Transcript shortening through alternative polyadenylation promotes gene expression during fracture healing

Reyad Elbarbary (relbarbary@pennstatehealth.psu.edu)
The Pennsylvania State University College of Medicine https://orcid.org/0000-0002-1006-6877
Deepak Kumar Khajuria
The Pennsylvania State University College of Medicine https://orcid.org/0000-0002-3737-8504
Irena Nowak
The Pennsylvania State University College of Medicine
Ming Leung
The Pennsylvania State University College of Medicine
Vengadeshprabhu Karuppagounder
The Pennsylvania State University College of Medicine
Yuka Imamura
Pennsylvania State University College of Medicine
Christopher Norbury
The Pennsylvania State University College of Medicine
Fadia Kamal
The Pennsylvania State University College of Medicine

Article

Keywords:

Posted Date: July 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1762249/v1

Transcript shortening through alternative polyadenylation promotes gene expression during fracture healing

Running title: Alternative polyadenylation in bone fracture healing

Deepak K. Khajuria[1,2], Irena Nowak[1,2], Ming Leung[3], Vengadeshprabhu Karuppagounder[1,2],

Yuka Imamura[3,4], Christopher C Norbury[5], Fadia Kamal[1,2,4], Reyad A. Elbarbary[1,2,6,7*]

[1]Department of Orthopaedics and Rehabilitation, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA
[2]Center for Orthopaedic Research and Translational Science (CORTS), The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA
[3]Institute for Personalized Medicine, Penn State College of Medicine, Hershey, PA 17033, USA
[4]Department of Pharmacology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA
[5]Department of Microbiology and Immunology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA
[6]Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA
[7]Center for RNA Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

Email addresses:
Deepak K. Khajuria: dkhajuria@pennstatehealth.psu.edu; Irena Nowak: inowak@pennstatehealth.psu.edu; Ming Leung: mleung@pennstatehealth.psu.edu; Vengadeshprabhu Karuppagounder: vkaruppagounder@pennstatehealth.psu.edu; Yuka Imamura: yimamura@pennstatehealth.psu.edu; Christopher C Norbury: ccn1@psu.edu; Fadia Kamal: fkamal@pennstatehealth.psu.edu

*Correspondence: Reyad A. Elbarbary (relbarbary@pennstatehealth.psu.edu; rae19@psu.edu).
Telephone: 717-531-4836.
Fax: 717-531-0349

Abstract

Maturation of the 3′ end of almost all eukaryotic messenger RNAs (mRNAs) requires cleavage and polyadenylation. Most mammalian mRNAs are polyadenylated at different sites within the last exon, generating alternative polyadenylation (APA) isoforms that have the same coding region but distinct 3′ untranslated regions (UTRs). The 3′UTR contains motifs that regulate mRNA metabolism; thus, changing the 3′UTR length via APA can significantly impact gene expression. Endochondral ossification is a central process in bone healing, and the impact of APA on gene expression during this process is unknown. Here, we report widespread utilization of APA that impacts multiple pathways with established roles in bone healing. Importantly, progression of endochondral ossification is typified by global 3′UTR shortening that is coupled with an increased abundance of shortened transcripts as compared to all other transcripts, underscoring the role of APA in promoting gene expression during endochondral bone formation. Our mechanistic studies of genes that undergo APA in the fracture callus uncover an intricate regulatory network in which APA boosts the expression of collagen, type I, alpha 1 (Col1a1) and Col1a2 genes, which encode the 2 subunits of the abundantly expressed protein collagen 1. APA does so via shortening the 3′UTRs of Col1a1 and Col1a2 mRNAs, which removes the binding sites of miR-29a-3p that otherwise potently triggered the degradation of both transcripts. Taken together, our study takes the lead in characterizing crucial roles of APA in tailoring the 3′UTR landscape and regulating gene expression during fracture healing.

Introduction

Endochondral ossification is the process through which long bones form during development and is the most common way for adult bone healing[1-3]. Most of cortical bone fractures heal via initial formation of a fibrocartilaginous (soft) callus filled with proliferating chondrocytes, which at later stages terminally differentiate to hypertrophic chondrocytes and secret extracellular matrix components, including type I and type 10 collagen, that mediate gradual mineralization of the soft callus and hardening of the fracture gap[1-3]. This enables invasion of new blood vessels and subsequent replacement of chondrocytes with newly formed woven bone via the endochondral ossification process[1-3].

The process of 3′ end formation of most of eukaryotic messenger RNAs (mRNAs) occurs through co-transcriptional recognition of a specific sequence motif called the polyadenylation signal (PA), followed by endonucleolytic cleavage and addition of a polyadenosine (poly A) tail at the cleavage and polyadenylation site (CPS) [4-7]. More than 70% of mammalian genes have multiple PAs, the utilization of which results in the expression of alternative polyadenylation (APA) isoforms[7,8]. As most of the PAs are localized within the last exon, APA isoforms have the same coding sequence (CDS) but differ in the size of the 3′ untranslated region (UTR)[9]. The 3′ UTR contains several sequence and structure motifs, including microRNA (miRNA) binding sites, that regulate mRNA metabolism at several levels[7,8,10]. miRNAs are short ~22-nucleotide RNAs that direct posttranscriptional repression of most mammalian genes via inhibition of mRNA translation and/or promotion of mRNA decay[11,12]. Accordingly, removal of miRNA binding sites and other important regulatory motifs from the 3′ UTR via APA can significantly impact mRNA translatability, stability, and/or localization[7,8,10]. Although less probable, PAs might also exist within introns, leading to APA isoforms with different CDSs[13,14]. APA is

regulated during different biological processes. For example, rapidly proliferating cells exhibit widespread shortening of 3′ UTRs, which is in stark contrast to the general trend of 3′ UTR lengthening that occurs in differentiating cells[8,10].

The increased interest in APA and its substantial roles in post-transcriptional regulation of gene expression has spurred the development of novel bioinformatic tools that globally identify and quantify APA isoforms in RNA-sequencing (RNA-seq) data. One of the powerful tools is Dynamic analyses of Alternative PolyAdenylation from RNA-Seq (DaPars) that quantifies dynamic APA events via analysis of localized changes in RNA-seq density near the 3′ end of mRNA[15]. DaPars has been adopted by many databases and used to analyze changes in the abundance of APA isoforms under different biological conditions, including analysis of 3′ UTR shortening in samples from The Cancer Genome Atlas[8,16]. Limitations of DaPars include that it can only handle two-group, but not one-group, datasets. In addition, the coverage of RNA-seq reads around tandem poly(A) sites is not uniform, which increases the rate of false positive detection of PA usage[8]. To address these limitations, other bioinformatic tools have been developed including the Identification of Novel alternative PolyAdenylation Sites (InPAS) algorithm. InPAS is based on and was developed from DaPars; however, it adjusts cleavage sites by utilizing annotation packages to correct 3′ UTR ranges and prevent false CPS detection[17,18]. Recently, the APAlyzer bioinformatic toolkit has been also developed to overcome the intrinsic limitation of de novo CPS prediction by utilizing the comprehensive CPS collection in the PolyA_DB database[19-21]. The APAlyzer also examines APA events in all genic regions, including 3′ UTRs and introns[18,19].

Although APA events and their roles in regulating gene expression have been identified in several tissues and disease conditions[8,15,18,22], they are understudied in musculoskeletal tissues

and uncharacterized during bone regeneration. Here, we analyze for the first time APA events in the healing callus and elucidate their roles in regulating gene expression post-transcriptionally during endochondral ossification.

Results

Replacement of cartilaginous callus by woven bone is associated with increased cell proliferation and immune-cell infiltration

We performed tibial diaphyseal osteotomy and stabilized the fracture using an intramedullary nail; an established model for secondary fracture healing that proceeds through the endochondral ossification process[1-3]. As we and others reported in this model[23-25], the fracture gap during the early repair phase, i.e., between days 7 and 10 post-fracture, was filled with and surrounded by a fibrocartilaginous callus (**Fig. S1a**). As healing proceeded to day 14 post-fracture (d14), the chondrocytes filling the soft callus progressed into the hypertrophic status (**Fig. 1a**; **Fig. S1b**) and secreted the hypertrophy markers type 10 collagen (Col X) (**Fig. 1b**), matrix metallopeptidase 13 (MMP13) (**Fig. 1c**), and type I collagen (Col I) (**Fig. 1d**); all are essential proteins for soft-callus mineralization. As healing further progressed from d14 to d21, endochondral ossification mediated complete replacement of the cartilaginous callus by newly formed woven bone, accompanied by re-establishment of the bone marrow population (**Fig. 1e, f**). Quantitative measurement of gene expression using reverse transcriptase combined with quantitative polymerase chain reaction (RT-qPCR) corroborated our histological analysis and demonstrated almost undetectable expression of chondrocyte and hypertrophic-chondrocyte markers at d21 (**Fig. 1g**), indicating complete resorption of the cartilaginous callus.

To gain insight into biological processes involved in the calcification and resorption of the cartilaginous callus as well as formation of woven bone, we performed RNA-seq using total RNA isolated from d14 and d21 callus tissues. We identified 9330 differentially expressed genes ($P_{adj} < 0.05$; DESeq2[26]) at d21 relative to d14 (**Fig. 1h**; **Fig. S1c**; **Table S1**). As expected, the expression of genes involved in cartilage development and ossification was downregulated at d21 as compared to d14 (gene ontology (GO) analysis[27,28]; **Fig. 1i**; **Figs S2** and **S3**; **Tables S2** and **S3**). On the other hand, d21 showed upregulation in several pathways involved in immune-cell responses (**Fig. 1i**; **Figs S2** and **S3**; **Tables S2** and **S3**), consistent with re-establishment of bone marrow population. Importantly, cell division, DNA replication, and chromosome segregation were among the most upregulated pathways at d21 (**Fig. 1i**; **Figs S2** and **S3**; **Tables S2** and **S3**). Accordingly, as endochondral ossification proceeds, rapidly proliferating cells, including immune cells, populate the callus to replace cells with a lower proliferation capacity. These observations are important to our studies as rapidly proliferating cells are reported to exhibit global 3′ UTR shortening[8,10] (see below).

The transcriptome of the healing callus exhibits widespread utilization of alternative PAs

When a gene contains multiple PAs in the last exon (**Fig. 2a**), utilization of the distal PA (dPA) during mRNA maturation results in incorporation of the alternative UTR (aUTR) sequences, generating an APA isoform with a longer 3′ UTRs (long APA isoform "lAPA") (**Fig. 2a**). On the other hand, utilization of the proximal PA (pPA) removes the aUTR region and limits the size of the 3′ UTR to the constitutive UTR (cUTR), generating a short APA isoform (sAPA) (**Fig. 2a**). To study the extent of APA in the healing callus, we first performed transcriptomic analysis of APA isoforms in the callus of d14 using the DaPars-based InPAS algorithm[17,18]. Quantitation of

the relative expression of APA isoforms is based on calculating the Percent of Distal Utilization Index (PDUI) (**Fig. 2a**), which indicates the fraction of each analyzed transcript that is expressed as lAPA isoform (**Fig. 2a**). Results indicated that ~14% of analyzed genes exclusively utilized the dPA (PDUI = 1; Group 1 in **Fig. 2b**; **Table S4**), while ~2% exclusively utilized the pPA (PDUI = 0, Group 7 in **Fig. 2b**; **Table S4**). Crucially, ~84% of the genes utilized both the dPA and pPA to variable extents (Groups 2-6 in **Fig. 2b**; **Table S4**). Henceforth, this group will be called APA genes/events. Performing the same analysis using the APAlyzer, which utilizes a different approach to quantitate APA isoforms, generated comparable data as APA events were identified in ~78% of genes, showing ~81% overlap with InPAS-identified APA events (**Fig. 2c; Table S5**). Importantly, in ~50% of APA genes, the minor APA isoform constituted ≥ 20% of the total gene reads (**Fig. 2b**), and GO analysis[29] of this subset of APA genes indicated significant enrichment in a vast array of biological pathways that play essential roles in fracture healing (**Fig. 2d**; **Fig. S4**; **Table S6**). In fact, a sizable number of transcripts with established roles in ossification, angiogenesis, osteoblast differentiation, endochondral ossification, and deposition of extracellular matrix generated PDUI of 40-50%, suggesting substantial utilization of both pPA and dPA (**Table S7**). Examples of these genes include the bone matrix component collagen, type I, alpha 1 (Col1a1) and alpha 2 (Col1a2), the angiogenic factor vascular endothelial growth factor A (Vegfa), the osteoclast marker acid phosphatase 5, tartarate resistant (Acp5; also known as TRAP), and the bone/cartilage development inducer bone morphogenic protein 5 (Bmp5) (**Table S7**). These results indicate that thousands of genes are expressed in the callus as multiple APA isoforms and suggest critical roles of APA in bone regeneration.

Utilization of pPA generates extremely shortened 3′ UTRs of Col1a1 and Col1a2 transcripts

Type I Collagen, which is the main component of organic bone matrix, is a heterotrimer comprising two α1 chains (encoded by Col1a1) and one α2 chain (encoded by Col1a2)[30]. According to our RNA-seq data analyses, Col1a1 and Col1a2 are the most abundant transcripts in d14 callus and are both expressed as 2 main APA isoforms (**Fig. 3a**; **Tables S4**, **5**, **7**). To confirm these results, we mapped the 3′ end of both Col1a1 and Col1a2 mRNAs by performing rapid amplification of cDNA 3′ ends (3′ RACE) (**Fig. S5a**, **b**) using RNA purified from d14 callus. Indeed, we identified two APA isoforms of Col1a1 and Col1a2 mRNAs (**Fig 3a-c)** and mapped their 3′ ends to CPSs that are in line with bioinformatic predictions (**Fig. S6 a, b**). Interestingly, we detected the same Col1a1 and Col1a2 APA isoforms upon performing 3′ RACE on RNA isolated from undifferentiated or differentiated murine osteoblastic MC-3T3-E1 cells (**Fig. 3d-f**).

Type I collagen is also abundantly secreted by mineralizing chondrocytes (**Fig. 1d**)[1,3,23]. To investigate whether the expression of Col1a1 and Col1a2 APA isoforms is specific to bone cells or conserved in hypertrophic chondrocytes, we differentiated the chondrogenic ATDC5 cells under conditions that induced chondrocyte hypertrophy and mineralization (**Fig. 3g, h**; **Fig. S7**). We detected the same APA isoforms in hypertrophic ATDC5 cells (**Fig. 3d, e**). Notably, we sequenced the 3′ RACE-PCR products of Col1a1 and Col1a2 in all these experiments and confirmed that the 3′ end of each of the lAPA and sAPA isoforms is conserved among the studied cell lines and the callus (**Fig. S6a**, **b**). Thus, utilization of pPAs by Col1a1 and Col1a2 genes generates extremely shortened 3′ UTRs and is widespread in bone cells as well as hypertrophic chondrocytes.

3′ UTR shortening de-represses the expression of Col1a1 and Col1a2

We constructed a reporter containing the 3′ UTR of Col1a1 lAPA isoform (lAPA 3′ UTR), which represents the full-length 3′ UTR, downstream to the firefly luciferase (FLuc) coding region (FLuc-lAPA 3′ UTR**; Fig. 4a**). As expected, when introduced into MC3T3 cells, this reporter expressed 2 APA isoforms of FLuc mRNA (**Figs. 4b**). The 3′ end of each of the 2 isoforms was identical to that of endogenous Col1a1 (**Fig. S6 a**). Similarly, a reporter containing the lAPA 3′ UTR of Col1a2 mRNA expressed 2 APA isoforms of FLuc mRNA identical to those of endogenous Col1a2 (**Figs. 4c; Fig. S6 b**). Accordingly, the FLuc-lAPA 3′ UTR reporters recapitulate alternative cleavage and polyadenylation of the corresponding endogenous transcripts. We next examined the 3′ UTRs of both Col1a1 and Col1a2 for putative PAs that might modulate proximal cleavage and polyadenylation. In each transcript we identified an A-rich simple sequence repeat (SSR) that contains overlapping canonical (AAUAAA) PAs located < 50 nt upstream to the proximal CPS (**Supp Fig. 6 a, b**). Deletion of the SSR (FLuc-ΔpPA 3′ UTR reporter; **Fig. 4a**) almost completely abolished 3′ UTR shortening in both Col1a1 and Col1a2 reporters, resulting in the expression of each reporter as a single APA isoform with a full-length 3′ UTR (**Fig. 4b, c**). These results define SSR-embedded PAs as the main driver of 3′ UTR shortening in both Col1a1 and Col1a2 genes.

To investigate how shortening of 3′ UTRs impacts gene expression, we constructed an additional FLuc reporter containing the sAPA 3′ UTR of Col1a1 or Col1a2 (FLuc-sAPA 3′ UTR**; Fig. 4d**). Comparing the reporter activity of FLuc-ΔpPA 3′ UTR and FLuc-sAPA 3′ UTR constructs indicated potent repression of FLuc activity (> 70%) by the lAPA 3′ UTR of either Col1a1 or Col1a2 (**Fig. 4e, f**), while the repression observed by the sAPA 3′ UTR was very moderate (~25%) (**Fig. 4e**) and was not significant (**Fig. 4f)** for Col1a1 and Col1a2, respectively.

These results demonstrate that the aUTR region, which is removed during 3′ UTR shortening, has a potent inhibitory effect on the expression of Col1a1 and Col1a2. Consistent with this observation, a reporter that contains the aUTR region of Col1a1 or Col1a2 (FLuc-aUTR**; Fig. 4d)** repressed FLuc activity to the same extent observed with the full-length 3′ UTR (**Fig. 4e, f**). To corroborate the reporter results and investigate the impact of 3′ UTR shortening on endogenous transcripts, we performed pulse-chase experiments and assessed the stability of endogenous Col1a1 and Col1a2 APA isoforms. Results indicated substantially faster degradation of the lAPA isoforms relative to the sAPA isoforms (**Fig. 4g**). Thus, 3′ UTR shortening enhances Col1a1 and Col1a2 expression via, at least in part, stabilization of the transcripts.

Shortening of the 3′ UTR enhances Col1a1 and Col1a2 expression via removal of miR-29a-3p binding sites

To define the inhibitory elements embedded within the aUTR of Col1a1 mRNA, we searched for, among others, putative miRNA response elements (MREs). Among the top scoring miRNAs with conserved MREs, four were identified by both TargetScan[31] and miRDB[32]: namely the three members of the miR-29 family and miR-6980-5p (**Fig. 5a; Tables S8, S9**). In fact, miR-29 family members have been reported to inhibit the expression of Col1a1 with variable efficiency in different tissues[33-36]. miRNA sequencing (miR-seq) on small RNA isolated from d14 callus identified miR-29a-3p as one of the top 15 expressed miRNAs, generating ~20,000 counts per million (CPM) (**Table S10**). The expression level of miR-29a-3p was ~8-fold and 34-fold higher than that of miR-29c-3p and miR-29b-3p, respectively (**Table S10**). On the other hand, miR-6980-5p was expressed at very low levels, generating < 10 CPM (**Table S10**). Accordingly, we focused our subsequent studies on miR-29a-3p. In situ hybridization confirmed miR-29a-3p

expression at high levels in woven bone osteoblasts and lining cells (**Fig. 5b**). Interestingly, miR-29a-3p has three MREs in the 3′ UTR of Col1a1 mRNA, all of them are in the aUTR region (**Fig. 5c**; **Fig. S8a**; **Tables S8, S9**). miR-29a-3p also has a single MRE in the aUTR of Col1a2 mRNA (**Fig. 5d**; **Fig. S8b**; **Tables S8, S9)**, suggesting that miR-29a-3p suppressed the expression of both Col1a1 and Col1a2 via aUTR-embedded MREs. To further investigate this hypothesis, we used the same FLuc reporters that we constructed to have different isoforms and regions of Col1a1 or Col1a2 3′ UTR (**Fig. 5e, f**). A miR-29a-3p mimic effectively repressed the reporters that contained ΔpPA 3′ UTR or aUTR of either Col1a1 or Col1a2 by ≥50% (**Fig. 5e, f**). In contrast, the miR-29a-3p mimic had no effect on the sAPA reporters (**Fig. 5e, f**). Mutating the single MRE in Col1a2 aUTR (FLuc-ΔMRE) abolished the inhibitory effect of miR-29a-3p (**Fig. 5f)**. These data confirm that miR-29a-3p effectively represses the expression of Col1a1 and Col1a2 via binding to the aUTR. Consist with this, the miR-29a-3p mimic downregulated the total level of endogenous Col1a1 mRNA by 30-40%, but exerted substantially more repression of the lAPA isoform and downregulated its level by 70-80% (**Fig. 5g**). Comparable results were obtained with endogenous Col1a2 mRNA (**Fig. 5h**). Taken together, these results underscore the role of 3′ UTR shortening in de-repressing the expression of Col1a1 and Col1a2 in biological contexts where miR-29a-3p, and possibly other members of miR-29 family, are highly expressed.

Endochondral ossification is accompanied by prevalent 3′ UTR shortening

Our RNA-seq data indicate that progression of endochondral ossification is associated with increased cell proliferation (**Fig. 1i; Figs S2-S4; Tables S2 and S3**). As rapidly proliferating cells exhibit widespread shortening of 3′ UTR[8,10], we hypothesized that 3′ UTR shortening is

prevalent during endochondral ossification. To investigate this, we used the InPAS to quantitate changes in the ratios of APA isoforms at d21 as compared to d14 (InPAS calculates changes as log_2fold-change in PDUI values; **Fig. 6a**; **Table S4**) and corroborated the results using the APAlyzer (calculates changes as relative expression difference (RED) values; **Fig. 6a**; **Table S5**). Positive PDUI log_2FC or RED values indicate higher ratios of lAPA isoforms (i.e., lengthening events), while negative values indicate shortening events. Both InPAS and APAlyzer identified a global trend toward 3′ UTR shortening at d21 relative to d14 **(Fig 6b)**. Limiting InPAS analysis to transcripts that exhibited significant changes in 3′ UTR length (*Padj* < 0.05 and > 20% change in the expression ratio of the 2 APA isoforms) demonstrated that the number of significant shortening events was ~24-fold higher than the number of significant lengthening events (**Fig. 6c**), providing further evidence of substantial 3′ UTR shortening. APAlyzer confirmed these results (**Fig. S9 a, b**) and corroborated ~95% of the InPAS-identified significant shortening events (**Fig. 6d**). We also identified APA events that occur in introns (intronic polyadenylation; IPAs) but did not detect any trend toward IPA activation or suppression at d21 relative to d14 (**Fig. S9c, d; Table S11**). Importantly, GO analysis[37] of genes exhibiting significant 3′ UTR shortening demonstrated enrichment in pathways associated with cell cycle and proliferation, cellular immune response, cellular growth, assembly, and organization, cell-to-cell signaling, and transcriptional regulation (**Tables S12 and S13**).

To investigate whether the length of the 3′ UTR correlates with the extent/frequency of pPA utilization, we classified transcripts that undergo significant 3′ UTR shortening into 5 categories based on the length of the 3′ UTR (**Fig. 6e, top**) or aUTR (**Fig. 6f, top**) and calculated the average PDUI log_2FC for each category. We found that genes displayed a greater extent of

pPA utilization (i.e., higher frequency of 3′ UTR shortening) as the length of the 3′ UTR (**Fig. 6e, bottom**) or aUTR (**Fig. 6f, bottom)** increases.

Transcripts that undergo 3′ UTR shortening exhibit higher expression

The 3′ UTR plays crucial roles in regulating the steady-state level of transcripts. We investigated the global impact of d14-to-d21 3′ UTR shortening on gene expression and found that genes whose expression was upregulated at d21 relative to d14 displayed a greater extent of 3′ UTR shortening than genes that showed no change in expression (**Fig. 7a**; **Fig. S10a**). In a striking contrast, downregulated genes displayed the opposite trend and exhibited 3′ UTR lengthening (**Fig. 7a**; **Fig. S10a**). Consistent with this global trend, when we limited our analysis to genes that displayed significant 3′ UTR shortening ($P_{adj} < 0.05$ and > 20% change in the ratio of APA isoforms; **Supp Table 4**), we found their expression to be significantly higher than all other genes (**Fig. 7b, Supp Fig. 10b**). These results highlight the impact of 3′ UTR shortening on boosting gene expression during endochondral ossification.

Interleukin-1 receptor-associated kinase 2 (Irak2), a serine/threonine kinase that associates interleukin-1 receptor upon activation, is one of the genes that was identified by InPAS and APAlyzer to undergo significant 3′ UTR shortening during endochondral ossification (**Fig. 7c**, **Tables S4** and **S5**). Shortening of the Irak2 3′ UTR was accompanied by a significant increase in gene expression as demonstrated by RNA-seq data (**Table S1**) and RT-qPCR quantitation (**Fig. 7d**), consistent with the global trend. 3′ RACE experiments confirmed expression of 2 main APA isoforms of Irak2 in the callus (**Fig. 7e**), mapped the 3′ end of each isoform (**Fig. S10c**), and confirmed shortening of the Irak2 3′ UTR at d21 relative to d14 (**Fig. 7e**). We then investigated whether the relative expression ratio of Irak2 APA isoforms varies

among cell types and found that while the lAPA isoform is the major isoform in differentiated MC3T3 and ATDC5 cells, it is expressed at very low levels in total bone marrow cells (**Fig. 7f, g)**. Expression of the lAPA isoform in both immune ($CD45^+$) and non-immune ($CD45^-$) cells fractionated from bone marrow was barely detectable (**Fig. 7g)**. Interestingly, primary monocytes (isolated from bone marrow) almost exclusively expressed the sAPA isoform (**Fig. 7h**), and *in vitro* differentiation of monocytes into macrophages resulted in mild 3′ UTR lengthening (**Fig. 7h, i**). Activation of mature macrophages using lipopolysaccharide (LPS) did not change the relative expression ratio of APA isoforms (**Fig. 7h, J**). Accordingly, bone marrow cells, which populate the re-established intra-medullary cavity at d21, are a major source of Irak2 sAPA isoform, explaining, at least in part, the observed d14-to-d21shortening of Irak2 3′ UTR.

Discussion

Secondary bone healing via endochondral ossification is the prevalent healing mechanism clinically[1-3,23]. During secondary healing, stromal progenitors differentiate into chondrocytes due to the hypoxic conditions that develop at the fracture gap as a result of the rupturing of blood vessels[1-3]. Chondrocytes proliferate to form the fibrocartilaginous (soft) callus, and then undergo terminal hypertrophic differentiation. Hypertrophic chondrocytes secret extracellular proteins, including Col I and Col X, that facilitate mineralization of the soft callus and metalloproteinases, including MMP13, that reorganize the extracellular matrix[1-3,23]. Mineralization of the soft callus enables formation of new blood vessels across the fracture site, and hypertrophic chondrocytes are gradually replaced by newly formed bone via the process of endochondral ossification[1-3]. This process resembles the endochondral formation of long bone from hypertrophic chondrocytes during development.[38] Our data provide a comprehensive comparison between the

transcriptome of d14 callus, which represents the terminal stages of soft-callus hypertrophy and mineralization, and that of d21 callus, the time point at which the fracture gap is completely bridged by woven bone. Genes expressed in the callus of d14 are enriched in pathways involved in cartilage and connective tissue development, ossification of the fibrocartilaginous callus, and extracellular matrix reorganization. Replacement of mineralizing hypertrophic chondrocytes by developing woven bone is accompanied by a significant increase in the immune and inflammatory response, including activation of leukocytes, adaptive immune response, cytokine production, cell cycle, nuclear division, and DNA replication. Accordingly, as endochondral ossification proceeds, the callus becomes populated by rapidly dividing cells. It is noteworthy that the intramedullary cavity is re-established and invaded by bone marrow at d21, which may partially account for the observed increase in immune response. However, some of the transcriptomic changes identified, particularly those related to the adaptive immune system, may also suggest active infiltration of immune cells into d21 callus. The roles of different immune cell populations during these late phases of healing are yet to be identified.

The boundaries of the 3′UTRs of most RNA polymerase-derived transcripts are generated co-transcriptionally by recognition of PAs and endonucleolytic cleavage at CPSs, followed by addition of poly (A) tails[8]. As most mammalian genes have multiple PAs that are tandemly located in the last exon, mRNAs are expressed as APA isoforms that have the same CDS but distinct 3′UTRs. Notably, during evolution the size of 3′ UTRs has expanded from ~140 nts in Caenohabditis elegans to 1-2 kilobases in humans[10], producing an increased number of 3′ UTR-embedded elements that regulate mRNA processing and metabolism[7,8,10]. Thus, APA has significant functional consequences and contributes to diversification of the transcriptome and establishment of cell identity[39]. Global reprogramming of APA typifies several types of cancer,

and specific APA events are implicated in malignancies and autoimmune diseases[8,15,40]. Despite increased interest in APA, its roles in bone homeostasis and repair remain unknown. Our data highlight the dynamic nature of the callus 3′ UTR landscape among different healing phases and identify thousands of genes expressed in the healing callus as APA isoforms. Rapidly proliferating cells that gradually populate the callus during endochondral ossification shows a global trend of 3′ UTR shortening, consistent with the established association between cell proliferation and APA control[8,10]. Shortening of the 3′ UTR clearly correlated with increased expression of hundreds of genes with critical roles in bone repair, which emphasized the role of APA in post-transcriptional regulation of gene expression during endochondral ossification. Crucially, APA can occur also in introns (IPA). Although IPA has been reported to be prevalent in immune cells[41], we did not detect any trend toward IPA activation at d21 relative to d14, despite an obvious increase in immune cell recruitment and activity over this time period. However, we identified several IPA events at each timepoint, which are expected to stimulate mRNA degradation via nonsense mediated decay or result in expression of truncated proteins. The biological significance of these IPA events in fracture healing remains uncharacterized and will be interesting to address in future studies.

Regulation of APA is stimulus- and cell/tissue-specific. An example of the latter is Irak2 that undergoes significant 3′ UTR shortening as healing proceeds from d14 to d21. sAPA is the major Irak2 APA isoform in bone marrow, while lAPA is the major isoform in differentiated MC3T3 and ATDC5 cells. The major Irak2 APA isoform switches from lAPA at d14 to sAPA at d21 as the bone marrow population is re-established at d21. In myeloid cells, sAPA is the sole isoform in monocytes and the major isoform in macrophages and activated macrophages. Given that Irak2 mRNA exhibits a clear correlation between 3′ UTR shortening and increased steady-

state abundance, 3′ UTR shortening might be one of the factors that boost Irak2 expression post-transcriptionally in monocytes and macrophages, which in turn directs the expression of inflammatory cytokines[42].

The growing interest in APA as an important mechanism for diversification of the transcriptome spurred technical advances to quantify APA isoforms. Various experimental approaches have been developed that enabled deep sequencing and annotation of Poly(A) sites in human and mouse genomes, leading to the establishment of several PolyAsite databases that integrate results from different 3′ end sequencing approaches[8]. However, the volume of data generated so far using 3′ end sequencing, and the scope of cell/tissue types covered by these data, are very limited compared to generic RNA-seq data. Therefore, subsequent bioinformatic efforts have been made to identify and quantitate PAs usage from RNA-seq data. Algorithms have been developed to identify PAs as positions that exhibit a distinct drop in RNA-seq reads[15]. Different RNA-seq sets have been analyzed using these algorithms, including analysis of The Cancer Genome Atlas using DaPars, which led to identification of global 3′ UTR shortening in cancer cells[15,16]. These different algorithms perform the same task using different approaches, and direct comparison of their relative accuracy in identifying and quantitating APA isoforms is lacking. One limitation of these algorithms is their reliance on identifying variations in the RNA-seq coverage profile around tandem PAs, which might be caused by other factors beyond APA. Furthermore, RNA-seq read coverage is inherently non-uniform along transcripts, which increases the rate of false positive identification of de novo PAs. These limitations can be addressed by focusing the analysis on PAs that are already annotated; an approach that has been applied by different APA analysis tools, including InPAS[17,18] and APAlyzer[19] that we utilized in this work. InPAS and APAlyzer identified the same pattern of changes in 3′ UTR length in our

data during endochondral ossification and agreed on the major findings. In general, APAlyzer identified more APA events than InPAS did, but we focused our studies on APA events detected by both tools as a means to increase robustness. It is also noteworthy that limiting APA analysis to already annotated PAs limits the coverage of APA events and restricts detection of de novo PAs. In conclusion, we used different filters and approaches to increase accuracy of PA detection, even though this might have resulted in underestimating the number of callus APA events.

PA sequences have likely been selected during evolution to establish specific ratios of expression of APA isoforms in each cell/tissue. Consistent with this, during cell differentiation, genes with multiple APA isoforms undergo changes in isoform ratios to achieve patterns that maintain the required tissue-specific expression levels[39]. Our data indicate that the ratio of sAPA/lAPA for Col1a1 or Col1a2 in mineralizing osteoblasts was comparable to that in mineralizing chondrocytes, indicating the importance of establishing a particular sAPA/lAPA ratio to maintain the required expression level of Col I protein in mineralizing cells. Shortening of the 3′ UTR in Col1a1 and Col1a2 was primarily mediated by one or more of the SSR-embedded (AAUAAA) hexamers that follow the general roles of canonical PAs: they are embedded in an A/U-rich region located ≥ 21 nucleotides upstream of the identified proximal CPSs. Luciferase assays and analyses of the decay rate of endogenous transcripts indicate that 3′ UTR shortening substantially enhanced the stability of Col1a1 and Col1a2 mRNAs and induced gene expression via, at least in part, removal of miR-29a-3p MREs that are embedded in the aUTR. These similar mechanisms that regulate the expression of both Col1a1 and Col1a2 mRNAs point to an evolutionary path that culminated in parallel regulatory elements embedded in the transcripts of two different genes that encode subunits of the same protein.

Members of miR-29 family (miR-29), which have overlapping targets, stimulate the differentiation of MC3T3 cells via direct inhibition of the expression of the osteogenesis inhibitory factors HDAC4, TGFβ3, ACVR2A, CTNNBIP1, and DUSP2[35]. Consistent with the osteogenic roles of miR-29, our data indicate high expression of miR-29a-3p in newly formed osteoblasts and bone lining cells. Further, the expression of miR-29b has been reported to increase during MC3T3 differentiation[35]. miR-29 also inhibits the expression of Col1a1 to a variable extent in different tissue/cell types[33,34,36,43]. However, the repressive effect of miR-29 on Col1a1 expression is mild or undetectable in MC3T3 cells, especially during early phases of differentiation[35]. This raised a long-standing question of how Col1a1 evades the repressive effects of miR-29 during osteogenic differentiation[35]. In fact, it is counterintuitive that miR-29 is pro-osteogenic and, at the same time, inhibits the expression of Col I; an essential protein for differentiation and mineralization. Our data provide a likely answer to this long standing question, and propose APA-mediated 3′ UTR shortening as the mechanism that enables Col1a1 and Col1a2 genes to achieve high expression levels in the presence of high concentration of the pro-osteogenic miR-29a-3p. Importantly, pathological over-expression of Col I induces fibrosis, and a miR-29 mimic can act as anti-fibrotic in different tissues through repressing the expression of Col I and other fibrotic proteins[33,34,36,43]. More than one miR-29 mimic are now in clinical trials as investigative next-generation treatment for idiopathic pulmonary fibrosis[44]. However, the anti-fibrotic effects of miR-29 and its impact on Col I expression vary substantially among tissues[33,34,36,43], suggesting that tissue-specific APA-mediated shortening of Col1a1 and Col1a2 3′UTRs might contribute to blunting of the anti-fibrotic effects of miR-29 in certain tissues.

In summary, our data unravel novel roles of APA as a post-transcriptional regulatory mechanism that participates in shaping the transcriptome and regulating gene expression during bone regeneration.

Materials and Methods

Animals

All animal protocols were approved by the University Committee on Animal Resources (IACUC) at the Pennsylvania State University College of Medicine. Adult, male C57BL/6J mice at 14-week age were purchased from The Jackson Laboratory (Bar Harbor, ME, USA) and allowed to acclimate for 2 weeks. Mice were housed in ventilated cages and provided with ad libitum access to pelleted food and water. The mouse facility had a 12-hour light/dark cycle at 21.1°C to 22.8°C and 30% to 70% humidity.

Mid-diaphysis tibial fracture surgery

Mid-diaphyseal tibial fractures were induced in the right hindlimb and stabilized by intramedullary nail as described previously[23,24]. X-ray images were collected post-operatively and at harvest time to confirm proper alignment at the fracture site. Harvest time was determined based on published studies on the healing time course of this model[23,24].

Safranin-O/fast green and IF staining

Sagittal sections of 5 μm thickness spanning the centre of the fracture callus were obtained and stained with Hematoxylin/Safranin-O/Fast Green as described previously[23]. Imaging was performed using the OsteoMeasure system (OsteoMetrics Inc.). For IF staining, Col I and Col II were stained according to the 2-step staining protocol, while MMP13 and Col X were stained

according to the 3-step protocol as we detailed previously[23]. Mounting and nuclear staining was performed using ProLongTM Gold antifade reagent with DAPI (Invitrogen). Images were captured using Zeiss Axio Observer 7 upright wide-field microscope (Carl Zeiss Microscopy GmbH).

RNA purification, reverse transcriptase (RT)-qPCR, and ISH

Callus tissue was isolated and snap frozen in liquid nitrogen, and RNA was purified as we previously described45. RNA was purified from cell lines, primary monocytes, and macrophages using RNeasy Kit (Qiagen). cDNA was prepared from < 1 μg RNA using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) and qPCR-amplified using TaqMan Fast Advanced Master Mix and TaqMan Gene-Expression Assays (Thermo Fisher Scientific). For lAPA isoforms, qPCR was performed using PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) and primers specified in Table S14. ISH was performed using miRCURY LNA miRNA ISH kit (Qiagen) according to manufacturer's instructions.

Library preparation and RNA sequencing

RNA-seq was performed on three biological replicates (i.e., callus tissues) for each timepoint. cDNA libraries were prepared using the Illumina® Stranded mRNA Prep and Ligation kit (Illumina) as per manufacturer's instructions. Briefly, poly(A) RNA was purified from 200 ng total RNA using oligo (dT) beads. The purified fraction was then subjected to fragmentation, reverse transcription, end repair, 3′ adenylation, and adaptor ligation, followed by PCR amplification and SPRI bead purification (Beckman Coulter). The unique dual index sequences (IDT® for Illumina® RNA UD Indexes Set A, Ligation, Illumina) were incorporated in the adaptors for multiplexed

high-throughput sequencing. The final product was assessed for size distribution and concentration using BioAnalyzer High Sensitivity DNA Kit (Agilent Technologies). The libraries were pooled, diluted to 3 nM using 10 mM Tris-HCl, pH 8.5, and denatured using the Illumina protocol. The denatured libraries were loaded onto an S1 flow cell (Illumina NovaSeq 6000) and run for 2x53 cycles according to manufacturer's instructions.

Analysis of RNA-seq data and APA

De-multiplexed and adapter-trimmed sequencing reads were generated using Illumina bcl2fastq (released version 2.20.0.422) allowing no mismatches in the index read. BBDuk (ourceforge.net/projects/bbmap/) was used to trim/filter low quality sequences using "qtrim=lr trimq=10 maq=10" option. Filtered reads were aligned to the mouse reference genome (GRCm38) using HISAT2 (version 2.1.0)[46] applying --no-mixed and --no-discordant options. Read counts were calculated using HTSeq[47] by supplementing Ensembl gene annotation (GRCm38.78). EdgeR[48] was used to fit the read counts to the negative binomial model along with generalized linear model (GLM), and differentially expressed genes were determined using DESeq2[26]. GO and functional analyses were performed by Ingenuity Pathway Analysis[37], DAVID[29], iDEP[28], and/or GAGE[27]. Heatmaps and volcano plots were prepared using iDEP[28].

APA was analyzed using two different software: InPAS v3.14 and APAlyzer v4.1. Default settings were used for each program. For InPAS, PDUI for each gene was calculated as the ratio of lAPA RPM to the total RPM of the two main APA isoforms, and PDUI log2FC was calculated as the difference in $\log_2$PDUI between d21 and d14. Significant APA events were considered those with adjusted *P* value < 0.05 and > 20% change in the ratio of the two main APA isoforms. For APAlyzer, RED value was calculated as difference in log2 (RPM ratio) of

the two APA isoforms (i.e. log2(aUTR read number/cUTR read number)) between d21 and d14. The aUTR length was calculated as the difference between the two defined CPSs in the 3′UTR. For comparison of gene expression of different APA gene sets (**Fig. 7a, b**; **Fig. S10a, b**), genes were classified to significantly upregulated, no change, or significantly downregulated sets based on counting the reads mapped to CDS only. Differential expression of CDS reads were analyzed using the APAlyzer package. This approach was used to avoid confounding the data by including reads from the 3′ UTR region that is different among APA isoforms. However, using total RNA reads (generated by the DESeq2) instead of CDS reads generated correlations comparable to those shown in **Fig. 7a, b** and **Fig. S10 a, b** (data not shown). CDF curves were created using R package ggplot2.

Library preparation and miRNA sequencing

Small RNA-sequencing libraries were prepared from 250 ng total RNA using the QIAseq miRNA Library Kit (QIAGEN) as per the manufacturer's instructions. This system offers a built-in Unique Molecular Identifier (UMI) application that eliminates possible PCR duplicates in sequencing datasets and, therefore, facilitates unbiased gene expression profiling. Unique barcode sequences were incorporated in the adaptors for multiplexed high-throughput sequencing. Size distribution and concentration of the final product were assessed using BioAnalyzer High Sensitivity DNA Kit (Agilent Technologies). Pooled libraries were diluted to 2 nM in EB buffer (Qiagen), denatured using the Illumina protocol, diluted to 10 pM using pre-chilled hybridization buffer, loaded onto a TruSeq v2 Rapid flow cell (Illumina HiSeq 2500), and run for 75 cycles using a single-read recipe according to the manufacturer's instructions. De-multiplexed sequencing reads were generated using Illumina bcl2fastq (released version

2.20.0.422, Illumina) allowing no mismatches in the index read. Primary read mapping and UMI analysis were conducted using the GeneGlobe Data Analysis Center (QIAGEN).

Cell culture and transfection

MC3T3-E1 cells were obtained from ATCC, cultured in α-Modified Eagle Medium (α-MEM) supplemented with 10% fetal bovine serum (FBS), and differentiated by adding 50 μg/ml ascorbic acid, 10 mM β-glycerophoshate, and 0.1 μM dexamethasone. ATDC5 cells were obtained from the European Collection of Authenticated Cell Cultures (ECACC) and cultured in a 1:1 mixture of Dulbecco's Modified Eagle Medium (DMEM) and Ham's F12 supplemented with 5% FBS, 1% insulin, transferrin and selenium (Thermo Fisher). ATDC5 cells were differentiated in the presence of 10 mM β-Glycerophosphate (Sigma) and 50 μg/ml 2-phospho-L-ascorbic acid (Sigma). MC3T3-E1 cells were transfected with pmirGLO Dual-Luciferase (Promega) alone or in combination with control or miR-29a-3p mimic (mirVana mimics; Thermo Fisher Scientific) using Lipofectamine 3000 (Thermo Fisher Scientific), and were transfected with control or miR-29a-3p mimic alone using Lipofectamine RNAiMAX (Thermo Fisher Scientific). Primary monocytes were isolated from bone marrow using the EasySep Mouse Monocytes Isolation Kit (StemCell Technologies). Monocytes were differentiated to macrophages in DMEM supplemented with 10% FBS and 15 % L929 cell supernatant.

3′ RACE-PCR

cDNA was prepared from ≤ 1 μg RNA using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) and the Universal Adapter primer (**Table S14**). cDNA was PCR-amplified using a gene-specific forward primer (**Table S14**), the Universal Amplification Reverse primer

(**Table S14**), and the proofreading, thermostable Platinum SuperFi II DNA Polymerase (Thermo Fisher Scientific). The resultant blunt-end PCR product was cloned into the Zero Blunt® TOPO® vector using Zero Blunt® TOPO® PCR Cloning Kit (Thermo Fisher Scientific). Clones were then sequenced using Sanger DNA Sequencing (GENEWIZ).

Luciferase assay and construction of pmirGLO Dual Luciferase reporters

Different isoforms and region of Col1a1 and Col1a2 3′ UTRs were cloned into the pmirGLO Dual Luciferase vector (Promega) downstream to the FLuc CDS as follows. The pmirGLO Dual Luciferase vector was linearized using PmeI restriction enzyme (NEB). Each 3′ UTR region/isoform was amplified from cDNA using the Platinum SuperFi II DNA Polymerase (Thermo Fisher Scientific) and the In-Fusion primer pairs specified in **Table S14**. The resultant PCR product was then cloned into the linearized pmirGLO Dual Luciferase vector using the In-Fusion HD Cloning Kit (Takara). Luciferase assays were performed using the Dual-Glo® Luciferase Assay System (Promega) as per manufacturer's instructions.

Labeling of nascent RNA and analysis of RNA decay

We used the Click-iT Nascent RNA Capture Kit (Thermo Fisher Scientific) as per manufacturer's instructions. Briefly, cells were incubated with EU-containing medium for 2 hours (pulse step). Cells were then divided into two groups: the first group was collected to determine the total level of EU-labelled RNA, while the second group was switched to EU-free medium and incubated for 24 hours (chase step). In each group, total RNA was isolated from collected cells and subjected to a copper-catalyzed click reaction with an azide-modified biotin. EU RNA was then captured using streptavidin magnetic beads and used for cDNA preparation as

described above. qPCR quantitation of total mRNA and lAPA isoform was performed as described above. The % remaining EU mRNA (**Fig. 4g**) was calculated by normalizing the level of EU mRNA measured after the 24-hour chase (measured in the second group) to the total level of EU mRNA synthesized during the 2-hour pulse (measured in the first group).

Statistical Analysis

The K–S test was used to compare distributions between gene sets. For qPCR and luciferase assays, the unpaired Student's t-test was used to determine statistical significance between two groups, whereas ANOVA (followed by Tukey's post hoc test) was used to determine statistical significance among three or more groups. Statistical analyses were performed using the GraphPad Prism software. The following symbols were used to indicate significance: (ns) $P \geq 0.05$; (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; (****) $P < 0.0001$.

Data Availability

Raw RNA-seq data have been deposited in NCBI's Gene Expression Omnibus[49] and are accessible through GEO Series accession number GSE205053.

Acknowledgments

This work was supported by National Institutes of Health (NIH) R01 DK121327 to R.A.E.

Conflict of interests

The authors declare no conflict of interest.

Contributions: Study design: RAE. Data collection: DKK, IN, ML, VK, and RAE. Data analysis and interpretation: YI, CCN, FK, and RAE. Writing the manuscript: CCN, FK, and RAE.

References

1 Claes, L., Recknagel, S. & Ignatius, A. Fracture healing under healthy and inflammatory conditions. *Nat Rev Rheumatol* **8**, 133-143, doi:10.1038/nrrheum.2012.1 (2012).

2 Bahney, C. S. *et al.* Cellular biology of fracture healing. *J Orthop Res* **37**, 35-50, doi:10.1002/jor.24170 (2019).

3 Einhorn, T. A. & Gerstenfeld, L. C. Fracture healing: mechanisms and interventions. *Nat Rev Rheumatol* **11**, 45-54, doi:10.1038/nrrheum.2014.164 (2015).

4 Richard, P. & Manley, J. L. Transcription termination by nuclear RNA polymerases. *Genes Dev* **23**, 1247-1269, doi:10.1101/gad.1792809 (2009).

5 Shi, Y. & Manley, J. L. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* **29**, 889-897, doi:10.1101/gad.261974.115 (2015).

6 Proudfoot, N. J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926, doi:10.1126/science.aad9926 (2016).

7 Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**, 18-30, doi:10.1038/nrm.2016.116 (2017).

8 Gruber, A. J. & Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nat Rev Genet* **20**, 599-614, doi:10.1038/s41576-019-0145-z (2019).

9 Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**, 1173-1183, doi:10.1101/gr.132563.111 (2012).

10 Mayr, C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol* **26**, 227-237, doi:10.1016/j.tcb.2015.10.012 (2016).

11 Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20-51, doi:10.1016/j.cell.2018.03.006 (2018).

12 Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509-524, doi:10.1038/nrm3838 (2014).

13 Tian, B., Pan, Z. & Lee, J. Y. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res* **17**, 156-165, doi:10.1101/gr.5532707 (2007).

14 Singh, I. *et al.* Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat Commun* **9**, 1716, doi:10.1038/s41467-018-04112-z (2018).

15 Xia, Z. *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun* **5**, 5274, doi:10.1038/ncomms6274 (2014).

16 Cancer Genome Atlas Research, N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-1120, doi:10.1038/ng.2764 (2013).

17 Sheppard, S., Lawson, N. D. & Zhu, L. J. Accurate identification of polyadenylation sites from 3' end deep sequencing using a naive Bayes classifier. *Bioinformatics* **29**, 2564-2571, doi:10.1093/bioinformatics/btt446 (2013).
18 Zhang, Y. *et al.* Alternative polyadenylation: methods, mechanism, function, and role in cancer. *J Exp Clin Cancer Res* **40**, 51, doi:10.1186/s13046-021-01852-7 (2021).
19 Wang, R. & Tian, B. APAlyzer: a bioinformatics package for analysis of alternative polyadenylation isoforms. *Bioinformatics* **36**, 3907-3909, doi:10.1093/bioinformatics/btaa266 (2020).
20 Wang, R., Nambiar, R., Zheng, D. & Tian, B. PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res* **46**, D315-D319, doi:10.1093/nar/gkx1000 (2018).
21 Wang, R., Zheng, D., Yehia, G. & Tian, B. A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res* **28**, 1427-1441, doi:10.1101/gr.237826.118 (2018).
22 Nourse, J., Spada, S. & Danckwardt, S. Emerging Roles of RNA 3'-end Cleavage and Polyadenylation in Pathogenesis, Diagnosis and Therapy of Human Disorders. *Biomolecules* **10**, doi:10.3390/biom10060915 (2020).
23 Khajuria, D. K. *et al.* Aberrant structure of fibrillar collagen and elevated levels of advanced glycation end products typify delayed fracture healing in the diet-induced obesity mouse model. *Bone* **137**, 115436, doi:10.1016/j.bone.2020.115436 (2020).
24 Brown, M. L. *et al.* Delayed fracture healing and increased callus adiposity in a C57BL/6J murine model of obesity-associated type 2 diabetes mellitus. *PLoS One* **9**, e99656, doi:10.1371/journal.pone.0099656 (2014).
25 Kaback, L. A. *et al.* Osterix/Sp7 regulates mesenchymal stem cell mediated endochondral ossification. *J Cell Physiol* **214**, 173-182, doi:10.1002/jcp.21176 (2008).
26 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
27 Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161, doi:10.1186/1471-2105-10-161 (2009).
28 Ge, S. X., Son, E. W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534, doi:10.1186/s12859-018-2486-6 (2018).
29 Jiao, X. *et al.* DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **28**, 1805-1806, doi:10.1093/bioinformatics/bts251 (2012).
30 San Antonio, J. D., Jacenko, O., Fertala, A. & Orgel, J. Collagen Structure-Function Mapping Informs Applications for Regenerative Medicine. *Bioengineering (Basel)* **8**, doi:10.3390/bioengineering8010003 (2020).
31 McGeary, S. E. *et al.* The biochemical basis of microRNA targeting efficacy. *Science* **366**, doi:10.1126/science.aav1741 (2019).
32 Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* **48**, D127-D131, doi:10.1093/nar/gkz757 (2020).

33 van Rooij, E. *et al.* Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc Natl Acad Sci U S A* **105**, 13027-13032, doi:10.1073/pnas.0805038105 (2008).
34 Dey, S. *et al.* Loss of miR-29a/b1 promotes inflammation and fibrosis in acute pancreatitis. *JCI Insight* **6**, doi:10.1172/jci.insight.149539 (2021).
35 Li, Z. *et al.* Biological functions of miR-29b contribute to positive regulation of osteoblast differentiation. *J Biol Chem* **284**, 15676-15684, doi:10.1074/jbc.M809787200 (2009).
36 Ulrich, V. *et al.* Chronic miR-29 antagonism promotes favorable plaque remodeling in atherosclerotic mice. *EMBO Mol Med* **8**, 643-653, doi:10.15252/emmm.201506031 (2016).
37 Kramer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530, doi:10.1093/bioinformatics/btt703 (2014).
38 Aghajanian, P. & Mohan, S. The art of building bone: emerging role of chondrocyte-to-osteoblast transdifferentiation in endochondral ossification. *Bone Res* **6**, 19, doi:10.1038/s41413-018-0021-z (2018).
39 Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S. & Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev* **27**, 2380-2396, doi:10.1101/gad.229328.113 (2013).
40 Gruber, A. J. *et al.* Discovery of physiological and cancer-related regulators of 3' UTR processing with KAPAC. *Genome Biol* **19**, 44, doi:10.1186/s13059-018-1415-3 (2018).
41 Gruber, A. J., Gypas, F., Riba, A., Schmidt, R. & Zavolan, M. Terminal exon characterization with TECtool reveals an abundance of cell-specific isoforms. *Nat Methods* **15**, 832-836, doi:10.1038/s41592-018-0114-z (2018).
42 Zhou, H. *et al.* IRAK2 directs stimulus-dependent nuclear export of inflammatory mRNAs. *Elife* **6**, doi:10.7554/eLife.29630 (2017).
43 Cushing, L. *et al.* miR-29 is a major regulator of genes associated with pulmonary fibrosis. *Am J Respir Cell Mol Biol* **45**, 287-294, doi:10.1165/rcmb.2010-0323OC (2011).
44 Hanna, J., Hossain, G. S. & Kocerha, J. The Potential for microRNA Therapeutics and Clinical Research. *Front Genet* **10**, 478, doi:10.3389/fgene.2019.00478 (2019).
45 Le Bleu, H. K. *et al.* Extraction of high-quality RNA from human articular cartilage. *Anal Biochem* **518**, 134-138, doi:10.1016/j.ab.2016.11.018 (2017).
46 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
47 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
48 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
49 Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210, doi:10.1093/nar/30.1.207 (2002).

50 Leclerc, G. J., Leclerc, G. M. & Barredo, J. C. Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines. *Cancer Cell Int* **2**, 1, doi:10.1186/1475-2867-2-1 (2002).

Figure legends

Figure 1. Cells with higher proliferation capacity populate the callus as the cartilaginous callus that bridges the fracture gap is replaced by woven bone. (**a-d**) Staining of d14 callus. (**a**) Safranin O/fast green staining. The cartilaginous extracellular matrix of the soft callus is stained reddish orange. (**b**, **c**) IF staining of Col X (purple) (b) and MMP13 (green) (c). Both proteins are secreted by hypertrophic chondrocytes. (**d**) Co-staining of Col I (red) and Col II (green). Areas where both proteins are expressed (yellow) surround mineralizing chondrocytes. (**e**) As in (**a**) except d21 callus was stained. Absence of reddish orange stain indicates complete resorption of the cartilaginous callus. (**f**) As in (d) except d21 callus was stained. The trabecular structure of woven bone bridging the fracture gap is obvious. In all IF images, DAPI stains nuclei (blue) and the scale bar = 200 μm. All images are representative of n = 5. (**g**) RT-qPCR quantitation of the indicated transcripts in RNA purified from d14 and d21 callus tissues. Both Col1a2 and Aggrecan (Acan) are chondrocyte markers, while Col10a1 is a hypertrophic-chondrocyte marker. The expression level of each transcript was normalized to that of β-actin mRNA, and the normalized level at d14 is defined as 100. N = 3. Bar graphs represent average ± SEM. (***) P < 0.001using unpaired Student's t-test. (**h**) A volcano plot of RNA-seq data displaying fold change in gene expression values at d21 relative to d14 (**Table S1**). Significantly differentially expressed genes (FDR < 0.05) are highlighted in red and blue for upregulated and downregulated genes, respectively. RNA-seq data were generated from three biological replicates of each timepoint. (**i**) GO analysis of significantly differentially expressed genes shown in (h). Representative examples of upregulated (red) or downregulated (blue) enriched functional categories are shown, and the adjusted *P* value (P_{adj}) for each pathway is given (see also **Fig. S2**, **Fig. S3**, **Table S2**, and **Table S3** for comprehensive lists of significantly enriched functional categories).

Figure 2. APA is widely utilized by genes expressed at the callus of day 14 post-fracture. (**a**) Schematic of the short APA (sAPA) and long APA (lAPA) isoforms of the same gene. Utilization of the proximal polyadenylation site (pPA) confines the size of the 3′ UTR to the constitutive UTR (cUTR), generating the sAPA isoform. On the other hand, utilization of the distal PA (dPA) results in the inclusion of the alternative UTR (aUTR) within the 3′ UTR, generating the lAPA isoform. InPAS-based calculation of PDUI values is provided. (**b**) Pie chart presentation of the distribution of the indicated 7 PDUI groups in the transcriptome of d14 callus. (**c**) Presentation of the overlap between InPAS- and APAlyzer-identified APA events in the transcriptome of d14 callus. Events identified by both programs are shown in red; events identified by InPAS only are shown in blue. (**d**) GO analysis of APA events at d14 callus. Significantly enriched functional categories are shown, and the enrichment false discovery rate (FDR) for each pathway is given (comprehensive lists of significantly enriched functional categories are provided in **Fig. S4** and **Table S6**).

Figure 3. Long and short APA isoforms of Col1a1 and Col1a2 mRNAs are expressed in the fracture callus as well as osteogenic and chondrogenic cell lines. (**a**) Schematic of the lAPA and sAPA isoforms of Col1a1 (left) and Col1a2 (right). The identified size of the 3′ UTR of each isoform is shown. (**b**) Representative gel image of Col1a1 3′ RACE-PCR products showing the 2 APA isoforms. 3′ RACE was performed using RNA purified from d14 callus (3′ RACE procedure is depicted in **Fig. S5a, b**). The 3′ end of each APA isoform was mapped by excising and sequencing the corresponding PCR band (**Fig. S6a**). (**c**) As in (b) except 3′ RACE-PCR was performed on Col1a2 mRNA. Sequencing of the PCR products confirmed the InPAS-identified lAPA and sAPA isoforms. (**d, e**) As in (b) and (c), respectively, except RNA was purified from the indicated cell lines. ATDC5 cells were differentiated under conditions that promoted

hypertrophic differentiation and mineralization (**Fig. S7**). (**f**) RT-qPCR quantitation of the osteoblast and mineralization marker bone gamma carboxyglutamate protein (Bglap) in undifferentiated and differentiated MC3T3 cells. The expression level of Bglap mRNA was normalized to that of β-actin mRNA, and the normalized level in undifferentiated cells is defined as 100. (**g**) Alizarin red staining of ATDC5 before (-) and after (+) a 3-week differentiation course. Positive staining of differentiated cells indicates chondrocyte mineralization (**Fig. S7**). (**h**) as in (f) except RNA was purified from ATDC5 cells shown in (g) and quantitation was performed on the specified transcripts. Results demonstrate high expression of chondrocyte-hypertrophy and mineralization markers in differentiated ATDC5 cells.
Gel images are representative of three independent replicates. Bar graphs represent the average of three independent replicates ± SEM. (**) $P < 0.01$; (****) $P < 0.0001$ using unpaired Student's t-test.

Figure 4. Shortening of the 3′ UTR promotes the expression of Col1a1 and Col1a2. (**a**) Schematic of FLuc reporters that contain either the 3′ UTR of the lAPA isoform (i.e., full-length 3′ UTR) of Col1a1or Col1a2 (FLuc-lAPA 3′ UTR) or the deletion mutant where the SSR was deleted (FLuc-ΔpPA 3′ UTR; X denotes the position of the deleted SSR). CDS: coding region. (**b**) 3′ RACE-PCR of FLuc mRNA performed on RNA isolated from MC-3T3 cells transfected with Col1a1 3′ UTR reporters shown in (a). PCR was performed using an FLuc-specific primer. Results indicate that deletion of the SSR in FLuc-ΔpPA 3′ UTR reporter abrogated 3′ UTR shortening, resulting in expression of FLuc mRNA as a single, long APA isoform. (**c**) As in (b) except Col1a2 3′ UTR reporters were used. (**d**) Schematic of FLuc reporters that contain different isoforms or regions of Col1a1 or Col1a2 3′ UTR. (**e**) Bar graph presentation of the

results of dual luciferase assays performed on MC3T3 cells transfected with FLuc-Col1a1 3′ UTR reporters shown in (**d**). FLuc activity was normalized to Renilla luciferase (RLuc) activity, and the FLuc/RLuc ratio in cells transfected with the "empty" FLuc (FLuc-) reporter is defined as 100. (**f**) As in (e) except FLuc-Col1a2 3′ UTR reporters were used. N = 4-5 independent transfections, each measured in triplicate. (**g**) Bar graph presentation of the results of pulse-chase experiments used to measure RNA decay. MC3T3 cells were incubated with 5-ethynyl uridine (EU), which is an analog of uridine, for 2 hours to label nascent RNA "pulse step". Cells were then either collected to determine the total level of EU-labeled mRNA (EU mRNA) or grown for 24 hours in EU-free medium "chase step". At each timepoint, EU RNA was purified (see Materials and Methods) and the level of each of the specified EU mRNA was determined using RT-qPCR. Primers that bind to the aUTR were used to specifically quantitate the lAPA isoform of Col1a1 or Col1a2 (**Table S14**). The level of EU mRNA measured after the pulse step is defined as 100%, and the graph shows the % remaining after the 24-hour chase. Results demonstrate faster loss of Col1a1 and Col1a2 lAPA isoforms as compared to total mRNA. β-actin mRNA was used as a positive control for EU mRNA decay, and consistent with its reported half-life (~6.6 hours)[50], < 10% remained after the 24-hour chase.

Gel images are representative of three independent replicates. Bar graphs represent the average ± SEM. (ns) $P > 0.05$; (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; (****) $P < 0.0001$ using one-way analysis of variance (ANOVA) followed by Tukey's post hoc test for (e) and (f), and unpaired Student's t-test for (g)

Figure 5. miR-29a-3p mediates the degradation of lAPA, but not sAPA, isoforms of Col1a1 and Col1a2 mRNAs. (**a**) List of miRNAs that have conserved MREs in the 3′ UTR of mouse

Col1a1 mRNA (identified by both TargetScan and miRDB) (**Tables S8, S9**). miR-29a-3p (highlighted in red) exhibited the highest expression among the 4 shown miRNAs in the callus of d14 post-fracture (**Table S10**). (**b**) ISH using either a control (left) or miR-29a-3p (right) probe. Results show intense staining (blue) of miR-29a-3p in the woven bone area of d14 callus. (**c**) Schematic of the lAPA isoform of Col1a1 mRNA showing the putative miR-29a-3p MREs in the aUTR region. (**d**) As in (c) except Col1a2 mRNA is shown. (**e**) Bar graph presentation of the results of dual luciferase assays performed on MC3T3 cells transfected with the indicated Col1a1 FLuc reporters along with either a control (blue) or miR-29a-3p (red) mimic. FLuc activity was normalized to RLuc activity, and the FLuc/RLuc ratio in cells transfected with control mimic is defined as 100. N = 4-5 independent transfections, each measured in triplicate. (**f**) as in (e) except cells were transfected with Col1a2 FLuc reporters. In the FLuc-ΔMRE reporter, we replaced the binding region of miR-29a-3p seed sequence by a random sequence (denoted by "X"). (**g**) RT-qPCR quantitation of either total mRNA or lAPA isoform of Col1a1 in MC3T3 cells transfected with control (blue) or miR-29a-3p (red) mimic. Primers that bind within the aUTR region were used to specifically amplify and quantitate the lAPA isoform (**Table S14**). Expression level was normalized to that of β-actin mRNA, and the normalized level in control mimic-transfected cells is defined as 100. N = 5. (**h**) As in (g) except total mRNA and lAPA isoform of Col1a2 were quantitated.

Bar graphs represent the average ± SEM. (ns) $P > 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; (****) $P < 0.0001$. Significance of difference was calculated using unpaired Student's t-test in (e) and (f), and two-way ANOVA in (g) and (h).

Figure 6. Prevalent 3′ UTR shortening accompanies endochondral ossification. (**a**) General schematic presentation of APA isoform showing representative proximal and distal PAs (pPA and dPA, respectively). PDUI Log_2FC and RED values are calculated by InPAS and APAlyzer, respectively, as indicated. (**b**) Average and median PDUI log_2FC (left) and RED (right) values calculate by InPAS and APAlyzer, respectively, for all analyzed genes (**Tables S4, S5**). The negative average and median values indicate global trend toward 3′ UTR shortening at d21 relative to d14. (**c**) A Scatter plot showing changes in 3′ UTR length at d21 as compared to d14. Each dot represents a gene, and significantly lengthened or shortened 3′ UTRs are shown in red and blue, respectively. Significance is defined as *Padj* < 0.05 and > 20% change in the expression ratio of the 2 APA isoforms. (**d**) Correlation between InPAS-generated PDUI log_2FC and APAlyzer-generated RED values for genes that exhibited significant 3′ UTR shortening at d21 relative to d14 (shown in blue in "c"). (**e**) (Top): genes that exhibited significant 3′ UTR shortening were categorized into 5 similarly sized bins based on their 3′ UTR length. (bottom): Scatter plot showing the relationship between 3′ UTR length and the extent of 3′ UTR shortening (presented as PDUI log_2FC). Each dot represents a gene. Data are presented as average ± SEM. (**f**) As in (e) except genes were categorized based on their aUTR length. (*) $P < 0.05$; (***) $P < 0.001$, (****) $P < 0.0001$ using one-way ANOVA followed by Tukey's post hoc test.

Figure 7. 3′ UTR shortening is associated with increased gene expression. (**a**) Cumulative distribution function (CDF) curves of PDUI log_2FC values (generated by InPAS) for genes whose expression was significantly upregulated (red line), significantly downregulated (blue line), or unchanged (black line) at d21 relative to d14. Significant change in gene expression is defined as *Padj* < 0.05. All InPAS-identified APA events are included in the analysis (See **Fig.**

S10a for APA events identified by APAlyzer). *P*-values (K-S test) for significance in difference between red or blue and black genes are indicated. (**b**) CDF curves comparing gene expression of genes identified by InPAS to show significant 3′ UTR shortening (blue line) and all other genes (black line). *P*-values (K-S test) are based on comparing blue and black genes (See **Fig. S10b** for APAlyzer data). (**c**) Schematic of the lAPA and sAPA isoforms of Irak2. (**d**) RT-qPCR quantitation of Irak2 expression in d14 and d21 callus. The expression level of Irak2 mRNA was normalized to that of β-actin mRNA, and the normalized level in d14 callus is defined as 100. (**e-h**) Representative gel image of Irak2 3′ RACE-PCR products in d14 and d21 callus (e), differentiated MC3T3 and ATDC5 cells (f), Total, $CD45^+$, and $CD45^-$ bone marrow cells (g), or monocytes, macrophages (MΦ), and LPS-stimulated MΦ (h). (**i**) as in (d) except the MΦ marker Adgre1 was quantitated. Results indicate successful *in-vitro* differentiation of monocytes to MΦ. (**J**) As in (d) except Il1b mRNA was quantitated. Results indicate upregulated expression of the pro-inflammatory Il1b gene in response to LPS stimulation.

Bar graphs represent the average ± SEM. (**) $P < 0.01$; (***) $P < 0.001$; (****) $P < 0.0001$ using unpaired Student's t-test.

Figures

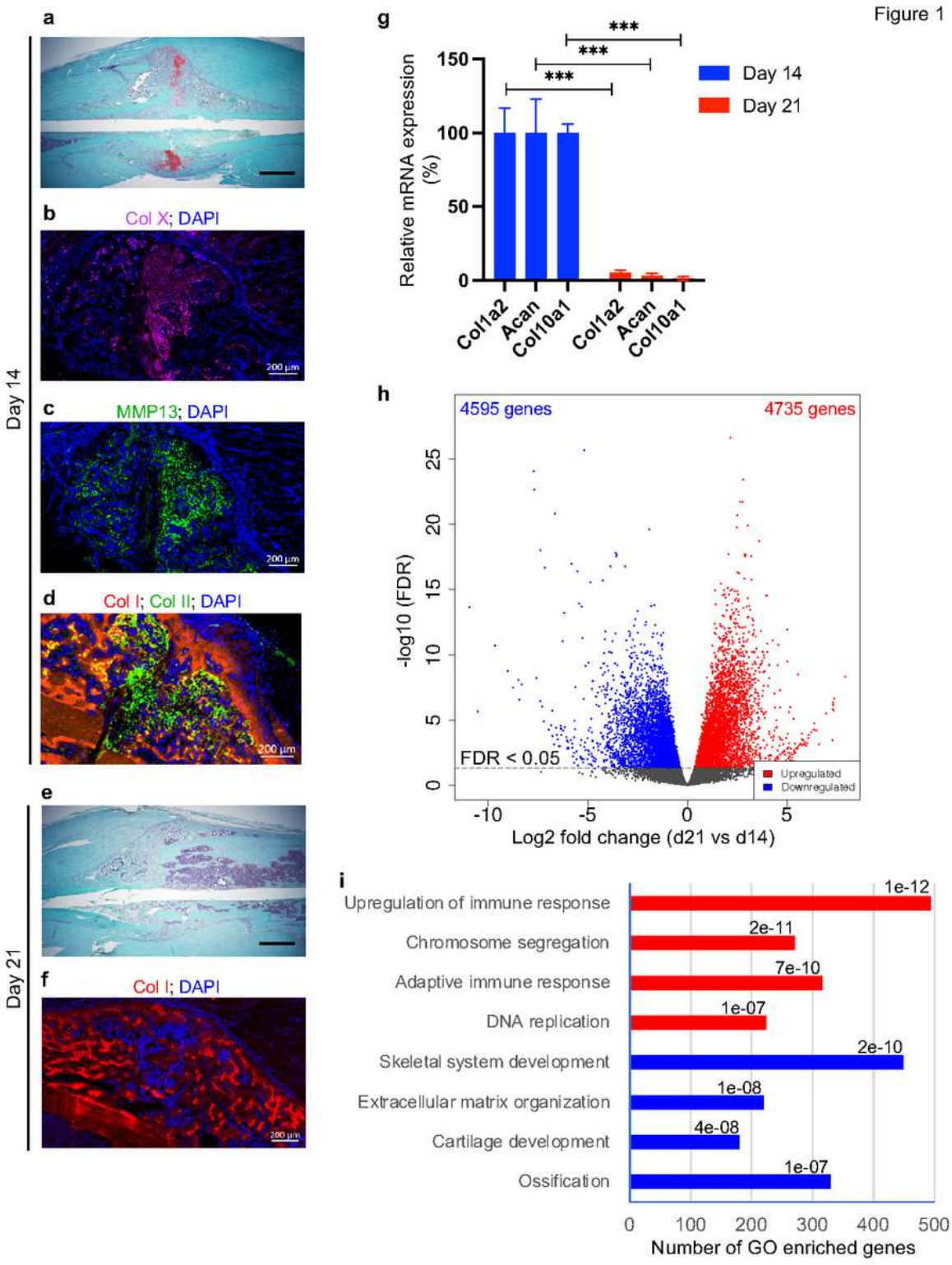

Figure 1

Cells with higher proliferation capacity populate the callus as the cartilaginous callus

that bridges the fracture gap is replaced by woven bone. (a-d) Staining of d14 callus. (a) Safranin O/fast green staining. The cartilaginous extracellular matrix of the soft callus is stained reddish orange. (b, c) IF staining of Col X (purple) (b) and MMP13 (green) (c). Both proteins are secreted by hypertrophic chondrocytes. (d) Co-staining of Col I (red) and Col II (green). Areas where both proteins are expressed (yellow) surround mineralizing chondrocytes. (e) As in (a) except d21 callus was stained. Absence of reddish orange stain indicates complete resorption of the cartilaginous callus. (f) As in (d) except d21 callus was stained. The trabecular structure of woven bone bridging the fracture gap is obvious. In all IF images, DAPI stains nuclei (blue) and the scale bar = 200 μm. All images are representative of n = 5. (g) RT-qPCR quantitation of the indicated transcripts in RNA purified from d14 and d21 callus tissues. Both Col1a2 and Aggrecan (Acan) are chondrocyte markers, while Col10a1 is a hypertrophic-chondrocyte marker. The expression level of each transcript was normalized to that of b-actin mRNA, and the

normalized level at d14 is defined as 100. N = 3. Bar graphs represent average ± SEM. (***) P < 0.001using unpaired Student's t-test. (h) A volcano plot of RNA-seq data displaying fold change in gene expression values at d21 relative to d14 (Table S1). Significantly differentially expressed genes (FDR < 0.05) are highlighted in red and blue for upregulated and downregulated genes, respectively. RNA-seq data were generated from three biological replicates of each timepoint. (i) GO analysis of significantly differentially expressed genes shown in (h). Representative examples of upregulated (red) or downregulated (blue) enriched functional categories are shown, and the adjusted P value (Padj) for each pathway is given (see also Fig. S2, Fig. S3, Table S2, and Table S3 for comprehensive lists of significantly enriched functional categories).

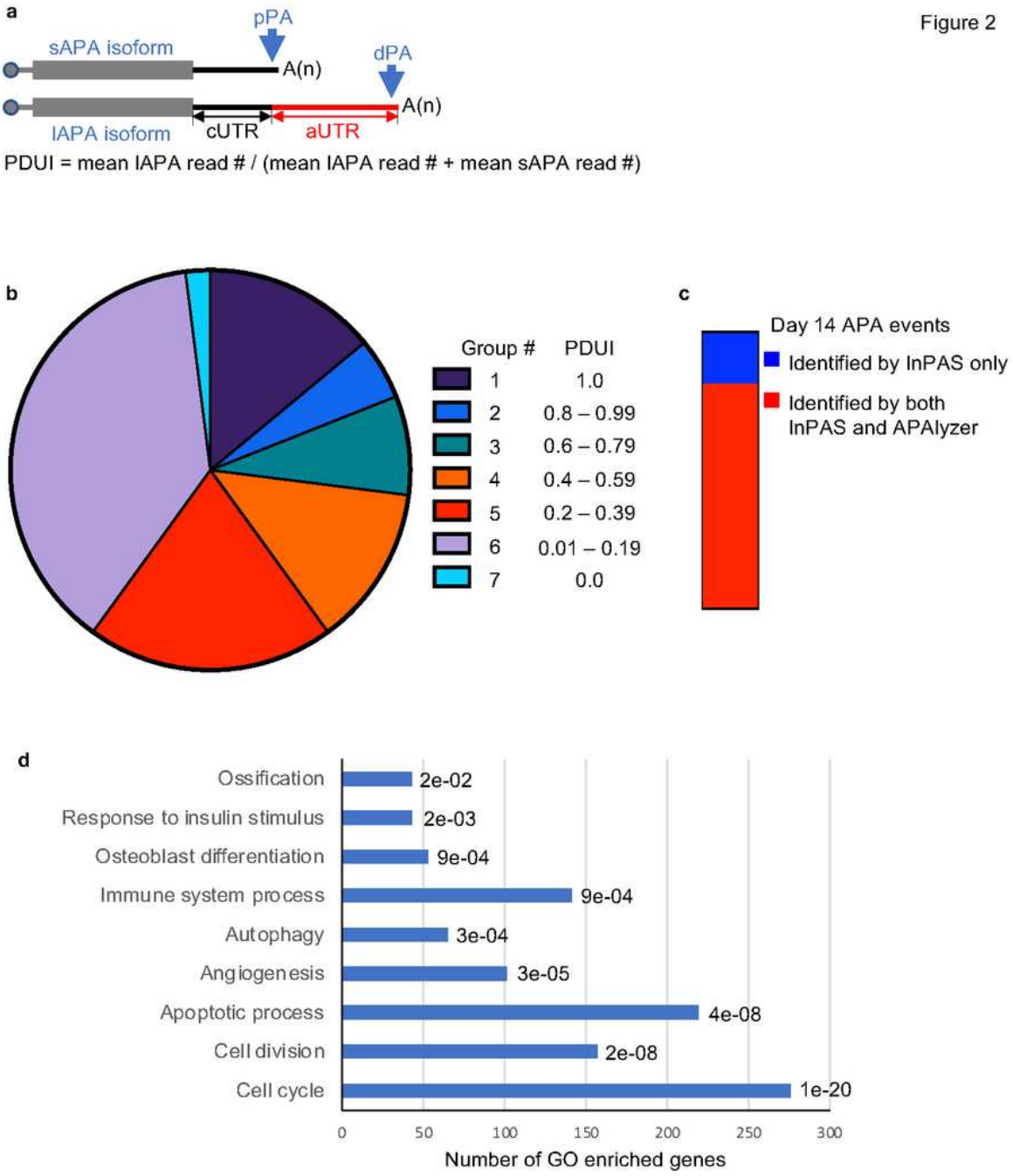

Figure 2

Figure 2

APA is widely utilized by genes expressed at the callus of day 14 post-fracture.

(a) Schematic of the short APA (sAPA) and long APA (lAPA) isoforms of the same gene.

Utilization of the proximal polyadenylation site (pPA) confines the size of the 3'UTR to the constitutive UTR (cUTR), generating the sAPA isoform. On the other hand, utilization of the distal PA (dPA) results in the inclusion of the alternative UTR (aUTR) within the 3'UTR, generating the lAPA isoform. InPAS-based calculation of PDUI values is provided. (b) Pie chart presentation of the distribution of the indicated 7 PDUI groups in the transcriptome of d14 callus. (c) Presentation of the overlap between InPAS- and APAlyzer-identified APA events in the transcriptome of d14 callus. Events identified by both programs are shown in red; events identified by InPAS only are shown in blue. (d) GO analysis of APA events at d14 callus. Significantly enriched functional categories are shown, and the enrichment false discovery rate (FDR) for each pathway is given (comprehensive lists of significantly enriched functional categories are provided in Fig. S4 and Table S6).

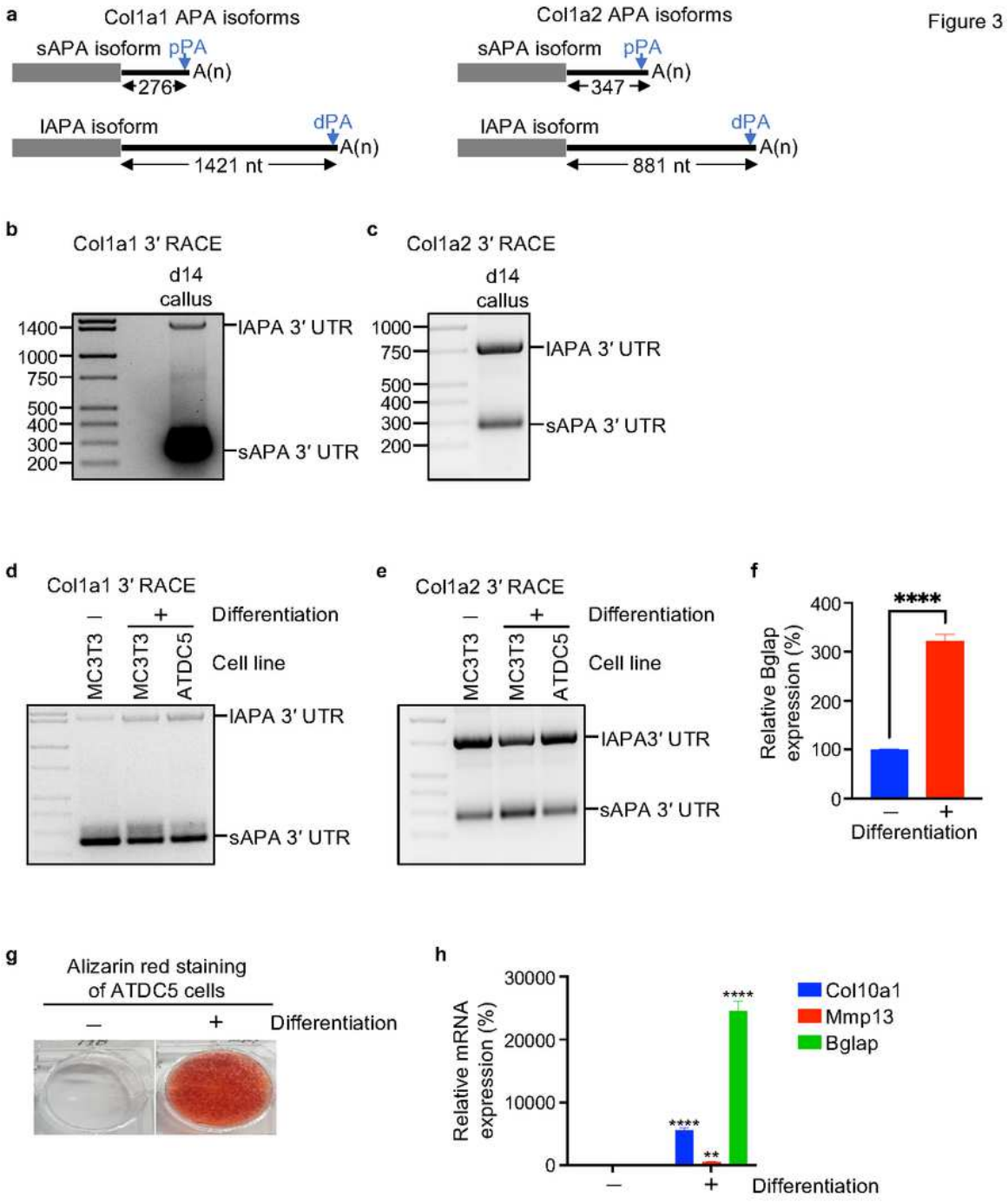

Figure 3

Figure 3

Long and short APA isoforms of Col1a1 and Col1a2 mRNAs are expressed in the

fracture callus as well as osteogenic and chondrogenic cell lines. (a) Schematic of the lAPA and

sAPA isoforms of Col1a1 (left) and Col1a2 (right). The identified size of the 3'UTR of each isoform is shown. (b) Representative gel image of Col1a1 3'RACE-PCR products showing the 2 APA isoforms. 3'RACE was performed using RNA purified from d14 callus (3'RACE procedure is depicted in Fig. S5a, b). The 3'end of each APA isoform was mapped by excising and sequencing the corresponding PCR band (Fig. S6a). (c) As in (b) except 3'RACE-PCR was performed on Col1a2 mRNA. Sequencing of the PCR products confirmed the InPAS-identified lAPA and sAPA isoforms. (d, e) As in (b) and (c), respectively, except RNA was purified from the indicated cell lines. ATDC5 cells were differentiated under conditions that promoted hypertrophic differentiation and mineralization (Fig. S7). (f) RT-qPCR quantitation of the osteoblast and mineralization marker bone gamma carboxyglutamate protein (Bglap) in undifferentiated and differentiated MC3T3 cells. The expression level of Bglap mRNA was normalized to that of b-actin mRNA, and the normalized level in undifferentiated cells is defined

as 100. (g) Alizarin red staining of ATDC5 before (-) and after (+) a 3-week differentiation course. Positive staining of differentiated cells indicates chondrocyte mineralization (Fig. S7). (h) as in (f) except RNA was purified from ATDC5 cells shown in (g) and quantitation was performed on the specified transcripts. Results demonstrate high expression of chondrocytehypertrophy and mineralization markers in differentiated ATDC5 cells.

Gel images are representative of three independent replicates. Bar graphs represent the average of three independent replicates ± SEM. (**) $P < 0.01$; (****) $P < 0.0001$ using unpaired Student's t-test.

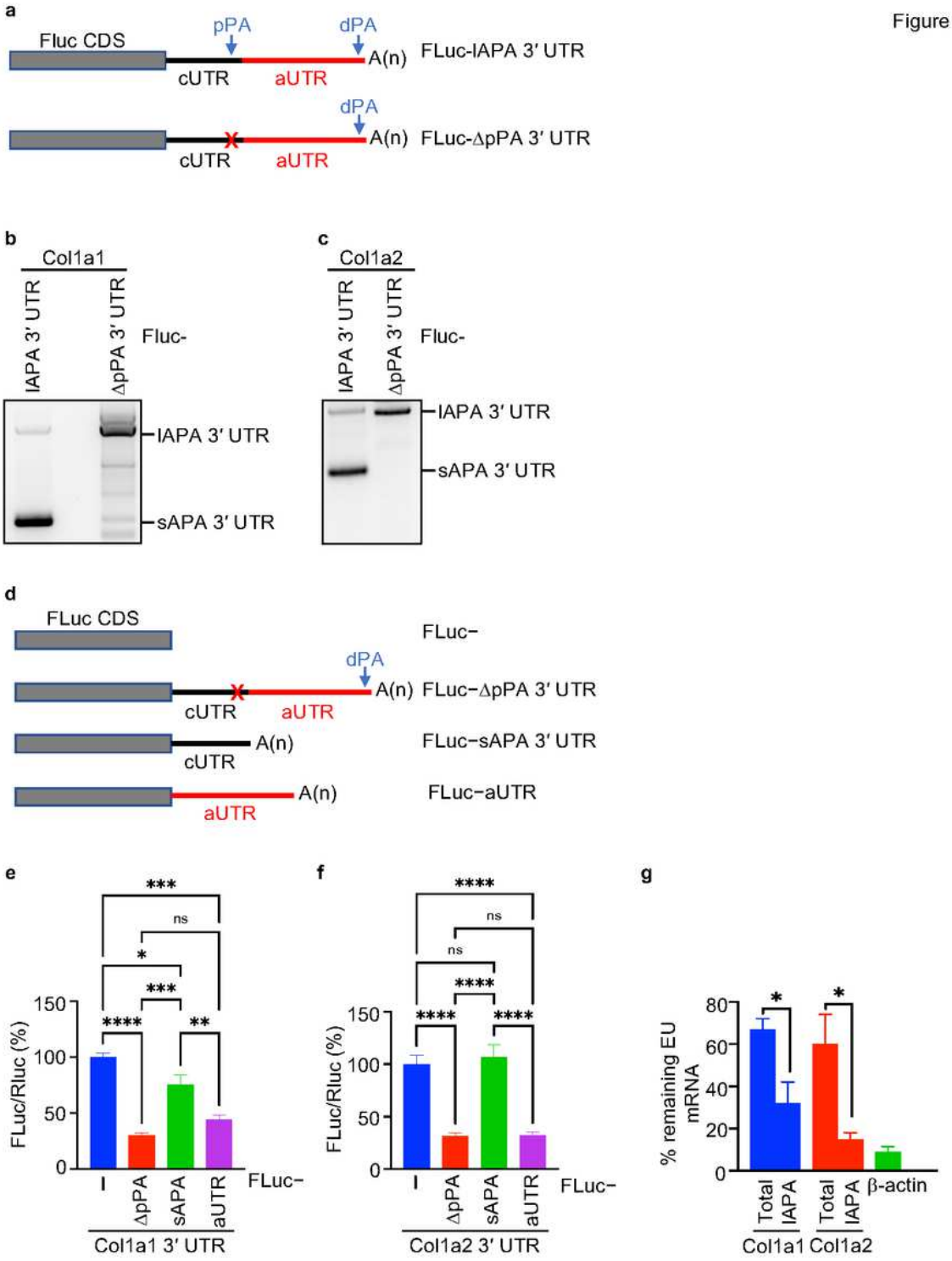

Figure 4

Shortening of the 3'UTR promotes the expression of Col1a1 and Col1a2. (a)

Schematic of FLuc reporters that contain either the 3'UTR of the lAPA isoform (i.e., full-length

3'UTR) of Col1a1or Col1a2 (FLuc-lAPA 3'UTR) or the deletion mutant where the SSR was deleted (FLuc-DpPA 3'UTR; X denotes the position of the deleted SSR). CDS: coding region. (b) 3'RACE-PCR of FLuc mRNA performed on RNA isolated from MC-3T3 cells transfected with Col1a1 3'UTR reporters shown in (a). PCR was performed using an FLuc-specific primer. Results indicate that deletion of the SSR in FLuc-DpPA 3'UTR reporter abrogated 3'UTR shortening, resulting in expression of FLuc mRNA as a single, long APA isoform. (c) As in (b) except Col1a2 3'UTR reporters were used. (d) Schematic of FLuc reporters that contain different isoforms or regions of Col1a1 or Col1a2 3'UTR. (e) Bar graph presentation of the results of dual luciferase assays performed on MC3T3 cells transfected with FLuc-Col1a1 3¢ UTR reporters shown in (d). FLuc activity was normalized to Renilla luciferase (RLuc) activity, and the FLuc/RLuc ratio in cells transfected with the “empty” FLuc (FLuc-) reporter is defined as 100. (f) As in (e) except FLuc-Col1a2 3'UTR reporters were used. N = 4-5 independent

transfections, each measured in triplicate. (g) Bar graph presentation of the results of pulse-chase experiments used to measure RNA decay. MC3T3 cells were incubated with 5-ethynyl uridine (EU), which is an analog of uridine, for 2 hours to label nascent RNA “pulse step”. Cells were then either collected to determine the total level of EU-labeled mRNA (EU mRNA) or grown for 24 hours in EU-free medium “chase step”. At each timepoint, EU RNA was purified (see Materials and Methods) and the level of each of the specified EU mRNA was determined using RT-qPCR. Primers that bind to the aUTR were used to specifically quantitate the lAPA isoform of Col1a1 or Col1a2 (Table S14). The level of EU mRNA measured after the pulse step is defined as 100%, and the graph shows the % remaining after the 24-hour chase. Results demonstrate faster loss of Col1a1 and Col1a2 lAPA isoforms as compared to total mRNA. bactin mRNA was used as a positive control for EU mRNA decay, and consistent with its reported half-life (~6.6 hours)50, < 10% remained after the 24-hour chase.

Gel images are representative of three independent replicates. Bar graphs represent the average ± SEM. (ns) $P > 0.05$; (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; (****) $P < 0.0001$ using oneway analysis of variance (ANOVA) followed by Tukey's post hoc test for (e) and (f), and unpaired Student's t-test for (g)

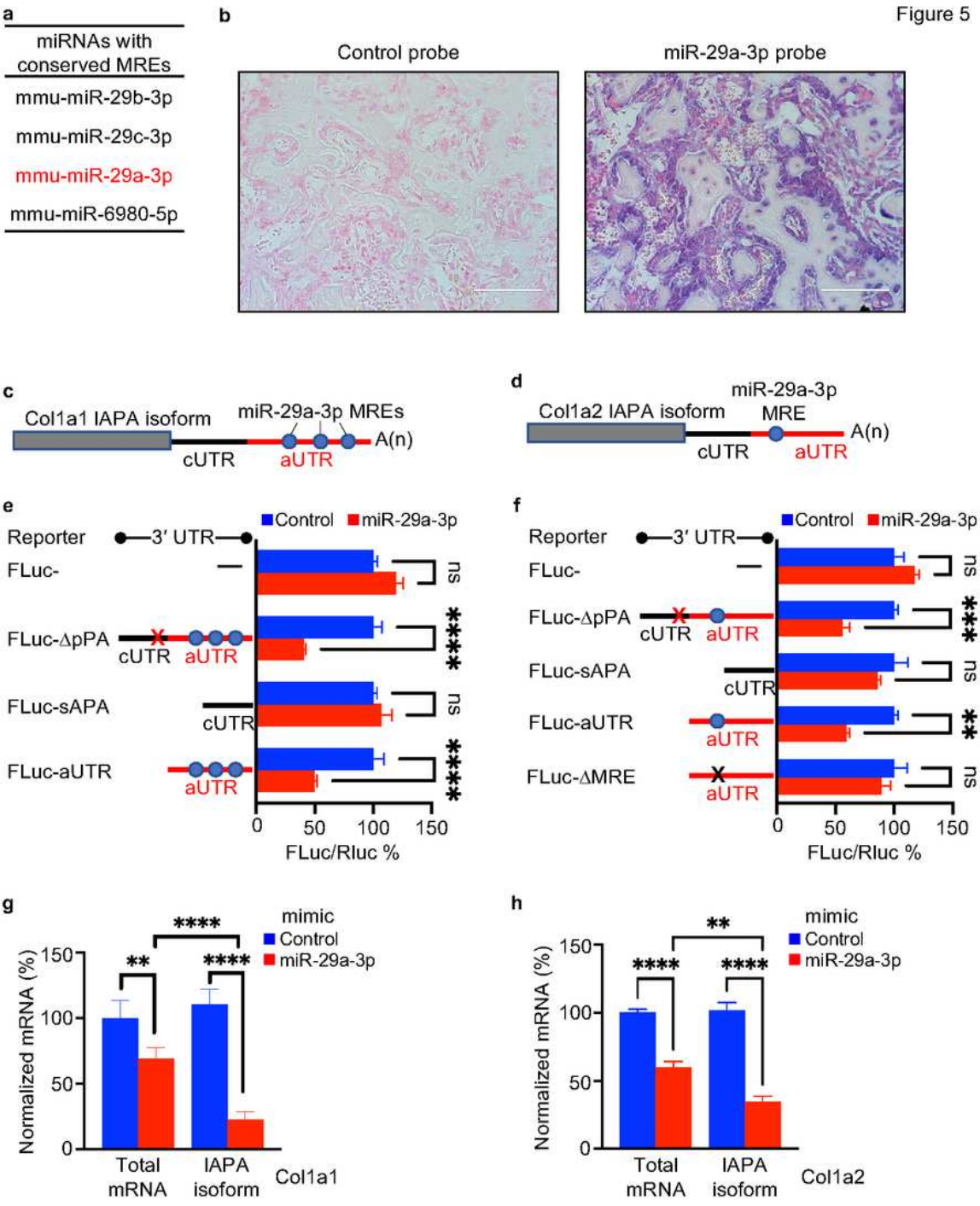

Figure 5

miR-29a-3p mediates the degradation of IAPA, but not sAPA, isoforms of Col1a1

and Col1a2 mRNAs. (a) List of miRNAs that have conserved MREs in the 3'UTR of mouse

Col1a1 mRNA (identified by both TargetScan and miRDB) (Tables S8, S9). miR-29a-3p (highlighted in red) exhibited the highest expression among the 4 shown miRNAs in the callus of d14 post-fracture (Table S10). (b) ISH using either a control (left) or miR-29a-3p (right) probe. Results show intense staining (blue) of miR-29a-3p in the woven bone area of d14 callus. (c) Schematic of the lAPA isoform of Col1a1 mRNA showing the putative miR-29a-3p MREs in the aUTR region. (d) As in (c) except Col1a2 mRNA is shown. (e) Bar graph presentation of the results of dual luciferase assays performed on MC3T3 cells transfected with the indicated Col1a1 FLuc reporters along with either a control (blue) or miR-29a-3p (red) mimic. FLuc activity was normalized to RLuc activity, and the FLuc/RLuc ratio in cells transfected with control mimic is defined as 100. N = 4-5 independent transfections, each measured in triplicate. (f) as in (e) except cells were transfected with Col1a2 FLuc reporters. In the FLuc-DMRE reporter, we replaced the binding region of miR-29a-3p seed sequence by a random sequence (denoted by

“X”). (g) RT-qPCR quantitation of either total mRNA or lAPA isoform of Col1a1 in MC3T3 cells transfected with control (blue) or miR-29a-3p (red) mimic. Primers that bind within the aUTR region were used to specifically amplify and quantitate the lAPA isoform (Table S14). Expression level was normalized to that of b-actin mRNA, and the normalized level in control mimic-transfected cells is defined as 100. N = 5. (h) As in (g) except total mRNA and lAPA isoform of Col1a2 were quantitated.

Bar graphs represent the average ± SEM. (ns) $P > 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; (****) $P < 0.0001$. Significance of difference was calculated using unpaired Student’s t-test in (e) and (f), and two-way ANOVA in (g) and (h).

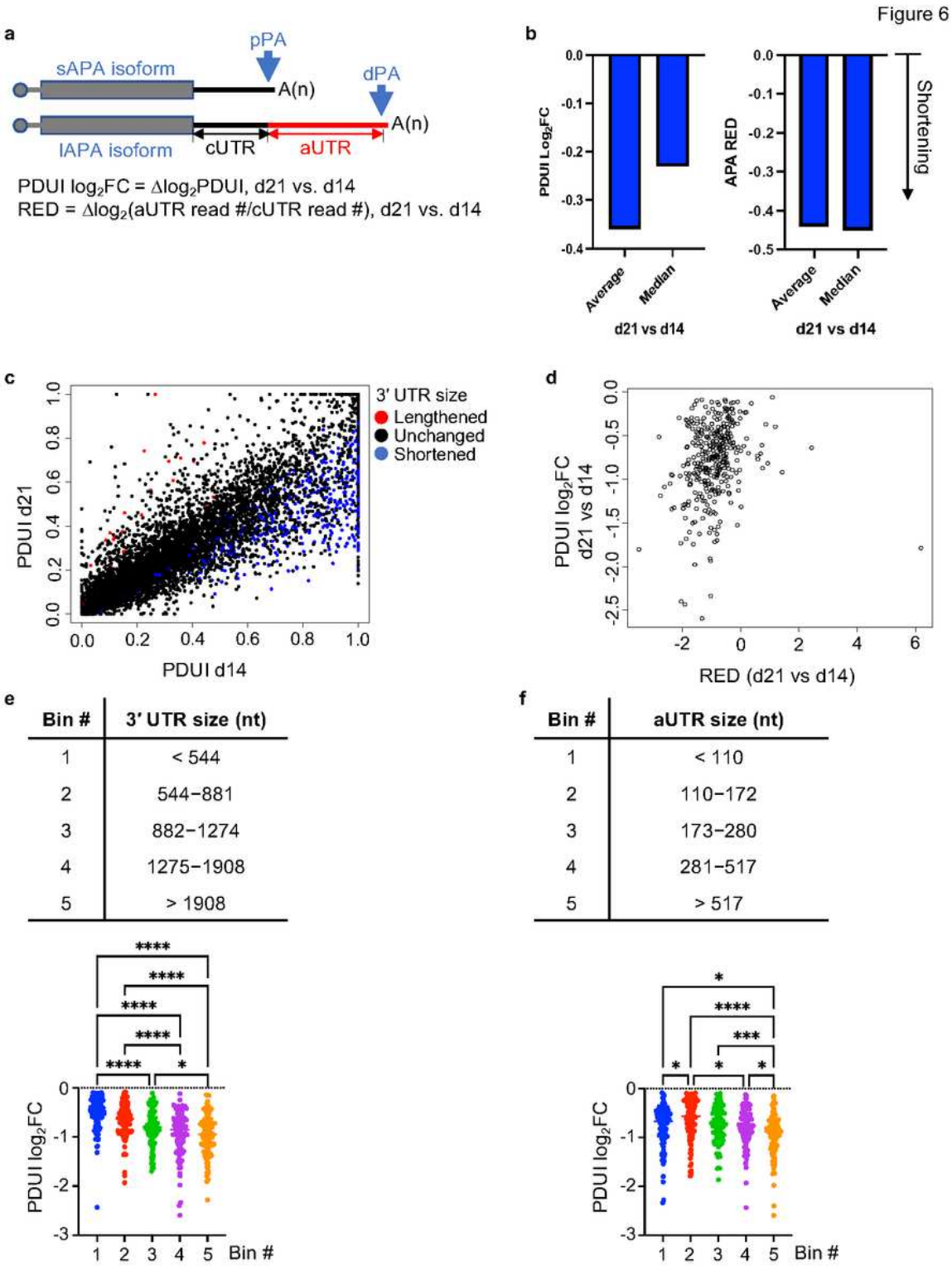

Figure 6

Prevalent 3'UTR shortening accompanies endochondral ossification. (a) General

schematic presentation of APA isoform showing representative proximal and distal PAs (pPA

and dPA, respectively). PDUI Log2FC and RED values are calculated by InPAS and APAlyzer, respectively, as indicated. (b) Average and median PDUI log2FC (left) and RED (right) values calculate by InPAS and APAlyzer, respectively, for all analyzed genes (Tables S4, S5). The negative average and median values indicate global trend toward 3'UTR shortening at d21 relative to d14. (c) A Scatter plot showing changes in 3'UTR length at d21 as compared to d14. Each dot represents a gene, and significantly lengthened or shortened 3'UTRs are shown in red and blue, respectively. Significance is defined as Padj < 0.05 and > 20% change in the expression ratio of the 2 APA isoforms. (d) Correlation between InPAS-generated PDUI log2FC and APAlyzer-generated RED values for genes that exhibited significant 3'UTR shortening at d21 relative to d14 (shown in blue in “c”). (e) (Top): genes that exhibited significant 3'UTR shortening were categorized into 5 similarly sized bins based on their 3'UTR length. (bottom): Scatter plot showing the relationship between 3'UTR length and the extent of 3'UTR shortening

(presented as PDUI log2FC). Each dot represents a gene. Data are presented as average ± SEM.

(f) As in (e) except genes were categorized based on their aUTR length. (*) P < 0.05; (***) P <

0.001, (****) P < 0.0001 using one-way ANOVA followed by Tukey’s post hoc test.

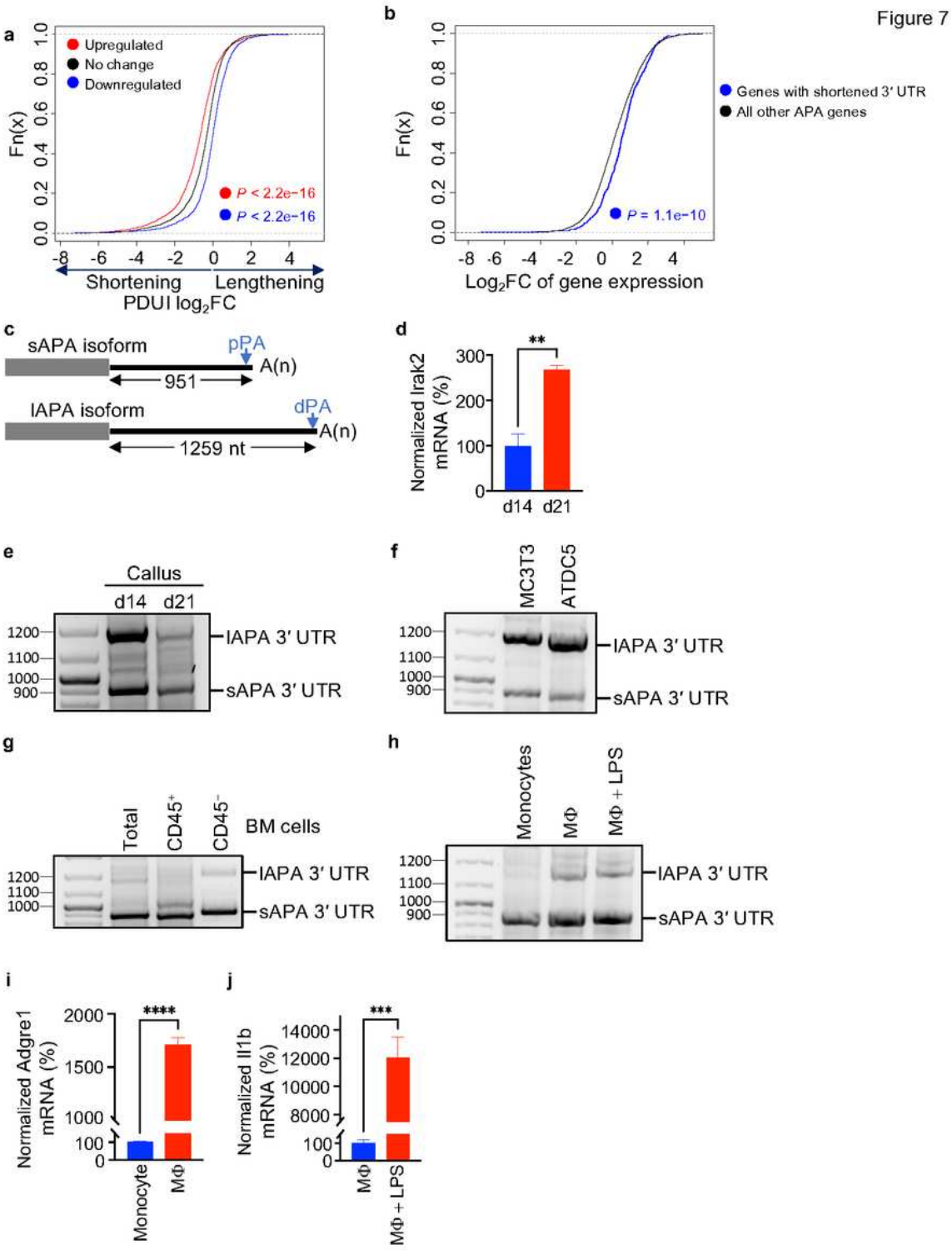

Figure 7

3'UTR shortening is associated with increased gene expression. (a) Cumulative

distribution function (CDF) curves of PDUI log2FC values (generated by InPAS) for genes

whose expression was significantly upregulated (red line), significantly downregulated (blue line), or unchanged (black line) at d21 relative to d14. Significant change in gene expression is defined as Padj < 0.05. All InPAS-identified APA events are included in the analysis (See Fig. S10a for APA events identified by APAlyzer). P-values (K-S test) for significance in difference between red or blue and black genes are indicated. (b) CDF curves comparing gene expression of genes identified by InPAS to show significant 3'UTR shortening (blue line) and all other genes (black line). P-values (K-S test) are based on comparing blue and black genes (See Fig. S10b for APAlyzer data). (c) Schematic of the lAPA and sAPA isoforms of Irak2. (d) RT-qPCR quantitation of Irak2 expression in d14 and d21 callus. The expression level of Irak2 mRNA was normalized to that of b-actin mRNA, and the normalized level in d14 callus is defined as 100. (eh) Representative gel image of Irak2 3'RACE-PCR products in d14 and d21 callus (e), differentiated MC3T3 and ATDC5 cells (f), Total, CD45+, and CD45- bone marrow cells (g), or

monocytes, macrophages (MF), and LPS-stimulated MF (h). (i) as in (d) except the MF marker

Adgre1 was quantitated. Results indicate successful in-vitro differentiation of monocytes to MF.

(J) As in (d) except Il1b mRNA was quantitated. Results indicate upregulated expression of the

pro-inflammatory Il1b gene in response to LPS stimulation.

Bar graphs represent the average ± SEM. (**) P < 0.01; (***) P < 0.001; (****) P < 0.0001

using unpaired Student's t-test.

Supplementary Files

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

- ElbarbarySuppFigures.pdf
- ElbarbaryTableS1.xls
- ElbarbaryTableS2.xls
- ElbarbaryTableS3.xls
- ElbarbaryTableS4.xlsx
- ElbarbaryTableS5.xls
- ElbarbaryTableS6.xlsx
- ElbarbaryTableS7.xlsx
- ElbarbaryTableS8.xlsx
- ElbarbaryTableS9.xlsx
- ElbarbaryTableS10.xlsx
- ElbarbaryTableS11.xls
- ElbarbaryTableS12.xls
- ElbarbaryTableS13.xls

- ElbarbaryTableS14.xlsx
- ElbarbarySuppfiglegends.pdf