Osteogenesis is driven by a kinase and alternative splicing supernetwork

Results

**Temporal profiling of the dynamic transcriptome during osteogenesis.** To detect changes in transcript expression and LSVs during osteogenic differentiation, we used the hMSC-TERT4 cell line, which was previously validated ⚠️ as a faithful and potent model for bone-marrow derived stromal cells. To obtain temporal mRNA expression profiles, we induced hMSC-TERT4 cells to differentiate (see methods) and conducted paired-end RNA-Seq at timepoints 0, 6, 12 hrs, as well as 1, 3, 6, 9 and 12 days post-induction. Alternative splicing (AS) and differential splicing patterns were detected and quantified according to Percent Spliced-In (PSI, ψ) values reported by JUM⚠️ and PSI-Sigma⚠️. JUM uses an event-centric model of alternative splicing where LSVs, termed “AS structures”, are classified as any one of the 5 classic splice modes: alternative 3’ splice site (A3SS), alternative 5’ splice site (A5SS), cassette exon, mutually exclusive exon (MXE) or intron retention (IR) events. In addition, JUM reports a “composite” category which may be any combination of the 5 classic splice modes. On the other hand, PSI-Sigma natively quantifies multiple PSI values for every exon, reflecting membership of one exon with multiple LSVs. We summed up all PSI values for each unique exon to obtain one exon-centric PSI value per exon. We noted there were negligible differences in results compared to using LSV-centric PSI values (supplementary figure ❓ ). Overall, 12693👽 unique AS structures (corresponding to ❗splice junctions from ❗ unique genes, 1.83 👽 unique AS structures per gene) were detected by JUM and 43895 👽 exons (from ❗ unique genes, 5.68 👽 unique AS structures per gene) were detected by PSI-Sigma. Of these, 1062 👽 AS structures (from 983 👽 unique genes, 1.08 👽 unique AS structures per gene) and 3272 👽 exons (from 1899 👽 unique genes, 1.72 👽 unique exons per gene) were found to be differentially spliced, with significant changes in PSI levels between any two timepoints. For these differential splicing events, JUM detected 115 A3SS, 310 A5SS, 234 cassette exon, 1 intron retention and 402 composite events 👽 👽 👽 👽 👽 while PSI-Sigma detected 442 A3SS, 618 A5SS, 688 IR, 238 MES, 241 MXS and 1045 SES events 👽 👽 👽 👽 👽(figure ❓ ). There were 7594 👽 unique AS structures (from ❗ unique genes, 1.83 👽 unique AS structures per gene) and 41466 👽 unique exons exons (from ❗ unique genes, 3.23 👽 unique exons per gene) found to be constitutive, meaning there were no significant changes over all pairwise comparisons between timepoints. Of those differentially spliced, 448 👽 unique AS structures (from 430 👽 unique genes) and 1048 👽 exons (from ❗ genes) were found to undergo junction or exon switching over the time course, defined as a change in the PSI ranking of constituent events (JUM) or exons (PSI-Sigma) (figure ❓ ).

JUM and PSI-Sigma were comparable in their detection of AS events, as 6344 👽 AS structures or exons (32.3% 👽 of those in JUM, 10.5% 👽 of those in PSI-Sigma) detected by both tools were also found to be overlapped or adjacent. Of those identified as differentially spliced in both tools, 343 👽 AS structures or exons (32.3% 👽 of all differential in JUM, 10.5% 👽 of all differential in PSI-Sigma) were found to be overlapped or adjacent. When we considered the AS events only occurring in genes with AS detected by both tools, these percentages increased to 74.9% 👽 of all JUM AS structures and 19.6% 👽 of all PSI-Sigma exons, while the increase was more dramatic for differentially spliced events, at 82.6% 👽 of JUM AS structures and 44.4% 👽 of PSI-Sigma exons (figure ❓ ).

PCA analysis??? ❓❓

**Alternative splicing is coordinated at both the gene and transcript level.** We hypothesized that there exist concerted changes in AS patterns, hence AS could be over-represented in genes which play important roles in osteoblast fate selection. This hypothesis appears to be well supported by the spliceomic RNA-Seq data, as we observed that longer parent transcript lengths ⚠️ were not associated with a greater number of differential junctions or exons per gene, suggesting that AS patterns are likely to be non-random (figure ❓ ). To explore the functional ontology of genes exhibiting differential splicing over the time-series, we performed Gene Set Enrichment Analysis (GSEA) in KEGG pathways, PFAM protein families and GO Terms. We found 398 👽 genes had differential JUM AS structures when comparing only the first (0h) and last (12d) timepoints, and these were enriched for the KEGG pathways “Vascular smooth muscle contraction” and “Focal adhesion” (figure ❓ ). The common functional link between the two pathways appeared to be *KCNMA1*, which encodes a mechanosensitive potassium ion channel with known roles in the context of osteoblast differentiation involving integrin signalling pathways⚠️⚠️. On the other hand, all 983 genes with differential AS structures over the time course were enriched for the “Focal adhesion” and “ECM-receptor interaction” pathways. For the PSI-Sigma gene set, there were 253 genes with differential exons when only the first and last time points were compared, however, no relevant nor significant KEGG pathway enrichment was observed. When all 1899 genes with differential exons were enriched, a few pathways relevant to osteoblast differentiation was observed (figure ❓ ) in addition to multiple distinct groups of genes (supplementary figure ❓). We noticed that many genes encoding protein kinases were differentially spliced in both the JUM and PSI-Sigma gene sets. We further explored this observation by conducting GSEA of all differentially spliced genes across the time-series for PFAM protein families. Surprisingly, amongst the 30 most significant enrichments, the greatest number of hits belonged to protein kinase domain as well as RNA recognition motif (RRM) domain containing proteins, and this result was consistent for both JUM and PSI-Sigma(figure ❓ ). Proteins of the RRM domain-containing family are well characterised RNA-binding proteins (RBPs) which play important roles in the regulation of the spliceosome and AS by acting as splicing factors (SFs)⚠️. In addition to the RRM and kinase domains, the zinc finger C−x8−C−x5−C−x3−H type family of SFs, calponin homology domain containing proteins (associated with actin cytoskeleton binding) ⚠️ and the lysyl oxidase family (associated with collagen crosslinking) ⚠️ were also enriched in the JUM gene set. These enrichments for PFAM families were reflected in the GO term analysis, where we observed “mRNA splicing, via spliceosome” (BP), “protein kinase activity” (MF), “actin binding” (MF) and “extracellular matrix organization” (BP) to be significantly enriched (figure ❓ ). Notably, significant enrichments for the GO terms “cell cycle” (BP), “transcription coregulator activity” (MF) and “chromatin remodeling” (BP) suggest that changing patterns in AS may modulate lineage restriction and specification. Additionally, we identified many more classes of genes relevant to osteoblast differentiation including K+/Ca2+/Cl– ion channels, lysosome/golgi components and members of the Rho-GTPase signalling network (supplementary table ❓ ). Together, this evidence suggests that there are coordinated changes in AS across the transcriptome which are notably include genes encoding SFs and kinases.

AS changes relative usage of transcript regions rather than overall gene expression, hence junction or exon ontology analyses allow greater insight into the regional regulation of AS compared to standalone gene ontology analyses ⚠️. To that end, we sought to understand whether AS occurred in certain transcript regions. Extracting the differential regions from each differential AS junction structure from JUM as well as differential exons from PSI-Sigma, we matched each of these regions to their counterpart region or exon in the Ensembl reference annotation ⚠️. For the JUM data, 976 AS structures (❗% of all differential AS structures detected) were able to be matched to the reference. For the PSI-Sigma data, 2570 exons had a reference counterpart (❗% of all differential exons detected). Of the JUM differential regions matched to the reference, 95.6% matched to a protein-coding transcript, 78.4% to a CDS region, 45.2% to a lncRNA transcript, 52.9% to a 5’ UTR region and 37.2% to a 3’ UTR region. On the other hand, out of the PSI-Sigma differential exons matched to the reference, 72.6% matched to a protein-coding transcript, 64.6% to a CDS region, 30.5% to an lncRNA transcript, 31.1% to a 5’ UTR region and 21.2% to a 3’ UTR region. (figure ❓ ) ❌ ❌ ❌ (These look wrong. They are NOT out of those matched, but rather of all exons in total. FIX THIS.). Notably the ranking of transcript region types matched by differential regions were identical for JUM and PSI-Sigma, confirming the robustness of both tools. Differential regions did not match to reference miRNA transcripts in either tool. By shuffling the positions of the differential regions in a permutation test with 5000 iterations, we found that the differential regions were significantly over-represented for the 5’ UTR region compared to random for both tools ❓❗❗❗. This result suggests that changes in AS occur in non-random transcript regions.

Nonsense-mediated decay (NMD) naturally occurring process in the mammalian cell whereby ‘junk’ transcripts containing a Premature Termination Codon (PTC) are degraded by XRN after its recruitment by the NMD complex⚠️. It was previously reported that the number of transcripts harbouring PTCs and hence subject to NMD could be very high⚠️, up to ❗ of transcripts in humans compared to the reference value of 6.95% 👽 (⚠️cite ensembl). To explore whether transcripts subject to NMD (or NMD Transcripts, NMDTs) comprised a major proportion of our reported differential splicing events, we considered whether differential junctions or exons were unique to known NMDTs in the reference transcriptome (Ensembl v 👽). We also supplemented the reference annotation with our reconstructed hMSC-TERT4 transcriptome by predicting NMD transcripts *de novo* using the “51-nt rule” (see Methods). The “51-nt rule”, previously described elsewhere, was reported to yield an optimum balance between false positives and false negatives ⚠️.

Transcript truncations due to AS may be translated into functionally significant protein region truncations⚠️. Accordingly, we reasoned that differentially spliced regions may be overrepresented in areas which encode for functionally significant protein domains or regions. After re-matching the differential regions to annotated CDS regions, differential CDS-relative coordinates were converted to protein-relative coordinates and matched with protein domain annotations obtained from Biomart ⚠️. We only considered overlaps with protein domains (Interpro⚠️), Low Complexity Regions (LCRs) (SEG algorithm⚠️), Intrinsically Disordered Regions (IDRs) (MobiDB-lite⚠️), coiled-coiled domains (ncoils⚠️), signal peptide cleavage sites (SignalP ⚠️),

**Proteomic validation of splice junctions and exons.** negatively regulate osteoblast differentiation, we conducted an RNA interference screen of 719 kinase genes, ❗ of which encode

**Specificity of alternative splicing is achieved by a multilayered upstream network of RNA binding proteins and kinases.**

Gene expression was quantified using RSEM⚠️ and differential gene expression analysis was done using the RUV/EdgeR workflow⚠️. Overall, we detected a total of ❗ genes, of which❗ had significant change in expression between any two timepoints.

**Kinase signalling networks are extensively modulated by alternative splicing.**

To determine whether kinases positively or negatively regulate osteoblast differentiation, we conducted an RNA interference screen of 719 kinase genes, ❗ of which encode for protein kinases and ❗ encoding for small molecule kinases.

**Reconstruction of the kinase and splicing regulatory networks in the differentiating osteoblast.** fddd

???????????????????

**Single cell alternative splicing patterns reflect a defined osteogenic trajectory.** Fddd

**Aberrant splicing factor binding is associated with osteoblastic disease.** Fddd