Elucidating dynamics and regulation of alternative splicing in osteogenic differentiation

Yuan Yuan Wang^{1,4}, Rene F Chun^{2,3}, Samir Adhikari⁴, Christopher M Lopez³, Mason Henrich³, Vahe Yacoubian³, Lan Lin⁵, John S. Adams^{2,3}, Yi Xing^{4,5}

¹Bioinformatics Interdepartmental Graduate Program, University of California, Los Angeles

²Department of Orthopaedic Surgery, University of California, Los Angeles

³Department of Molecular, Cell & Developmental Biology, University of California, Los Angeles

⁴Center for Computational and Genomic Medicine, Children's Hospital of Philadelphia

⁵Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania

Abstract

Nearly all human multi-exonic genes undergo alternative splicing (AS) via regulation by RNA-binding proteins (RBPs), but few studies have examined the temporal dynamics of AS and its regulation during cell differentiation in the bone niche. We sought to evaluate how AS, under the control of RBPs, affects cell fate commitment during induced osteogenic differentiation of human bone marrow-derived multipotent stem/stromal progenitor cells (MSPCs). We generated a time-course RNA sequencing (RNA-seq) dataset representative of induced MSPC differentiation to osteoblasts. Our analysis revealed widespread AS changes, coordinated with differential RBP expression, at multiple time points, including many AS changes in non-differentially expressed genes. We also developed a computational approach to profile the dynamics and regulation of AS by RBPs using time-course RNA-seq data, by combining temporal patterns of exon skipping and RBP expression with RBP binding sites in the vicinity of regulated exons. In total we identified nine RBPs as potential key splicing regulators during MSPC osteogenic differentiation. Perturbation of one candidate, *KHDRBS3*, inhibited osteogenesis and bone formation *in vitro*, validating our computational prediction of “driver” RBPs. Overall, our work highlights a high degree of complexity in the splicing regulation of MSPC osteogenic differentiation. Our computational approach may be applied to other time-course data to explore dynamic AS changes and associated regulatory mechanisms in other biological processes or disease trajectories.

Introduction

Alternative splicing (AS) is a prevalent mechanism for generating regulatory and functional complexity in eukaryotes (Nilsen and Graveley, 2010). The most common mode of AS in human cells is exon skipping. Less common modes include use of alternative 5' splice sites, alternative 3' splice sites, mutually

exclusive exons, and intron retention (McManus and Graveley, 2011; Park et al., 2018). In each case, the resultant transcript isoforms might produce proteins with altered structure and function. Not only does AS have a vital role in normal cellular processes (Cieply et al., 2016; Song et al., 2017; Zhang et al., 2016), it is implicated in a variety of diseases (Dvinge et al., 2016; Pagani and Baralle, 2004; Rahman et al., 2020; Scotti and Swanson, 2016; Zhang and Manley, 2013).

Splicing is coordinately regulated by the interaction of *cis*-acting RNA elements and *trans*-acting splicing factors. The *cis*-acting elements include pre-mRNA sequence motifs in exons or flanking introns that function as splicing enhancers or silencers when bound by RNA-binding proteins (RBPs) acting as splicing factors (Gerstberger et al., 2014). Recently, the study of AS has been accelerated by the development of high-throughput sequencing methods and novel bioinformatic tools that identify and quantitate AS (Park et al., 2018; Shen et al., 2014).

As humans age the balance of bone and fat content in the skeleton shifts more toward fat, leading the way to skeletal fragility (Duque, 2008; Meunier et al., 1971). Age-related bone loss resulting in osteoporosis and increased fractures represents a major health problem with increased morbidity and decreased quality of life (Blume and Curtis, 2011; Kapinos et al., 2018). Human bone-forming cells or osteoblasts are derived from bone marrow-derived multipotent stem/stromal progenitor cells (MSPCs). As their name implies, MSPCs are capable of undergoing osteogenesis, adipogenesis, chondrogenesis and myogenesis (Augello et al., 2010; Pittenger et al., 1999) with a major differentiation branch point between osteogenesis and adipogenesis. This branch point represents a binary fate choice for MSPC. The signals that specify MSPC to specific cell differentiation pathways are not fully understood but transcription factors are considered by most to be the dominant signal. *RUNX2* (Komori et al., 1997; Otto et al., 1997) and *PPAR γ* (Barak et al., 1999; Rosen et al., 1999) are well known transcription factors that have roles in osteogenesis and adipogenesis, respectively. Our understanding of osteogenic differentiation from MSPC has made tremendous progress in the last several decades. This includes: 1) establishing criteria for MSPC characterization (Chan et al., 2018; Dominici et al., 2006; Robey and Riminucci, 2020); 2) optimizing procedures for isolation (Soleimani and Nadri, 2009; Wagey and Short, 2013) and culture of MSPC (Ciuffreda et al., 2016; Jaiswal et al., 1997); and 3) determining some of the key regulatory molecules, particularly those that control transcription, that modulate MSPC differentiation patterns (Chen et al., 2016; Han et al., 2019).

In the mouse, recent studies in transcription and epigenetic mechanisms regulating MSPC differentiation using single-cell RNA sequencing (scRNA-seq) approaches have yielded important findings with regard

to the fate decisions in MSPC differentiation (Rauch et al., 2019; Wolock et al., 2019; Zhong et al., 2020). For example, Zhong et al identified a lineage commitment progenitor population that could differentiate to adipocytes or osteocytes (Zhong et al., 2020). Additionally, two groups (Rauch et al., 2019; Wolock et al., 2019) found that the expression profiles of transcription factors revealed that adipogenesis required a much larger number of expressed transcription factors compared to osteogenesis, and that there existed a subset of transcription factors that acted in both pro-osteogenic and anti-adipogenic manner, supporting a net gain in osteoblasts over adipocytes. One report demonstrated that adipogenesis invoked more chromatin remodeling relative to osteogenesis (Rauch et al., 2019). In each report noted above, selection of either the osteogenesis or adipogenesis differentiation path appeared binary.

In this work, we studied the role of AS during the MSPC osteogenic differentiation. We conducted a time course study with primary MSPCs derived from the marrow space of human femurs cultured in osteogenic media. At selected time points, RNA was obtained for RNA sequencing (RNA-seq). Extensive and dynamic transcriptomic alterations are orchestrated during induced osteogenic differentiation, including remodeling of the gene expression and splicing programs. We developed a multi-step bioinformatic strategy to identify RBPs that may be key regulators of AS in the process of osteogenesis.

Results

Extensive transcriptomic alterations characterize the induced, stepwise differentiation of primary MSPCs to osteoblasts.

In order to evaluate gene expression and AS regulation during osteogenesis, we cultured MSPCs in osteogenic differentiation media over 12 days to obtain temporal MSPC osteogenic differentiation datasets. Cells were stained for alkaline phosphatase activity and mineralization by alizarin red every two days (**Figure 1A**). The anticipated increase in staining intensity was observed qualitatively (**Figure 1A**) and quantitatively (**Figure 1B**). RNAs were isolated in triplicate on day 0, day 2, day 4, day 6, day 8 and day 12, followed by high-throughput RNA-seq (**Figure S1A**). The RNA-seq-derived gene expression data showed that markers of immature osteoblasts (*MYC* and *SOX9*) were gradually downregulated as markers of osteoblast differentiation (*RUNX2* and *ALPL*) were upregulated during the time course (**Figure 1C**), consistent with osteogenic differentiation of the MSPCs. These sequential changes in histological staining and RNA-seq-derived marker gene expression confirmed MSPC progression toward a mature osteoblast.

A substantial remodeling of the gene expression and AS profiles occurred in osteogenic media-induced osteogenic differentiation of MSPCs. Using rMATS-turbo (rMATS 4.0.2) (Shen et al., 2014), we

uncovered ~23,000 AS events across all 18 samples (**Figure S1D**), including exon skipping, alternative 5' splice sites, alternative 3' splice sites and intron retention. Because exon skipping is the most prevalent and most well-characterized type of AS events in human transcriptomes (Park et al., 2018) and because it represented 72% of the total AS events in this study, we focused on exon skipping events (**Figure S1D**). Temporal profiling of induced changes in the MSPC transcriptome from RNA-seq was assessed globally by principal component analysis (PCA). The PCA plots of total gene expression, RBP gene expression, and exon skipping showed similar temporal patterns (**Figure 1D**), suggesting coordination between gene expression, RBP gene expression and exon skipping. Moreover, the genes contributing most to the separation of samples on PC1 in total gene expression PCA plot were highly enriched in RNA splicing-related gene ontology terms (**Figure S1B, S1C**), further indicating the coupling of gene expression and splicing programs during induced MSPC osteogenic differentiation. Interestingly, genes contributing most to the separation of samples on PC2 in exon skipping PCA plot were enriched in regulation of transmembrane calcium ion transport term as well as regulation of RNA splicing term (**Figure S1E, S1F**). This is consistent with the observation that the net flux of calcium to the extracellular environment in osteoblasts is closely related to mineralization of the collagenous extracellular matrix (Boonrungsiman et al., 2012), again highlighting the importance of AS in osteogenic differentiation. Together, this transcriptome-wide analyses of osteogenic differentiation identifies interplay of gene expression, splicing and bone development related biological processes.

Pair-wise differential analysis identifies temporal patterns of gene expression and exon skipping during MSPC-to-osteoblast differentiation

It has been well established that coordinated splicing networks play a vital role in cell fate determination and result in physiological consequences in various developmental and tissue remodeling processes in humans (Baralle and Giudice, 2017). In the field of skeletal modeling and remodeling the focus has been on changes in transcription factor gene expression (Rauch et al., 2019). On the other hand, there is a lack of systematic assessment of splicing networks in skeletal development and maintenance. To fill this knowledge gap and elucidate the dynamics and regulation of AS during osteogenic differentiation, we performed pair-wise comparisons on both gene expression and exon skipping in MSPC over 12 days of induced osteogenesis (**Figure 2A**). Beginning in the early stages of osteogenesis (day 0-2), there was a robust and dynamic alteration in the splicing program of MSPCs as they matured to an osteoblast phenotype (**Figure 2A**).

To further decipher the relationship between gene expression and exon skipping in the process of induced MSPC osteogenic differentiation, we investigated the expression patterns of splicing-regulated genes.

Interestingly, most of those splicing-regulated genes were not differentially expressed (**Figure 2B**), suggesting that splicing, in and of itself, represents another layer of transcriptome remodeling during osteogenesis. For example, the filamin A gene (*FLNA*), of which missense point mutations are associated with a range of X-linked skeletal dysplasias (Feng and Walsh, 2004), was persistently highly expressed in both the MSPC and mature osteoblast population of cells with no change in gene expression over the 12-days of induced osteogenesis (**Figure 2B, 2C**). Despite no change in expression of *FLNA*, the inclusion level (percent spliced in, PSI) of the alternatively spliced exon 30 (ENST00000369850), which maps to the functional filamin domain repeat 15, steadily decreased from 0.72 to 0.52 over 12 days of osteogenic induction (**Figure 2C**).

Notably, of 604 exon skipping events (from 488 genes) with significant PSI value changes in at least 3 pair-wise comparisons, 53 were located in 47 genes encoding transcription factors (**Supplementary Table 1**). Those exon skipping-regulated transcription factors were often affected by 1) loss of a portion or the entirety of a functional domain encoded by the alternatively spliced exon, 2) a frame shift resulting in disruption or presence of downstream functional domains of the translated protein or 3) nonsense-mediated decay (**Supplementary Table 1**); either of these scenarios can lead to loss/gain of transcription factor function with a corresponding global reconstruction of the gene expression network. Exon skipping events residing in transcription factor genes without frameshift, NMD induction or functional domain encoding exons can also exert functional consequences. An example is differential inclusion/exclusion of exon 4 in the transcriptional co-activator gene, paired mesoderm homeobox protein 1 gene (*PRRX1*; also known as *PRX1*) (**Figure S2**). Two isoforms were produced by alternative splicing of the cassette exon 4 in *PRRX1*: the exon 4 skipping isoform, PRRX1A and exon 4 inclusion isoform, PRRX1B. Interestingly, the inclusion of exon 4 in the PRRX1 gene introduced an earlier stop codon, shifting exon 5 to the 3' untranslated region and encodes a shorter protein isoform (PRRX1b) lacking the functional OAR (otp, aristaless, and rax) domain (**Figure S2A, S2B**). The OAR domain is involved in DNA binding and inhibition of transcription activation (Norris and Kern, 2001; Reichert et al., 2013). Previous studies have established that PRRX1a and PRRX1b act differentially to regulate progenitor cell proliferation and differentiation (Takano et al., 2016; Wang et al., 2018). Indeed, the sashimi plot in **Figure S2C** shows that the inclusion level of exon 4 decreased from 63% on day 0 to 41% on day 12, resulting in an isoform switch from the OAR-absent PRRX1B to the OAR-containing PRRX1A during induced MSPC osteogenesis. This suggests that OAR-containing PRRX1A favors an osteogenic differentiation fate more than does the OAR-absent PRRX1B. This result parallels with previous observations in the mouse system where overexpression of *Prrx1b*, but not *Prrx1a*, interferes with *Osx*- and *Runx2*-directed mRNA expression and inhibits pre-osteoblast-to-osteoblast differentiation (Lu et al., 2011).

Many AS events are coordinately regulated by *trans*-acting RBPs in a developmental stage-specific manner (Erhard et al., 2019; Fu and Ares, 2014). The global changes in RBP gene expression (**Figure 2D**) and exon skipping PSI values (**Figure 2F**) are displayed as heatmaps, each defined by 5 clusters with distinct inter-cluster temporal patterns determined by unsupervised hierarchical clustering. Overall, the most pronounced changes for both RBP gene expression and exon skipping were observed in the early stages of induced osteogenesis (from day 0 to day 2, **Figure 2D, 2F**). The temporal RBP gene expression patterns and temporal exon skipping PSI value patterns were further delineated into lines (**Figure 2E, 2G**). Distinct patterns of RBP gene expression emerged from different clusters (**Figure 2E**): steady upregulation (e.g., cluster 3); steady downregulation (e.g., cluster 2); early stage change (e.g., cluster 1) or late stage change (e.g., cluster 5). Distinct patterns of exon skipping PSI values were also observed (**Figure 2G**). This included clusters of exons with: progressive inclusion (e.g., cluster 5); progressive exclusion (e.g., clusters 3) and early stage change (e.g., clusters 1 and 2).

Computational screening identifies RBP candidates for regulation of exon skipping during osteogenic differentiation

To further identify RBPs that could control the splicing program in induced MSPC osteogenic differentiation, we developed a computational screening method searching for potential ‘key splicing regulators’ among 129 RBPs with known motif position weight matrix (Dominguez et al., 2018; Ray et al., 2013). *Trans*-acting regulatory RBPs usually bind to *cis*-acting RNA elements in the precursor mRNA (pre-mRNA) in a sequence-specific manner with the resulting RBP-directed splicing behavior frequently dependent on the location of the RBP binding site relative to the regulated exon (Fu and Ares, 2014; Gabut et al., 2008). Taking this *cis-trans* locational information into consideration, the method described here combines both correlation analysis and region-specific motif enrichment analysis to estimate the region-specific regulatory potential of RBPs to influence pre-mRNA splicing (**Figure 3A**). Detected exon skipping events not significantly changed in any of the 5 pair-wise comparisons with day 0 were classified as background events, while exon skipping events that were significantly changed in at least three of the five comparisons were retrieved as foreground events. For each differentially expressed RBP, foreground events were assigned into a ‘positive’ or ‘negative’ foreground event set based on the correlation coefficient between its PSI value and the RBP gene expression if highly correlated ($r^2 > 0.5$) (**Figure 3A, top**). Sequences extracted from each exon skipping event were subsequently scanned for the presence of an RBP binding site based on the motif scores calculated from matching the position weight matrix of the RBP motifs to possible binding positions (**Figure 3A, middle**). Finally, one-tailed Fisher’s exact tests were performed to estimate the region-specific enrichment of motifs in the foreground event

sets compared to the background event set (**Figure 3A, bottom**). Eleven RNA binding motifs for nine RBPs were determined to be significantly enriched in at least one region for at least one foreground event set (**Figure 3B, 3C**).

KH RNA Binding Domain Containing, Signal Transduction Associated 3B, (*KHDRBS3*, *SLM2*, *T-STAR*) was one of the identified RBP splicing regulators with its RNA binding motif enriched in exon body region from negatively-correlated foreground event set and showed a temporal increase in expression during induced osteogenic differentiation (**Figure 3B, 3C**). Previous studies reported that *KHDRBS3* and another member from the STAR family, *KHDRBS1* (*SAM68*), are involved in splicing regulation in a variety of developmental and disease processes (Baralle and Giudice, 2017; Danilenko et al., 2017; Farini et al., 2020; Feracci et al., 2016; Traunmuller et al., 2016; Zhang et al., 2020). Examples of two significantly changing exon skipping events, unaccompanied by a difference in gene expression, are shown in **Figure S3**. During induced osteogenesis two potential *KHDRBS3* pre-mRNA targets, the transcription factor coregulator gene *AIRD4B* (*RBP1L1*, *RBBP1L1*) and the periostin (*POSTN*) gene were alternatively spliced but not differentially expressed. In both instances a *KHDRBS3* RNA binding motif was present in the body of the regulated exon with the PSI value of the target exon being negatively correlated with *KHDRBS3* gene expression (**Figure S4, S3A, S3C**).

ARID4B and *ARID4A* (*RBP1*, *RBBP1*) are two homologous members of the AT-rich interaction domain (ARID) gene superfamily (Wilsker et al., 2005); they encode subunits of the SIN3 Transcription Regulator Family Member A (SIN3A)/HDAC (histone deacetylase) transcriptional corepressor complex which functions in various cellular processes including proliferation, differentiation and cell fate decision (Clark et al., 2015; Fleischer et al., 2003). As depicted in **Figure S3B**, the alternatively spliced cassette exon 16 in *ARID4B*, which encodes part of the Tudor-knot domain, is significantly more often skipped (PSI diminished) in day-12 compared to day-0 differentiating MSC. Compared to the traditional Tudor domain, which is involved in protein-protein interaction (Lasko, 2010), the Tudor-knot contains crucial configurations needed for RNA binding activity (Shimojo et al., 2008). Therefore, partial disruption of the Tudor-knot during induced osteogenesis might result in changes in the assembly or stability of *ARID4B* involved supramolecular complexes. For example, in the mouse *Arid4a* is reported to be a *Runx2* coactivator that promotes osteoblastic differentiation (Monroe et al., 2010). Considering the fact that *ARID4A* and *ARID4B* can also physically interact with each other (Wu et al., 2013), the alteration of their protein-protein or protein-RNA interaction might squelch *ARID4A*'s function as a *RUNX2* coactivator.

Periostin is a secreted extracellular matrix protein that was originally identified in cells from the mesenchymal lineage in the skeleton (e.g., osteoblasts and osteoblast-derived cells). Periostin expression promotes bone anabolism partially through its ability to regulate osteoblast differentiation from mesenchymal progenitors (Bonnet et al., 2012); deletion of *Postn* gene impairs fracture consolidation in mice (Duchamp de Lageneste et al., 2018). As was the case with *ARID4B*, *POSTN* expression does not change significantly over the course of induced osteogenesis but exon skipping in *POSTN* pre-mRNA does, indicating that *POSTN* has the potential to contribute to osteogenic differentiation without a change in expression level. There are multiple isoforms of periostin reported, all differing in their C-terminal sequences (Litvin et al., 2004). This is consistent with our observation that the inclusion level of exon 18 which resides in the C-terminal region of *POSTN* decreased from 76% to 50% over the 12-day course of MSPC osteogenesis (**Figure S3D**).

siRNA knockdown of KHDRBS3 reduces osteogenesis in vitro

As noted above, *KHDRBS3* was one of the nine RBPs that emerged from our computational screening pipeline (**Figure 3** and **Figure S4**). *KHDRBS3* showed a robust increase in expression over time during induced osteogenesis. This made *KHDRBS3* an ideal candidate for siRNA knockdown to test the biological significance of *KHDRBS3* during osteogenesis. siRNA knockdown of *KHDRBS3* relative to negative control siRNA was observed over seven days of induced osteogenic differentiation (**Figure 4**). As assessed by qPCR, a statistically significant knockdown of *KHDRBS3* to levels at or below initial levels of expression at all time points was achieved (**Figure 4A**). Diminished osteogenic differentiation was confirmed by significant reduction of the osteogenic maturation marker genes *RUNX2* (**Figure 4B**) and *BGLAP* (**Figure 4C**) in three and four time points, respectively. A significant reduction in functional readout of osteogenesis, alkaline phosphatase staining (**Figure 4D**) was also observed. These data indicate that *KHDRBS3* expression normally supports osteogenesis perhaps via its action on downstream target genes (e.g., *ARID4B* and *POSTN*) to promote progression of MSPC maturation down the osteogenic pathway.

Discussion

While there have been some studies of AS events that occur during tissue-specific adipogenesis (Aprile et al., 2018; Huot et al., 2012; Li et al., 2014; Vernia et al., 2016), there is much less known about AS during the differentiation of mesenchymal progenitors down the osteogenic pathway in the human bone marrow niche. Studies of AS during osteogenesis have focused primarily on exon skipping in *RUNX2*, the master regulator gene for osteogenesis (Makita et al., 2008). Skipping exon 5 and/or 7 in *RUNX2* pre-mRNA produces isoforms of *RUNX2* incapable of DNA binding and downstream transactivation of

genes required for normal bone formation, including osterix (*OSX*), *OCN* (osteocalcin), *OPN* (osteopontin) and *COL1A1*. Some of these genes can also be regulated by AS. For example, *COL1A1* encodes alpha-1 type I collagen, which is the most plentiful collagen in bone. Aberrant AS of the *COL1A1* gives rise to a form of the human skeletal disease osteogenesis imperfecta (Johnson et al., 2000; Wang et al., 1996; Xia et al., 2008).

In this study, we generated a 12-day time-course RNA-seq dataset from primary cultures of MSCs harvested from a human femur after they were induced to differentiate to bone-forming osteoblasts *in vitro*. This dataset not only permits a comprehensive examination of stepwise changes for both gene expression and AS, but also opens the opportunity to examine for the first time in a completely unbiased mode the potential regulatory role of RBP-directed changes in AS in MSC differentiation. We found a high degree of similarity between the temporal patterns of overall gene expression, RBP gene expression and exon skipping (**Figure 1D**), suggesting that these events are mechanistically linked. We also observed that: 1) genes with the greatest variance in expression were significantly enriched for splicing-related gene ontology terms (**Figure S1B, S1C**); and 2) a large proportion of the differentially spliced genes encode transcription factors (**Supplementary Table 1**).

By combining temporal correlation of exon skipping and RBP expression with RBP binding site enrichment in the vicinity of regulated exons (**Figure 3**), we present a computational approach to identify key RBPs that drive AS changes in osteogenic differentiation. Considering the fact that AS is often regulated by binding of *trans*-acting RBPs to *cis*-acting RNA elements in a position-dependent manner (Fu and Ares, 2014; Park et al., 2018; Yee et al., 2019), it is important to understand how the region-specific RBP binding influences nearby AS events. Instead of using the whole set of differential AS events, for each candidate RBP two sets of AS events were determined based on positive or negative temporal correlation with RBP expression in osteogenic differentiation. RBP motif enrichment was assessed in three regions (upstream intron, exon body, downstream intron) as a proxy for region-specific RBP association. It should be noted that this computational strategy is generic and can be applied to any time-course RNA-seq dataset and to any type of AS patterns to elucidate RBP regulation of AS.

In this work, nine RBPs were identified as potential key splicing regulators in the process of MSC to osteoblast differentiation. Among those was *KHDRBS3* that exhibited increased expression during osteogenesis (**Figure 3C**) and has been found to control cell fate in development or disease (Baralle and Giudice, 2017; Traunmuller et al., 2016; Zhang et al., 2020). As depicted in **Figure 4**, successful siRNA knockdown of *KHDRBS3* resulted in a commensurate decrease in functional readouts of osteogenesis,

including alkaline phosphatase-directed bone formation. This result indicated that: 1) an increase in KHDRBS3 expression influences normal MSPC osteogenesis; and 2) our computational strategy was successful in identifying a key splicing regulator in the dynamic setting of MSPC-to-osteoblast differentiation. Further evidence is needed to prove that the function of KHDRBS3 on osteogenesis results from its regulation on splicing.

METHODS

MSPC culture

Primary cultures of MSPCs were obtained from PromoCell (C-12974, Heidelberg, Germany). Cells were characterized by the vendor according to criteria proposed by the International Society for Cellular Therapy (Dominici et al., 2006). Lot 402Z027 (47, male, Caucasian) was used in RNA-seq study. Lot 429Z013.1 (56, male, Caucasian) was used in siRNA knockdown study. MSPCs were initially cultured in recommended growth media (PromoCell, C28009) and differentiated in MSPC osteogenic differentiation medium (PromoCell, C-28013) on plates coated with human fibronectin (PromoCell, C-43060). Media was changed every two or three days. Samples for osteogenic differentiation RNA-seq data were obtained at day 0, 2, 4, 6, 8, and 12; and daily samples from day 0 to 7 for siRNA knockdown.

Cell staining for biomarkers of osteogenic differentiation

Alkaline phosphatase staining reagent (5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium) was prepared from BCIP/NBT tablet (Sigma B-5655, St. Louis, MO) in 10 ml water and incubated on cell monolayers after PBS wash for 10 minutes. BCIP/NBT reagent was removed by washing with PBS-Tween 0.05% followed by a PBS wash. Alkaline phosphatase staining was quantified by spectrophotometry at 620 nm. Alizarin Red S (ARS; Sigma A-5533) was prepared at 2% in water and adjusted to pH 4.1 and filtered before usage. Cells were fixed with 10% buffered formalin (Fisher) and washed with water prior to addition of 2% ARS, pH 4.1 for 20 minutes. Excess ARS stain was washed from cells by water four times. Staining was quantified by spectrophotometry at 405 nm.

siRNA knockdown

KHDRBS3 SMART pool On-Target Plus siRNA (Dharmacon) and On-Target Plus non-targeting pool siRNA (Dharmacon) was used as negative control for experiments. Transfection of 7000 cells per well (96-well plates) was conducted with X-treme GENE siRNA transfection reagent (Sigma) at 160 ng siRNA to 1 ul reagent ratio. After eight hours, siRNA and transfection reagent containing media was removed and replaced with MSPC osteogenic differentiation medium (day 0). Media was changed every two or three days for a total of seven days. RNA was isolated from 96-well plates using RNeasy 96 (Qiagen). For qPCR gene expression analysis, cDNA was synthesized by SuperScript IV reverse transcriptase (ThermoFisher) and qPCR performed with TaqMan Fast Advanced Master Mix (ThermoFisher) with eukaryotic 18S rRNA endogenous control probe/primer (ThermoFisher) and gene specific probe/primers: *RUNX2* (Hs01047973_m1), *BGLAP* (Hs01587814_g1), and *KHDRBS3* (Hs00938827_m1). Staining was performed as described as above.

RNA isolation and sequencing library preparation

For RNA-sequencing (RNA-seq), RNA was extracted from 24-well plate MSPC cultures at 0, 2, 4, 6, 8 and 12 days of induced osteogenesis with Trizol (ThermoFisher) and purified with Direct-zol RNA microprep columns (Zymo Research). Three biological replicates were isolated at each time point. RNA-seq libraries were prepared with TruSeq Stranded mRNA Library Kit (Illumina) after which RNA was assessed for quality by Tape Station (Agilent) and quantified by Qubit (ThermoFisher). RNA-seq libraries were pooled, quantified by Qubit 3.0, diluted accordingly and committed to Illumina Paired End 101 base sequencing at the UCLA Broad Stem Cell Research Center High Throughput Sequencing Facility.

RNA-seq read alignment

High-quality raw sequencing reads were obtained and assigned to a corresponding sample by demultiplexing with a maximum of 1 mismatch allowed in the barcode sequence (barcode sequence length 7). Alignment was done using *Hisat2* (v2.0.3-beta) (Kim et al., 2019) with default parameters and a pre-built index for reference plus transcripts based on genome assembly GRCh37 (hg19) annotation (grch37_tran, ftp://ftp.ccb.jhu.edu/pub/infphilo/hisat2/data/grch37_tran.tar.gz).

Gene expression quantification and differential gene expression analysis

Gene expression/transcript abundance were measured in both raw counts and TPM (Transcripts Per Million) using the alignment tool kallisto (v0.43.1) (Bray et al., 2016). Ensemble v75 GRCh37 (hg19) cDNA annotation was used as the guiding reference for kallisto. Transcript-level estimates from kallisto were summarized into gene expression matrices by tximport (v1.6.0, R package) (Soneson et al., 2015) for downstream gene-level analysis. Differential expression analysis was conducted with the count-based tool DESeq2 (v1.18.1, R package) (Love et al., 2014). Technical replicates were collapsed, and lowly expressed genes (TPM ≤ 5 in all samples) were filtered out before performing differential expression analysis. For each comparison, genes with an absolute log₂ fold change $> \log_2(1.5)$ and an FDR (false discovery rate)-adjusted p-value < 0.01 were assumed to be differentially expressed genes. The differentially expressed gene list for the entire osteogenic differentiation pathway was defined as genes differentially expressed in all comparisons between time point day 0 and other time points (day 2, 4, 6, 8, 12).

Alternative splicing analysis to identify significantly changing foreground events and background events

AS events were detected and quantified by rMATS-turbo (Shen et al., 2014), with Ensemble v75 GRCH37 (hg19) GTF annotation. Exon inclusion levels, measured as PSI values, were calculated by

junction reads (reads spanning the splicing junctions) normalized by effective junction length. AS events with low junction read support (≤ 10 average junction reads, ≤ 10 total inclusion junction reads or ≤ 10 total skipping junction reads over all 18 samples), or with extreme PSI value ranges ($\text{PSI} \leq 0.05$ or ≥ 0.95 in all 18 samples) were excluded from downstream analysis. Differential exon skipping analysis was then performed using rMATS-turbo (with default parameter $-c 0.0001$) for five pair-wise comparisons between time point day 0 and other available time points (day 2, 4, 6, 8, 12). Exon skipping events for each comparison were considered differential if they meet the following criteria 1) >10 average junction reads (inclusion and skipping junction reads) in both groups; 2) does not have extreme PSI values ($\text{PSI} \leq 0.05$ or $\text{PSI} \geq 0.95$ for all 6 samples in the comparison); 3) $\text{FDR} < 0.01$; and 4) absolute change in PSI ($|\Delta\text{PSI}| > 0.05$). The significant event set for the whole osteogenic differentiation pathway was composed of events that were identified to be differentially spliced in at least 3 of the 5 comparisons. The background event set for the whole osteogenic differentiation pathway was defined as events with no significant change during MSPC osteogenic differentiation which meet the following cutoffs in all 5 comparisons: 1) > 10 average junction reads in both groups; 2) does not have extreme PSI values ($\text{PSI} \leq 0.05$ or $\text{PSI} \geq 0.95$ for all 6 samples in the comparison; and 3) $\text{FDR} > 0.5$.

Principle component analysis (PCA)

A total of 129 RBPs (including many well-characterized splicing factors) were curated from two different sources (Dominguez et al., 2018; Ray et al., 2013) and included in PCA analysis of RBP expression. For total gene or RBP expression, a pseudo-count of 1 was added to each TPM value before $\log_2$ transformation to avoid arithmetic error and large negative values. PCA was then performed after removing genes/RBPs/exon skipping events with no variance among samples. Samples were projected to their PC1-PC2 space by PCA score. LOESS (locally estimated scatterplot smoothing) regression lines with 95% confidence intervals were added to the PC1-PC2 plot using R package ggplot2 (v3.1.0).

Gene set enrichment analysis (GSEA)

GSEA (v3.0) software (Subramanian et al., 2005) was utilized on pre-ranked gene lists based on absolute values of principle component loadings from PCA. Ranked lists from total gene expression PCA included only the top 10,000 genes; for ranked lists from exon skipping, duplicated genes with lower rank were removed. All gene ontology gene sets (c5, v7.0, https://www.gsea-msigdb.org/gsea/msigdb/download_file.jsp?filePath=/msigdb/release/7.0/c5.all.v7.0.symbols.gmt) were used as gene sets database with 1000 permutations to calculate the enrichment score and p-values. Top gene ontology terms from the GSEA analysis served as input for REVIGO webserver (<http://revigo.irb.hr/>) to account for the semantic similarities and dispersibilities of gene sets.

Representative gene ontology terms were visualized in semantic similarity-based scatter plots after removing redundant ones.

Hierarchical clustering of time course datasets and heatmaps

Hierarchical clustering was performed on Z-score transformed TPMs of differentially expressed RBPs (from a total of 1542 RBPs) (Gerstberger et al., 2014) or PSI values of differentially spliced exons detected in the whole osteogenesis pathway as described above (hclust function in R package stats, v3.4.4).

Protein family domain analysis

Pfam domain scanning was conducted to search for potential functional domains affected by exon skipping events. Preprocessed Pfam annotation data, which maps HMM predicted high-quality Pfam-A domains to UCSC hg19 coordinates, were downloaded from the UCSC hg19 annotation database (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/ucscGenePfam.txt.gz>). Sequences of target exons or frameshifted downstream exons from differential exon skipping events were extracted and then scanned against the Pfam annotation data using bedtools (v2.25.0) (Quinlan and Hall, 2010).

RBP candidate screening method

189 RBP binding motifs with position weight matrix information for 129 RBPs (including many well-characterized splicing factors) were curated from two different sources and screened in this analysis. This includes 78 6-mer motifs for 78 RBPs from RNA Bind-n-Seq (RBNS) (Dominguez et al., 2018) and 111 7-mer motifs for 82 RBPs from RNAcompete (Ray et al., 2013).

Significant/background event lists and differentially expressed RBPs were identified as described before (see AS analysis section and gene expression analysis section). For each motif of the differentially expressed RBPs, significant exon skipping events were further assigned to two foreground event sets. They were composed of events ($n > 50$) whose PSI value was positively or negatively correlated ($R^2 > 0.5$, rlm function in R package MASS, v7.3-49) with differential RBP gene expression across different time points.

To identify region-specific RBP regulatory patterns for exon skipping events, we evaluated three regions around the alternatively spliced exons: 1) 300 nt of intronic sequence upstream of the target exon; 2) the exon body sequences; and 3) 300 nt of intronic sequence downstream the target exon. Scores for each motif were calculated by sliding window scanning of the position weight matrix at each possible binding

position. Region-specific motif occurrence was then determined by comparing the calculated motif scores with a threshold score (80% of the maximum PWM score). If there was any position with a calculated motif score $\geq$ the threshold score for a particular exon skipping event, then the motif occurrence was marked as “True” for this event in the corresponding region; otherwise it was marked “False”.

To determine whether a motif occurred in a specific region more often in foreground event sets than in the background event set, we used a one-tailed Fisher’s exact test to test the null hypothesis that the number of events with motif occurrence at a specific region was not different between the foreground (either positive or negative set) and the background event set. If an RBP motif was of significantly enriched occurrence ($p < 0.05$) in any region for any foreground event set, it was considered a key RBP for exon skipping regulation in the MSPC osteogenic differentiation process.

References

- Aprile, M., Cataldi, S., Ambrosio, M.R., D’Esposito, V., Lim, K., Dietrich, A., Bluher, M., Savage, D.B., Formisano, P., Ciccodicola, A., *et al.* (2018). PPARgammaDelta5, a Naturally Occurring Dominant-Negative Splice Isoform, Impairs PPARgamma Function and Adipocyte Differentiation. *Cell Rep* 25, 1577-1592 e1576.
- Augello, A., Kurth, T.B., and De Bari, C. (2010). Mesenchymal stem cells: a perspective from in vitro cultures to in vivo migration and niches. *Eur Cell Mater* 20, 121-133.
- Barak, Y., Nelson, M.C., Ong, E.S., Jones, Y.Z., Ruiz-Lozano, P., Chien, K.R., Koder, A., and Evans, R.M. (1999). PPAR gamma is required for placental, cardiac, and adipose tissue development. *Mol Cell* 4, 585-595.
- Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* 18, 437-451.
- Blume, S.W., and Curtis, J.R. (2011). Medical costs of osteoporosis in the elderly Medicare population. *Osteoporos Int* 22, 1835-1844.
- Bonnet, N., Conway, S.J., and Ferrari, S.L. (2012). Regulation of beta catenin signaling and parathyroid hormone anabolic effects in bone by the matricellular protein periostin. *Proc Natl Acad Sci U S A* 109, 15048-15053.
- Boonrungsiman, S., Gentleman, E., Carzaniga, R., Evans, N.D., McComb, D.W., Porter, A.E., and Stevens, M.M. (2012). The role of intracellular calcium phosphate in osteoblast-mediated bone apatite formation. *Proc Natl Acad Sci U S A* 109, 14170-14175.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34, 525-527.

- Chan, C.K.F., Gulati, G.S., Sinha, R., Tompkins, J.V., Lopez, M., Carter, A.C., Ransom, R.C., Reinisch, A., Wearda, T., Murphy, M., *et al.* (2018). Identification of the Human Skeletal Stem Cell. *Cell* *175*, 43-56 e21.
- Chen, Q., Shou, P., Zheng, C., Jiang, M., Cao, G., Yang, Q., Cao, J., Xie, N., Velletri, T., Zhang, X., *et al.* (2016). Fate decision of mesenchymal stem cells: adipocytes or osteoblasts? *Cell Death Differ* *23*, 1128-1139.
- Cieply, B., Park, J.W., Nakauka-Ddamba, A., Bebee, T.W., Guo, Y., Shang, X., Lengner, C.J., Xing, Y., and Carstens, R.P. (2016). Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins. *Cell Rep* *15*, 247-255.
- Ciuffreda, M.C., Malpasso, G., Musaro, P., Turco, V., and Gneccchi, M. (2016). Protocols for in vitro Differentiation of Human Mesenchymal Stem Cells into Osteogenic, Chondrogenic and Adipogenic Lineages. *Methods Mol Biol* *1416*, 149-158.
- Clark, M.D., Marcum, R., Graveline, R., Chan, C.W., Xie, T., Chen, Z., Ding, Y., Zhang, Y., Mondragon, A., David, G., *et al.* (2015). Structural insights into the assembly of the histone deacetylase-associated Sin3L/Rpd3L corepressor complex. *Proc Natl Acad Sci U S A* *112*, E3669-3678.
- Danilenko, M., Dalgliesh, C., Pagliarini, V., Naro, C., Ehrmann, I., Feracci, M., Kheirollahi-Chadegani, M., Tyson-Capper, A., Clowry, G.J., Fort, P., *et al.* (2017). Binding site density enables paralog-specific activity of SLM2 and Sam68 proteins in Neurexin2 AS4 splicing control. *Nucleic Acids Res* *45*, 4120-4130.
- Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A., *et al.* (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell* *70*, 854-867 e859.
- Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., Deans, R., Keating, A., Prockop, D., and Horwitz, E. (2006). Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* *8*, 315-317.
- Duchamp de Lageneste, O., Julien, A., Abou-Khalil, R., Frangi, G., Carvalho, C., Cagnard, N., Cordier, C., Conway, S.J., and Colnot, C. (2018). Periosteum contains skeletal stem cells with high bone regenerative potential controlled by Periostin. *Nat Commun* *9*, 773.
- Duque, G. (2008). Bone and fat connection in aging bone. *Curr Opin Rheumatol* *20*, 429-434.
- Dvinge, H., Kim, E., Abdel-Wahab, O., and Bradley, R.K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* *16*, 413-430.
- Erhard, F., Baptista, M.A.P., Krammer, T., Hennig, T., Lange, M., Arampatzi, P., Jürges, C.S., Theis, F.J., Saliba, A.E., and Dolken, L. (2019). scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* *571*, 419-423.

- Farini, D., Cesari, E., Weatheritt, R.J., La Sala, G., Naro, C., Pagliarini, V., Bonvissuto, D., Medici, V., Guerra, M., Di Pietro, C., *et al.* (2020). A Dynamic Splicing Program Ensures Proper Synaptic Connections in the Developing Cerebellum. *Cell Rep* 31, 107703.
- Feng, Y., and Walsh, C.A. (2004). The many faces of filamin: a versatile molecular scaffold for cell motility and signalling. *Nat Cell Biol* 6, 1034-1038.
- Feracci, M., Foot, J.N., Grellscheid, S.N., Danilenko, M., Stehle, R., Gonchar, O., Kang, H.S., Dalglish, C., Meyer, N.H., Liu, Y., *et al.* (2016). Structural basis of RNA recognition and dimerization by the STAR proteins T-STAR and Sam68. *Nat Commun* 7, 10355.
- Fleischer, T.C., Yun, U.J., and Ayer, D.E. (2003). Identification and characterization of three new components of the mSin3A corepressor complex. *Mol Cell Biol* 23, 3456-3467.
- Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* 15, 689-701.
- Gabut, M., Chaudhry, S., and Blencowe, B.J. (2008). SnapShot: The splicing regulatory machinery. *Cell* 133, 192 e191.
- Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat Rev Genet* 15, 829-845.
- Han, L., Wang, B., Wang, R., Gong, S., Chen, G., and Xu, W. (2019). The shift in the balance between osteoblastogenesis and adipogenesis of mesenchymal stem cells mediated by glucocorticoid receptor. *Stem Cell Res Ther* 10, 377.
- Huot, M.E., Vogel, G., Zabarauskas, A., Ngo, C.T., Coulombe-Huntington, J., Majewski, J., and Richard, S. (2012). The Sam68 STAR RNA-binding protein regulates mTOR alternative splicing during adipogenesis. *Mol Cell* 46, 187-199.
- Jaiswal, N., Haynesworth, S.E., Caplan, A.I., and Bruder, S.P. (1997). Osteogenic differentiation of purified, culture-expanded human mesenchymal stem cells in vitro. *J Cell Biochem* 64, 295-312.
- Johnson, C., Primorac, D., McKinstry, M., McNeil, J., Rowe, D., and Lawrence, J.B. (2000). Tracking COL1A1 RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *J Cell Biol* 150, 417-432.
- Kapinos, K.A., Fischer, S.H., Mulcahy, A., Hayden, O., and Barron, R. (2018). Medical Costs for Osteoporosis-Related Fractures in High-Risk Medicare Beneficiaries. *J Am Geriatr Soc* 66, 2298-2304.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37, 907-915.
- Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R.T., Gao, Y.H., Inada, M., *et al.* (1997). Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* 89, 755-764.

- Lasko, P. (2010). Tudor domain. *Curr Biol* 20, R666-667.
- Li, H., Cheng, Y., Wu, W., Liu, Y., Wei, N., Feng, X., Xie, Z., and Feng, Y. (2014). SRSF10 regulates alternative splicing and is required for adipocyte differentiation. *Mol Cell Biol* 34, 2198-2207.
- Litvin, J., Selim, A.H., Montgomery, M.O., Lehmann, K., Rico, M.C., Devlin, H., Bednarik, D.P., and Safadi, F.F. (2004). Expression and function of periostin-isoforms in bone. *J Cell Biochem* 92, 1044-1061.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Lu, X., Beck, G.R., Jr., Gilbert, L.C., Camalier, C.E., Bateman, N.W., Hood, B.L., Conrads, T.P., Kern, M.J., You, S., Chen, H., *et al.* (2011). Identification of the homeobox protein Prx1 (MHox, Prrx-1) as a regulator of osterix expression and mediator of tumor necrosis factor alpha action in osteoblast differentiation. *J Bone Miner Res* 26, 209-219.
- Makita, N., Suzuki, M., Asami, S., Takahata, R., Kohzaki, D., Kobayashi, S., Hakamazuka, T., and Hozumi, N. (2008). Two of four alternatively spliced isoforms of RUNX2 control osteocalcin gene expression in human osteoblast cells. *Gene* 413, 8-17.
- McManus, C.J., and Graveley, B.R. (2011). RNA structure and the mechanisms of alternative splicing. *Curr Opin Genet Dev* 21, 373-379.
- Meunier, P., Aaron, J., Edouard, C., and Vignon, G. (1971). Osteoporosis and the replacement of cell populations of the marrow by adipose tissue. A quantitative study of 84 iliac bone biopsies. *Clin Orthop Relat Res* 80, 147-154.
- Monroe, D.G., Hawse, J.R., Subramaniam, M., and Spelsberg, T.C. (2010). Retinoblastoma binding protein-1 (RBP1) is a Runx2 coactivator and promotes osteoblastic differentiation. *BMC Musculoskelet Disord* 11, 104.
- Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457-463.
- Norris, R.A., and Kern, M.J. (2001). The identification of Prx1 transcription regulatory domains provides a mechanism for unequal compensation by the Prx1 and Prx2 loci. *J Biol Chem* 276, 26829-26837.
- Otto, F., Thornell, A.P., Crompton, T., Denzel, A., Gilmour, K.C., Rosewell, I.R., Stamp, G.W., Beddington, R.S., Mundlos, S., Olsen, B.R., *et al.* (1997). *Cbfa1*, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* 89, 765-771.
- Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* 5, 389-396.
- Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am J Hum Genet* 102, 11-26.

- Pittenger, M.F., Mackay, A.M., Beck, S.C., Jaiswal, R.K., Douglas, R., Mosca, J.D., Moorman, M.A., Simonetti, D.W., Craig, S., and Marshak, D.R. (1999). Multilineage potential of adult human mesenchymal stem cells. *Science* 284, 143-147.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Rahman, M.A., Krainer, A.R., and Abdel-Wahab, O. (2020). SnapShot: Splicing Alterations in Cancer. *Cell* 180, 208-208 e201.
- Rauch, A., Haakonsson, A.K., Madsen, J.G.S., Larsen, M., Forss, I., Madsen, M.R., Van Hauwaert, E.L., Wiwie, C., Jespersen, N.Z., Tencerova, M., *et al.* (2019). Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. *Nat Genet* 51, 716-727.
- Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., *et al.* (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172-177.
- Reichert, M., Takano, S., von Burstin, J., Kim, S.B., Lee, J.S., Ihida-Stansbury, K., Hahn, C., Heeg, S., Schneider, G., Rhim, A.D., *et al.* (2013). The Prrx1 homeodomain transcription factor plays a central role in pancreatic regeneration and carcinogenesis. *Genes Dev* 27, 288-300.
- Robey, P.G., and Riminucci, M. (2020). Chapter 2 - Skeletal stem cells: Tissue-specific stem/progenitor cells of cartilage, bone, stroma, and marrow adipocytes. In *Principles of Bone Biology* (Fourth Edition), J.P. Bilezikian, T.J. Martin, T.L. Clemens, and C.J. Rosen, eds. (Academic Press), pp. 45-71.
- Rosen, E.D., Sarraf, P., Troy, A.E., Bradwin, G., Moore, K., Milstone, D.S., Spiegelman, B.M., and Mortensen, R.M. (1999). PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. *Mol Cell* 4, 611-617.
- Scotti, M.M., and Swanson, M.S. (2016). RNA mis-splicing in disease. *Nat Rev Genet* 17, 19-32.
- Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* 111, E5593-5601.
- Shimojo, H., Sano, N., Moriwaki, Y., Okuda, M., Horikoshi, M., and Nishimura, Y. (2008). Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. *J Mol Biol* 378, 987-1001.
- Soleimani, M., and Nadri, S. (2009). A protocol for isolation and culture of mesenchymal stem cells from mouse bone marrow. *Nat Protoc* 4, 102-106.
- Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4, 1521.

- Song, Y., Botvinnik, O.B., Lovci, M.T., Kakaradov, B., Liu, P., Xu, J.L., and Yeo, G.W. (2017). Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Mol Cell* 67, 148-161 e145.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545-15550.
- Takano, S., Reichert, M., Bakir, B., Das, K.K., Nishida, T., Miyazaki, M., Heeg, S., Collins, M.A., Marchand, B., Hicks, P.D., *et al.* (2016). Prrx1 isoform switching regulates pancreatic cancer invasion and metastatic colonization. *Genes Dev* 30, 233-247.
- Traunmuller, L., Gomez, A.M., Nguyen, T.M., and Scheiffele, P. (2016). Control of neuronal synapse specification by a highly dedicated alternative splicing program. *Science* 352, 982-986.
- Vernia, S., Edwards, Y.J., Han, M.S., Cavanagh-Kyros, J., Barrett, T., Kim, J.K., and Davis, R.J. (2016). An alternative splicing program promotes adipose tissue thermogenesis. *Elife* 5.
- Wagey, R., and Short, B. (2013). Isolation, enumeration, and expansion of human mesenchymal stem cells in culture. *Methods Mol Biol* 946, 315-334.
- Wang, J., Saraswat, D., Sinha, A.K., Polanco, J., Dietz, K., O'Bara, M.A., Pol, S.U., Shayya, H.J., and Sim, F.J. (2018). Paired Related Homeobox Protein 1 Regulates Quiescence in Human Oligodendrocyte Progenitors. *Cell Rep* 25, 3435-3450 e3436.
- Wang, Q., Forlino, A., and Marini, J.C. (1996). Alternative splicing in COL1A1 mRNA leads to a partial null allele and two In-frame forms with structural defects in non-lethal osteogenesis imperfecta. *J Biol Chem* 271, 28617-28623.
- Wilsker, D., Probst, L., Wain, H.M., Maltais, L., Tucker, P.W., and Moran, E. (2005). Nomenclature of the ARID family of DNA-binding proteins. *Genomics* 86, 242-251.
- Wolock, S.L., Krishnan, I., Tenen, D.E., Matkins, V., Camacho, V., Patel, S., Agarwal, P., Bhatia, R., Tenen, D.G., Klein, A.M., *et al.* (2019). Mapping Distinct Bone Marrow Niche Populations and Their Differentiation Paths. *Cell Rep* 28, 302-311 e305.
- Wu, R.C., Jiang, M., Beaudet, A.L., and Wu, M.Y. (2013). ARID4A and ARID4B regulate male fertility, a functional link to the AR and RB pathways. *Proc Natl Acad Sci U S A* 110, 4616-4621.
- Xia, X.Y., Cui, Y.X., Huang, Y.F., Pan, L.J., Yang, B., Wang, H.Y., Li, X.J., Shi, Y.C., Lu, H.Y., and Zhou, Y.C. (2008). A novel RNA-splicing mutation in COL1A1 gene causing osteogenesis imperfecta type I in a Chinese family. *Clin Chim Acta* 398, 148-151.
- Yee, B.A., Pratt, G.A., Graveley, B.R., Van Nostrand, E.L., and Yeo, G.W. (2019). RBP-Maps enables robust generation of splicing regulatory maps. *RNA* 25, 193-204.

Zhang, D., Hu, Q., Liu, X., Ji, Y., Chao, H.P., Liu, Y., Tracz, A., Kirk, J., Buonamici, S., Zhu, P., *et al.* (2020). Intron retention is a hallmark and spliceosome represents a therapeutic vulnerability in aggressive prostate cancer. *Nat Commun* 11, 2089.

Zhang, J., and Manley, J.L. (2013). Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov* 3, 1228-1237.

Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F., *et al.* (2016). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell* 166, 1147-1162 e1115.

Zhong, L., Yao, L., Tower, R.J., Wei, Y., Miao, Z., Park, J., Shrestha, R., Wang, L., Yu, W., Holdreith, N., *et al.* (2020). Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *Elife* 9.

Figure 1. Extensive transcriptomic alterations characterize the induced stepwise differentiation of MSPC to osteoblasts.

(A) Photographs of blue alkaline phosphatase and red hydroxyapatite (bone calcification) staining of MSPCs cultured in osteoblast differentiation medium over 10 days.

(B) Measurement of optical density in photographed wells in panel A.

(C) From-left-to-right are normalized expression measures from RNA-seq at indicated time points, showing a temporal decrease in immature marker (*MYC* and *SOX9*) expression and increase in osteoblast marker (*RUNX2* and *ALPL*) expression. Error bars represent the MEAN±SEM (n=3).

(D) From left-to-right, principal component analysis (PCA) plots of total gene expression (**left**), RBP gene expression (**middle**), and exon skipping (**right**). Samples were projected to the space of the first two principal components (PCs) with percentage of variation explained shown in x- and y-axis labels. Local regression lines are added by LOESS (locally estimated scatterplot smoothing) method, with 95% confidence intervals shown in grey.

Z-score transformed PSI values for 604 significantly changed exon skipping events; panel G, line plots revealing the patterns of change in PSI value for corresponding clusters in panel F. Red line represents the median value for each cluster.

Supplementary Figure 2. Exon 4 inclusion/exclusion of *PRRX1* results in an isoform switch.

(A) Gene structure representation of the *PRRX1* gene. Bars and dashed lines represent exons and introns, respectively. Untranslated regions (UTR) are denoted in grey. Two stop codons residing in exon 4 and 5 are depicted as STOP signs. The OAR (otp, aristaless, and rax) domain is located in the green region in exon 5.

(B) Protein structure representation of the two *PRRX1* isoforms. The top PRRX1b isoform has exon 4 incorporated and the bottom PRRX1a isoform has exon 4 excluded. The stop codon in exon 4 of the inclusion isoform renders exon 5 and OAR domain untranslated.

(C) Sashimi plot showing AS changes of the *PRRX1* gene on day 0 (upper panel) and day12 (lower panel). The black bars and dashed lines on the bottom represent exons and introns, respectively. Solid peaks represent RPKM of reads mapped to each region. Arches represent splice junctions and the numbers represent number of reads mapped to each splice junction.

GeneID	geneSymbol	PFAM	frame shift	NMD	PSI _{00-PSI₁₂}	FDR(day0-day12)	chrom	strand	exon start	exon end	upstreamES	upstreamEE	downstreamES	downstreamEE
ENSG00000039560	RAI14	-	NO	-	0.168606812	0	chr5	+	34813678	34813765	34812284	34812313	34814687	34814774
ENSG00000116132	PRRX1	-	NO	-	0.21618986	1.19E-13	chr1	+	170699417	170699489	170695360	170695542	170705188	170705374
ENSG00000122034	GTF3A	zf_C2H2	YES	-	0.097629224	2.21E-13	chr13	+	28006867	28006941	28004669	28004758	28008275	28008287
ENSG00000001167	NFYA	-	NO	-	-0.267664243	3.48E-11	chr6	+	41048549	41048636	41046767	41046903	41051784	41051931
ENSG00000068903	SIRT2	-	YES	-	0.154433978	6.01E-08	chr19	-	39389018	39389065	39384458	39384507	39390145	39390351
ENSG00000108312	UBTF	HMG_box	NO	-	0.171796905	1.37E-07	chr17	-	42289711	42289822	42289240	42289374	42290186	42290307
ENSG00000157216	SSBP3	SSDP	NO	-	0.092628716	2.68E-07	chr1	-	54723741	54723822	54722799	54722859	54747110	54747200
ENSG00000135111	TBX3	T-box	NO	-	0.27274122	6.55E-07	chr12	-	115117717	115117777	115117309	115117456	115118683	115118951
ENSG00000125945	ZNF436	-	NO	-	0.350190291	1.01E-06	chr1	-	23694465	23694558	23693534	23693661	23695858	23695935
ENSG00000006455	EYA2	-	YES	-	0.218708723	1.97E-06	chr20	+	45700823	45700891	45644819	45644936	45717877	45718020
ENSG00000166888	STAT6	STAT_int*	YES	NMD*	0.09131425	2.06E-06	chr12	-	57503825	57503893	57501945	57502082	57504911	57505157
ENSG00000153786	ZDHC7	-	NO	-	-0.192762388	3.03E-06	chr16	-	85022368	85022479	85015475	85015600	85023909	85024241
ENSG000000082641	NFE2L1	-	NO	-	-0.136495916	3.44E-06	chr17	+	46134393	46134483	46133747	46133960	46134705	46134864
ENSG00000113649	TCERG1	FF	YES	NMD	-0.16991098	6.02E-06	chr5	+	145889629	145889723	145888707	145888808	145890003	145890328
ENSG00000005801	ZNF195	KRAB*	YES	-	0.276500981	6.31E-06	chr11	-	3394806	3394886	3392806	3392933	3400267	3400355
ENSG00000102103	PQB1	-	YES	-	-0.127281594	1.30E-05	chrX	+	48759495	48759794	48759206	48759319	48760008	48760072
ENSG00000198453	ZNF568	-	NO	-	-0.263769264	1.58E-05	chr19	+	37413487	37413748	37408480	37408550	37416101	37416160
ENSG00000074219	TEAD2	TEA	NO	-	-0.237218087	1.65E-05	chr19	-	49859215	49859227	49858559	49858676	49860508	49860571
ENSG00000196498	NCOR2	-	NO	-	-0.093681883	2.67E-05	chr12	-	124825147	124825240	124824839	124824989	124826368	124826501
ENSG00000139546	TARBP2	dsrm	YES	NMD*	-0.119554016	3.19E-05	chr12	+	53895843	53895968	53895089	53895245	53896810	53896996
ENSG00000006455	EYA2	-	YES	-	0.191078217	4.14E-05	chr20	+	45702796	45702974	45644819	45644936	45717877	45718020
ENSG00000139546	TARBP2	dsrm	YES	NMD*	-0.06649497	6.33E-05	chr12	+	53895798	53895968	53894704	53895245	53896810	53896913
ENSG00000101849	TBL1X	LisH*	YES	-	-0.273448155	6.54E-05	chrX	+	9621584	9621729	9608312	9608400	9622254	9622328
ENSG00000198176	TFDP1	DP	YES	-	0.075586604	8.95E-05	chr13	+	114292132	114292211	114290848	114291015	114294434	114294999
ENSG00000005801	ZNF195	-	NO	-	0.194226144	0.000115099	chr11	-	3394806	3394905	3392806	3392933	3400267	3400355
ENSG00000005889	ZFX	Zfx_Zfx_act*	YES	-	0.149878128	0.000137914	chrX	+	24193505	24193560	24190872	24190917	24197299	24197667
ENSG00000102103	PQB1	-	YES	-	-0.08218082	0.000148055	chrX	+	48759509	48759661	48759206	48759319	48760008	48760072
ENSG000000085274	MYNN	BTB*	YES	-	0.183656738	0.000159271	chr3	+	169491818	169491885	169491214	169491250	169492052	169492349
ENSG00000198176	TFDP1	DP	YES	-	0.090569578	0.000308031	chr13	+	114292132	114292199	114290848	114291015	114294434	114294713
ENSG00000115207	GTF3C2	-	YES	-	0.129756297	0.000350349	chr2	-	27573180	27573499	27566197	27566445	27579605	27579866
ENSG00000111653	ING4	ING	NO	NMD	-0.060823976	0.000376274	chr12	-	6764803	6765079	6762395	6762562	6765892	6765964
ENSG00000100084	HIRA	-	NO	-	-0.06751115	0.000488248	chr22	-	19365391	19365589	19363153	19363315	19371142	19371228
ENSG00000134852	CLOCK	-	YES	-	-0.135910675	0.000584389	chr4	-	56376078	56376232	56355540	56355632	56412648	56412813
ENSG00000177981	ASB8	Ank_3	NO	-	0.067965541	0.00060343	chr12	-	48543869	48543923	48543311	48543781	48544983	48545088
ENSG00000116731	PRDM2	-	NO	-	0.203212682	0.000658775	chr1	+	14099572	14099683	14075855	14075982	14142921	14143065
ENSG00000090060	PAPOLA	-	NO	-	-0.054834451	0.000925408	chr14	+	97026985	97027048	97022511	97022750	97029155	97029230
ENSG000000054267	ARID4B	Tudor-knot	NO	-	0.173855012	0.000945713	chr1	+	235377083	235377341	235359345	235359430	235383107	235383283
ENSG00000112983	BRD8	-	YES	-	-0.057979735	0.001008484	chr5	-	137499775	137499822	137498818	137499033	137500008	137500102
ENSG00000187079	TEAD1	TEA	NO	-	-0.205878924	0.001083303	chr11	+	12900435	12900447	12886384	12886447	12901254	12901389
ENSG00000179094	PER1	-	NO	-	-0.073428664	0.001235771	chr17	-	8049275	8049455	8047372	8048311	8049689	8049806
ENSG00000106459	NRF1	Nrf1_DNA-bind	YES	-	-0.105166536	0.001602051	chr7	+	129311268	129311383	129297182	129297414	129317471	129317598
ENSG00000185024	BRF1	TFIIB	NO	-	-0.126536098	0.001684363	chr14	-	105707601	105707751	105695156	105695250	105718843	105718916
ENSG00000198176	TFDP1	DP	YES	-	0.139016818	0.0018896	chr13	+	114291934	114292211	114290848	114291015	114294434	114294625
ENSG00000102081	FMR1	-	NO	-	0.162169694	0.002627816	chrX	+	147019617	147019680	147018984	147019119	147022094	147022181
ENSG00000171456	ASXL1	ASXH	NO	-	-0.145845325	0.003703502	chr20	+	31017703	31017856	31017140	31017234	31019123	31019287
ENSG00000186448	ZNF197	KRAB*	YES	-	-0.086552457	0.003889162	chr3	+	44673596	44673688	44672553	44672713	44673964	44674091
ENSG00000137309	HMGAI1	-	NO	-	-0.055388456	0.004489747	chr6	+	34204980	34205094	34204649	34204740	34208513	34208659
ENSG00000100429	HDAC10	Hist_deacetyl	NO	-	-0.235606061	0.005589482	chr22	-	50687756	50687945	50687531	50687597	50688063	50688132
ENSG00000111206	FOXK1	-	NO	-	0.072458342	0.010560464	chr12	-	2970464	2970578	2966848	2966829	2973485	2973661
ENSG000000067066	SP100	SAND	NO	-	0.168979088	0.023504136	chr2	+	231372707	231372746	231371017	231371160	231375839	231375881
ENSG00000152784	PRDM8	-	YES	-	0.203431373	0.039352881	chr4	+	81115266	81115393	81112625	81112723	81117166	81117278
ENSG00000138785	INTS12	-	NO	-	0.113845432	0.085586846	chr4	+	106624042	106624180	106621006	106621171	106624804	106624966
ENSG00000134982	APC	Arm	YES	-	-0.017115841	0.825678682	chr5	+	112170647	112170862	112164531	112164669	112173249	112173419

* affected PFAM/NMD is related to frame-shifted downstream exon (downstream exon which is frame-shifted by inclusion/exclusion of target exon).

Supplementary Table 1. Transcription factors with exon skipping events are often affected by frame shift and/or disruption of functional domains. Shown are transcription factors with significant exon skipping changes during osteogenic differentiation. Column 3 reports the protein family domain (Pfam) affected by inclusion/exclusion of target exon, due to the presence/absence of the domain, encoded by the target exon or the downstream exon (*) which is frame-shifted by incorporation/exclusion of target exon. Column 4 indicates whether the target exon incorporation results in a frameshift. Column 5 indicates whether the target exon or the frameshifted downstream exon (*) is involved in nonsense mediated decay.