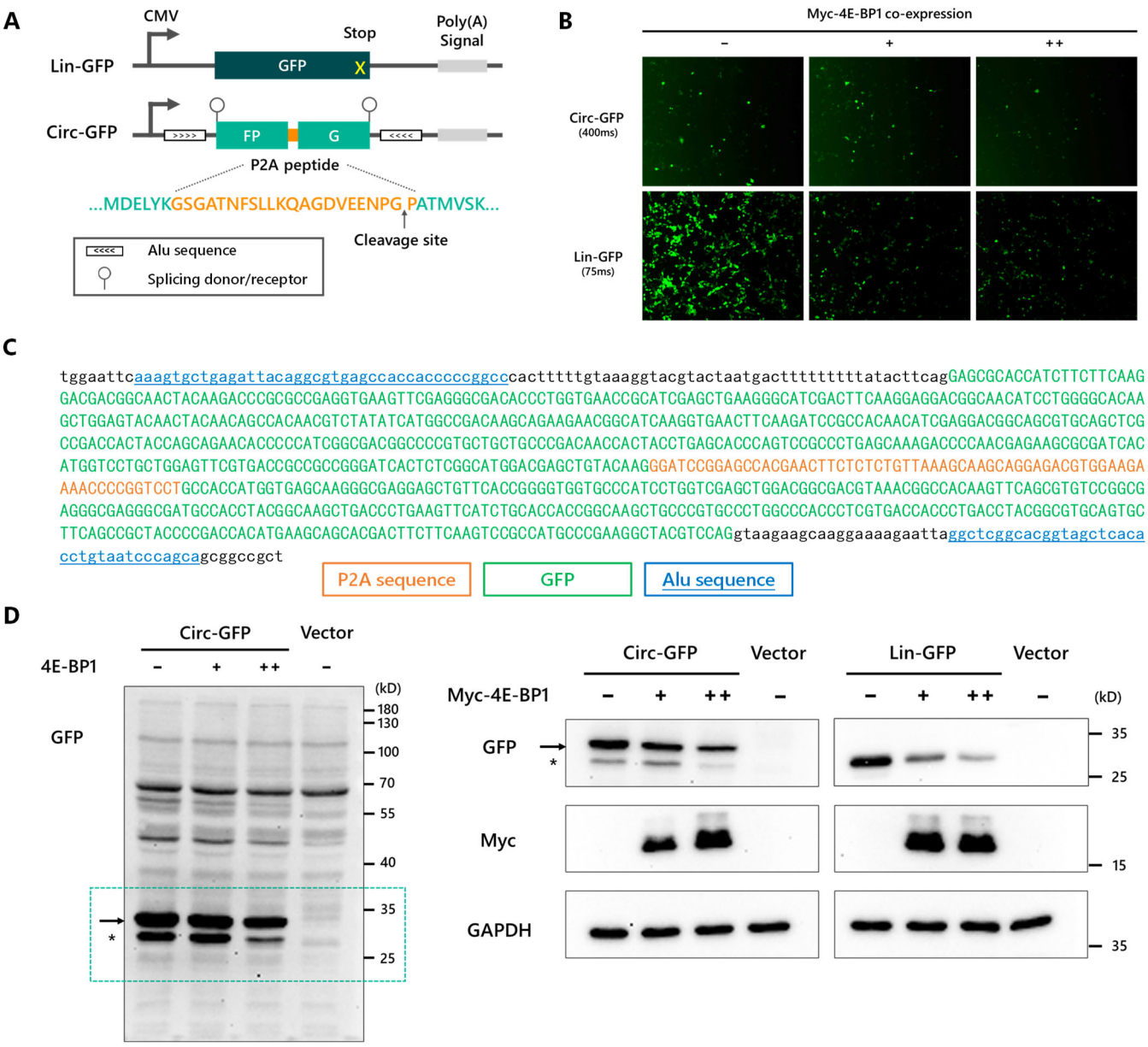


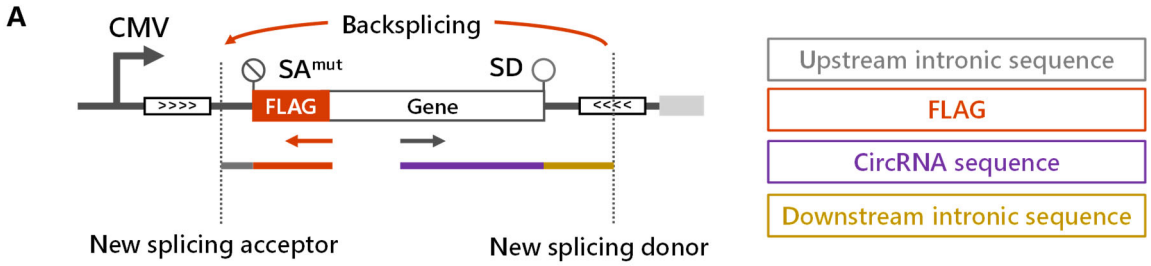
Figure S1. The functional translation of circular GFP RNA

A. Schematic diagram illustrates construct design of circular and linear GFP overexpression.

B. Functional GFP proteins expressed by the cir-GFP overexpression are resistant to the activity of 4E-BP1. HEK293T cells were transfected with 1.5 ng/well plasmids of circ-GFP or lin-GFP in 6-well plate, and 0, 0.5 and 1ng/well Myc-4E-BP1 was used for co-transfection. The fluorescent signal from cells were detected 40h after transfection by fluorescent microscopy, and the exposure time was 400 and 75ms for circ-GFP and lin-GFP, respectively.

C. The sequence of circ-GFP constructs. Intronic sequence is in lower case.

D. Western blot from samples in Fig. S1B. The overview of anti-GFP immunoblotting at long exposure time is presented at left, and the GFP monomer was used for analysis selected by green dotted box. The asterisk denotes non-specific bands from GFP fragments.



BBX SA-mut BSJ seq	TGCTACCAAGATACTAGCTGATTGGTGGGCTGTTCTTGATCCAAAGGAAAAGCAGAAATACACAGACATGGCCAAGGAGgtaagaagcaaggaaaagaattaggctcggcacggtagctcacacct/gtaaaggtacgtactaatgactttttttttatacttcacGATTACAAGGATGACGACGATAAGA
SYMPK SA-mut BSJ seq	GGCATCGATGGCATGACCACCTCAGAGAGGGTGGTGGATCTTCTGAACCAGGCGGCGCTGATCACCAATGACTCAAAGATCACAGTG(9bp_missing)/taatgactttttttttatacttcacGATTACAAGGATGACGACGATAAGA
ZFAND6 SA-mut BSJ seq	CAGTCACTTTTATCAGAATCTGTAGCATCTTCTCAATTGGACAGTACATCTGTGGACAAAGCAGTACCTGAAACAGAAGATGTGCAGgtaagaagcaaggaaaagaattaggctcggcacggtagctcacacct/gtaaaggtacgtactaatgactttttttttatacttcacGATTACAAGGATGACGACGATAAGAT

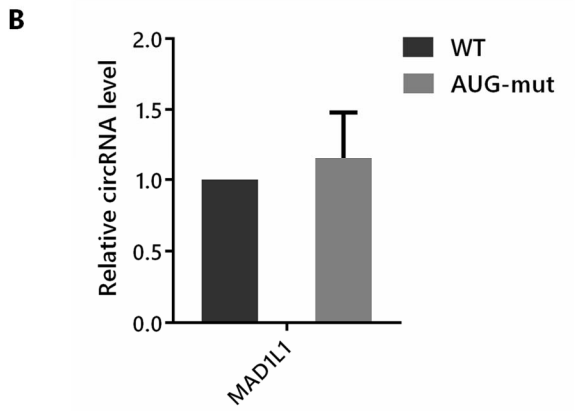
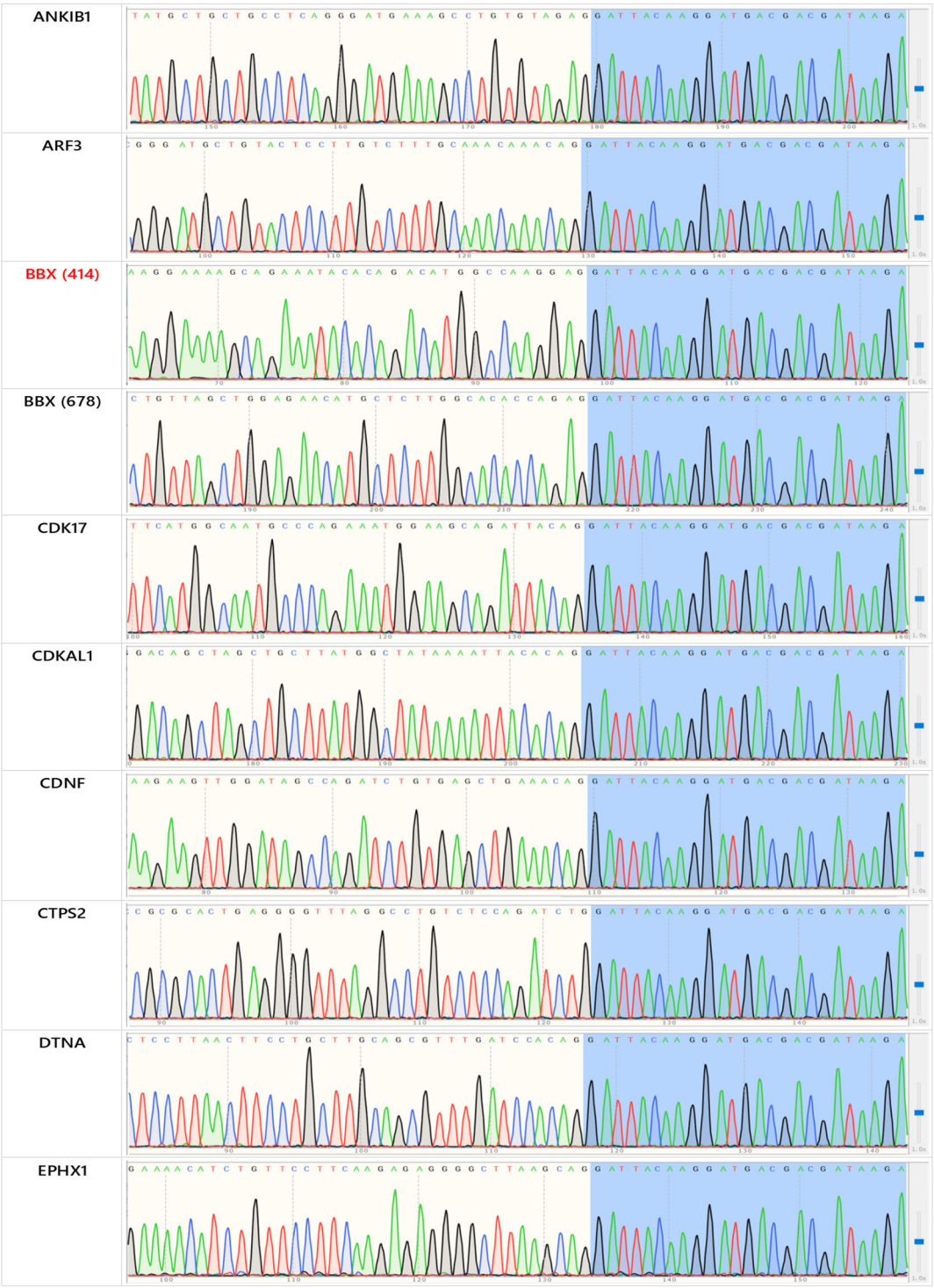


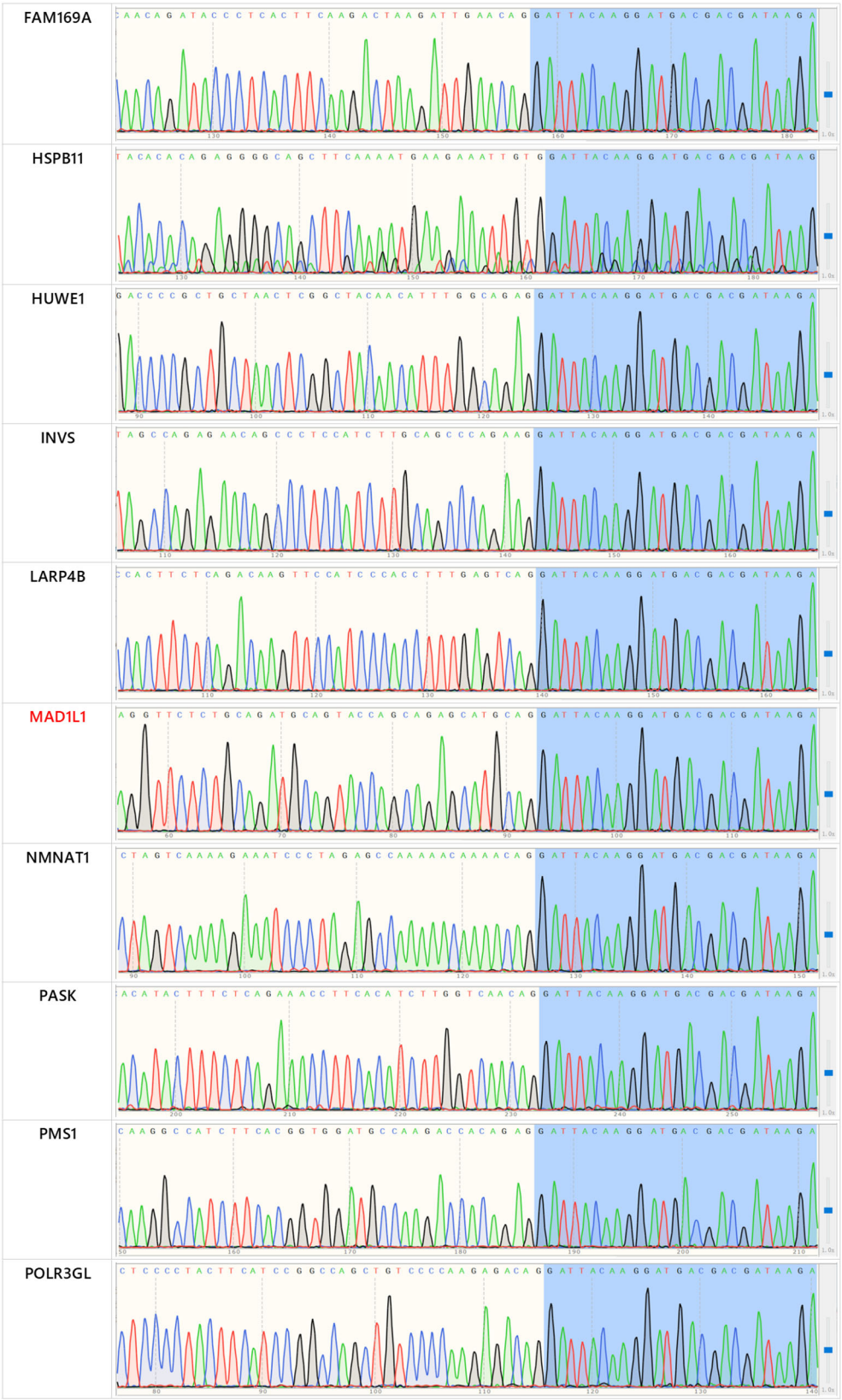
Figure S2. Analysis of transcripts from corresponding mutants (related to Fig. 2D and 3B)

A. Sequencing analysis of new BSJ formed from splicing acceptor mutation. An illustration describing new backsplicing happens from other splicing sites on vector when the putative splicing acceptor is mutated from AG to AC. Intronic sequence is in lower case, the mutated cytosine is highlighted in cyan, and the proposed new BSJ is presented by slash.

B. CircRNA levels of MAD1L1 were measured by q-RT-PCR with FLAG and gene specific divergent primers. CircRNA expression was normalized to the WT group. GAPDH was used for calibration. Data is presented as mean ±SD (n=3)

Table S1. BSJ validation by Sanger sequencing. The FLAG sequence is marked in blue. The BSJ is represented by the junction of two background colors.




FAM169A
HSPB11
HUWE1
INVS
LARP4B
MAD1L1
NMNAT1
PASK
PMS1
POLR3GL

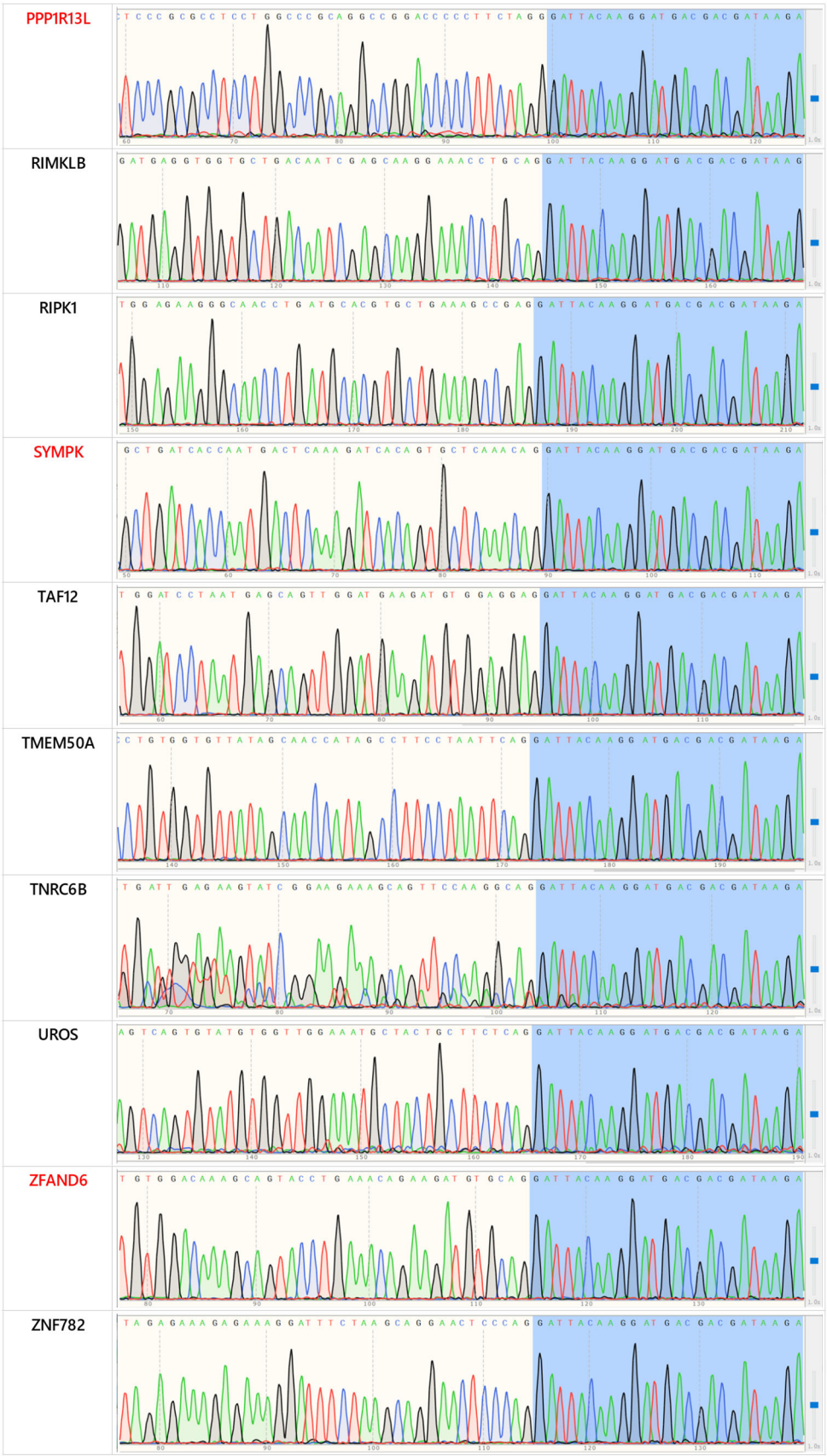

PPP1R13L
RIMKLB
RIPK1
SYMPK
TAF12
TMEM50A
TNRC6B
UROS
ZFAND6
ZNF782

Table S2. Sequence information of primers used in this study.

Primer	Sequence (5'-3')	Application
FLAG_R	CTTATCGTCGTCATCCTTG	FLAG specific primer for qRT-PCR, RT-PCR
ANKIB1_F	CAAATTATGATATCTGAAGGAGC	RT-PCR (with FLAG_R) and BSJ sequencing
ARF3_F	CTCTGGAGACACTACTTCC	RT-PCR (with FLAG_R) and BSJ sequencing
BBX(414)_F	GCCATCGCTCTCTTGTAC	RT-PCR, qRT-PCR (with FLAG_R) and BSJ sequencing
BBX(678)_F	GAAAGCAAATCCTGGCTAC	RT-PCR (with FLAG_R) and BSJ sequencing
CDK17_F	CCTATTGTGAAGAATGGCAG	RT-PCR (with FLAG_R) and BSJ sequencing
CDKAL1_F	GGAAGATTCAAAACCACAAG	RT-PCR (with FLAG_R) and BSJ sequencing
CDNF_F	TGCTATTATCTAGGAGCCAC	RT-PCR (with FLAG_R) and BSJ sequencing
CTPS2_F	GGCGTTTAGACAATTCCAG	RT-PCR (with FLAG_R) and BSJ sequencing
DTNA_F	CACTGAACTCAACGTGTC	RT-PCR (with FLAG_R) and BSJ sequencing
EPHX1_F	GAGATCCACGACTTACAC	RT-PCR (with FLAG_R) and BSJ sequencing
FAM169A_F	AATGTAGGCTTTGTACCTC	RT-PCR (with FLAG_R) and BSJ sequencing
HSPB11_F	GACCACCACAGGAATGT	RT-PCR (with FLAG_R) and BSJ sequencing
HUWE1_F	GTTCCATAGAGCATTTGAC	RT-PCR (with FLAG_R) and BSJ sequencing
INVS_F	ATCGTAGGAAACTCTGCTC	RT-PCR (with FLAG_R) and BSJ sequencing
LARP4B_F	CTGAGGTTAGCATTACACTTC	RT-PCR (with FLAG_R) and BSJ sequencing
MAD1L1_F	GGTTTTATCCACCCTGAG	RT-PCR, qRT-PCR (with FLAG_R) and BSJ sequencing
NMNAT1_F	GACTCTGAAGGTGCTAAG	RT-PCR (with FLAG_R) and BSJ sequencing
PASK_F	GCTCCTATTGTCTATCATCAC	RT-PCR (with FLAG_R) and BSJ sequencing
PMS1_F	ATTGAGGTGCGAGATAAC	RT-PCR (with FLAG_R) and BSJ sequencing
POLR3GL_F	GCCTTCTCCACTCTTCC	RT-PCR (with FLAG_R) and BSJ sequencing
PPP1R13L_F	TGAAACACATGGATCTGAAG	RT-PCR, qRT-PCR (with FLAG_R) and BSJ sequencing
RIMKLB_F	GATCAAGATGTGTAGTTCTG	RT-PCR (with FLAG_R) and BSJ sequencing
RIPK1_F	TTTCCACAGAACCCAGG	RT-PCR (with FLAG_R) and BSJ sequencing
SYMPK_F	GCATCACAGTTTTTCACTC	RT-PCR, qRT-PCR (with FLAG_R) and BSJ sequencing
TAF12_F	TAAAGATACCAGGCACTCC	RT-PCR (with FLAG_R) and BSJ sequencing
TMEM50A_F	TGGATTTCTAGAGGGCTTG	RT-PCR (with FLAG_R) and BSJ sequencing
TNRC6B_F	TCATCTTCACGGGCTTG	RT-PCR (with FLAG_R) and BSJ sequencing
UROS_F	GTTTGAGTTTTTGTCTCTTCC	RT-PCR (with FLAG_R) and BSJ sequencing
ZFAND6_F	CCCAGTCAGCATTAGAC	RT-PCR, qRT-PCR (with FLAG_R) and BSJ sequencing
ZNF782_F	TACAGAGATGTGATGCTGG	RT-PCR (with FLAG_R) and BSJ sequencing
GAPDH_F	ACAACTTTGGTATCGTGGAAGG	qRT-PCR and RT-PCR in RNase R assay
GAPDH_R	GCCATCACGCCACAGTTTC	qRT-PCR and RT-PCR in RNase R assay
copGFP_F	CTACGAGAACCCCTTCCTGC	qRT-PCR
copGFP_R	TGAAGATCACGCTGTCCTCG	qRT-PCR
FLAG_F	GATTACAAGGATGACGACGA	RT-PCR (with MYC_R)
MYC_F	CTCATCTCAGAAGAGGATCTG	RT-PCR (with FLAG_R)
MYC_R	CAGATCCTCTTCTGAGATGAG	RT-PCR (with FLAG_F)