

Progressive lengthening of 3' untranslated regions of mRNAs by alternative cleavage and polyadenylation in cellular senescence of mouse embryonic fibroblasts

Miao Han^{1,#}, Guoliang Lv^{2,#}, Hongbo Nie^{1,#}, Ting Shen¹, Yichi Niu¹, Xueping Li, Meng Chen¹, Xia Zheng³, Wei Li³, Chen Ding⁴, Gang Wei¹, Jun Gu², Xiao-Li Tian⁵, Yufang Zheng⁶, Xinhua Liu⁷, Jinfeng Hu⁸, Wei Tao^{2,*}, Ting Ni^{1,*}

¹State Key Laboratory of Genetics Engineering & Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200438 China

²School of Life Sciences, Peking University, Beijing, 100871 China

³Division of Biostatistics, Dan L. Duncan Cancer Center and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas, 77030

⁴State Key Laboratory of Genetics Engineering, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200438 China

⁵Institute of Molecular Medicine, Peking University, Beijing, 100871 China

⁶School of Life Sciences, Fudan University, Shanghai, 200438 China

⁷Department of Pharmacology, School of Pharmacy and Institute of Biomedical Sciences, Fudan University, Shanghai, 201203 China

⁸Department of Natural Products Chemistry and ‡Department of Pharmacology,

School of Pharmacy, Fudan University, No. 826 Zhangheng Road, Shanghai 201203,

People's Republic of China

[#]These authors contribute equally to this work.

*Correspondence should be addressed to W.T. (email: weitao@pku.edu.cn) or T.N.

(email: tingni@fudan.edu.cn).

ABSTRACT

Background

Cellular senescence has historically been viewed as an irreversible cell cycle arrest that acts to prevent cancer. Recent discoveries demonstrated that cellular senescence also played a vital role in normal embryonic development, tissue renewal and senescence-related diseases. Alternative cleavage and polyadenylation (APA) is an important layer of post-transcriptional regulation, which has been found playing an essential role in development, activation of immune cells and cancer progression. However, the role of APA in the process of cellular senescence remains unclear.

Materials and Methods

We applied high-throughput paired-end polyadenylation sequencing (PA-seq) and strand-specific RNA-seq sequencing technologies, combined systematic bioinformatics analyses and experimental validation to investigate APA regulation in different passages of mouse embryonic fibroblasts (MEFs) and in aortic vascular smooth muscle cells of rats (VSMCs) with different ages.

Results

Based on PA-seq, we found that genes in senescent cells tended to use distal pA sites and an independent bioinformatics analysis for RNA-seq drew the same conclusion. In consistent with these global results, both the number of genes significantly preferred to use distal pAs in senescent MEFs and VSMCs were significantly higher than genes tended to use proximal pAs. Interestingly, the expression levels of genes preferred to use distal pAs in senescent MEFs and VSMCs tended to decrease, while

genes with single pAs did not show such trend. More importantly, genes preferred to use distal pAs in senescent MFEs and VSMCs were both enriched in common senescence-related pathways, including ubiquitin mediated proteolysis, regulation of actin cytoskeleton, cell cycle and wnt signaling pathway. By cis-elements analyses, we found that the longer 3' UTRs of the genes tended to use distal pAs progressively can introduce more conserved binding sites of senescence-related miRNAs and RBPs. Furthermore, 375 genes with progressive 3' UTR lengthening during MEF senescence tended to use more strong and conserved polyadenylation signal (PAS) around distal pA sites and this was accompanied the observation that expression level of core factors involved in cleavage and polyadenylation complex was decreased.

Conclusions

Our finding that genes preferred distal pAs in senescent mouse and rat cells provide new insights for aging cells' posttranscriptional gene regulation in the view of alternative polyadenylation given senescence response was thought to be a tumor suppression mechanism and more genes tended to use proximal pAs in cancer cells. In short, APA was a hidden layer of post-transcriptional gene expression regulation involved in cellular senescence.

INTRODUCTION

Cellular senescence was originally described as a process that limited the proliferation of normal human fibroblasts in culture, and caused by the loss of telomeres after extensive proliferation in the absence of endogenous telomerase activity [1, 2]. In fact, besides telomere erosion, many stimuli and stress could also cause cellular senescence, including DNA double strand breaks, strong mitogenic signals, oxidative stress, loss of the PTEN tumor suppressor and ectopic expression of the cyclin-dependent kinase inhibitors (CDKIs) [3]. Morphological changes and molecular markers of senescent cells were identified in the past decades [4, 5]. These features include a flattened and enlarged cell morphology, absence of proliferative markers, senescence-associated β -galactosidase activity (SA- β -gal), expression of tumor suppressors, cell cycle inhibitors, and also DNA damage markers, including nuclear foci of constitutive heterochromatin and prominent secretion of growth factor, proteases, cytokines and other factors that have potent autocrine and paracrine activities [3, 6]. Noteworthy, no hallmark of senescence identified thus far is entirely specific to the senescent state, and not all senescent cells express all possible senescence markers [3].

Cellular senescence was generally viewed as an important mechanism for preventing the growth of cells at risk for neoplastic transformation [7, 8]. In recent years, it has become apparent that cellular senescence can also be involved in multiple physiological and pathological processes, such as normal embryonic development and tissue damage, suggesting that cellular senescence plays both beneficial and detrimental roles [6]. In younger individuals, cellular senescence blocks proliferation

of impaired cells, thus preventing them from transforming into cancer cells and leads to anti-cancer and anti-aging benefits [6, 9]. However, when people get old, a growing number of cells enter senescence, which causes decreased tissue function, elevated inflammation and exhaustion of stem cells, thus become pro-aging [6, 9]

Cellular senescence can be generally divided into replicative senescence and stress-induced premature senescence (SIPS) [6]. Primary cells including human diploid fibroblasts (HDFs) and mouse embryonic fibroblasts (MEFs) that cultured *in vitro* are often used as model systems to investigate the mechanism of replicative senescence[10, 11]. It is noteworthy that MEFs enter replicative senescence at regular culture condition where oxygen level is high, the senescence won't be established if given low oxygen tension[12-14]. Similarly, HDFs under more physiological O₂ concentration (3%) are achieved more numbers of population doublings than the O₂ concentration of air (20%)[15]. Further studies demonstrated that DNA damage response (DDR) contributes to *in vitro* replicative senescence for both HDFs and MEFs [12, 15]. Interestingly, MEFs have significant less numbers of population doublings (PD) than HDFs *in vitro*, making MEFs a time-saving model for investigation of replicative senescence.

Several signaling pathways have been functionally involved in the progression of cellular senescence and *in vivo* ageing including telomere attrition, epigenetic alternations, deregulated nutrient-sensing, mitochondrial dysfunction, altered intercellular communication, exhaustion of stem cells [9]. Notably, most of these pathways have not been evolved as direct regulators of ageing, such as nutrient

signaling is critical in promoting growth effects during embryogenesis and early development[16]. Intensive studies have attempted, through high-throughput genome-wide transcriptomics, to identify gene expression signatures that define cellular senescence. Many studies have found that dramatic changes in transcriptomes occur accompany the dramatic phenotype changes of senescent cells [10, 17-19]. Kim et al. profiled cellular senescence phenotype and mRNA expression patterns during replicative senescence in human diploid fibroblasts [10]. Interestingly, they found that the gene expression modules governing each stage supported the development of the associated senescence phenotypes [10]. Furthermore, Mazin et al. found widespread splicing changes during human brain development and aging [19]. Other studies also suggested that modification of mRNA processing may be a feature of aging [18, 20, 21]. However, our knowledge about regulation of the transcriptome in cellular senescence is still limited. RNA processing is a multi-step but tightly regulated pathway which includes transcription, splicing, editing, transport, stability, localization, translation and degradation [22]. What is more, co- and post-transcriptional regulation of gene expression is very complex and multifaceted, covering the complete RNA lifecycle from biogenesis to decay [23]. Increasing evidence supports that post-transcriptional regulation plays important roles during normal brain aging and cellular senescence through interacting between microRNAs and 3' untranslated region (3' UTR) of target genes, or interacting between RBPs (RNA binding proteins) and 3' UTR of cyclin-dependent kinase inhibitors, such as p16 [24, 25]. These results raised the possibility that changes of 3' UTR would also

contribute to cellular senescence and/or aging.

Cleavage and polyadenylation of nascent RNA is essential for maturation of the vast majority of eukaryotic mRNAs and determines the length of in the 3' UTR. Such process requires several cis-acting RNA elements and several dozens of core and auxiliary polypeptides [26]. The key cis-element that defines cleavage is a 6 nucleotide (nt) motif called the poly(A) signal (PAS), the canonical form of which is AAUAAA. However, it can adopt more than ten variants thereof and is mostly located in 40nt upstream of the cleavage site [26]. In addition to the PAS, additional U-rich sequence elements located upstream as well as U-rich and GU-rich sequence elements located downstream of the cleavage site also can enhance the efficiency of the 3' end processing reaction [26]. The 3' end-processing machinery, contained several sub-complexes, including cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I and II (CFI and CFII), as well as additional accessory such as Poly(A) polymerase (PAP), nuclear poly(A) binding protein 1 (PABPN1), Symplekin and RNA polymerase II (RNAAP II), is responsible for the cleavage and polyadenylation reaction[27]. With the surge of high-throughput sequencing technologies, genomic studies in the past few years have indicated that most eukaryotic mRNA genes have multiple polyadenylation sites (pAs) and thus multiple 3' UTRs, caused by alternative cleavage and polyadenylation (APA) [27]. For examples, about 72% genes in *Saccharomyces cerevisiae*, 70% in *Arabidopsis*, 30% in *C. elegans*, 43% in zebrafish, 79% in mouse, and 69% in human undergo APA [28-34]. In addition, 66% of mouse long noncoding RNAs have also been found to

have APA, and a large amount of APA events in long noncoding RNAs have also been found in different human tissues [29, 35].

Alternative pAs can reside in the 3'-most exon or upstream regions, leading to multiple mRNA isoforms that contain different coding sequences, 3' untranslated regions (3' UTRs) or both [26]. In fact, more than a hundred different conserved sequences have been identified in mRNA 3' UTRs alone and can be considered as putative regulatory elements [36]. These sequences are divided into two main classes depending on the nature of their trans-acting factors which can be either microRNAs (miRNA) or RNA binding proteins (RBPs). Both miRNA and RBPs have been reported to control gene expression by modulating the translational efficiency or degradation of their mRNA targets while the specific control of mRNA subcellular localization relies more exclusively on RBP interactions [37]. Therefore, APA can play a significant role in mRNA metabolism, such as stability, translation, and subcellular localization, through changing the interactions between miRNAs and/or RBPs with the 3' UTRs region by controlling the length of the 3' UTRs [38, 39].

Furthermore, dynamic regulation of 3' UTRs by APA has been reported in different tissue types, development and cellular differentiation, cell proliferation, cell reprogramming and cancer cell transformation [31, 35, 40-47]. Especially, a large scale analysis of five mammals showed that pA usage is more conserved in the same tissue across species than in different tissues within the same species, highlighting an essential role of APA in establishing tissue- specific gene expression profiles [28]. Sanberg et al. first reported the link between APA and cell proliferation, indicating

fast growing cells favored proximal pA sites [43]. Consistent with the association between cell proliferation and APA, cancer cells also preferred expression of mRNA isoforms with shorter connection 3' UTRs compared with normal cells[44, 48, 49]. Progressive 3' UTR lengthening coupled with weakening of mRNA polyadenylation activity was observed during mouse embryonic development [42]. Conversely, 3' UTRs become shorter and the expression of mRNAs encoding cleavage and polyadenylation (C/P) factors increased during reprogramming of somatic cells from several different cell types [47]. In addition, the regulation of APA is dictated by a combination of several features, including variations in the abundance or activity of *trans*-acting factors such as core 3' processing factors, tissue-specific RBPs, as well as splicing and transcription factors, combinations of *cis*-acting RNA elements, nucleosome positioning, DNA methylation and histone post-translational modifications [27, 50]. To sum up, alternative cleavage and polyadenylation of transcripts from the same gene enables cell-type- or condition-specific expression of 3' UTR isoforms, there by adding another layer of post-transcriptional regulation. However, the extent to which differential expression of these isoforms is used to regulate mRNA and protein levels in cellular senescence is completely unclear.

In this study, we applied our published polyadenylation sequencing (PA-seq) and strand-specific RNA-seq sequencing technologies, followed by systematic bioinformatics analyses and experiments to investigate APA regulation in different passages of mouse embryonic fibroblasts (MEFs) and in aortic vascular smooth muscle cells of rats (VSMCs) with different ages. Based on PA-seq and RNA-seq data,

this study found that in senescent cells, many genes tended to use distal pA sites, together with decreased expression levels and also preferred to enriched in common senescence-related pathways. Interestingly, we found that genes preferred to use distal pAs in cellular senescence tended to use proximal pAs in cancer cells [48]. This was consistent with the previous knowledge that senescence response was a potent tumor suppressive mechanism [8]. We further found that global 3' UTR lengthening in senescent cells might be explained by decreased expression of cleavage and polyadenylation related factors. These above results provide first evidence to suggest that APA is a hidden layer of gene's post-transcriptional regulation during cellular senescence.

RESULTS

Establishment of replicative senescence model for MEFs

To uncover the role of alternative polyadenylation (APA) in cellular senescence, we isolated and continuously subcultured mouse embryonic fibroblasts (MEFs) to serve as a typical replicative senescence model [11-14]. Population doubling curve showed that MEF cells underwent decreased growth rate (Figure 1A), a traditional marker for replicative senescence. Cells with population doubling time (PD) 6, 8, 10 and 11 were further analyzed for additional senescence marks. Mki67, an indicator for cell vitality and also a robust assessment of cellular senescence [51], showed decreased expression level that determined by quantitative reverse transcription PCR (qRT-PCR)

(Figure 1B). Flow cytometry revealed reduced percentage of S phase cells from PD6 to PD11 (Figure 1C). P16, a well-known cyclin-dependent kinase inhibitor that involved in multiple cellular senescence types, exhibited elevated expression at both RNA (Figure 1D) and protein level (Figure 1E). Senescence-associated β -galactosidase activity (SA- β -gal) also increased from PD6 to PD11 (Figure 1F). All these above evidences support that MEF cells from PD6 to PD11 underwent replicative senescence and can be further applied to investigate the potential regulation of APA.

Prevalent APA in protein-coding and noncoding transcripts in MEFs

To explore the potential role of APA in replicative senescence, we performed our published PA-seq protocol [35] to map the 3' end of mRNAs in 4 different passages of MEF cells (PD6, PD8, PD10, PD11), which showed multiple senescence markers in later population doubling, including slower growth rate and elevated SA- β -gal activity (Figure 1). Previous studies have demonstrated that PA-seq can not only accurately and comprehensively identify poly(A) sites (pAs) in mRNAs and noncoding RNAs, but also provide quantitative information on the relative abundance of polyadenylated RNAs [35]. As an independent survey of global APA regulation in cellular senescence, strand-specific RNA-seq for the same four MEF samples (PD6, PD8, PD10, PD11) was also carried out, followed by a dedicated data analysis pipeline using RUD index developed by Tian's lab [52]. In order to evaluate the influence of cell cycle phase, which also associated with cellular senescence, to APA regulation, we investigated the

APA of cells in G0 phase through serum starvation of PD6 cell (the sample is called G0 hereafter). These above experimental designs enable us to address how widespread is the APA events and what's the global changes of APA during replicative senescence.

In total, we obtained ~54 million strand-specific paired-end 101-mer reads for PA-seq. Out of these, ~63% (about ~34 millions) can be uniquely mapped to the mouse reference genome (see details in Supplementary Table S1). We applied F-seq, a feature density estimator [53] to compute PA cluster based on uniquely mapped reads. The PA-seq data from 4 samples (PD6, PD8, PD10, PD11) and also the G0 stage were combined so that a unified peak-calling scheme can be applied. A total of 29,144 PA clusters were obtained after applying the following two criteria: (i) To filter out any potential internal priming events during reverse transcription, we removed PA clusters with 15 'A' in the 20 nucleotides region downstream of peak mode (the most frequent position in the cluster) or with continuous 6 'A' in the downstream of peak mode; (ii) Any peak with less than 20 tags was then removed to obtain reliable PA clusters. We further evaluated the genomic distribution of these PA clusters. Among these 29,144 high quality peaks, majority (26,905 or 92.3%) were located within the genomic region of RefSeq gene annotation (also include the 250 bp upstream and the 5000 bp downstream of the gene body, which may be the potential region of the promoter or the distal pAs of the genes), suggesting most of the PA peaks belongs to annotated genes. The remaining 2,239 (or 7.7%) PA clusters were mapped to the intergenic regions, suggesting a considerable number of novel intergenic transcripts also contain

poly(A) tail. In order to focus the analysis on robustly expressed APA events, we kept clusters that accounted for $\geq 10\%$ of all the tags within respective genes in any sample. If a lowly expressed PA cluster contains $> 5\%$ of all the tags within respective genes in the majority of samples ($\geq 80\%$), the peak was also kept. By such filtering of lowly expressed peaks, we got 18,175 and 464 high confident PA clusters in coding genes and noncoding RNAs, respectively (Figure 2A, Supplementary Table S2). 65.8% of these 18,639 high confident PA clusters are included in the PolyA_DB 2 (Supplemental Figure S1A), which contains pA sites (pAs) for genes in several vertebrate species, including human, mouse, rat, chicken and zebrafish, using alignment between cDNA/ESTs and genome sequences [54]. We next examined the genomic distribution of identified PA clusters or sites located in protein-coding genes. 42% of the PA clusters are overlapped with known polyadenylation sites (Known pA), while 41% and 10% clusters are located in the annotated and extended 3' UTR regions, respectively (Figure 2B). In addition, a considerable proportion (4%) of the PA clusters fall into introns, and a small subset of the PA clusters is mapped to exons (2%) as well as the upstream regions of the transcripts ($<1\%$). Together, our result support the notion that 3'-end formation in the MEF transcriptomes is much more complex than previously appreciated as in the human transcriptomes [33, 35, 55]. We further evaluated the tag abundance in the top three categories of PA clusters (Known pA, 3' UTR and extended 3' UTR in Figure 2C). Consistent with previous knowledge that annotated pAs have the highest expression, 56% of all the tags covering the known polyadenylation sites (Known pA in Figure 2C). Further analysis showed that the

overall tag number distribution in known pA site also tended to be higher compared to those in 3' UTR or extended 3' UTR regions (Supplemental Figure S1B). In addition, the median number of covered tags for pAs located in tandem 3' UTR is significantly higher than in both exon and intron regions, respectively ($P < 0.001$ and $P < 0.001$, respectively, Mann-Whitney U test, see Supplemental Figure S1C). These above results indicate that pAs located in known pA sites, 3' UTR and extended 3' UTR regions are three major groups in MEF cells.

Nucleotide composition near pA sites was analyzed to further validate the reliability of those identified cleavage and polyadenylation sites. In agreement with previous studies [29], Adenine base (A) has higher fraction at -40 to -20 nucleotide (nt) before pA sites while Uridine base (U) peaks downstream of pA sites (Figure 2D). Motif analysis was further performed and classical poly(A) signal (AATAAA at DNA level, PAS) is expected to be the most significant one (Figure 2E). Consistent with the abundance of pA site categories, PAS showed a correlated value of significance (Figure 2E). Furthermore, the nucleotide preference at the site of cleavage is in the order of A>U>C>G for pAs located in different regions of genes (Supplemental Figure S1D). In addition, all pAs identified in this study, except for pAs located in exon, shown very similar features: canonical PAS could be found in the upstream of pAs (Supplemental Figure S1E and S1F) and there was relatively high U-content and lower for C/G-content surrounding the pAs (Supplemental Figure S1F). Such nucleotide preference was also observed in arabidopsis [34] and zebrafish [31]. What is more, frequencies of an additional set of cis elements, which defined by polyA_svm

[56] and may be involved in the regulation of APA [57], were relatively low in surrounding regions of pAs located in exons (Supplemental Figure S1G). The above results suggest a potential novel mechanism of choosing pAs in exons of eukaryotic genes. Collectively, these data provide strong evidence to support that the pA sites identified by PA-seq are most likely true.

Alternative polyadenylation sites were further investigated for both protein-coding genes and non-coding RNAs (ncRNAs). Our PA-seq analyses manifested that ~42% of all expressed protein-coding genes have two or more PA clusters (Figure 2F, Supplemental Figure S2A-B) and each gene, on average, has 1.66 APA isoforms, which is comparable with another study based on PAS-seq [33] and is significantly higher than the previous estimate based on the entire mouse EST database [55]. In addition, we identified 464 PA clusters covering 345 ncRNAs, and 26% of these ncRNAs have two or more PA clusters and the average number of PA clusters is 1.35 (Figure 2G, Supplemental Figure SC). Some genes with multiple 3' UTR isoforms may display biased expression to the top isoform. However, this is not generally the case. By excluding the expression percentage of the top isoform, more than 60% of the genes with two or more pAs express the remaining isoforms as a minimum of 10% of the total expression (Figure 2H). What's more, when excluding the top two isoforms, restricted to genes with three or more pAs, more than 40% of the genes express the remaining isoforms at an appreciable level of 10% of total expression (Figure 2H). Therefore, multiple pAs should be considered at the same time to avoid systematic bias that could be introduced if only two pA sites are

considered per gene at a time.

Based on above analyses, top three categories of PA clusters (Known pA, 3' UTR and extended 3' UTR) occupy 93% (Figure 2B) of all identified PA cluster numbers and 98% (Figure 2C) of tags inside all identified PA clusters. Therefore, combined tags in these three PA clusters (TPM, or the number of Tags Per Million reads) were used to evaluate the expression level of each gene with APA. To examine whether the PA-seq results reflect relative gene expression level, the same RNA samples were sequenced by strand-specific RNA-seq [58]. In total, we got ~104 million paired-end 101-mer reads, and about 92% of them (~96 millions) were mapped uniquely to mouse genome (see details in Supplementary Table S3). Specifically, we calculated the gene expression from the RNA-seq data with the FPKM (Fragment Per Kilobase of exon model per Million mapped paired reads) value. We only considered genes that expressed in all passages ($FPKM \geq 1$ and $TPM \geq 1$) for further analysis. PA-seq and RNA-seq profiles were considerably correlated ($\rho = 0.81, 0.83, 0.79, 0.78$ for PD6, PD8, PD10, and PD11, respectively, see Supplemental Figure 3, Supplementary Table S4). Therefore, the read count generated by the PA-seq approach can potentially be used to reflect transcript abundance. Totally, 3,165 genes which have multiple 3' UTRs caused by APA and with high Spearman's rank correlation coefficient from PD6 to PD11 between RNA-seq and PA-seq data were used for global APA regulation analyses in cellular senescence.

Progressive lengthening of 3' UTRs for genes with APA during replicative senescence of MEFs

Because our PA-seq data contain experimentally defined pAs as well as their steady-state expression level, it provides us a unique opportunity to interrogate global changes of 3' UTR length of APA genes among different passages of MEF cells. Based on our published paper [35], effective 3' UTR length (named Effective_UTR hereafter) was used to estimate the relative usage of pAs inside or downstream of 3' UTR regions for each gene based on a consolidated gene model by employing both the locations and the tag counts of the PA clusters identified in the 3' UTR of each transcribed locus (see Methods and Materials) [35]. It should be noted that, this is a considerable improvement over the previous strategies by using microarray or RNA-seq data, which only considered the most proximal and distal PA sites annotated for individual gene loci [52]. We are excited to find that effective 3' UTR showed a global progressive lengthening trend during the population doubling process of MEF cells (Figure 3A, Supplementary Table S4).

To further evaluate the changes of 3' UTR length at individual gene level, we compared the effective 3' UTR length in later passages (PD8, PD10, PD11) with earlier one (PD6) using different cutoff (20bp, 50bp, 100bp, 200bp and 500bp, Figure 3B). There were always more number of genes with longer effective 3' UTR than with shortened effective 3' UTR when compared to PD6 at different thresholds (Figure 3B). Noteworthy, the number of genes with lengthened effective 3' UTR was gradually increased from PD8 to PD11 when compared to PD6, while the number of genes with

shortened effective 3' UTR is gradually decreased from PD8 to PD11 when compared to PD6 (Figure 3B). These results further validated the global lengthening trend during the process of replicative senescence in Figure 3A.

In consistent with the PA-seq data, we also observed the same trend of global 3' UTR lengthening during replicative senescence in MEFs using the measurement of RUD index based on RNA-seq data by following the analysis pipeline of Zhe et al. [52] (Figure 3C, Supplementary Table S4). The change of RUD index between different passages was applied to evaluate the preference of longer or shorter 3' UTR for specific gene. In agreement with PA-seq results, more genes have longer 3' UTR in later passages (PD8, PD10 and PD11) than in earlier passage (PD6) (Figure 3D). In addition, the number of genes with higher RUD index (labeled lengthen in Figure 3D) is continuously elevated from PD8 to PD11 when compared to PD6, while the number of genes with decreased RUD index (labeled shorten in Figure 3D) is gradually reduced from PD8 to PD11 when compared to PD6 (Figure 3D). This independent RUD analysis confirmed the global lengthening of 3' UTRs during replicative senescence in MEF cells.

To ask whether cell cycle affects effective 3' UTR length, MEF cells at population doubling time 6 (PD6) were treated with serum starvation to force them enter G0 phase, a reversible cell cycle arrest state. We found the global 3' UTR pattern of G0 sample is more close to PD6 than to PD11, a more senescent status, by both PA-seq (Figure 3A) and RNA-seq (Figure 3C) data. This result implies that global lengthening of 3' UTR is more specific to cellular senescence (or irreversible cell

cycle arrest) than to reversible cell cycle arrest (G0 phase).

To validate that genes did undergo 3' UTR lengthening during cellular senescence, we randomly selected 10 genes for quantitative real-time Polymerase Chain Reaction (qRT-PCR) validation (Figure S4-S5, see details about primers information in Supplementary Table S5). Two genes that showed shortened 3' UTR in PD11 compared to PD6 were included as internal control (Figure S5). All of 10 genes showed an elevated trend of using longer 3' UTR (Figure S5), with 7 genes have reached a statistical significant level. As a confirmation that qRT-PCR approach did reflect the trend of 3' UTR length change, all of the two control genes showed a shortened 3' UTR usage in PD11 compared to PD6 with a statistical significance (Figure S5). The experimental validation supports the observation that senescent cells underwent global 3' UTR lengthening.

Longer 3' UTRs tend to have decreased mRNA abundance during cellular senescence

We next examine the consequence of 3' UTR lengthening during MEF cellular senescence. Longer 3' UTR is thought to provide more sites for regulation by microRNA and/or RNA binding protein (RBP) [42, 43, 59], which affects the mRNA abundance and/or translation at post-transcriptional level. In agreement with this hypothesis, a global decrease of gene expression was observed during cellular senescence for genes that have multiple pA sites (Figure 4A). Internal control genes with single pA site did not show such trend (Figure 4B), supporting the noting that

global down-regulation of expression level is specific for APA genes. Expression change at individual gene level was also analyzed. To simplify the conclusion, we compared PD11 MEF cell with PD6 cell for both 3' UTR usage and steady-state expression level. The genes that showed significant lengthened 3' UTR outnumbered genes containing shortened 3' UTR four to one (Figure 4C, Figure S6A, Supplementary Table S7). Within 3' UTR lengthened genes, there are more genes showed a significant downregulation than genes have elevated expression level ($P < 5.6 \times 10^{-12}$, Binomial test, Figure 4C, right half). No such difference was observed for genes with shorter 3' UTR (Figure 4C, left half). Together, decreased expression at individual gene level supports the observation of global down-regulation of APA genes during cellular senescence.

To address whether APA regulates gene expression in another aging system, we constructed PA-seq libraries for total RNA from vascular smooth muscle cells (VSMCs) derived from young and aged rats. PA-seq libraries were sequenced and analyzed using the same strategy (see read mapping information in Supplementary Table S6). To our surprise, a very close ratio of longer genes versus shorter genes regarding 3' UTR length between rat VSMC and mouse MEF cells was observed (Figure S6A, Supplementary Table S7). What's more, there is a significant overlap of genes showed longer 3' UTR usage between mouse MEFs and rat VSMCs (Figure S7-S8). Further, the correlation that more genes have down-regulated expression given lengthened 3' UTR also occurs in rat VSMCs ($P < 1.5 \times 10^{-5}$, Binomial test, Figure 4D). These above results support the notion that APA-mediated 3' UTR

lengthening may play a role in gene expression in multiple cellular senescence systems.

To further ask whether there is a continuous lengthening of 3' UTR at individual gene level and its consequence on gene expression, linear trend test was applied to identify genes with progressive 3' UTR length changes in MEFs [31, 60]. There were 375 genes showed a continuous lengthening of 3' UTR from PD6 to PD11, while only 73 genes had progressively shortened 3' UTR (Figure 4E, see details in Supplementary Table S7). Interestingly, a similar global expression decrease of these 375 genes correlated with their continuously lengthened 3' UTR (Figure 4F, Figure S6B). Interestingly, the 73 genes contained shortened 3' UTR during cellular senescence did not display trend of expression down-regulation (Figure S6C-D). Together, these above results indicated global 3' UTR lengthening may contribute to decreased mRNA steady-state expression level during cellular senescence.

Functional Enrichment analysis of genes with longer 3' UTR during cellular senescence

To address whether 3' UTR lengthening is related to regulation of the process of cellular senescence, genes with longer 3' UTR were analyzed for functional enrichment. We first selected genes that showed lengthening in PD11 compared to PD6 since these two passages of MEFs displayed dramatic phenotypic changes (Figure 1F). DAVID (The Database for Annotation, Visualization and Integrated Discovery) from NIAID (National Institute of Allergy and Infectious Diseases, NIH) was applied for the Gene Ontology (GO) and pathway enrichment analyses since this

tool is widely used [61]. For the 322 genes that have significant longer 3' UTR in PD11 compared with PD6 in MEFs, the top enriched GO terms are related to Ubiquitin mediated proteolysis, Wnt signaling pathway, Cell cycle, Regulation of actin cytoskeleton and others (Figure 5A, Supplementary Table S8), most of which have connections to cellular senescence ([62-65], see Discussions for details). Genes progressively tend to use distal pAs during replicative senescence of MEFs are also enriched in senescent-related pathways (Figure S9, Supplementary Table S8). Interestingly, most of the top enriched GO terms in rat VSMCs are also related to cellular senescence (Figure 5B, Supplementary Table S8). Strikingly, four of the GO terms (in red for Figure 5A and 5B) are shared between mouse MEFs and rat VSMCs. These shared GO terms or pathways all linked to cellular senescence (see Discussions for details). We further examined whether these two species (mouse and rat) only shared the same pathway but not the same gene or gene itself underwent APA regulation conservatively. Worth noting, six genes in the ubiquitin mediated proteolysis showed longer 3' UTR in both mouse and rat senescent cells (Figure 5C). There were 7, 2 and 5 genes in the regulation of actin cytoskeleton, cell cycle and wnt signaling pathway, respectively (Figure 5D-F), had the same trend of 3' UTR lengthening in both mouse and rat senescent cells. These above functional enrichment results supported the idea that an intrinsic link between APA genes and cellular senescence exists.

Longer 3' UTRs introduced more conserved miRNA binding sites and RNA binding protein recognition sequences

To explain why longer 3' UTR decreases mRNA expression level during cellular senescence, we analyzed the number and density of recognition sites by microRNA (miRNA) and RNA binding protein (RBP) in the alternative 3' UTR regions between proximal and distal pA sites. In consistent with previous studies, we found that the numbers of conserved miRNA binding sites in long 3' UTR (determined by distal pA site, named as 3' UTR_L hereafter) and alternative 3' UTR (between proximal and distal pA site, named as 3' UTR_A hereafter) were significantly higher than in short 3' UTR (determined by proximal pA site, named as 3' UTR_S hereafter) region ($P < 2.2 \times 10^{-16}$ and $P < 5.6 \times 10^{-7}$, respectively, two side Wilcoxon signed rank test, Figure6A) [43, 59]. Notably, 3' UTR_L and 3' UTR_A were also significantly longer than 3' UTR_S ($P < 2.2 \times 10^{-16}$ and $P < 2.2 \times 10^{-16}$, respectively, two side Wilcoxon signed rank test, Figure6B). Furthermore, the conserved miRNA binding density in 3' UTR_S was significantly higher than 3' UTR_L and 3' UTR_A ($P < 2.2 \times 10^{-16}$ and $P < 2.4 \times 10^{-16}$, respectively, two side Wilcoxon signed rank test, Figure6C). However, some conserved miRNA binding sites are exclusively located in alternative 3' UTR (Figure 6D). Interestingly, there was higher density of binding sites in 3' UTR_L than in 3' UTR_S for mir-20a-5p, which had been proved involved in cellular senescence of MEFs [66] ($P < 0.03$, two side Wilcoxon signed rank test, Figure 6E). In similarity, miR-290a-5p, which acted as a physiological effector of senescence in MEFs[67], also had more higher conserved binding density in 3' UTR_L and 3'

UTR_A than in 3' UTR_S ($P < 2.5 \times 10^{-3}$ and $P < 9.4 \times 10^{-3}$, two side Wilcoxon signed rank test, Figure 6F). These above results suggest that more potential miRNA binding sites can be introduced in longer 3' UTR, although the density of binding can be either high or low in the alternative region.

The RNA-Binding Protein DataBase (RBPDB) is a collection of experimentally validated RNA-binding sites both *in vitro* and *in vivo*, or manually curated from primary literature [68]. Interestingly, we found that the expression level of about 58% expressed RBPs (307 expressed RBPs with FPKM > 1 in PD6, PD8, PD10 and PD11), annotated by this database, was decreased during replicative senescence of MEFs. RBPmap can predict the binding sites of 94 human/mouse RBPs, 59 among them are also annotated by RBPDB [69]. Using this tool, we found that 30 expressed RBPs had significantly higher conserved binding densities in both 3' UTR_L and 3' UTR_A than in 3' UTR_S (FDR < 0.05, two side Wilcoxon signed rank test, see details in Supplemental Table S9). Specifically, HNRNPA1 [70] and IGF2BP2 [71] can be involved in mRNA translocation, YBX1 [72] can regulate mRNA stability and NOVA1 [72], SNRNP70 [73], HNRNPF [74], HNRNPH2 [75], HNRNPK [76] can regulate APA (Figure 7B-J). Taken together, these above data indicated that longer 3' UTRs introduced more sites for RNA binding proteins, which may further affect mRNA stability, localization and even contribute to alternative polyadenylation itself.

Decreased expression of cleavage and polyadenylation related factors may regulate global 3' UTR lengthening

There are three types of regulators that controlling alternative polyadenylation, including trans-acting RNA binding proteins (RBPs), cis-acting DNA elements and chromatin features near pA sites [38]. Cis elements around proximal or distal pA sites were first analyzed to get insights of the APA regulation during cellular senescence. We compared the sequence conservation around the regions of the proximal and distal pAs of genes that showed progressive lengthening of 3' UTR (named as LE_proximal and LE_distal, respectively) to other three groups of pA sites, including i) proximal and distal pAs of genes that showed progressive shortening of 3' UTR (named as SH_proximal and SH_distal, respectively), ii) proximal and distal pAs of genes without significant changes of pAs usage preference (named as NC_proximal and NC_distal, respectively) and iii) pA regions of genes only contain one pA site (named as Single pA). Interestingly, polyadenylation signal (PAS) upstream LE_distal pAs was more conserved than that of LE_proximal pAs, implying that reduced usage of shorter 3' UTR during cellular senescence may be mediated by weaker proximal PAS (Figure 8A). Supporting this notion, the percentage of 'AAUAAA', the strongest PAS, in LS_distal pA sites is the highest among all these seven types of pA sites (Figure 8B). In addition, the downstream sequence conservation scores of proximal pA sites (including LE_proximal, SH_proximal and NC_proximal) are obviously higher than that of distal pA sites (including LE_distal, SH_distal and NC_distal), this may attribute to lower frequencies of strong PAS in proximal pAs than distal pAs,

therefore needed to provide more binding sites for RBPs to help they be used (Figure 8B), and also may attribute to 3' UTR is more conserved than intergenic regions, given downstream sequence of proximal pA site is still in the alternative 3' UTR region (Figure 8A). Interestingly, among three types of proximal pA sites and three types of distal pAs, downstream sequence conservation score of pAs with significantly usage shift (LE_proximal, SH_proximal, LE_distal, SH_distal) are more conserved than pAs without significantly usage shift, this may also attribute to more binding sites for RBPs needed to be supply to help these pAs with significantly usage shift to be used (Figure 8A). Furthermore, among three types of proximal pA sites, downstream sequence conservation score of LE_proximal is higher than SH_proximal and NC_proximal (Figure 8A), raising the possibility that genes showed lengthening of 3' UTR during cellular senescence were regulated by providing more conserved binding sequence between proximal and distal pA sites.

We next examined the possibility that RBPs control APA during cellular senescence. While the core factors in the 3' processing machinery play essential roles in the cleavage and polyadenylation (C/P) reaction, regulation of their expression level had also been implicated in APA regulation [27]. Therefore, the gene expression of 22 important poly(A) trans-factors [81] were analyzed based on the MEF replicative senescence RNA-seq data (PD6, PD8, PD10, PD11). Consistent with the overall expression trend of RBPs, we observed global down regulation of most poly(A) trans-acting factors during replicative senescence (Figure 8C), in which more genes tend to use longer 3' UTR. This result is in line with a recently published study,

in which global upregulation of most poly(A) factors is associated with more 3'-UTR-shortening events in five tumor types compared with normal tissues [48]. Another survey of the constitutive components of the polyadenylation machinery and other candidates from the literature revealed several trans-acting factors that were significantly upregulated in the cancer cell lines compared with non-transforming cells [44]. Specifically, consistent with the idea that cellular senescence can be a defense system for cells on the way to becoming cancerous [8], we noticed mRNA level of *Cstf2*, which was the first core C/P factor that can regulate APA [77], was down-regulated during MEF cellular senescence (Figure 8C) and up-regulated in cancer cells [44, 48]. Furthermore, Yao et al. showed knockdown of CstF64 (coded by *Cstf2* gene) led to more genes favored distal pA sites, which contributed to lengthening of 3' UTR in human HeLa cells [78]. We thus validated the mRNA expression and protein production of *Cstf2* gene during MEF cell senescence. qRT-PCR (Figure 8D) and Western Blot (Figure 8E) confirmed that *Cstf2* did show a reduced trend of mRNA level and protein expression, which may contribute to the 3' UTR lengthening during cellular senescence. Notably, Li et al. found that *Pcf11* could enhance usage of proximal pAs and *Pabpn1* could promote usage of distal pAs by using siRNA knockdown coupled with deep sequencing [79]. Interestingly, the expression of *Pcf11* was down-regulated and *Pabpn1* was up-regulated during replicative senescence of MEFs, consistent with the global 3' UTR lengthening pattern. Taken together, we provided evidence that changed expression of cleavage and polyadenylation related factors such as CstF64, Pcf11 and PABPN1, may regulate

global 3' UTR lengthening during MEF cellular senescence, although the speculation deserves further experimental validation in MEF cells.

DISCUSSION

An increasing number of reports demonstrate the prevalence of global APA regulation in a wide range of biological processes and pathological diseases [26, 39, 48]. For instance, more genes favor distal pA site usage, leading to global 3' UTR lengthening in cellular differentiation and development [42]. However, upon cellular proliferation, neuron activation and malignant transformation, more genes tend to favor proximal pA site usage, resulting in global 3' UTR shortening [43, 44, 46, 80]. Cellular senescence is regarded as a cancer prevention mechanism and a contributing factor for individual ageing[8, 9]. Nevertheless, whether global APA regulation exists in this important process is completely unclear. We thus applied our PA-seq method to profile global APA events by using mouse embryonic fibroblast (MEF) replicative senescence as a model system [11-14]. If grown under standard tissue culture conditions, MEFs proliferate limited number of population doublings (PD) and eventually undergo replicative cellular senescence[12, 14]. DNA damage, losses of centrosomal integrity and accumulation of mutations due to oxidative stress may also contribute to the entry of MEFs into senescence in culture [14]. DNA damage response (DDR) is further proved to be activated when MEFs undergo replicative cellular senescence [12]. However, the genome-wide molecular consequence of such signal is unclear.

Interestingly, global 3' UTR lengthening during MEF replicative senescence was observed. What's more, we found the expression levels of genes with multiple pAs showed a decreased steady-state expression during cellular senescence. Furthermore, genes both preferred to use distal pAs in senescent MEFs and showed progressive lengthening of 3' UTR during MEF cellular senescence have a decreased trend of steady-state mRNA expression level (Figure 4). In contrast, this gene expression decreasing pattern in senescent cells cannot be observed for genes either with single pA site or tended to use proximal pAs (Figure 4B and Figure S6D). The observation that genes used proximal pA sites in senescent cells did not show a trend of increased expression is consistent with the result that genes preferred longer 3' UTR in cancer cells did not have a decreased trend of expression[48]. Since global lengthening of 3' UTR in senescent cells and global shortening of 3' UTR in cancer cells were observed, these above results support the notion that APA-mediated 3' UTR lengthening serves as a posttranscriptional regulation layer in MEF cellular senescence.

To address whether this phenomenon of 3' UTR lengthening in cellular senescence is specific for mouse or can extend to other species, we sequenced and analyzed PA-seq data derived from old and young rat's aortic vascular smooth muscle cells (VSMCs). Intriguingly, there are also more genes prefer distal pAs usage in elder rat than in younger rat (Figure 4D), suggesting global 3' UTR lengthening in older cells is evolutionary conserved. What is more, genes preferred distal pAs also tended to display a trend of decreased gene expression. Most strikingly, shared pathways (Figure 5A,B) and genes (Figure 5C,D,E,F) preferred to use distal pAs both in

senescent MEF cells and in rat VSMCs derived from older rat were also be found. Consistent with the above results, Xia et al. observed broad shortening of 3' UTRs across seven TCGA (The Cancer Genome Atlas) tumor types and that genes with shorter 3' UTRs in tumors were prone to be up-regulated [48]. Furthermore, many genes were found to have opposite pAs usage preference between senescent cells and cancer cells (Supplemental Figure S10). On the one hand, this was consistent with the previous knowledge that senescence response was a potent tumor suppressive mechanism [8]. On the other hand, this also implied that APA regulation may play a vital role in deciding cell destiny, although definitely worth further intensive investigation.

Excitingly, genes with longer 3' UTR in senescent MEF cells and rat cells derived from older rat's vascular smooth muscle are all enriched in common pathways such as ubiquitin mediated proteolysis[62, 81, 82], regulation of actin cytoskeleton[65, 83], cell cycle[64] and wnt signaling pathway[63, 84], which was proved to be involved in cellular senescence and cell cycle arrest (Figure 5 and Supplemental Figure S9). Interestingly, the genes which were preferred to use proximal pAs and with increased expression levels in cancer cells were also enriched in ubiquitin mediated proteolysis and regulation of actin cytoskeleton pathways[48]. This also implied that the APA regulation of genes involve in these two pathways may play some important role in cell fate decision, such as senescence or tumorous.

Proteostasis is maintained by the proteostasis network, which comprises pathways that control protein synthesis, folding, trafficking, aggregation,

disaggregation, and degradation [85]. Notably, the identification of shared “ubiquitin mediated proteolysis” pathways between mouse and rat old cells is of interest as these machineries play a vital role in protein homeostasis and their functionality declines with age [85-87], while perturbations of the ubiquitin-proteasome pathway have been involved in pathogenesis of various neurodegenerative disease including Alzheimer's disease (AD) and (Huntington disease) HD [88, 89]. Specifically, NEDD4, an E3 ubiquitin ligase regulated by APA both in replicative senescence of MEFs and senescent rat VSMCs (Figure 5C), was shown to delay cellular senescence by degrading PPAR γ and eventually upregulating SIRT1[90]. Interestingly, NEDD4 was also be proved to be a potential proto-oncogene that negatively regulated a conserved component of the IIS (insulin/insulin-like growth factor-1-signaling) cascade, the phosphatase PTEN via ubiquitination, a paradigm analogous to that of Madm2 and p53[91]. In addition, Cul1, which belong to a Skp1-Cul1-F-Box E3 ligase complex, also involved in regulating the lifespan of *Caenorhabditis elegans* [92]. Interestingly, a Cul1 complex has also been shown to inhibit the activity of FOXO proteins by catalyzing their degradation [93-95]. Cul1 also can promote cell-cycle progression at the G1-S phase transition through regulating the abundance of cyclin E [96]. Furthermore, the full-length 3' UTR of *Cul1* gene, which was proved to use distal pAs in senescent MEFs by qRT-PCR (Figure S11A-B), has significantly lower stability ($P= 3.5 \times 10^{-3}$, $P= 1.5 \times 10^{-3}$ and $P= 2.2 \times 10^{-3}$ for 4, 8, 12 hours after actinomycin D treatment, respectively, two-tailed t-test, Supplementary Figure S11C). To assay how different 3' UTR isoforms of *Cul1* influence protein expression, shorter,

full-length and full-length with depletion of the sequence containing the proximal PAS (named as 3' UTR_S, 3' UTR_L and 3' UTR_M, respectively) were fused to a dual-luciferase reporter assay system. Both 3' UTR_L and 3' UTR_M yielded lower luciferase activity than the construct containing only the common (or shorter) 3' UTR regions in NIH3T3 mouse fibroblast cell line ($P < 0.04$ and $P < 0.02$, respectively, two-tailed t-test, Supplementary Figure S11D). These above data further support the idea that APA may be involved in cellular senescence via affecting the mRNA stability and protein translation efficiency of genes in senescence-related pathways.

Gourlay et al. demonstrated that increasing actin dynamic, either by a specific actin allele or by deletion of a gene encoding the actin-bundling protein Scp1p, can increase lifespan of yeast by over 65%, which may due to lower ROS level was produced [97]. In addition, CDK5 activation was upregulated in senescent cells, which further accelerated actin polymerization and induced senescence and the senescent shape change by reducing GTPase Rac1 activity [98]. Florian et al. also found that an unexpected shift from canonical to non-conical Wnt signaling in mice due to elevated expression of Wnt5a in aged haematopoietic stem cells (HSCs) appeared to induce stem cell ageing [99]. Interestingly, we found both Rac1 and Wnt5a tended to use distal pAs both in replicative senescence of MEFs and senescent rat VSMCs. In conclusion, our presented studies have revealed that APA regulation may be involved in several signaling pathways which are, likely, implicated in the molecular phenotypes of cellular senescence. However, whether these alterations reflect responses to the progression of ageing or causal mechanisms should await

further experimentation.

In agreement with previous studies, we found that alternative 3' UTR regions between proximal and distal pA sites contained higher numbers of conserved miRNA seed matches than common (i.e. shorter) 3' UTR regions (Figure 6A) [43]. The alternative 3' UTRs also contain 70 unique miRNA binding sites that do not appear in the common region (Figure 6D), providing additional cis elements for posttranscriptional regulation. Interestingly, the average length of alternative 3' UTR is larger than common 3' UTR (Figure 6B), leading to the density of potential seed matches of miRNAs in alternative 3' UTR is lower than that of common 3' UTR (Figure 6C). It is worth noting that the overall number is more important than the density of miRNA binding sites in posttranscriptional regulation. Excitingly, a few senescence-associated miRNAs even have both larger number and higher density of potential miRNA binding sites in alternative 3' UTRs. As examples, miR-20a-5p and miR-290-5p, both of which have higher conserved miRNA binding density in alternative 3' UTR regions (Figure 6E,F), have been proved to induce senescence in MEFs [66, 67]. In addition, miR-93 and miR-494, which also expressed in MEFs based on Nidadavolu et al.'s data and contained higher seed binding site densities in long 3' UTRs based on our data, also involved in senescence of MFEs [100-102]. Besides miRNAs, many RBPs, which may regulate the stability or translocations of mRNA [70-72], have higher densities in full length (i.e., long) and alternative 3' UTRs than in short 3' UTR regions (Figure 7). Therefore, using distal pAs in senescent MEFs provide multiple opportunities for miRNAs and RBPs binding with 3'

UTR regions of mRNAs, which may regulate mRNA stability, translation and translocation.

In consistent with previous investigations, we found that the distal pAs which preferred to be used in senescent MEFs tend to have more conserved and more strong poly(A) signal (PAS) [26, 44]. Notably, the correlation of using distal pAs with generally down-regulated of core factors in C/P reactions is in agreement with the observation that up-regulation of core factors is associated with proximal pAs preference in cancer cells [44, 48]. In general, we speculate that a combination effect of the expression level changed RBPs eventually determined the global 3' UTR lengthening, although some of the RBPs may favor proximal pA site usage or have context dependent roles for different genes [103-105]. As a considerable number of gene's APA can be affected by certain RBPs, it is reasonable that some of these RBPs can have observable phenotypic changes upon knockdown. Taken together, all these above results raise an intriguing model for the regulation role of APA in cellular senescence. Upstream signals down-regulate most of the RBPs, including core RBPs for cleavage and polyadenylation, some of which have significant roles in controlling the APA of genes, thus leads to a global lengthening of 3' UTR; Longer 3' UTRs trend to contain more miRNA binding sites and/or more cis-elements that affect RNA stability and/or translation efficiency, eventually affect the protein level of genes that related to cellular senescence. However, more experimental validations are warranted to fully prove such hypothesis.

Notably, there are two major categories of cellular senescence including

replicative senescence and stress-induced premature senescence (SIPS) [6]. Here we proved that MEFs replicative senescence underwent global 3' UTR lengthening. VSMCs derived from old rat are likely a combination of replicative senescence and varieties of stress-induced senescence. Although we discovered more genes use longer 3' UTR, whether SIPS itself will induce global 3' UTR lengthening remains elusive. Thus more cellular senescent models are required to fully understand the prevalence and functional relevance of 3' UTR lengthening, which serves as an important posttranscriptional gene regulation mechanism. Importantly, PA-seq and RNA-seq only can manifest the steady-state level of mRNAs, nascent transcript sequencing such as GRO-seq[106] and Bru-seq[107], may tell us whether the genes tended to use distal pAs during transcription in senescent cells. In addition, in this study, we only predicted the potential binding sites for miRNAs and RBPs in 3' UTR regions. miRNA expression data and also CLIP-seq [108] or PAR-CLIP seq [109] data for RBPs can help us further investigate how miRNAs and RBPs shape the mRNA fate after transcription. Furthermore, TAIL-seq, which allows us to measure poly(A) tail length at the genomic scale, is also a potent tool that can be used to dissect dynamic regulation of mRNA turnover and translational control [110]. Recently, Lars et al. developed BAT-seq to analyze single-cell isoform data, and found that single cells from the same state differed in the choice of APA isoforms [111]. In this study, we only investigated the pA usages of genes from a cell population, whether a more senescent cell is more prone to use distal pA can be further investigated by using the same or comparable method.

In conclusion, these above results provide evidences to support the intrinsic link between APA and senescence in multiple species, although functional validation and underlying mechanism deserves further investigation.

Materials and Methods

Cell isolation and cultivation

For replicative senescence, primary MEFs isolated from embryos of 12.5-14 days' pregnant C57BL/6 mice according to the method described previously [62] were cultured in DME (GIBICO, Grand Island, NY, USA), supplemented with 10% fetal bovine serum (FBS, GIBICO, Grand Island, NY, USA) and 1% Penicillin/Streptomycin (GIBICO, Grand Island, NY, USA), at 37 °C in a humidified incubator with 5% CO₂. To develop replicative senescence, confluent MEFs were evenly transferred into new dishes and the cells were cultured until get confluent again to generate one population doubling (PD). MEFs were maintained in a 100-mm dish, and in this study, confluent status of cells means almost confluent culture condition without contact inhibition. The number (n) of PD was calculated using the equation ($n = \log_2 (N_e/N_s)$), where N_e and N_s are the numbers of cells at the end of cell culture and those seeded at the start of one passage, respectively. The numbers of population doublings (PD, times) and total culturing period were monitored one time every three days until MEFs completely stopped proliferating.

Cell cycle analysis by FACS

MEFs were grown in 100-mm dishes and harvested with 1 ml of 0.25% trypsin when reached 70-80% confluence. And then cells were washed twice with ice-cold 1×PBS, fixed with 70% ethanol for 2 hours in 4 °C and washed again with the same PBS before penetrated with 1% Triton X-100 plus RNase A (40 U/ml) and stained with propidium iodide (50 µg/ml) on shaking platform at 37 °C. As staining finished, DNA contents were measured on a FACS Calibur HTS (Becton Dickinson). The percentage of diploid cells in G1, S and G2/M phases were analyzed by Summit version 5.0 software (Dako, Campbellfield, Australia).

SA-β-gal staining

SA-β-gal activity, as a classic biomarker of senescent cells, was also monitored at the same time for all the growth stages of MEFs. Culture medium was removed from 12-well plates and cells were washed once with 1 ml of 1×PBS and fixed with 0.5 ml of fixative solution for 10-15 minutes at room temperature. While the cells were in the fixative solution, the staining solution mix (staining solution, staining supplement and 20 mg/ml X-gal in DMSO) was prepared according to the manufacturer's instructions (Biovision, Cat. #K320-250, USA). As washed twice with 1 ml of 1×PBS, cells were incubated with the aforementioned mix at 37 °C overnight or 12 hours. Then the cells were observed under an inverted microscope (Leica, Germany).

qRT-PCR and Western Blot

MEFs were harvested at 70-80% confluence. Total RNA and protein were isolated with TRIzol Reagent according to the provided manual (Invitrogen, USA). For qRT-PCR analysis, RNA was reverse-transcribed into cDNA using random primers, and mRNA levels of *Mki67*, *p16* and *Cstf2* were measured by qPCR (LightCycler, Roche, Swiss). See Supplementary Table S10 for primer information.

For western blot, protein was resolved in 1% SDS, subjected to SDS-PAGE, transferred to nitrocellulose membrane and incubated with primary antibodies of interest (p16, 1:500-1000, sc-1207, Santa Cruz; Cstf2, 1:1000, ab64942, abcam; GAPDH, 1:2000-3000, sc-32233, Santa Cruz) overnight at 4 °C. In the second day, the blots were washed three times in PBST (1 × PBS + 0.1% Tween-20) and incubated with diluted secondary antibodies (IRDye800CW Goat anti-Mouse (1:10000, Cat.926-32210, LI-COR) for 1-2 hours at room temperature. And then the protein blots were washed three times with PBST before visualized under odyssey infrared imaging system (Odyssey, LI-COR).

Validation of pAs usage preference by qRT-PCR

RNAs were reverse transcribed with oligo(dT), followed by PCR with two pairs of primers (proximal and distal) targeting different regions of the cDNAs. Specificity, the region targeted by the proximal pair is common to both APA isoforms and the region by the distal pair is unique to the longer isoform. qPCR signals from the

proximal and distal pairs of primers are compared to indicate the relative expression of the two isoforms. See Supplementary Table S3 for primer information.

mRNA stability

NIH3T3 Cell lines were treated with actinomycin D (5 µg/ml), and after 0, 4, 8, 12 and 24 hours total RNA was isolated and analyzed by qRT-PCR. Primer sequence can be found in Supplementary Table S5.

Luciferase Assays

psiCHECK-2 Vector (Promega, Cat. no. C8021) provides a quantitative and rapid approach for measuring the effect of 3' UTR on protein translation efficiency. Renilla luciferase is used as a primary reporter gene, and the 3' UTR of interest can be cloned downstream of the Renilla luciferase translational stop codon. Measurement of Renilla luciferase activity is a convenient indicator of 3' UTR influencing on protein translation efficiency while the transcription level is unchanged. This firefly reporter cassette has been specifically designed to be an intraplasmid transfection normalization reporter. Thus, a short 3' UTR, a long 3' UTR and a long 3' UTR with a mutation of proximal polyadenylation signal can be compared.

NIH3T3 cells were seeded into a 24-well plate at a density of 15,000 cells/well, 3 kinds of Mouse Cul1 3' UTR were separately subcloned into the psiCHECK-2 Vector

using the XhoI and PmeI restriction enzyme sites. After an overnight incubation, the cells were treated with a transfection mixture consisting of 50 μ l of serum-free medium, 1 μ l of Lipofectin reagent (life technologies, Cat. no. E2431), 0.5 μ g of psiCHECK-2 Vector:short; 0.5 μ g of psiCHECK-2 Vector:long; 0.5 μ g of psiCHECK-2 Vector:mutant each 4 well. After a 20-minute incubation, 100 μ l of serum-containing medium was added to the wells. 24 hours post transfection Renilla and firefly luciferase activities were measured using the Dual-Luciferase Reporter 1000 Assay System. The results reflect the influence of 3' UTR on protein translation efficiency indirectly.

PA-seq and RNA-seq library construction

The PA-seq libraries were constructed by following the protocol published before by Ni et al.[35]. The dUTP-based strand-specific RNA-seq libraries were constructed by following the protocol of Parkhomchuk et al [58]. Both library types were sequenced by Illumina HiSeq2000 platform in a paired-end 2 \times 101 bp manner.

Paired-end mapping and mRNA abundances estimation

Paired-end reads from PA-seq were subjected to strand corrected according to 'TTT' at the beginning of the reads as previously described [35]. FastQC software was used for quality control of PA-seq and RNA-seq data (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). Processed raw data were

then aligned to mouse genome (version mm9) or rat genome (version rn5) using Tophat2 [112]. Reads were mapped uniquely using the “-g 1” option. Bowtie was used to get a good approximation for -r and --mate-std-dev parameters, which were needed for Tophat2 [113]. Combined counts of all PA clusters mapped to annotated PA, 3' UTR and extended 3' UTR regions (TPM, or the number of Tags Per Million reads) were used to evaluate the expression level of these genes with APA regulation. RNA-seq reads were also aligned to the genome in the similar way and Cufflinks was used to assemble transcripts, estimate the mRNA abundances by following the instruction published before [114].

Peak calling

After mapping, the information of cleavage and polyadenylation sites (pAs) was extracted. All aligned reads were pooled together so that a unified peak-calling scheme can be applied. F-seq [53] was used for peak calling of pAs by default parameters setting expect that set the feature length as 30. We resized the PA clusters to the shortest distance that contained 95% of the reads according to our previous publication [35]. To filter internal priming, we removed PA clusters with 15 ‘A’ in the 20 nucleotides region downstream of peak mode or with continuous 6 ‘A’ in the downstream of peak mode. Peaks with fewer than 20 supporting reads were removed. RefSeq annotation was used for the genomic location analysis. We classified the genome into eight groups and mapped the pA peaks to these groups using the following hierarchy: PA > 3' UTR > Extended 3' UTR > Exon > Intron > 5' UTR >

TSS > Promoter. Especially, PA defined as annotated pA site plus upstream and downstream 10 bp. Extended_3' UTR denoted downstream 5 kb of 3' UTR by following the previous study [35]. Note that the peaks mapped to extended 3' UTR region should land in intergenic region and not overlapped with any known transcripts. TSS (Transcription Start Site) represented exact transcriptional start site plus upstream and downstream 10 bp. Promoter defined as upstream 250 bp of TSS, which should also not overlap with any known transcripts. Next, we also wanted to exclude the situation where a weak and seldom used APA signal generates an event that passes the expression threshold simply because the gene itself is highly expressed by following the previous study [115]. We therefore kept cleavage events that accounted for $\geq 10\%$ of the total reads mapped in the corresponding genes in ≥ 1 samples. If an event was lower expressed but was expressed by $> 5\%$ of the total reads mapped in the corresponding genes in the majority of samples ($\geq 80\%$), we also kept the pA site.

Comparison of PA-seq data with PolyA_DB

PolyA_DB2 was downloaded from Bin Tian's website (http://exon.umdj.edu/polya_db) [54]. They retrieved all cDNA/EST sequences listed in human, mouse, rat, chicken and zebrafish UniGene databases from NCBI (July and August 2005 versions), and then aligned them with genome sequences downloaded from the UCSC Genome Bioinformatics Site (<http://genome.ucsc.edu>) using BLAT [54]. The genome version be used here is mm5 for mouse, therefore we used the online tool LiftOver (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>), which can

convert genome coordinates and genome annotation files between different assemblies, to change the pAs to the version of mm9. Due to the imprecise nature of cleavage by CPSF-73 or fuzzy definition of the pAs by their surrounding *cis* elements, the cleavages may occur at different positions within a small sequence window [116]. Therefore, if the pAs identified by PolyA_DB2 located in the upstream 10 bp or downstream 10 bp region of the pAs identified by our data, we take them as the same pAs.

Motif analysis

MEME motif enrichment analysis was used to identify the enriched motifs in the upstream (-40 nt) of our identified pAs, with a motif length was set to 6 which is equal to the length of PAS [117]. In addition, frequencies of matches of 15 motif identified by Hu et al. [57] was defined by polyA_svm, only highly similar matches of motifs were considered (> 75th percentile of all possible positive scores) [56].

Define the effective 3' UTR length and RUD score

For each expressed genes, the effective 3' UTR length and RUD score based on the PA-seq and RNA-seq data were calculated, respectively. We used effective 3' UTR length to analyze the PA-seq data following our previous study [35]. Specifically, we used Entrez gene ID to cluster the RefSeq transcripts IDs. Overlapping genes on the

same strand were removed in order to avoid misassignment of pAs to genes. We also restricted our attention to genes whose annotated transcripts shared the same stop codon and also without junctions in the 3' UTR regions. Based on the consolidated gene model above, the pipeline considers both the corresponding 3' UTR length and the tag counts of the pA peaks identified in the tandem 3' UTR regions, including pA peaks located in known pAs, 3' UTR and also extended 3' UTR regions, of each transcribed locus to compute “effective 3' UTR length” [35]. For genes located in the plus strand, the 3' UTR length for each pA was calculated by the right boundary of the pA peak minus the stop codon. While, for genes located in the minus strand, the 3' UTR length for each pA was calculated by the stop codon minus the left boundary of the pA peak.

The calculation of RUD (the score for relative expression of isoforms using distal PolyA sites) based on RNA-seq was also following a previous study by Ji et al [52]. Only uniquely mapped and properly paired reads were used for the calculation of RUD. Properly paired reads are those with two pairing reads mapped to different strands of the same chromosome. RUD was based on the ratio of read density in aUTR (alternative UTR) to that in cUTR (constitutive UTR). The regions upstream and downstream of a proximal pA were named as cUTR and aUTR, respectively. The proximal pAs were defined by PA-seq data. Therefore, a high RUD value manifests higher abundance of long 3' UTR isoforms resulting from usage of promoter-distal pAs. In theory, the range of RUD should be from 0 to 1. However, the RUD may be more than 1 caused by either low or biased coverage of RNA-seq reads in 3' UTR

regions, therefore the genes with RUD more than 1 would be discard also. A standard score (Z score) of Effective 3' UTR length and RUD for each gene across different passages were calculated for normalization. In addition, a cutoff of both TPM = 1 (Tags of Per Million mapped reads) and FPKM = 1 (Fragments Per Kilobase of exon model per Million mapped reads) for PA-seq and RNA-seq in each passage respectively, was used to select expressed genes.

Comparison of APA profiles between different passage of MEF cells

The 3' UTR switching for each gene among the four passages was detected by measuring the linear trend alternative to independence similarly described as previously reported [31, 60]. To test a gene with totally J tandem pAs across I passages, we performed the test as follows: 1) calculate the tandem 3' UTR length for each of the J tandem pA; 2) put the number of reads for each tandem pA at each passage to an $I \times J$ contingency table: take tandem pAs as columns (from the site with shortest 3' UTR to that with longest); take the I stages by time series as rows; 3) if the total number of reads in the $I \times J$ table was less than 30, this gene was discarded for the test; otherwise 4) let the lengths of tandem 3' UTR denote the scores for the columns; let the row number denote the row score; 5) calculate Pearson correlation r (named as tandem 3' UTR isoform switch index, TSI, following the paper of Li et al [31]) by the following formula:

$$r = \frac{\sum_{i=1}^I \sum_{j=1}^J n_{ij}(u_i - \bar{u}_i)(v_j - \bar{v}_j)}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J n_{ij}(u_i - \bar{u}_i)^2 \sum_{i=1}^I \sum_{j=1}^J n_{ij}(v_j - \bar{v}_j)^2}}, \quad (1)$$

where $\bar{u}_i = \sum_{j=1}^J n_{ij} / n_{i+}$ and $\bar{v}_j = \sum_{i=1}^I n_{ij} / n_{+j}$.

6) calculate the test statistic $M^2 = (n_{++} - 1)r^2$; For large samples, this statistic is approximately chi-squared with $df = 1$ and P value can be obtained. The genes with significant P value corresponding to a false discovery rate (FDR) cutoff of 5% (Benjamini-Hochberg), estimated by Benjamini-Hochberge method with R software, were considered as significantly different among different passages. In specificity, a FDR value less than or equal to 0.05 with positive TSI implies a lengthening 3' UTR across the different passages of cellular senescence; FDR value less than or equal to 0.05 with negative TSI implies a shortening 3' UTR.

Pathways enrichment analysis

DAVID (Database for Annotation, Visualization and Integrated Discovery) [61] was used for the pathways enrichment analysis and KEGG (Kyoto Encyclopedia of Genes and Genomes) database was chose.

Prediction conserved binding sites of miRNA and RBP in 3' UTRs

The sequences of all mature miRNAs of mouse were downloaded from the miRBase database (Release 21). 8mer, 7mer-m8 and 7mer-1A targets denote conventional miRNA seed matches [118]. Conserved target sites were identified using multiple genome alignments obtained from UCSC Genome Browser (mm9, 30-way alignment) requiring a perfectly aligned match in mouse, rat, human and dog, same as the paper of Sandberg et al[43]. Highly conserved RBP binding sites were predicted by

RBPmap online tools.

Cluster analysis of RBPs

Clustering of genes based on their FPKMs was done by using the Cluster software (<http://rana.lbl.gov/EisenSoftware.htm>). We first log transformed FPKM of each gene in each passage, centered them by the mean and then normalized each gene. K-means clustering approach with euclidean distance was used to classify the various types of gene profiles. The cluster result was then visualized by Treeview software (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>).

Conservation analysis surrounding pAs

Conservation scores in the 400nt surrounding different types of pAs were calculated using phastCons tracks from 30 vertebrates and was downloaded from UCSC genome browser (<http://hgdownload.soe.ucsc.edu/goldenPath/mm9/phastCons30way/>)[119].

Data Access

Both PA-seq and RNA-seq raw data can be found at the NCBI Sequence Read Archive (SRA) with submission number SRP065821.

SUPPLEMENTAL DATA

ACKNOWLEDGEMENTS

We thank Dr. Jun Zhu for critical reading and insightful comments of the manuscript.

This work was supported by National Key Basic Research Program of China (973 program: 2013CB530700, 2015CB943000), National Science Foundation of China (31271348, 31471192), Shanghai Pujiang Talent Plan (13PJ1400700), the Thousand Talents Plan and Research and Innovation Project of Shanghai Municipal Education Commission (14ZZ007). We thank Genenergy Biotech (Shanghai) Co., Ltd. for the deep sequencing service.

Conflict of interest statement

The authors state no conflict of interest.

References:

- [1]. HAYFLICK, L. and P.S. MOORHEAD, The serial cultivation of human diploid cell strains. *Exp Cell Res*, 1961. 25: p. 585-621.
- [2]. Olovnikov, A.M., Telomeres, telomerase, and aging: origin of the theory. *Exp Gerontol*, 1996. 31(4): p. 443-8.
- [3]. Rodier, F. and J. Campisi, Four faces of cellular senescence. *J Cell Biol*, 2011. 192(4): p. 547-56.
- [4]. Alster, O. and Z. Korwek, [Markers of cellular senescence]. *Postepy Biochem*, 2014. 60(2): p. 138-46.
- [5]. Carnero, A., Markers of cellular senescence. *Methods Mol Biol*, 2013. 965: p. 63-81.
- [6]. Munoz-Espin, D. and M. Serrano, Cellular senescence: from physiology to pathology. *Nat Rev Mol Cell Biol*, 2014. 15(7): p. 482-96.
- [7]. Kuilman, T., et al., The essence of senescence. *Genes Dev*, 2010. 24(22): p. 2463-79.
- [8]. Campisi, J., Cellular senescence as a tumor-suppressor mechanism. *Trends Cell Biol*, 2001. 11(11): p. S27-31.

- [9]. López-Otín, C., et al., The Hallmarks of Aging. *Cell*, 2013. 153(6): p. 1194-1217.
- [10]. Kim, Y.M., et al., Implications of time-series gene expression profiles of replicative senescence. *Aging Cell*, 2013. 12(4): p. 622-34.
- [11]. Manning, J.A. and S. Kumar, A potential role for NEDD1 and the centrosome in senescence of mouse embryonic fibroblasts. *Cell Death Dis*, 2010. 1: p. e35.
- [12]. Di Micco, R., et al., DNA damage response activation in mouse embryonic fibroblasts undergoing replicative senescence and following spontaneous immortalization. *Cell Cycle*, 2008. 7(22): p. 3601-6.
- [13]. Busuttil, R.A., et al., Oxygen accelerates the accumulation of mutations during the senescence and immortalization of murine cells in culture. *Aging Cell*, 2003. 2(6): p. 287-94.
- [14]. Parrinello, S., et al., Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts. *Nat Cell Biol*, 2003. 5(8): p. 741-7.
- [15]. Chen, Q., et al., Oxidative DNA damage and senescence of human diploid fibroblast cells. *Proc Natl Acad Sci U S A*, 1995. 92(10): p. 4337-41.
- [16]. Land, S.C., C.L. Scott and D. Walker, mTOR signalling, embryogenesis and the control of lung development. *Semin Cell Dev Biol*, 2014. 36: p. 68-78.
- [17]. Abdelmohsen, K., et al., Senescence-associated lncRNAs: senescence-associated long noncoding RNAs. *Aging Cell*, 2013. 12(5): p. 890-900.
- [18]. Harries, L.W., et al., Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell*, 2011. 10(5): p. 868-78.
- [19]. Mazin, P., et al., Widespread splicing changes in human brain development and aging. *Mol Syst Biol*, 2013. 9: p. 633.
- [20]. Yannarell, A., D.E. Schumm and T.E. Webb, Age-dependence of nuclear RNA processing. *Mech Ageing Dev*, 1977. 6(4): p. 259-64.
- [21]. Meshorer, E. and H. Soreq, Pre-mRNA splicing modulations in senescence. *Aging Cell*, 2002. 1(1): p. 10-6.
- [22]. Pan, K., et al., Coupling and Coordination in Gene Expression Processes with Pre-mRNA Splicing. *Int J Mol Sci*, 2015. 16(3): p. 5682-5696.
- [23]. Bahrami-Samani, E., et al., Computational challenges, tools, and resources for analyzing co- and post-transcriptional events in high throughput. *Wiley Interdiscip Rev RNA*, 2015. 6(3): p. 291-310.
- [24]. Wei, Y.N., et al., Transcript and protein expression decoupling reveals RNA binding proteins and miRNAs as potential modulators of human aging. *Genome Biol*, 2015. 16: p. 41.
- [25]. Wang, W., et al., Loss of HuR Is Linked to Reduced Expression of Proliferative Genes during Replicative Senescence. *Molecular and Cellular Biology*, 2001. 21(17): p. 5889-5898.
- [26]. Elkon, R., A.P. Ugalde and R. Agami, Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet*, 2013. 14(7): p. 496-506.
- [27]. Zheng, D. and B. Tian, RNA-Binding Proteins in Regulation of Alternative Cleavage and Polyadenylation. *Adv Exp Med Biol*, 2014. 825: p. 97-127.
- [28]. Derti, A., et al., A quantitative atlas of polyadenylation in five mammals. *Genome Res*, 2012. 22(6): p. 1173-83.
- [29]. Hoque, M., et al., Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods*, 2013. 10(2): p. 133-9.
- [30]. Jan, C.H., et al., Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature*, 2011. 469(7328): p. 97-101.

- [31]. Li, Y., et al., Dynamic landscape of tandem 3' UTRs during zebrafish development. *Genome Res*, 2012. 22(10): p. 1899-906.
- [32]. Ozsolak, F., et al., Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*, 2010. 143(6): p. 1018-29.
- [33]. Shepard, P.J., et al., Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*, 2011. 17(4): p. 761-72.
- [34]. Wu, X., et al., Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. *Proc Natl Acad Sci U S A*, 2011. 108(30): p. 12533-8.
- [35]. Ni, T., et al., Distinct polyadenylation landscapes of diverse human tissues revealed by a modified PA-seq strategy. *BMC Genomics*, 2013. 14: p. 615.
- [36]. Xie, X., et al., Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, 2005. 434(7031): p. 338-45.
- [37]. Vindry, C., et al., RNA-binding protein-mediated post-transcriptional controls of gene expression: integration of molecular mechanisms at the 3' end of mRNAs? *Biochem Pharmacol*, 2014. 89(4): p. 431-40.
- [38]. Di Giammartino, D.C., K. Nishida and J.L. Manley, Mechanisms and consequences of alternative polyadenylation. *Mol Cell*, 2011. 43(6): p. 853-66.
- [39]. Lutz, C.S. and A. Moreira, Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip Rev RNA*, 2011. 2(1): p. 22-31.
- [40]. Zhang, H., J.Y. Lee and B. Tian, Biased alternative polyadenylation in human tissues. *Genome Biol*, 2005. 6(12): p. R100.
- [41]. Wang, E.T., et al., Alternative isoform regulation in human tissue transcriptomes. *Nature*, 2008. 456(7221): p. 470-6.
- [42]. Ji, Z., et al., Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A*, 2009. 106(17): p. 7028-33.
- [43]. Sandberg, R., et al., Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, 2008. 320(5883): p. 1643-7.
- [44]. Mayr, C. and D.P. Bartel, Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 2009. 138(4): p. 673-84.
- [45]. Singh, P., et al., Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res*, 2009. 69(24): p. 9422-30.
- [46]. Flavell, S.W., et al., Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron*, 2008. 60(6): p. 1022-38.
- [47]. Ji, Z. and B. Tian, Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One*, 2009. 4(12): p. e8419.
- [48]. Xia, Z., et al., Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3' -UTR landscape across seven tumour types. *Nature Communications*, 2014. 5: p. 5274.
- [49]. Lai, D.P., et al., Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis. *Hum Mol Genet*, 2015.
- [50]. Millevoi, S. and S. Vagner, Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res*, 2010. 38(9): p. 2757-74.

- [51]. Correia-Melo, C., D. Jurk and J.F. Passos, Robust multiparametric assessment of cellular senescence. *Methods Mol Biol*, 2013. 965: p. 409-19.
- [52]. Ji, Z., et al., Transcriptional activity regulates alternative cleavage and polyadenylation. *Mol Syst Biol*, 2011. 7: p. 534.
- [53]. Boyle, A.P., et al., F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics*, 2008. 24(21): p. 2537-8.
- [54]. Lee, J.Y., et al., PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res*, 2007. 35(Database issue): p. D165-8.
- [55]. Tian, B., et al., A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res*, 2005. 33(1): p. 201-12.
- [56]. Cheng, Y., R.M. Miura and B. Tian, Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics*, 2006. 22(19): p. 2320-5.
- [57]. Hu, J., et al., Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA*, 2005. 11(10): p. 1485-93.
- [58]. Parkhomchuk, D., et al., Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res*, 2009. 37(18): p. e123.
- [59]. Legendre, M., et al., Differential repression of alternative transcripts: a screen for miRNA targets. *PLoS Comput Biol*, 2006. 2(5): p. e43.
- [60]. Fu, Y., et al., Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res*, 2011. 21(5): p. 741-7.
- [61]. Huang, D.W., B.T. Sherman and R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 2009. 4(1): p. 44-57.
- [62]. Deschenes-Simard, X., et al., Cellular senescence and protein degradation: breaking down cancer. *Cell Cycle*, 2014. 13(12): p. 1840-58.
- [63]. Hofmann, J.W., et al., The effects of aging on the expression of Wnt pathway genes in mouse tissues. *Age (Dordr)*, 2014. 36(3): p. 9618.
- [64]. Chandler, H. and G. Peters, Stressing the cell cycle in senescence and aging. *Curr Opin Cell Biol*, 2013. 25(6): p. 765-71.
- [65]. Amberg, D., et al., Cellular ageing and the actin cytoskeleton. *Subcell Biochem*, 2012. 57: p. 331-52.
- [66]. Poliseno, L., et al., The proto-oncogene LRF is under post-transcriptional control of MiR-20a: implications for senescence. *PLoS One*, 2008. 3(7): p. e2542.
- [67]. Pitto, L., et al., miR-290 acts as a physiological effector of senescence in mouse embryo fibroblasts. *Physiol Genomics*, 2009. 39(3): p. 210-8.
- [68]. Cook, K.B., et al., RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res*, 2011. 39(Database issue): p. D301-8.
- [69]. Paz, I., et al., RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res*, 2014. 42(Web Server issue): p. W361-7.
- [70]. Roy, R., et al., hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGF-2/S6K2 signalling. *Nucleic Acids Res*, 2014. 42(20): p. 12483-97.
- [71]. Nielsen, F.C., J. Nielsen and J. Christiansen, A family of IGF-II mRNA binding proteins (IMP) involved in RNA trafficking. *Scand J Clin Lab Invest Suppl*, 2001. 234: p. 93-9.
- [72]. Capowski, E.E., et al., Y box-binding factor promotes eosinophil survival by stabilizing granulocyte-macrophage colony-stimulating factor mRNA. *J Immunol*, 2001. 167(10): p. 5970-6.

- [73]. Gunderson, S.I., M. Polycarpou-Schwarz and I.W. Mattaj, U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell*, 1998. 1(2): p. 255-64.
- [74]. Decorsiere, A., et al., Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes Dev*, 2011. 25(3): p. 220-5.
- [75]. Bagga, P.S., G.K. Arhin and J. Wilusz, DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. *Nucleic Acids Res*, 1998. 26(23): p. 5343-50.
- [76]. Naganuma, T., et al., Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J*, 2012. 31(20): p. 4020-34.
- [77]. Takagaki, Y., et al., The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell*, 1996. 87(5): p. 941-52.
- [78]. Yao, C., et al., Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci U S A*, 2012. 109(46): p. 18773-8.
- [79]. Li, W., et al., Systematic profiling of poly(A)⁺ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet*, 2015. 11(4): p. e1005166.
- [80]. Lau, A.G., et al., Distinct 3'UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF). *Proc Natl Acad Sci U S A*, 2010. 107(36): p. 15945-50.
- [81]. Hinman, J.D., et al., Age-dependent accumulation of ubiquitinated 2',3'-cyclic nucleotide 3'-phosphodiesterase in myelin lipid rafts. *Glia*, 2008. 56(1): p. 118-33.
- [82]. Vernace, V.A., T. Schmidt-Glenewinkel and M.E. Figueiredo-Pereira, Aging and regulated protein degradation: who has the UPPER hand? *Aging Cell*, 2007. 6(5): p. 599-606.
- [83]. Gourlay, C.W. and K.R. Ayscough, A role for actin in aging and apoptosis. *Biochem Soc Trans*, 2005. 33(Pt 6): p. 1260-4.
- [84]. Lezzerini, M. and Y. Budovskaya, A dual role of the Wnt signaling pathway during aging in *Caenorhabditis elegans*. *Aging Cell*, 2014. 13(1): p. 8-18.
- [85]. Koga, H., S. Kaushik and A.M. Cuervo, Protein homeostasis and aging: The importance of exquisite quality control. *Ageing Res Rev*, 2011. 10(2): p. 205-15.
- [86]. Ott, C. and T. Grune, Protein oxidation and proteolytic signalling in aging. *Curr Pharm Des*, 2014. 20(18): p. 3040-51.
- [87]. Powers, E.T., et al., Biological and chemical approaches to diseases of proteostasis deficiency. *Annu Rev Biochem*, 2009. 78: p. 959-91.
- [88]. Bossy-Wetzel, E., R. Schwarzenbacher and S.A. Lipton, Molecular pathways to neurodegeneration. *Nat Med*, 2004. 10 Suppl: p. S2-9.
- [89]. Li, X.J. and S. Li, Proteasomal dysfunction in aging and Huntington disease. *Neurobiol Dis*, 2011. 43(1): p. 4-8.
- [90]. Han, L., et al., Upregulation of SIRT1 by 17beta-estradiol depends on ubiquitin-proteasome degradation of PPAR-gamma mediated by NEDD4-1. *Protein Cell*, 2013. 4(4): p. 310-21.
- [91]. Wang, X., et al., NEDD4-1 is a proto-oncogenic ubiquitin ligase for PTEN. *Cell*, 2007. 128(1): p. 129-39.
- [92]. Ghazi, A., S. Henis-Korenblit and C. Kenyon, Regulation of *Caenorhabditis elegans* lifespan

- by a proteasomal E3 ligase complex. *Proc Natl Acad Sci U S A*, 2007. 104(14): p. 5947-52.
- [93]. Huang, H., et al., Skp2 inhibits FOXO1 in tumor suppression through ubiquitin-mediated degradation. *Proc Natl Acad Sci U S A*, 2005. 102(5): p. 1649-54.
- [94]. Plas, D.R. and C.B. Thompson, Akt activation promotes degradation of tuberin and FOXO3a via the proteasome. *J Biol Chem*, 2003. 278(14): p. 12361-6.
- [95]. Aoki, M., H. Jiang and P.K. Vogt, Proteasomal degradation of the FoxO1 transcriptional regulator in cells transformed by the P3k and Akt oncoproteins. *Proc Natl Acad Sci U S A*, 2004. 101(37): p. 13613-7.
- [96]. Dealy, M.J., et al., Loss of Cul1 results in early embryonic lethality and dysregulation of cyclin E. *Nat Genet*, 1999. 23(2): p. 245-8.
- [97]. Gourlay, C.W., et al., A role for the actin cytoskeleton in cell death and aging in yeast. *J Cell Biol*, 2004. 164(6): p. 803-9.
- [98]. Alexander, K., H.S. Yang and P.W. Hinds, Cellular senescence requires CDK5 repression of Rac1 activity. *Mol Cell Biol*, 2004. 24(7): p. 2808-19.
- [99]. Florian, M.C., et al., A canonical to non-canonical Wnt signalling switch in haematopoietic stem-cell ageing. *Nature*, 2013. 503(7476): p. 392-6.
- [100]. Li, N., et al., Increased expression of miR-34a and miR-93 in rat liver during aging, and their impact on the expression of Mgst1 and Sirt1. *Mech Ageing Dev*, 2011. 132(3): p. 75-85.
- [101]. Comegna, M., et al., Identification of miR-494 direct targets involved in senescence of human diploid fibroblasts. *FASEB J*, 2014. 28(8): p. 3720-33.
- [102]. Nidadavolu, L.S., L.J. Niedernhofer and S.A. Khan, Identification of microRNAs dysregulated in cellular senescence driven by endogenous genotoxic stress. *Aging (Albany NY)*, 2013. 5(6): p. 460-73.
- [103]. Licatalosi, D.D., et al., HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, 2008. 456(7221): p. 464-9.
- [104]. Moreira, A., et al., The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev*, 1998. 12(16): p. 2522-34.
- [105]. Castelo-Branco, P., et al., Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol*, 2004. 24(10): p. 4174-83.
- [106]. Core, L.J., J.J. Waterfall and J.T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 2008. 322(5909): p. 1845-8.
- [107]. Paulsen, M.T., et al., Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods*, 2014. 67(1): p. 45-54.
- [108]. Yeo, G.W., et al., An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*, 2009. 16(2): p. 130-7.
- [109]. Hafner, M., et al., Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 2010. 141(1): p. 129-41.
- [110]. Chang, H., et al., TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell*, 2014. 53(6): p. 1044-52.
- [111]. Velten, L., et al., Single-cell polyadenylation site mapping reveals 3' isoform choice variability. *Mol Syst Biol*, 2015. 11(6): p. 812.
- [112]. Kim, D., et al., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 2013. 14(4): p. R36.
- [113]. Langmead, B., Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics*,

2010. Chapter 11: p. Unit 11.7.

- [114]. Trapnell, C., et al., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 2012. 7(3): p. 562-78.
- [115]. Lianoglou, S., et al., Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev*, 2013. 27(21): p. 2380-96.
- [116]. Pauws, E., et al., Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res*, 2001. 29(8): p. 1690-4.
- [117]. Bailey, T.L., et al., MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, 2009. 37(Web Server issue): p. W202-8.
- [118]. Hansen, T.B., et al., Natural RNA circles function as efficient microRNA sponges. *Nature*, 2013. 495(7441): p. 384-8.
- [119]. Thomas, J.W., et al., Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, 2003. 424(6950): p. 788-93.

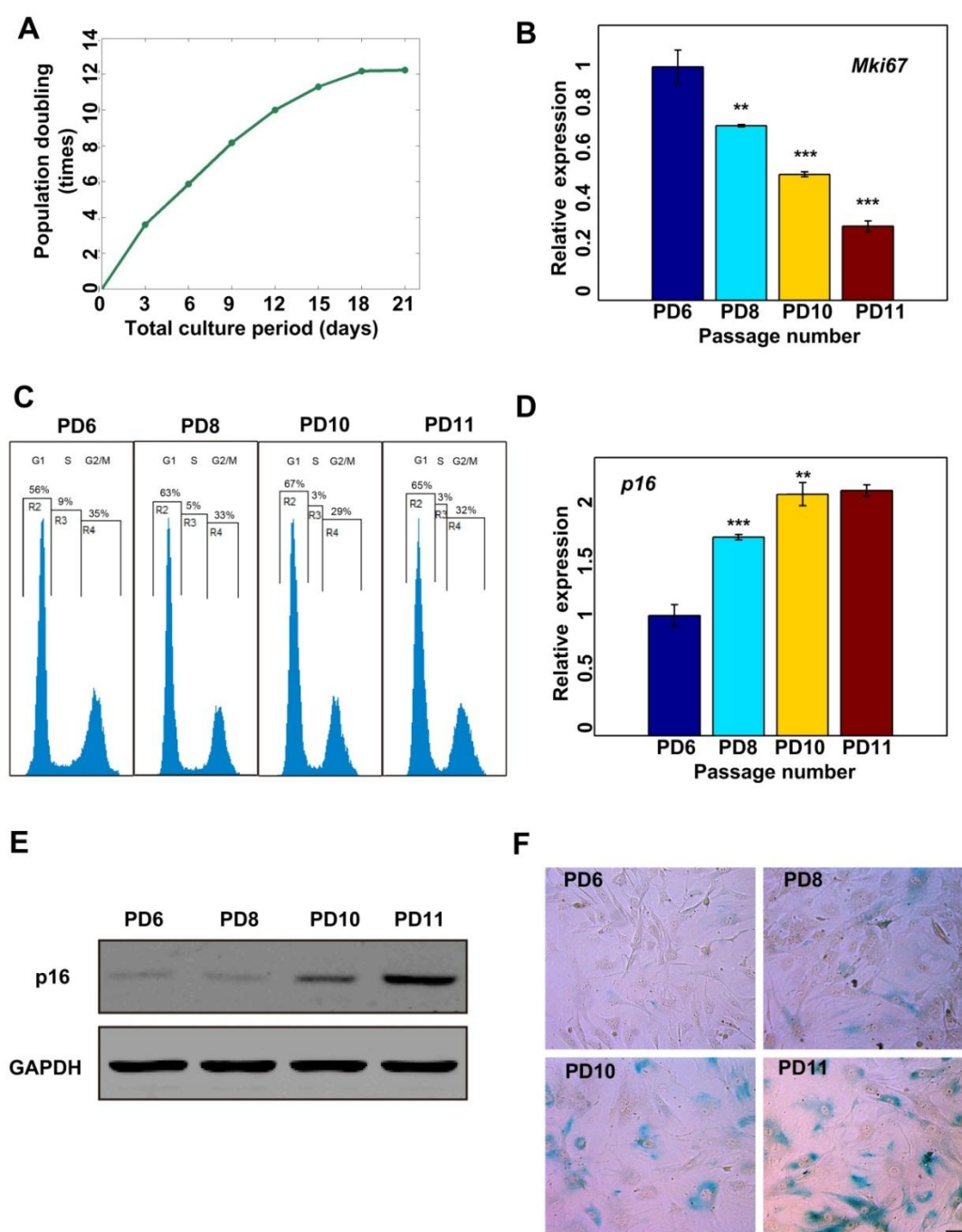


Figure1. Establishment of a replicative senescence model for MEFs. (A) MEFs were continuously subcultured in DMEM medium containing 10% FBS (fetal bovine serum) and their population doubling numbers (passaged numbers of the cells obtained by sequential population doubling) were recorded. (B) Detection expression levels of *Mki67* in PD6, PD8, PD10 and PD11 passages of MEFs by qRT-PCR. (C) MEFs from PD6, PD8, PD10, PD11 passages were subjected to FACS (Fluorescence-activated cell sorting) analysis. (D) Expression levels of p16 (encoded by *Cdkn2a*) in PD6, PD8, PD10 and PD11 passages of MEFs were detected by RT-qPCR. (E) Detection of p16 protein levels in PD6, PD8, PD10 and PD11 passages of MEFs by Western Blot. (F) MEFs from PD6, PD8, PD10 and PD11 passages were subjected to SA-β-gal (senescence-associated β-galactosidase) staining. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Student's t-test.

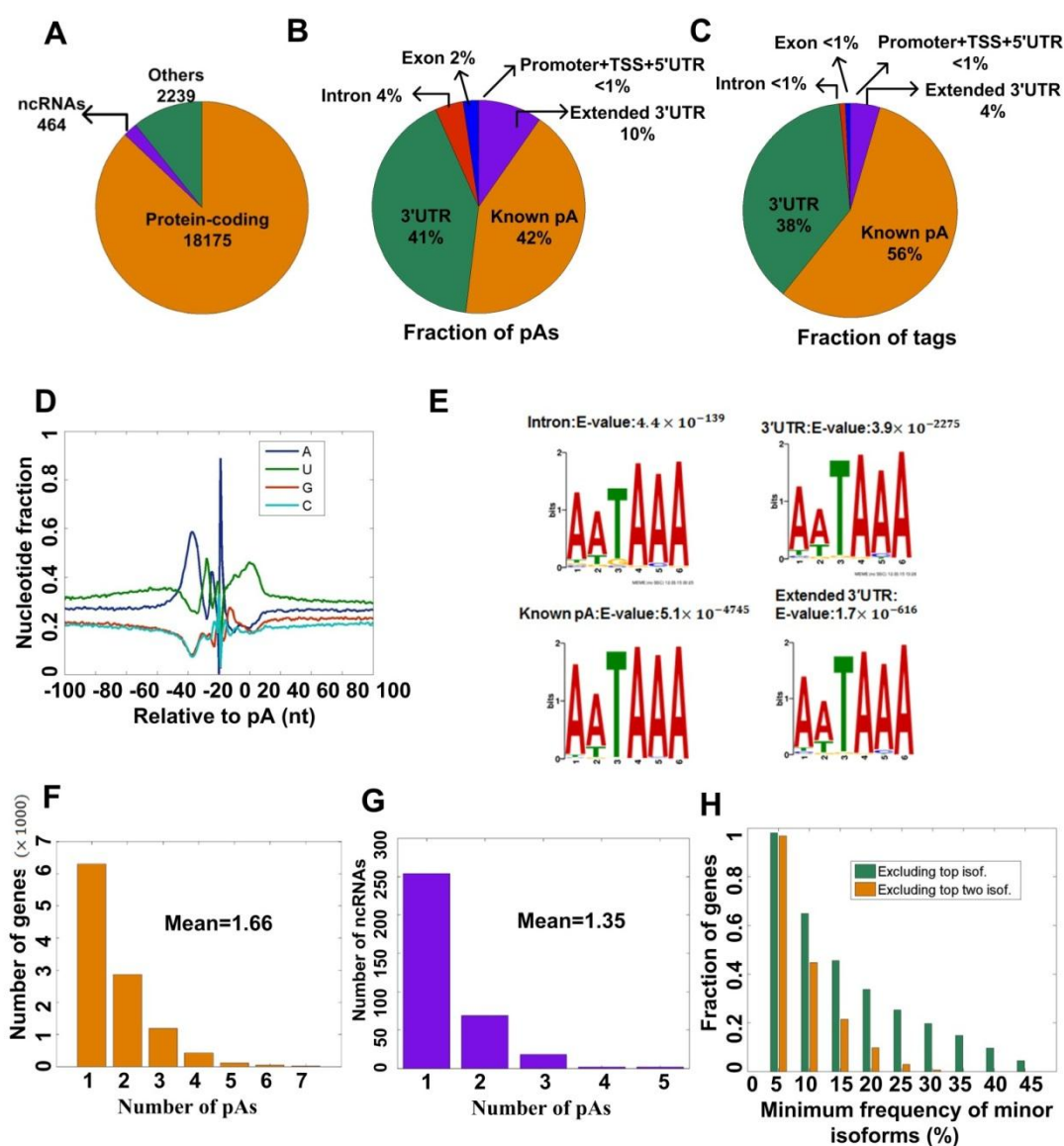


Figure2 Summary of pAs in MEFs identified by PA-seq. (A) Gene categories of pAs (poly(A) sites) based on RefSeq annotation. (B) The relative locations of the identified pAs in respective genes. The pAs classification scheme is described in materials and methods. (C) The relative locations of the identified pA tags in respective genes. (D) Nucleotide sequence composition around the pAs identified in this study. (E) MEME program identifies the canonical PAS (poly(A) signal) motif AATAAA with a very significant E-value from the upstream (-40bp) of the pAs located in intron, 3' UTRs, known pAs and extended 3' UTR regions. (F) The number of genes with different number of pAs. (G) The number of ncRNAs with different number of pAs. (H) Fraction of genes with the minimum frequency of minor 3' UTR isoforms exceeding given thresholds (0%, 5%, 15%, 20%, 25%, 30%, 35%, 40% and 45%). Frequencies of minor isoforms are calculated by excluding the top isoform (green bars) and the top two isoforms (orange bars), respectively.

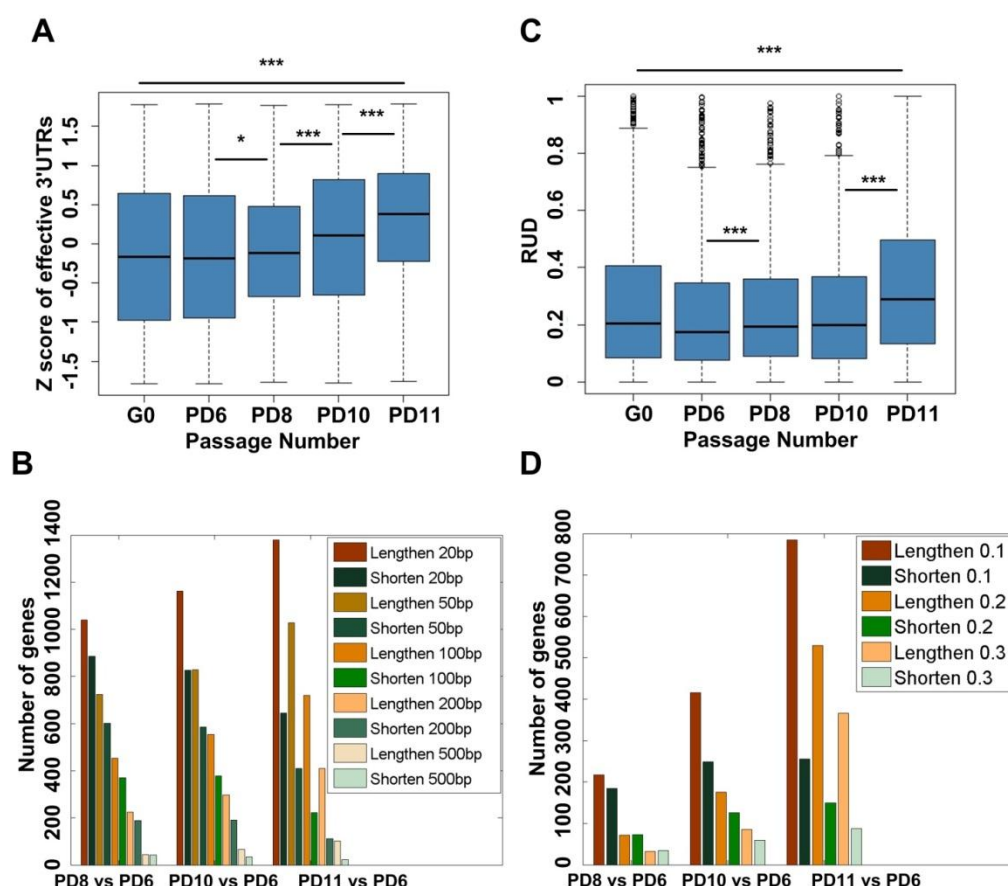


Figure3. Progressive lengthening of 3' UTRs for genes with APA regulation during replicative senescence of MEFs. (A) Box plot for Z scores of effective 3' UTRs across G0, PD6, PD8, PD10 and PD11 of MEFs. (B) Number of genes with lengthened effective 3' UTRs and number of genes with shortened effective 3' UTRs by comparing PD11, PD10, PD8 to PD6, respectively, given different thresholds based on PA-seq data. (C) Box plot for RUD scores (relative expression of mRNA isoforms using distal pA sites) across G0, PD6, PD8, PD10 and PD11 of MEFs. (D) Number of genes with higher RUD scores and number of genes with lower RUD scores by comparing PD11, PD10, PD8 to PD6, respectively, given different thresholds based on RNA-seq data. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Wilcoxon signed rank test. See Materials and methods for more information about Z score of effective 3' UTRs and RUD score.

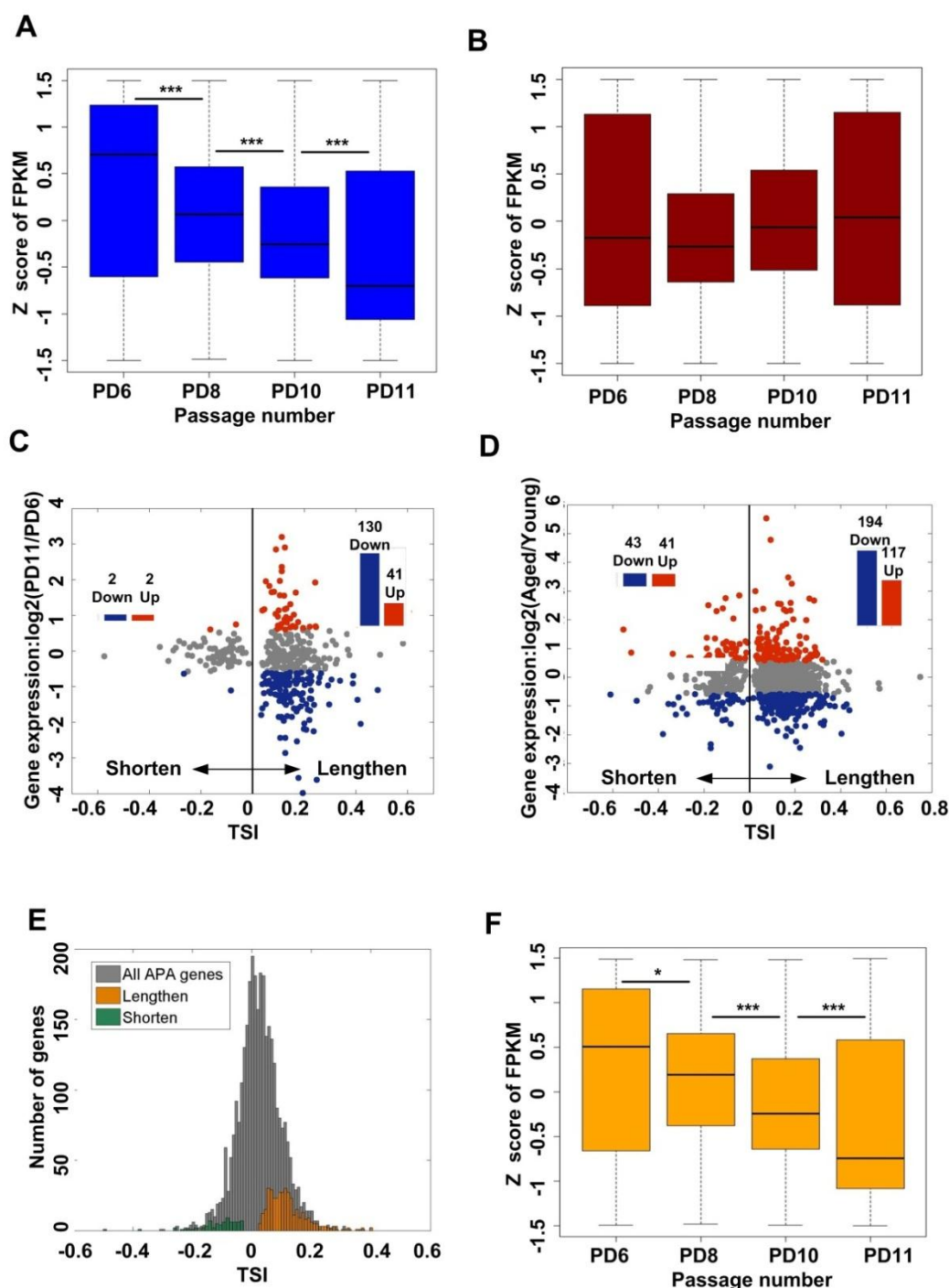


Figure4. Genes preferred to use distal pAs during cellular senescence tended to have decreased mRNA abundance. (A) Box plot for normalized expression (FPKM, Fragments Per Kilobase of exon model per Million mapped reads) across PD6, PD8, PD10 and PD11 for genes with multiple pAs. (B) Box plot for normalized expression (FPKM, Fragments Per Kilobase of exon model per Million mapped reads) across PD6, PD8, PD10 and PD11 for genes with single pA site. (C) For genes with significantly longer or shorter 3' UTRs in senescent MEFs (PD11), their fold-change (FC) expression between senescent MEFs and young MEFs (PD6) were plotted against their TSI values. Red and blue represented up- and down-regulated in senescent cells, respectively, which was identified by fold-change more than 1.5 and FPKM>1 in both samples.

The genes with significantly longer or shorter 3' UTRs in senescent MEFs were identified by linear trend test with Benjamini-Hochberg (BH) false-discovery rate at 5%. (D) For genes with significantly longer or shorter 3' UTRs in VSMCs derived from old rat (2 years old), their fold-change expression between VSMCs of old and young rat are plotted against their TSI values. Red and blue represented up- and down-regulated in senescent cells, respectively, which was identified by fold-change more than 1.5 and FPKM>1 in both samples. The genes with significantly longer or shorter 3' UTRs in VSMCs of old rat were identified by linear trend test with Benjamini-Hochberg (BH) false-discovery rate (FDR) at 5%. (E) Distribution of TSI values for all APA regulation genes. Orange bar showed distribution of TSI values for genes using a lengthened 3' UTR during replicative senescence of MEFs (with a positive Pearson Correlation r and FDR<0.05, Linear trend test). Green bar shown distribution of TSI values for genes using a shortened 3' UTR during replicative senescence of MEFs (with a negative Pearson Correlation r and FDR<0.05, linear trend test). Pearson Correlation r named as tandem UTR isoform switch index (TSI). (F). Box plot for normalized expression (FPKM) across PD6, PD8, PD10 and PD11 for genes with 3' UTR lengthening during replicative senescence of MEFs. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Wilcoxon signed rank test.

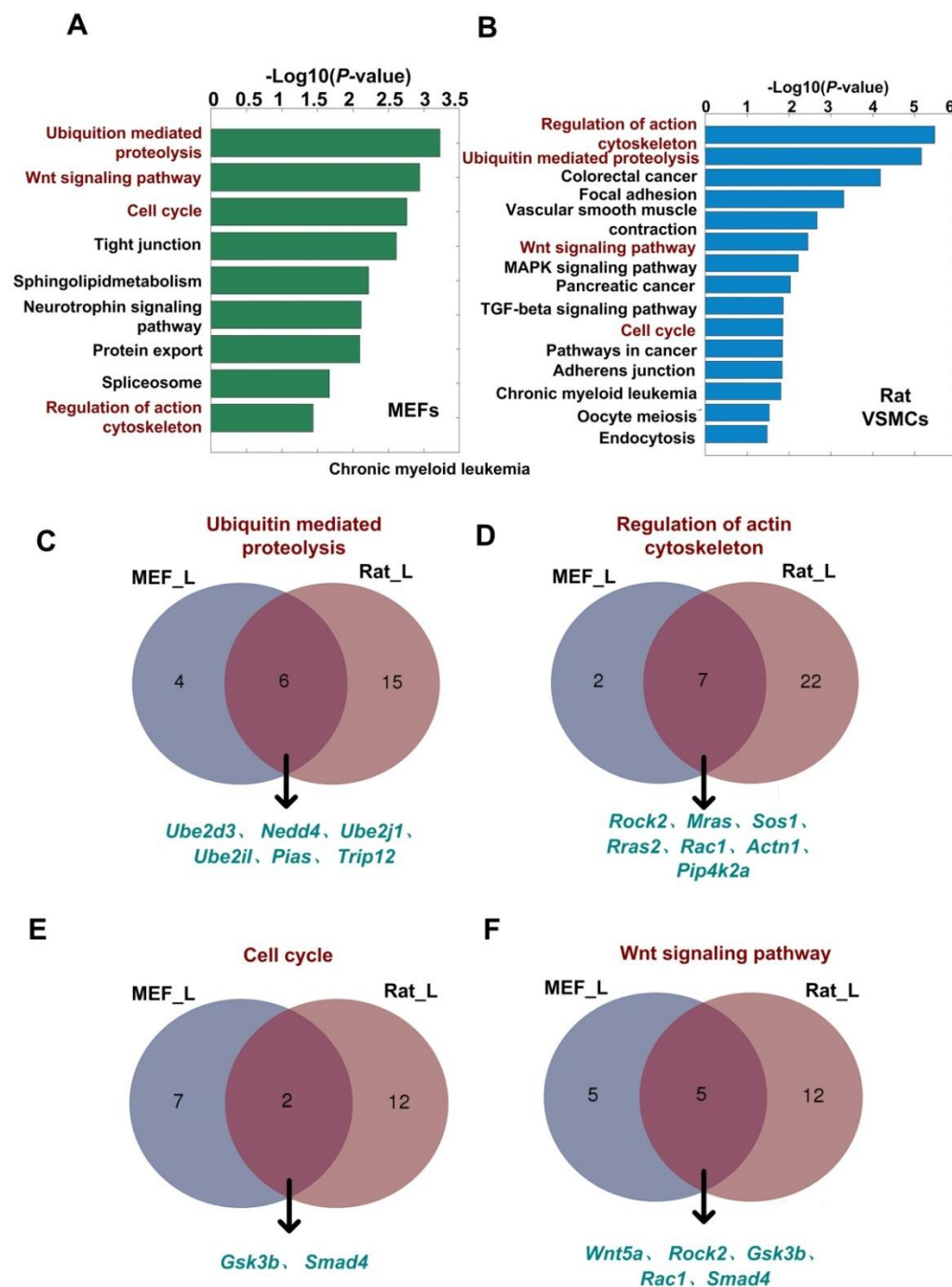


Figure 5. Genes preferred to use distal pAs both in senescent MEFs and VSMCs are enriched in common senescence-related pathways. (A-B) Significantly enriched ($P < 0.01$; Fisher's exact test) pathways of the genes tended to use distal pAs in senescent MEFs and VSMCs derived from old rat, respectively. The common senescence-related pathways between them are marked by red color. (C-F) Venn diagram comparison between mouse and rat for genes involved in ubiquitin mediated proteolysis, regulation of actin cytoskeleton, cell cycle and wnt signaling pathway and preferred to use distal pAs in senescent MEFs and VSMCs of old rat.

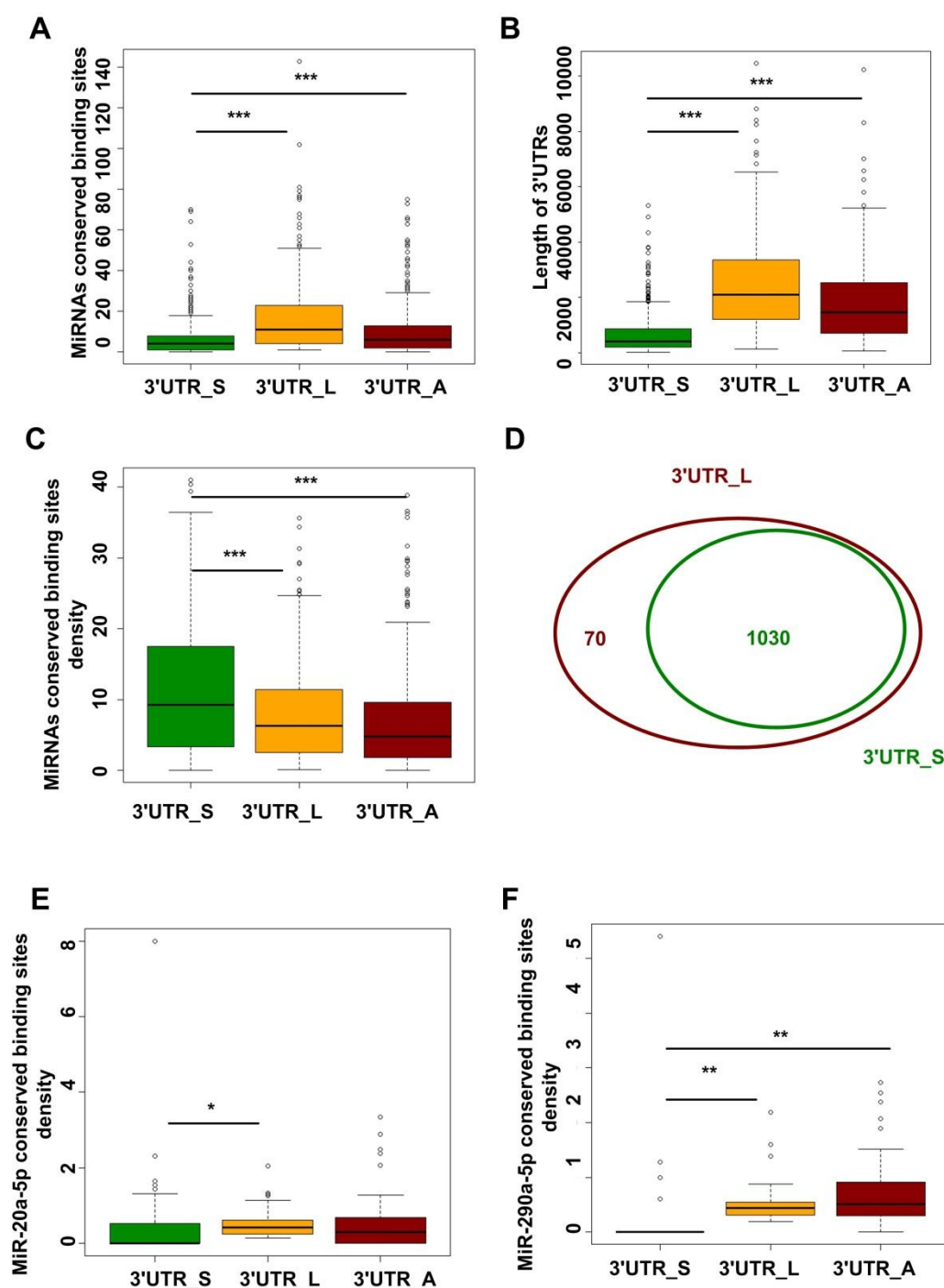


Figure6. Genes progressively tended to use distal pAs during replicative senescence of MEFs can introduce more conserved miRNA binding sites when use longer 3' UTRs. (A) Box plot comparison for the number of conserved miRNA binding sites among the shortest 3' UTRs (named as 3'UTR_S), the longest 3' UTRs (named as 3'UTR_L) and alternative 3' UTRs (named as 3'UTR_A) in genes progressively tended to use distal pAs during replicative senescence of MEFs. (B) Box plot for length comparison among 3' UTR_S, 3'UTR_L and 3'UTR_A in genes progressively tended to use distal pAs during replicative senescence of MEFs. (C) Box plot comparison for density of conserved miRNA binding sites among 3' UTR_S, 3'UTR_L and 3'UTR_A in genes progressively tended to use distal pAs during replicative senescence of MEFs.

(D). Venn diagram comparing conserved miRNA binding sites in the 3' UTR_S with 3' UTR_L for genes progressively tended to use distal pAs during replicative senescence of MEFs. (E) Comparing the density of miR-20a-5p potential binding sites in 3' UTR_S with 3' UTR_L and 3' UTR_A for genes progressively tended to use distal pAs during replicative senescence of MEFs. (F) Comparing the density of miR-290a-5p potential binding sites in 3' UTR_S with 3' UTR_L and 3' UTR_A for genes progressively tended to use distal pAs during replicative senescence of MEFs. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Wilcoxon signed rank test.

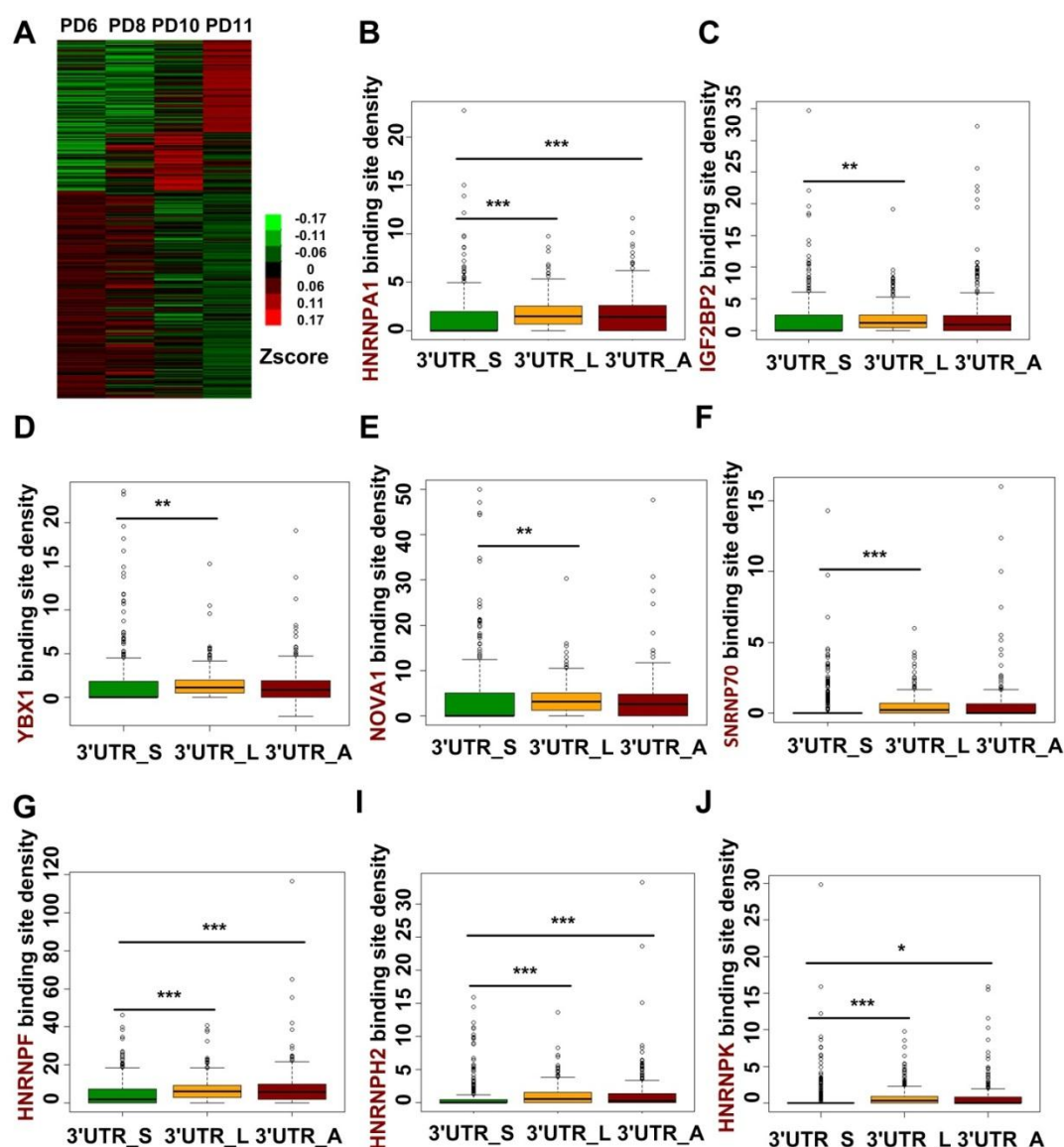


Figure7. Genes progressively tended to use distal pAs during replicative senescence of MEFs can introduce more conserved RBPs binding sites. (A) Heatmap of RBPs expression during replicative senescence of MEFs. (B-J) Comparing the density of HNRNPA1, IGF2BP2, YBX1, NOVA1, SNRNP70, HNRNPF, HNRNPH2 and HNRNPK potential binding sites in the shortest 3' UTRs (named as 3' UTR_S) with the longest 3' UTRs (named as 3' UTR_L) and alternative 3' UTRs (named as 3' UTR_A) for genes progressively tended to use distal pAs during replicative senescence of MEFs. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Wilcoxon signed rank test.

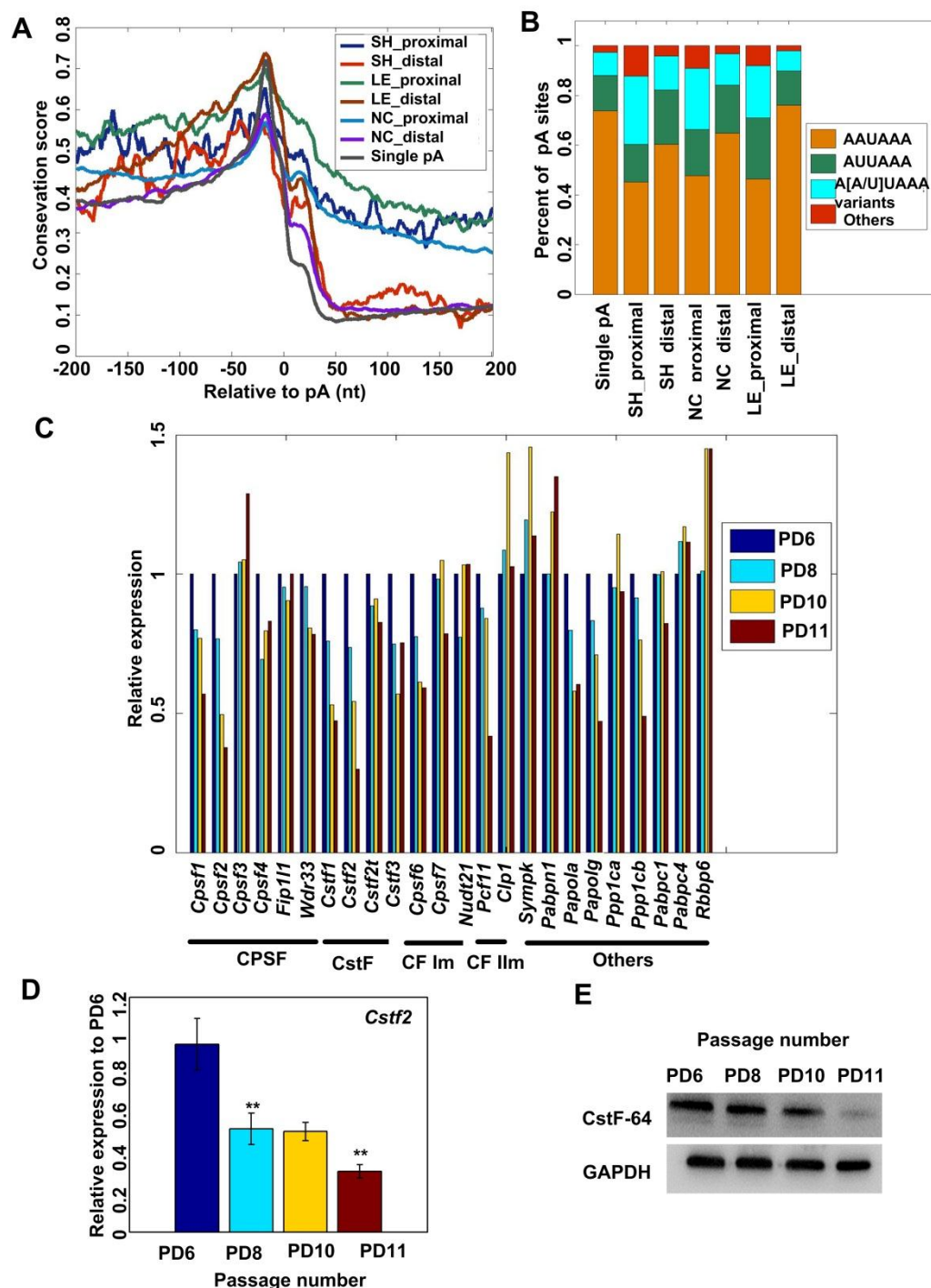


Figure8. Potential mechanisms for APA regulation during replicative senescence of MEFs.

(A) Comparison of conservation score in the 400nt region surrounding the proximal pAs and distal pAs of genes showed continuous 3' UTR shortening during replicative senescence of MEFs (named as SH_proximal and SH_distal, respectively), the proximal pAs and distal pAs of genes showed continuous 3' UTR lengthening during replicative senescence of MEFs (named as LE_proximal and LE_distal, respectively), the proximal pAs and distal pAs of genes without significant pAs usage preference during replicative senescence of MEFs (named as NC_proximal and NC_distal, respectively) and pAs of genes with only single pA site. (B) Distribution of PAS sequences in the -40 to -1 nt region for different types of pAs. (C). Gene expression of known

polyadenylation factors during replicative senescence of MEFs based on RNA-seq data. (D) Expression levels of *Cstf2* in PD6, PD8, PD10 and PD11 passages of MEFs were detected by qRT-PCR. (E) Detection protein levels of CstF-64 (gene symbol is *Cstf2*) in PD6, PD8, PD10 and PD11 passages of MEFs by Western Blot.

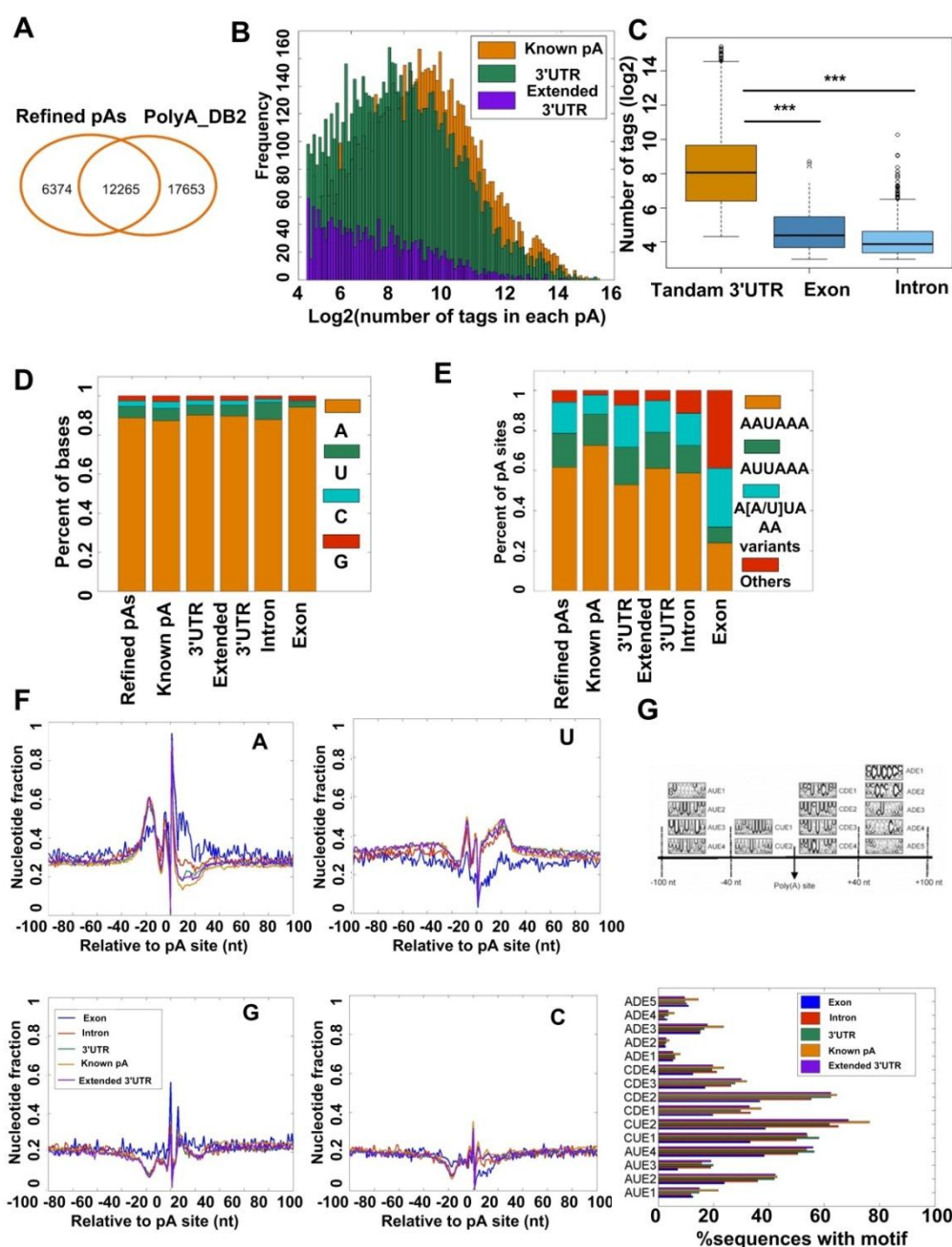


Figure S1, related to Figure 2. Base composition comparison of different types of pAs. (A) Venn diagram comparing pAs in the PolyA_DB2 database with those identified in this study. (B) Distribution of the number of tags from different types of tandem pAs. (C) Box plot for comparing the number of pA tags in tandem 3' UTRs with pA tags located in exon and intron. (D) Comparison of base composition of the 3' base next to pAs from different genomic locations. (E) Distribution of PAS sequences in the -40 to -1 nt region of pAs from different genomic locations. (F) Base compositions in the 200nt surrounding pAs from different genomic locations. (G) Frequencies of matches of different motifs defined by Hu et al. in the 200nt surrounding pAs from different genomic locations was predicted by polyA_svm.[1, 2]. Only highly similar matches of motifs were considered (>75th percentile of all possible positive scores). (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two Mann-Whitney U test.

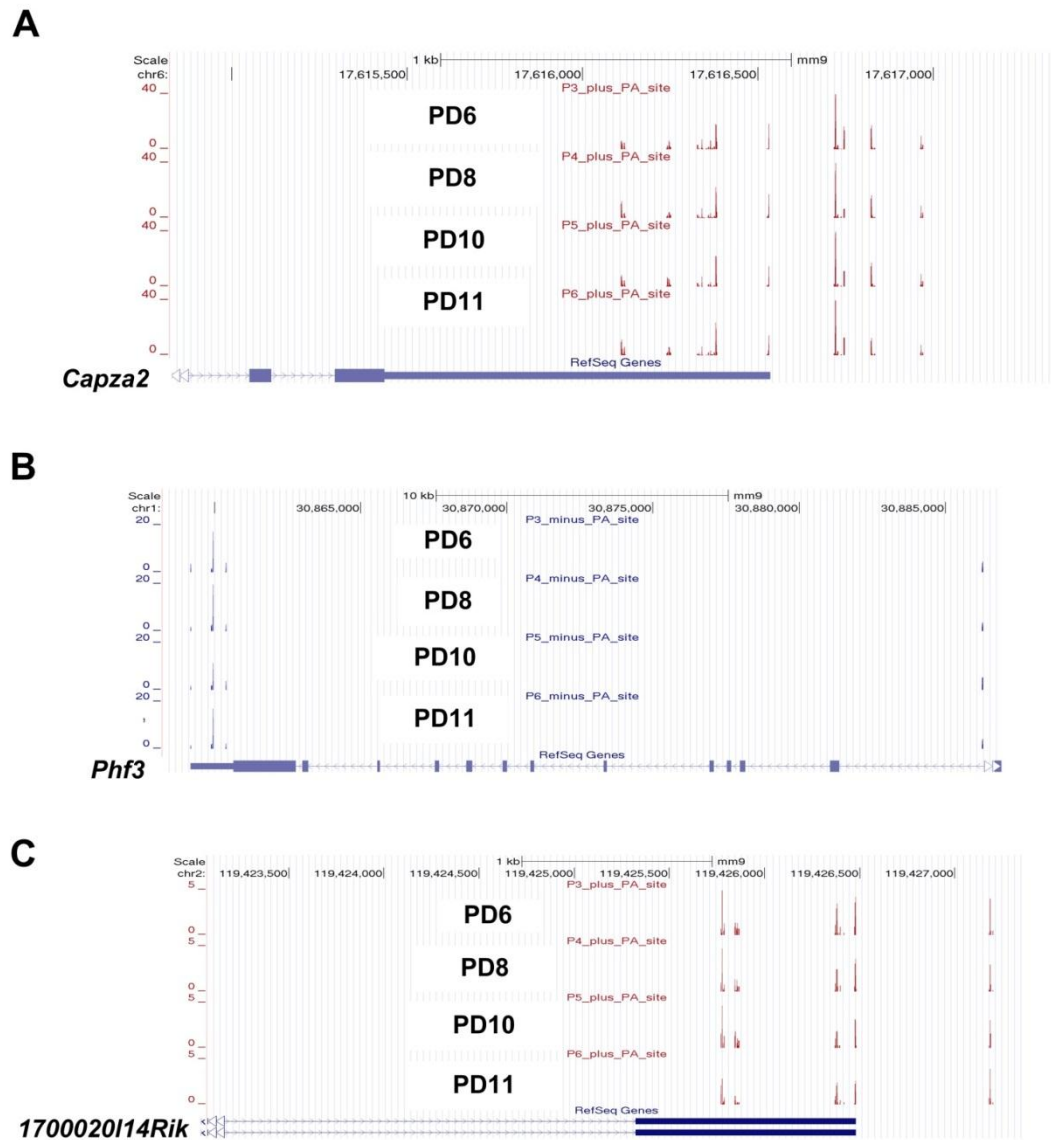


Figure S2. Examples of dynamic APA. (A-C) PA-seq tracks of three genes (*Capza2*, *1700020114Rik* and *Phf3*) with multiple pAs..

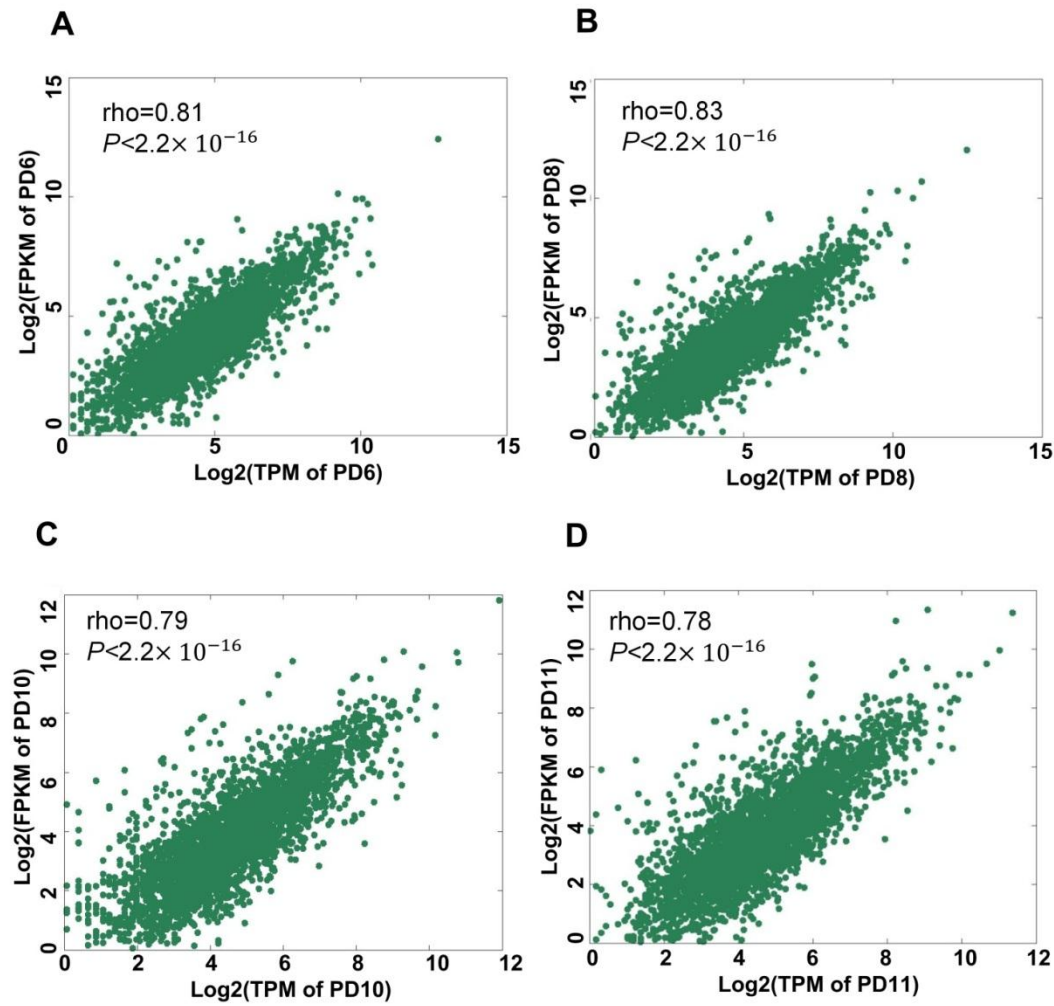


Figure S3. Correlation of expression measured by PA-seq and RNA-seq for genes with APA regulation. (A-D) Correlation of expression measured by PA-seq and RNA-seq for genes with APA regulation in PD6, PD8, PD10 and PD11, respectively. Rho means Spearman's rank correlation coefficient.

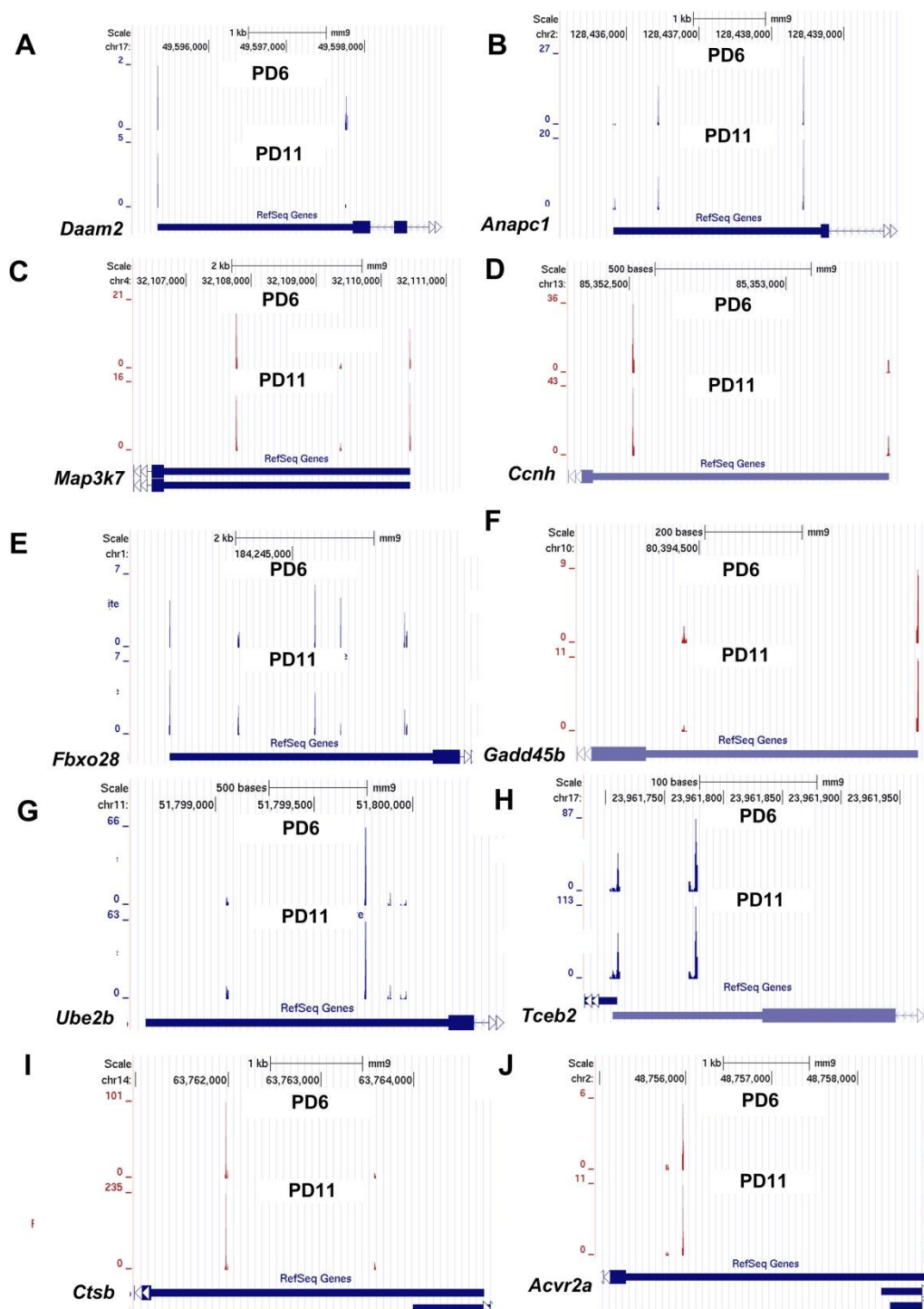


Figure S4. PA-seq track of genes randomly chosen to be validated if they tended to use distal pA sites significantly in senescent MEFs by qRT-PCR. (A-J) PA-seq tracks of *Daam2*, *Anapc1*, *Map3k7*, *Ccnh*, *Fbxo28*, *Gadd45b*, *Ube2b*, *Tceb2*, *Ctsb* and *Acvr2a* in senescent and young MEFs, respectively.

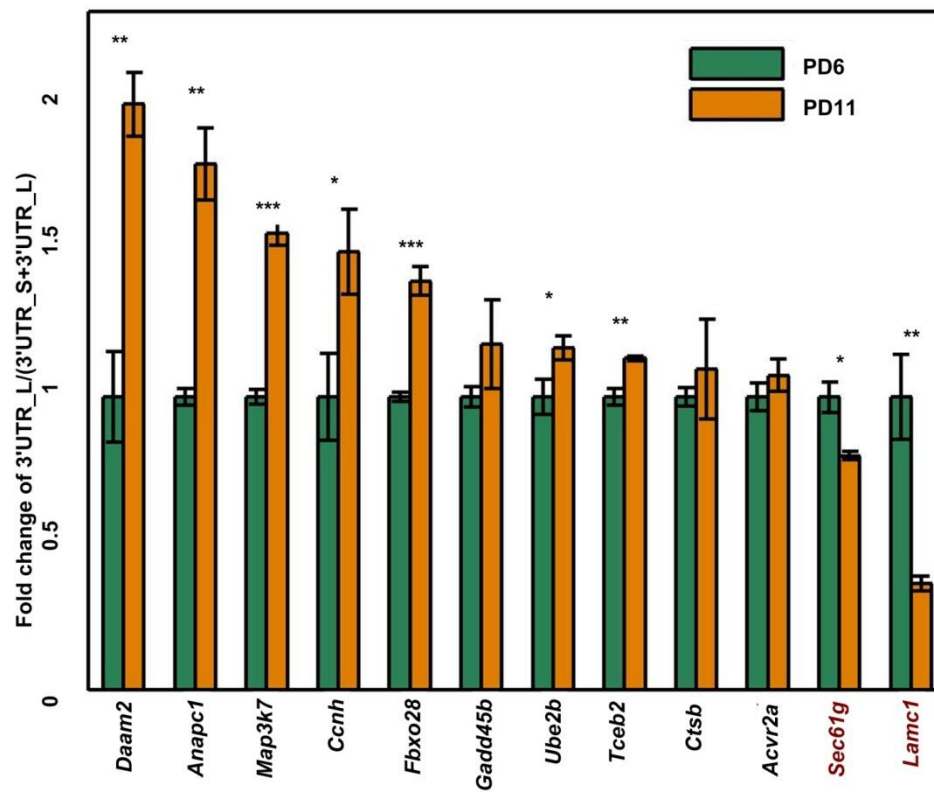


Figure S5. Validation results of genes tended to use distal pAs in senescent MFEs by qRT-PCR. Two genes (*Sec61g* and *Lamc1*) preferred proximal pAs in senescent cells served as negative controls. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Student's t-test.

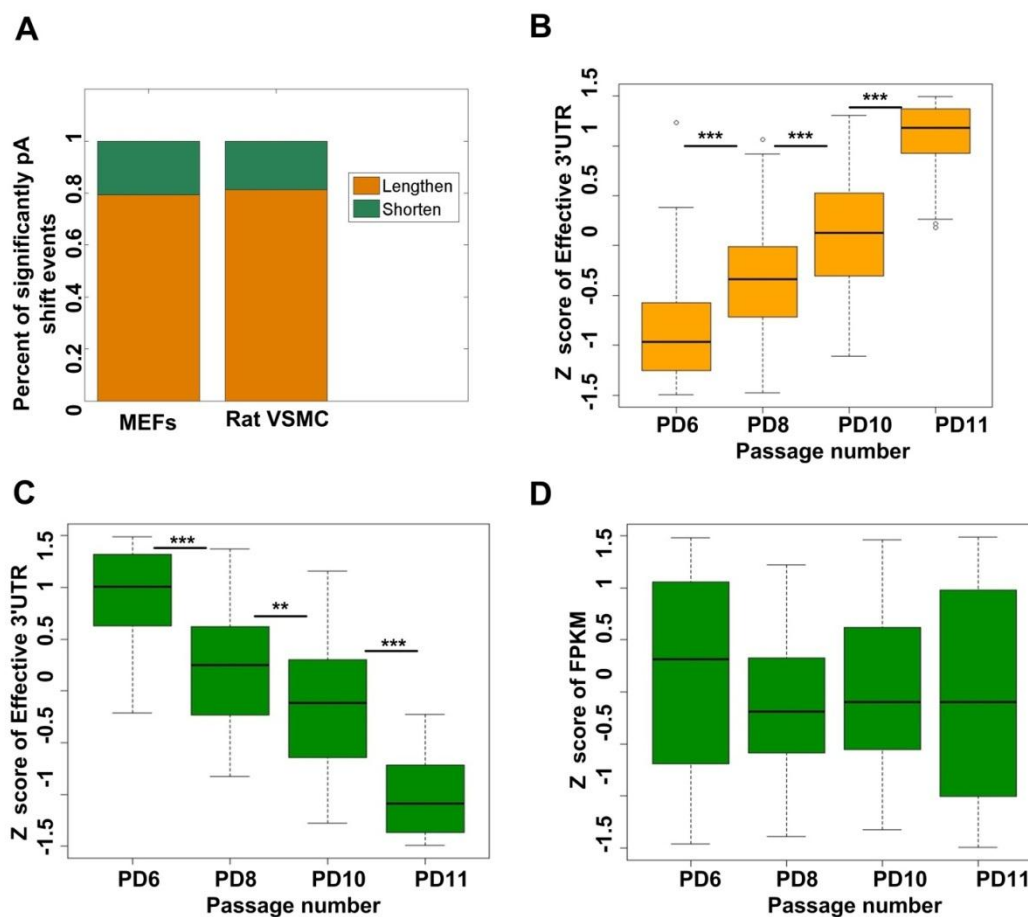


Figure S6, related to Figure 4. Genes tended to use proximal pAs did not have decreased mRNA abundance. (A) Fraction of genes significantly tended to use distal pAs (Lengthen) and proximal pAs (Shorten) when comparing senescent MEFs (PD11) with early passage of MEFs (PD6) and also fraction of genes significantly tended to use distal pAs (Lengthen) and proximal pAs (Shorten) when comparing VSMC (Vascular smooth muscle cells) from old rat (2 years old) with young rat (2 weeks old). (B) Comparing distribution of Z score of effective 3' UTRs across PD6, PD8, PD10 and PD11 for genes with lengthened 3' UTRs during replicative senescence of MEFs. (C) Comparing distribution of Z score of effective 3' UTRs length across PD6, PD8, PD10 and PD11 for genes with shortened 3' UTRs during replicative senescence of MEFs. (D). Comparing distribution of Z score of expression (FPKM) across PD6, PD8, PD10 and PD11 for genes with shortened 3' UTRs during replicative senescence of MEFs. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Wilcoxon signed rank test.

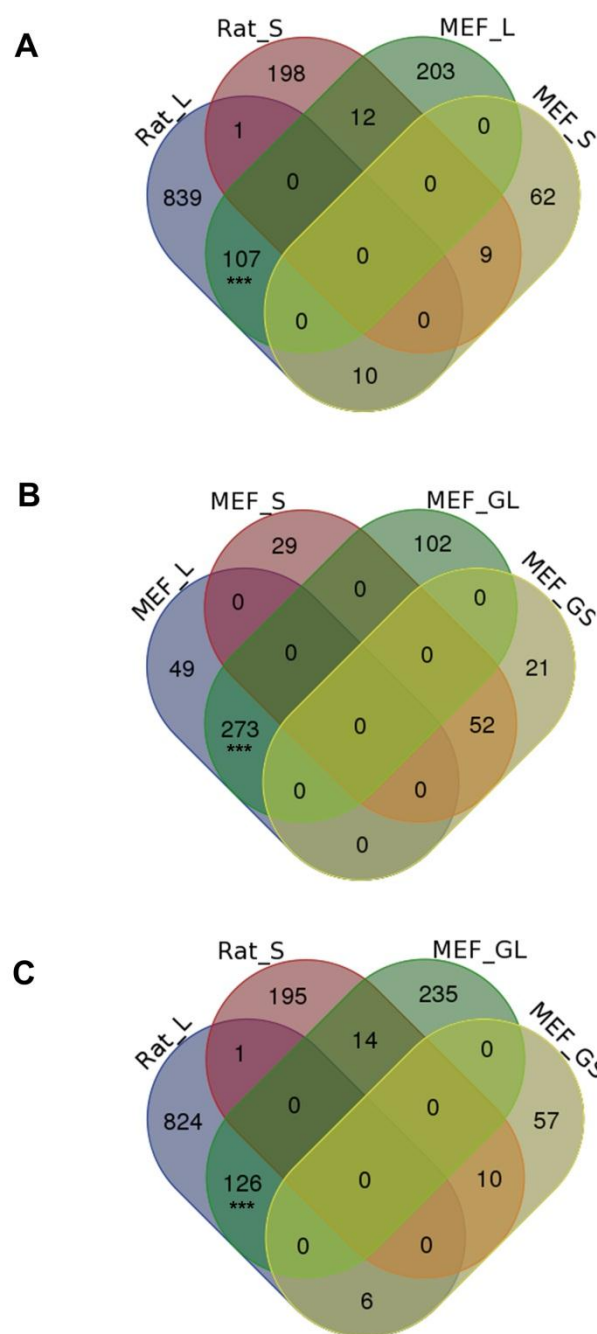


Figure S7. Comparison among genes significantly tended to use distal pAs and proximal pAs in senescent MEFs and VSMCs of old rat. (A) Venn diagram for genes preferring distal (named Rat_L) and proximal (named Rat_S) pAs when comparing VSMCs of old rat with young rat (named as Rat_L and Rat_S, respectively) and genes preferring distal (MEF_L) and proximal (MEF_S) pAs when comparing senescent MEFs (PD11) with Young MEFs (PD6). (B) Venn diagram comparison among MEF_L, MEF_S and genes gradually preferred to use distal (MEF_GL) and proximal (MEF_GS) pAs during replicative senescence of MEFs. (C) Venn diagram comparison among Rat_L, Rat_S, MEF_GL and MEF_GS. MEF_L, MEF_S, Rat_L, Rat_S, MEF_GL and MEF_GS were identified by linear trend test with Benjamini-Hochberg (BH) false-discovery rate (FDR) at 5%. (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, Fishers' exact test.

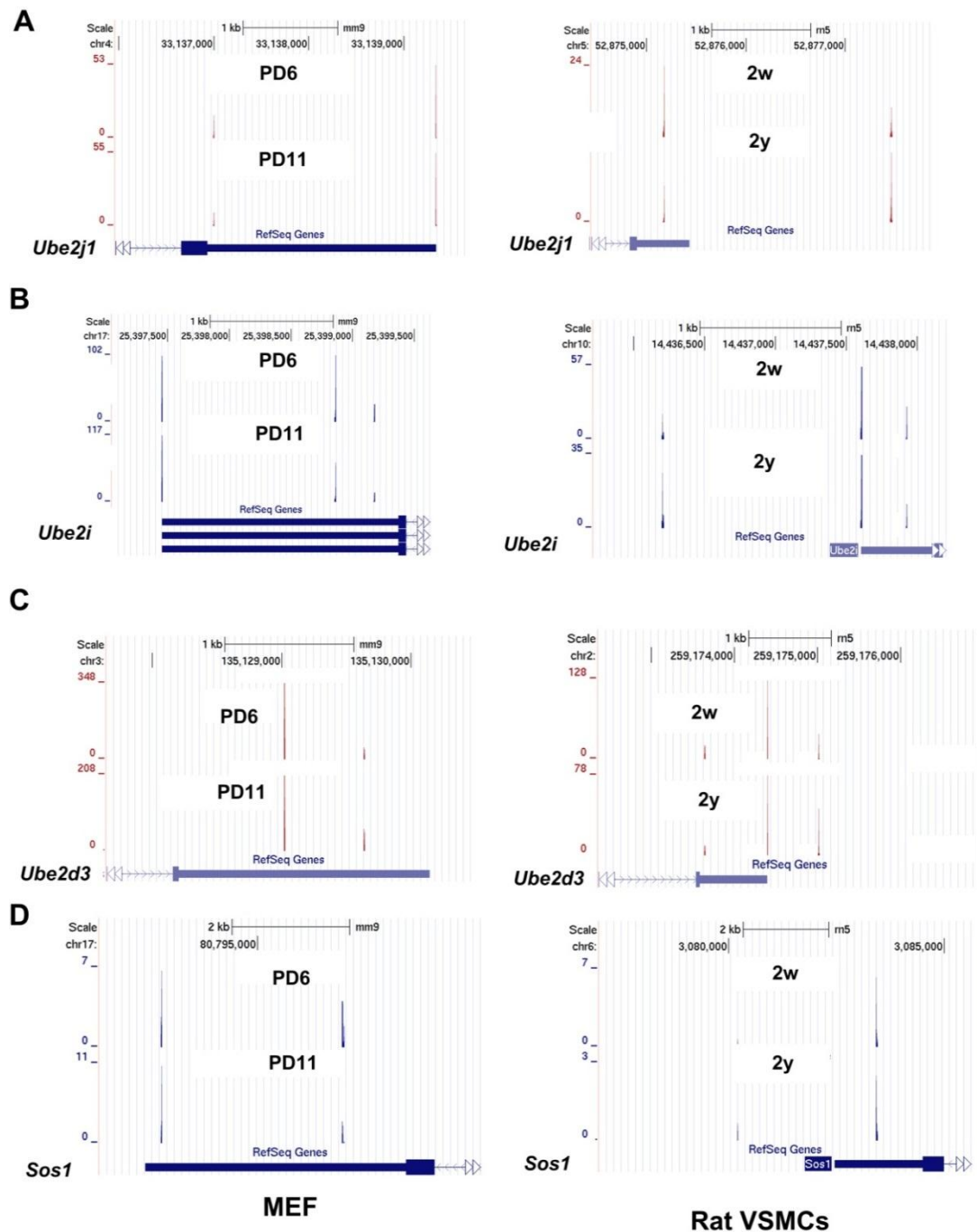


Figure S8. Examples of genes tended to use distal pAs in senescent MEFs and VSMCs of aged rat. (A-D) PA-seq tracks of *Ube2j1*, *Ube2i*, *Ube2d3*, and *Sos1* in senescent and young MEFs and VSMCs of old rat and young rat, respectively.

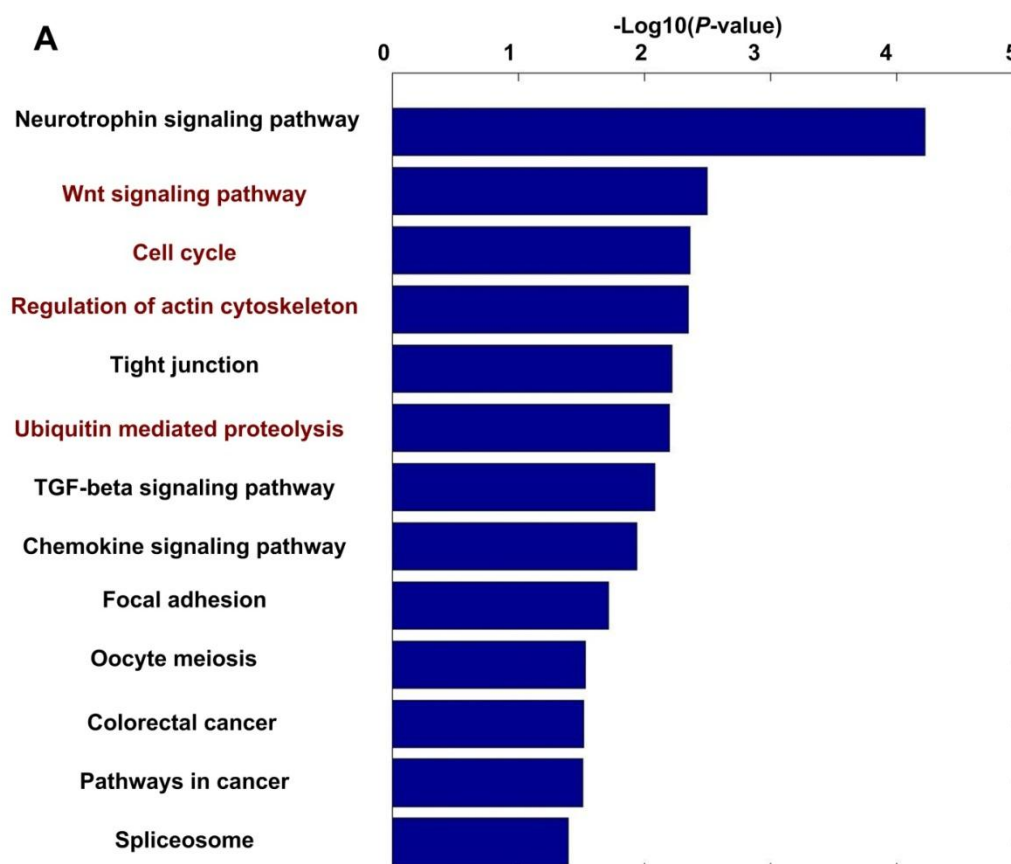


Figure S9. Genes progressively tended to use distal pAs during replicative senescence of MEFs are also enriched in senescence-related pathways. The common senescence-related pathways between Figure S9 and Figure 5A are marked by red color.

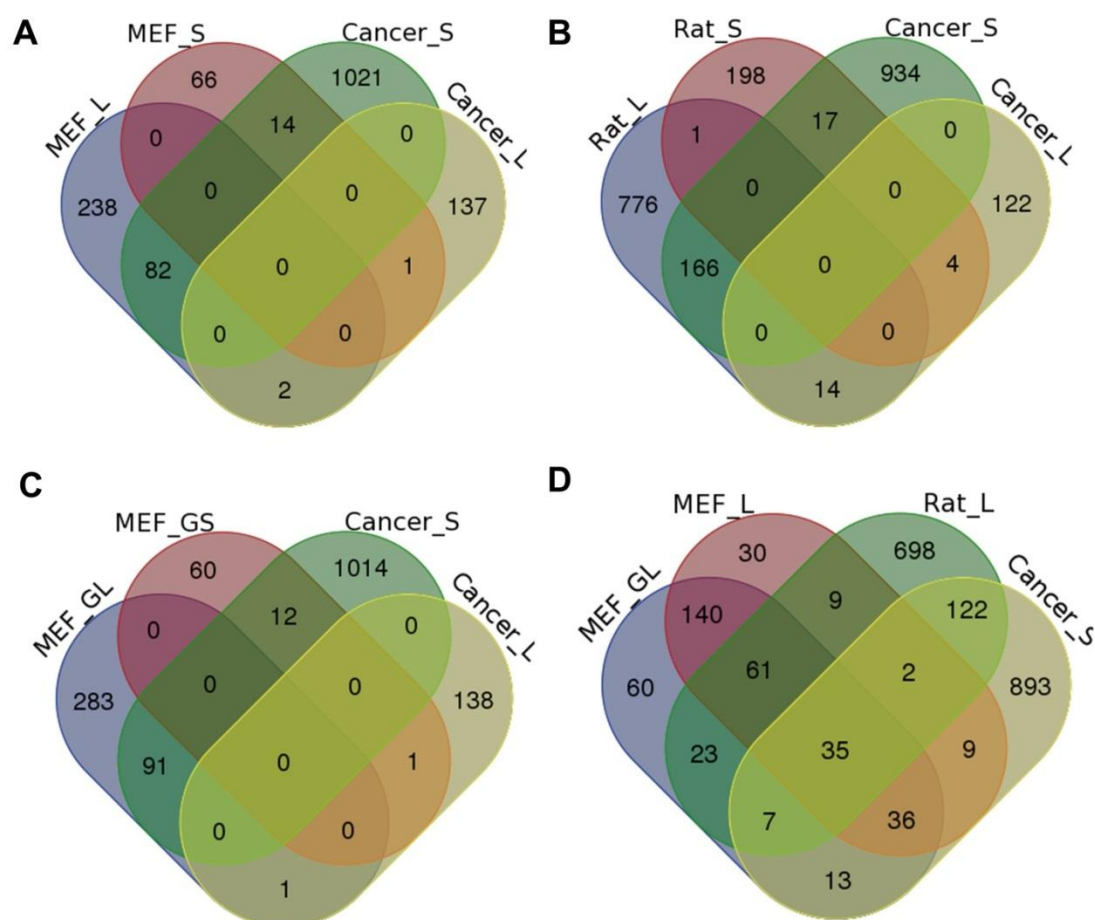


Figure S10. Comparison between genes significantly tended to use distal pAs and proximal pAs in senescent MEFs and VSMCs of old rat with genes significantly tended to use distal pAs and proximal pAs in cancer. (A) Venn diagram comparison among genes preferring distal (MEF_L) and proximal (MEF_S) pAs when comparing senescent MEFs (PD11) with Young MEFs (PD6) and genes preferring distal (Cancer_L) and proximal (Cancer_S) pAs when comparing tumors and normal tissues identified by Xia et al[3]. (B) Venn diagram comparison among genes preferring distal (Rat_L) and proximal (Rat_S) pAs when comparing VSMCs of old rat with young rat, Cancer_L and Cancer_S. (C) Venn diagram comparison among genes gradually preferred to use distal (MEF_GL) and proximal (MEF_GS) pAs during replicative senescence of MEFs, Cancer_L and Cancer_S. (D). Venn diagram comparison among MEF_L, Rat_L, MEF_GL and Cancer_S. MEF_L, MEF_S, Rat_L, Rat_S, MEF_GL and MEF_GS were identified by linear trend test with Benjamini-Hochberg (BH) false-discovery rate (FDR) at 5%.

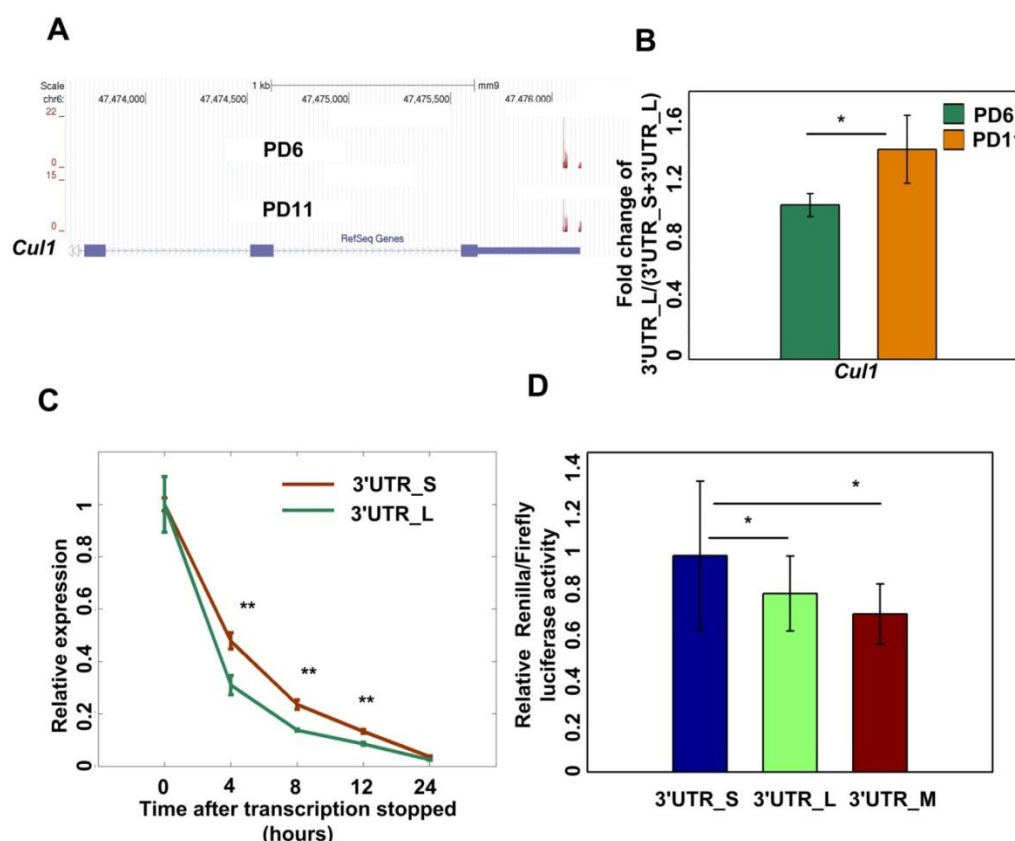


Figure S11. *Cull1* with longer 3' UTR has weaker stability and produces less protein. (A) PA-seq tracks of *Cull1* in senescent and young MEFs. (B) qRT-PCR validation of pAs usage preference for *Cull1* in senescent and young MEFs. (*) $P < 0.05$, two tailed Wilcoxon signed rank test. (C) Comparing stability of the short 3' UTR and the long 3' UTR of *Cull1* (named as 3' UTR_L and 3' UTR_S, respectively) in mouse fibroblast cell line NIH3T3. (**) $P < 0.01$, two tailed Wilcoxon signed rank test. (D) Luciferase expression from a reporter containing the short 3' UTR, compared to that from the reporter containing the long 3' UTR and the long 3' UTR with mutated PAS of proximal pA of *Cull1* (named as 3' UTR_L, 3' UTR_S and 3' UTR_M, respectively). (***) $P < 0.001$, (**) $P < 0.01$ and (*) $P < 0.05$, two tailed Student's t-test.

References:

- [1]. Hu, J., et al., Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA*, 2005. 11(10): p. 1485-93.
- [2]. Cheng, Y., R.M. Miura and B. Tian, Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics*, 2006. 22(19): p. 2320-5.
- [3]. Xia, Z., et al., Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3' -UTR landscape across seven tumour types. *Nature Communications*, 2014. 5: p. 5274.

SupplementaryTables

Supplementary Table S1. MEFs PA-seq reads mapping statistics.

Sample	Total Reads	Strand specific reads	mappable Read1	Read1 mapping rate	mappable Read2	Read2 mapping rate
G0	15,411,083	14,266,123	9,442,546	66.2%	11,471,857	80.4%
PD6	9,423,643	7,784,321	4,478,586	57.5%	5,689,182	73.1%
PD8	13,386,797	11,470,447	7,822,028	68.2%	9,472,369	82.6%
PD10	8,779,338	7,307,736	3,818,641	52.3%	4,799,825	65.7%
PD11	14,425,358	13,168,353	8,848,238	67.2%	10,736,847	81.5%

Supplemental Table S2: 18,639 Refined pAs identified in this study and their assigned categories (available as a separate Excel file).

Supplementary Table S3. MEFs RNA-seq reads mapping statistics.

Sample	Total Reads	Strand specific reads	mappable Read1	Read1 mapping rate	mappable Read2
G0	15,809,150	13,997,188	88.5%	15,809,150	88.6%
PD6	28,132,487	27,048,742	96.1%	26,876,587	95.5%
PD8	25,659,880	24,321,086	94.8%	24,175,499	94.2%
PD10	17,860,083	16,254,722	91.0%	16,060,448	89.9%
PD11	16,396,313	14,101,035	86.0%	13,980,835	85.3%

Supplemental Table S4: Summary 3,165 genes with APA regulation (available as a separate Excel file).

Supplementary Table S5. qRT-PCR primer sequences for genes be validated if

they have significant APA usage switch.

Gene	Common region primers (from 5' to 3')	Alternative region primers (from 5' to 3')
<i>Daam2</i>	GAGCGGGCCAATAAACAGG CCTTTCTTCTAATCCCTGTCTCC	TTCAGGGGTTGGAAGGACAA TTGCTTGTCTTCTGCAGCAG
<i>Anapc1</i>	CCTAGAAATGACTGCGGCAC CATATGACACGTGGACAGCA	TGCCATGGGAGGACTTGAAA GCTGTGCATGGTAAGACTGG
<i>Map3k7</i>	ACAGCAGGCTAATCAGGAGG GGCAACAGACTCAGGAAAGG	GTTGGCACTCACTCACTTGG GACCAGAGCTCACTTCCTGT
<i>Ccnh</i>	AGGAATGGACTGATGACGACC CAGACATGCTTCCTACTTCCG	GCAAACCATGTCTCCTGTG TTGTAAGGGCTTCTGGAGGA
<i>Fbxo28</i>	TTCAGTTTTCTTTGGGGCCG TGACAACTCTCTCCCGACAG	TGCCGATCCACACAGTCATA AGACTCCACCATTCTCCAGC
<i>Gadd45b</i>	AGGTGGCCAGTTACTGTGAA TTAGGGGACAGCAACTCGA	GGAGACTGAGACTTTAGAGCCA CTCCGCTGACTTATGCACAG
<i>Ube2b</i>	TCCAAACAGTCCAGCCAACA TCCTTAAAACCCGTGGCACT	CCAGCTCTGTATTTGGGGCCA TCTCTGCCTTGCTCAAACCA
<i>Tceb2</i>	CATCGAGCCCTTTTCCAGC GGGAAATGGGTCTCTAGGGG	AGGATTCTGGAGGCAGTGC ACTTGGGGTTAAGAGTTCTGTG
<i>Ctsb</i>	AGGACAAATGCCACCTCTCA GAGGACAGGACCAAGGAGAG	AGTATGAGTGCCAGGCCTTT GGCTCTGTGGTAGTGGAAGT
<i>Acvr2a</i>	TGGAAAGCATGGATCTGGGA TCCTTGATTTGGAGAGGGCC	CCTTGCCCAAATCTCCCATG GTGCAAGTTCATGGGACCAA
<i>Sec61g</i>	TGCTTAAACGTGACTGCTTTTC TTACTTCATGCCCTTTCCCC	GGGTGGCTGAGTCCTTCT AGAAACAAAACACCACACAGC
<i>Lamc1</i>	TCCATCGAGAAGCCCTAGTG AGTGATGGAGAGCAGCAGAG	ACTGTGGCCCTTTTCAATGT AAAATGCCAAGTGTTGCGCT
<i>Cul1</i>	AACAGACGCCAATGCCATTT TCCATTCAGACTCGCTCTCG	TAACAGCTGTCGTCTGAGGC ACCAAATGCAACTTGTACAGAAA

Supplementary Table S6. Rat PA-seq reads mapping statistics.

Sample	Total Reads	Strand specific reads	mappable Read1	Read1 mapping rate	mappable Read2	Read2 mapping rate
2 weeks	26,122,246	20,468,514	8,818,888	43.1%	11,694,741	57.1%
2 years	35,087,625	29,152,131	12,443,319	42.7%	16,098,918	55.2%

Supplementary Table S7. Gene list of genes with significant pA usage shift in replicative senescence of MEFs and aortic vascular smooth muscle cells of rats

(VSMCs) with different ages. (Available as a separate Excel file)..

Supplementary Table S8. Functional enrichment analysis for genes with significantly pA usage shift in replicative senescence of MEFs and aortic vascular smooth muscle cells of rats (VSMCs) with different ages. (Available as a separate Excel file).

Supplementary Table 9. Summary of RBPs with higher binding site density in longer 3' UTR than in shorter 3' UTR regions.

S_vs_L_FDR means the false discovery rate obtained by comparing the RNA binding site density between longer 3' UTR and shorter 3' UTR. S_vs_A_FDR means the false discovery rate obtained by comparing the RNA binding site density between longer 3' UTR and alternative 3' UTR. PD6, PD8, PD10, PD11 means the FPKM value of RBPs in the corresponding passage. Wilcox signed rank test was used to get the *P* value and Benjamini-Hochberg was used for multiple test correction.

RBPs	S_vs_L_FDR	S_vs_A_FDR	PD6	PD8	PD10	PD11
RBM28	3.38E-09	6.14E-05	25.27	21.05	22.52	20.60
HNRNPA2B1	3.38E-09	1.91E-04	228.74	171.85	170.35	109.61
HNRNPA1	4.34E-09	4.43E-05	378.49	266.20	202.91	98.76
MSI1	1.31E-08	4.43E-05	1.68	1.65	1.48	1.46
HNRNPF	8.13E-08	1.18E-04	70.76	63.89	64.20	49.49
DAZAP1	9.61E-08	7.68E-04	43.21	35.36	28.37	23.75
SFPQ	2.50E-06	3.60E-02	130.52	113.63	105.04	94.99
HNRNPH2	3.07E-06	3.51E-03	55.14	64.08	63.91	47.25
RBM38	1.25E-05	2.22E-03	10.68	8.95	6.80	5.66
RBM42	1.10E-04	0.03	41.66	40.49	51.28	63.90
HNRPLL	1.63E-04	0.10	47.28	40.30	40.34	18.07
G3BP2	1.63E-04	0.10	51.98	52.83	59.32	41.22
HNRNPK	6.75E-04	0.11	1176.23	928.64	491.28	349.95
PCBP2	6.93E-04	0.10	172.54	174.87	185.19	213.93
TUT1	1.08E-03	0.10	11.82	11.64	11.17	10.66
HNRNPM	1.26E-03	0.08	118.24	110.65	115.05	108.10
RBM41	5.63E-03	0.13	5.03	4.07	3.80	2.51
SNRNP70	5.63E-03	0.44	137.00	101.95	106.26	178.55
IGF2BP3	6.26E-03	0.14	19.24	17.08	15.81	10.54
YBX1	8.55E-03	0.49	599.60	705.81	913.46	794.75
RBM6	0.01	0.61	14.09	11.05	12.62	17.54
PCBP1	0.01	0.59	115.90	133.09	150.31	165.89
RBM5	0.01	0.63	42.58	37.10	33.38	30.13
PABPC4	0.02	0.18	24.29	27.14	28.40	27.08
TARDBP	0.02	0.07	150.22	97.48	117.86	103.30
NOVA1	0.02	0.69	2.25	1.36	3.16	2.97
IGF2BP2	0.03	0.18	39.66	29.70	25.23	24.83
HNRNPU	0.03	0.41	234.72	175.61	133.18	100.88

FMR1	0.03	0.44	28.05	23.50	15.18	9.74
SNRPA	0.05	0.74	84.84	81.36	73.36	102.15

Supplementary Table10. Primer sequences for genes which expression needed to be validated by qRT-PCR.

Gene	Sense primers (from 5' to 3')	Anti-sense primers (from 5' to 3')
<i>Gapdh</i>	GTGTTCTACCCCAATGTGT	ATTGTCATACCAGGAAATGAGCTT
<i>Mki67</i>	ATCATTGACCGCTCCTTTAGGT	GCTCGCCTTGATGGTTCCT
<i>p16</i>	CGCAGGTTCTTGGTCACTGT	TGTTACGAAAGCCAGAGCG
<i>Cstf2</i>	GCACCTCCTATGATGCAAGC	GCAGGGGCTAAGGAACCAG