Supplemental Material

Phosphoproteomic profiling reveals a defined genetic program for osteoblastic lineage commitment of human bone marrow derived stromal stem cells

Inigo Barrio-Hernandez[1], Abbas Jafari[2,3], Kristoffer T. G. Rigbolt[1], Philip Hallenborg[1], Virginia Sanchez-Quiles[1], Ida Skovrind[1], Vyacheslav Akimov[1], Irina Kratchmarova[1], Joern Dengjel[4], Moustapha Kassem[2,3], Blagoy Blagoev[1,*]

Table of Content:

Supplemental Figures

Supplemental Tables are provided as separate Excel files

Supplemental Methods

Supplemental References

Supplemental Figure S1
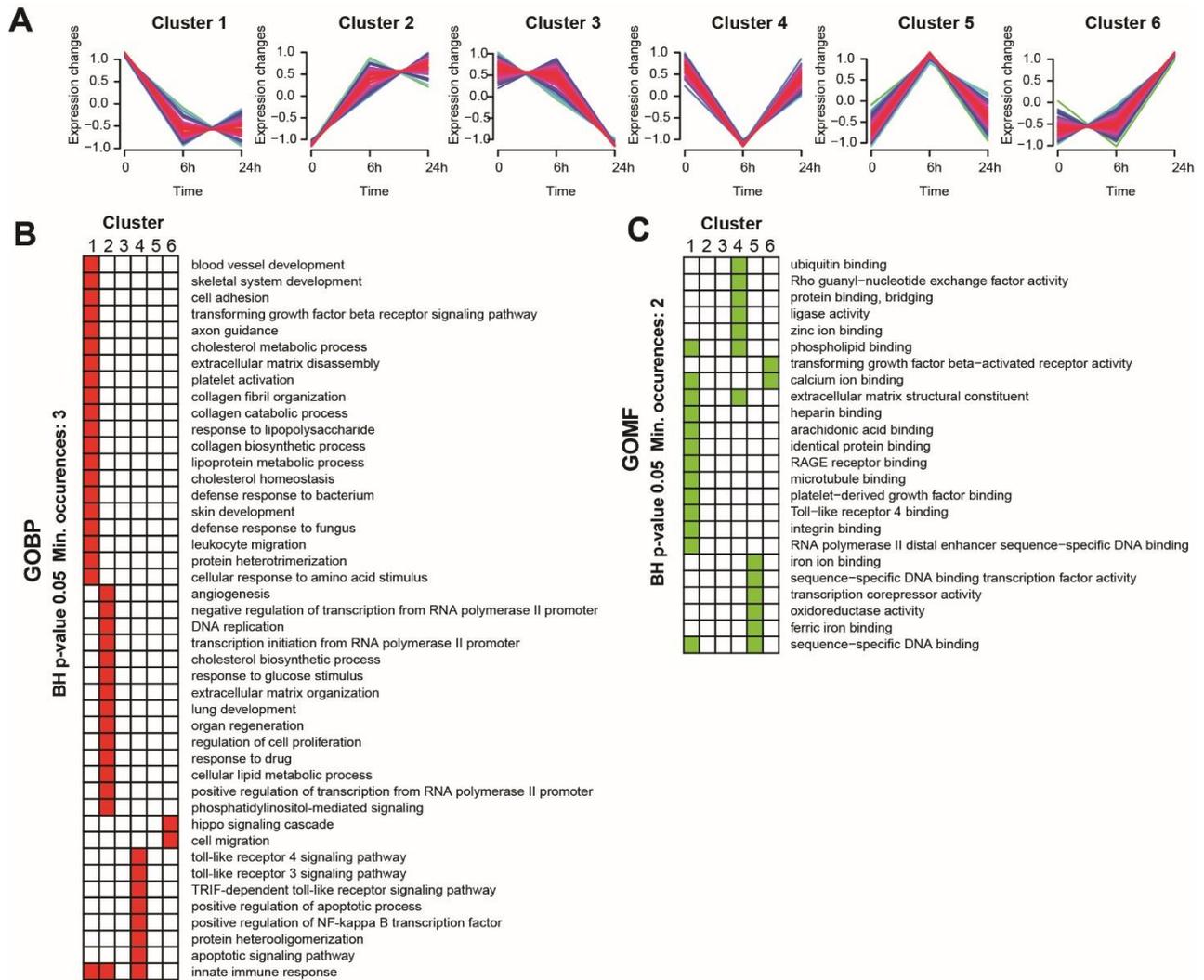
Supplemental Figure S1. Protein expression changes associated with osteoblast commitment of hMSCs. (A) Clustering of proteins regulated after 6 or 24 hours after induction of osteoblastic differentiation, using fuzzy c-means. (B, C) Significantly changed GO terms Biological Processes (B) and Molecular Functions (C). Only terms, that are significant, are colored.

Supplemental Figure S2
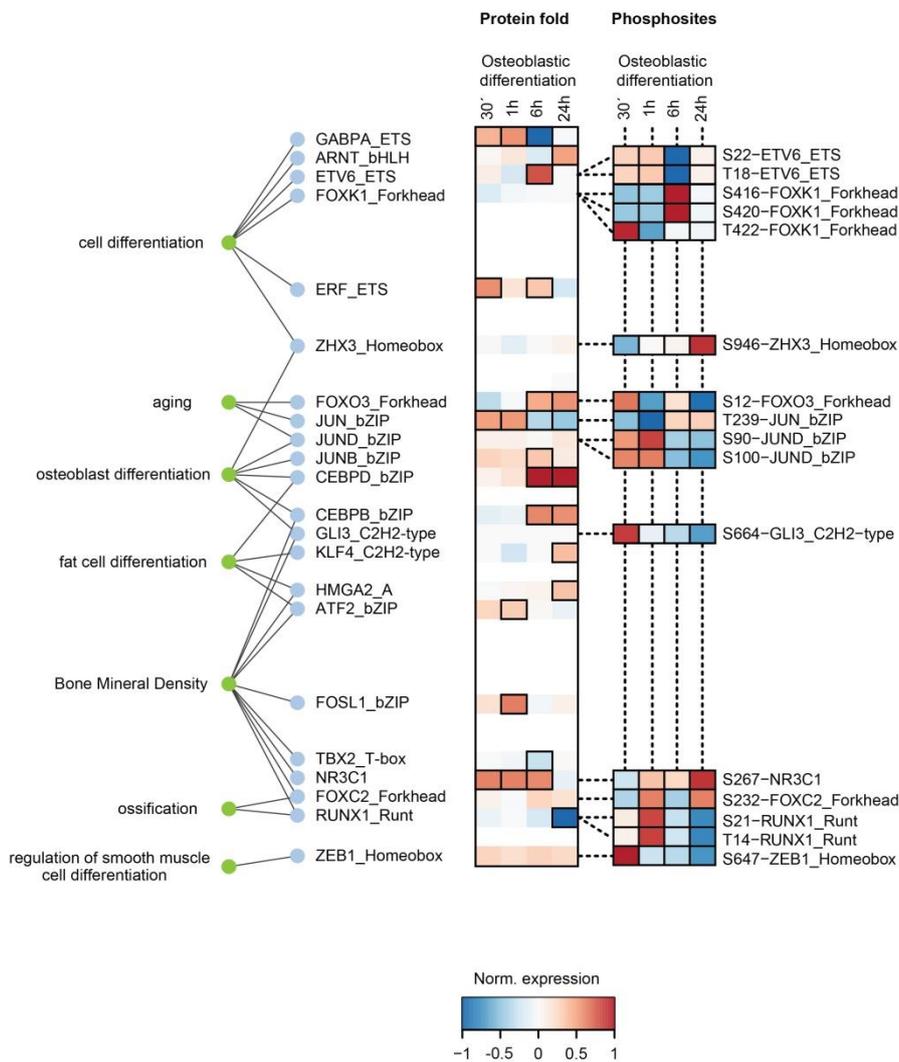
**Supplemental Figure S2. Osteoblastic differentiation induces changes in protein level and phosphorylation status of transcription factors associated with cellular differentiation and bone function.** Heat map showing regulation at protein and phosphorylation levels during the first 24 hours of osteoblast differentiation of quantified transcription factors related to commitment or bone homeostasis.

Supplemental Figure S3

A

B

C
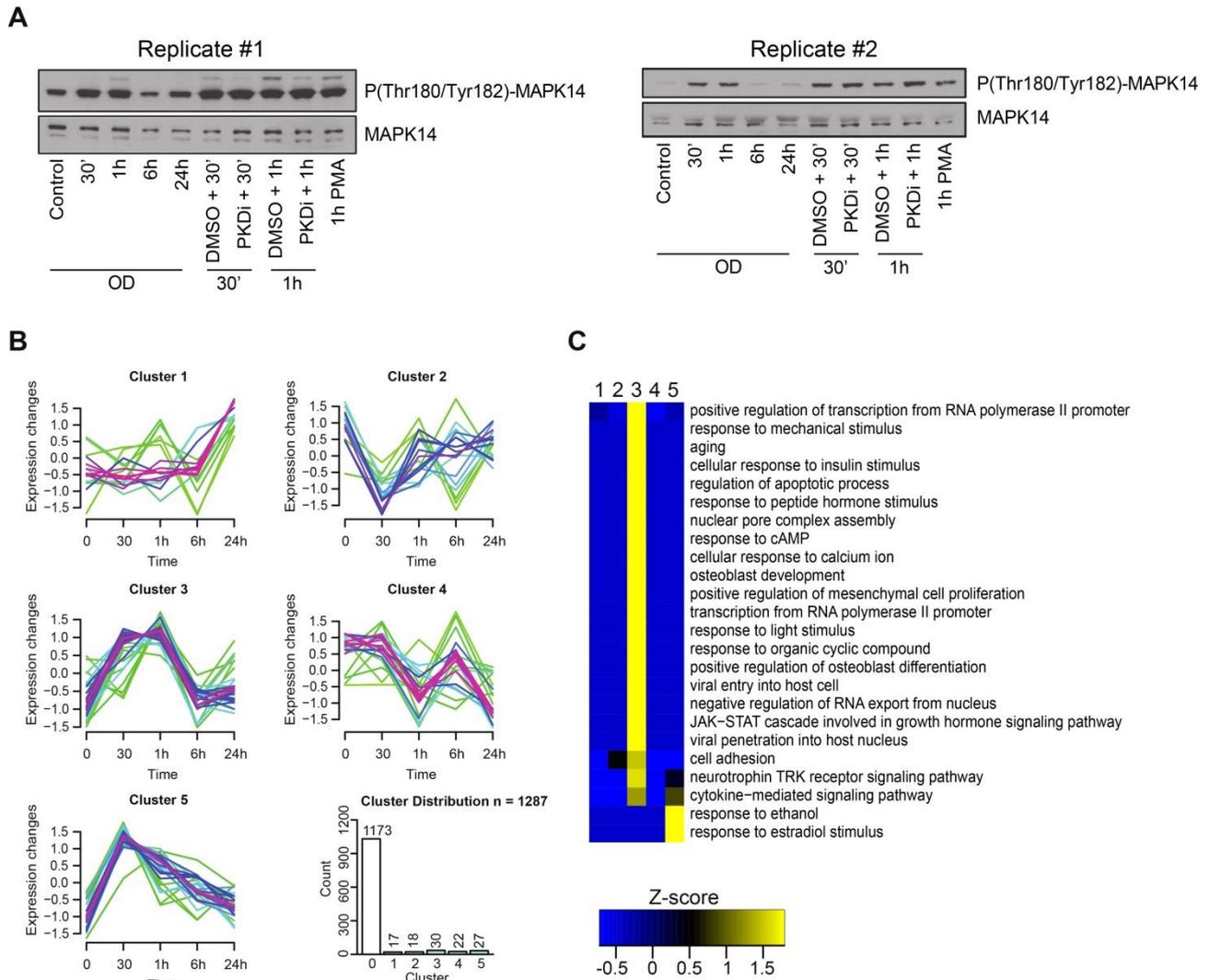
**Supplemental Figure S3. MAPK14 is not a direct target downstream of PRKD1 during the first hour of osteoblast lineage commitment.** (A) Western blot of total and phosphorylated MAPK14 during osteoblast differentiation and with or without inclusion of the PRKD1 inhibitor (PKDi). PMA was included as a positive control for activation of MAPK14. (B) Clustering output of the sites predicted to be phosphorylated by MAPK14. (C) Gene Ontology Biological Processes enrichment of sites predicted to be phosphorylated by MAPK14 and showing regulation, using unregulated ones as background. Significance is shown with a standardized value where blue means none significance.

Supplemental Figure S4
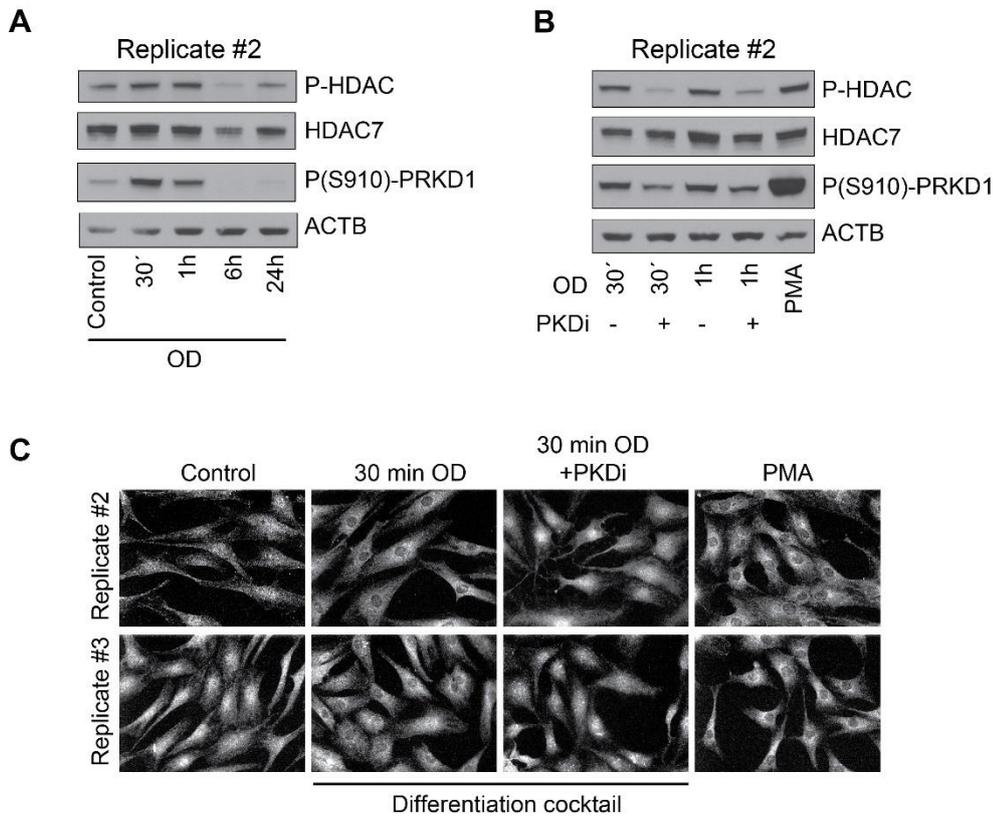
Supplemental Figure S4. PRKD1 regulates HDAC7 localization in immortalized hMSCs. (A) Replicate of Western blot images presented in Figure 7A of HDAC and PRKD1 phosphorylation levels during osteoblast differentiation (OD). (B) Replicate of western blots shown in Figure 7B of HDAC and PRKD1 phosphorylation 30 min and 1h after induction of osteoblast differentiation, with and without PRKD1 inhibitor (PKDi). (C) Replicates of immunofluorescent experiments presented in Figure 7C of HDAC7 in unstimulated hMSCs (Control), 30 min after osteoblast induction (30 min OD) with and without inhibitor of PRKD1 (PKDi) or PMA treated hMSCs.

Supplemental Figure S5
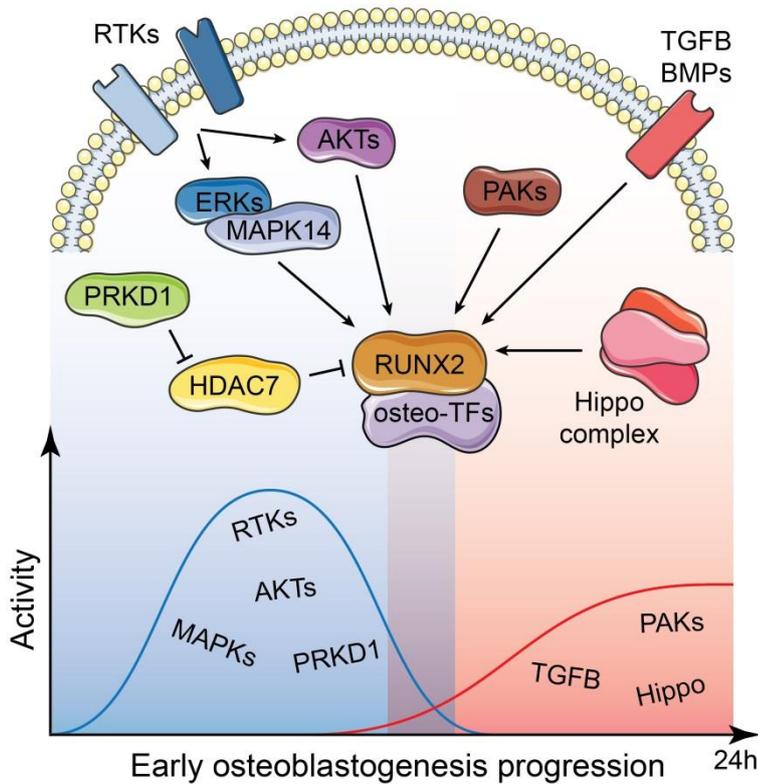
Supplemental Figure S5. Two distinct signaling waves act during early osteoblast differentiation. The first wave, initiated immediately after induction of differentiation, involves direct activation of MAPKs, AKT and PRKD1 downstream of receptor tyrosine kinases (RTKs). At this stage, RUNX2 is stimulated by PRKD1-dependent nuclear export of its repressor HDAC7. The second wave relies on de novo transcription and subsequent signaling through the PAK, Hippo, and TGFB/BMP pathways. In general, the signaling at the initial 24 hours of osteoblast differentiation converge on and activate osteogenic transcription factors (osteo-TFs), most notably RUNX2.

Supplemental Tables Legends

Supplemental Table S1. Protein identifications. List of all proteins identified and quantified from the three biological replicates.

Supplemental Table S2. Phosphorylation site identifications and kinase substrate predictions. Sheet 1: List of all phosphorylation sites identified and quantified from three biological replicates. Sheet 2: Kinase substrate predictions based on the NetworKIN. Sheet 3: Predicted phosphorylation sites for the MAP kinases.

Supplemental Table S3. KEGG enrichments. KEGG-based pathway enrichment analysis of the regulated phosphorylation sites.

Supplemental Methods

Cell culture, SILAC labeling and differentiation

As a model for human primary MSCs, we employed telomerized hMSC-TERT cell line that has been immortalized by ectopic expression of telomerase reverse transcriptase gene (hTERT) (Simonsen et al. 2002). The cells exhibit all typical characteristics of MSCs including heterotopic bone and bone marrow formation when implanted ectopically in immune deficient mice (Simonsen et al. 2002; Abdallah et al. 2005; Al-Nbaheen et al. 2013). For simplicity, we will refer to the hMSC-TERT as hMSCs throughout the manuscript. The cells were cultured in MEM media supplemented with 10% Fetal Bovine serum (FBS), 2 mM L-Glutamine, 100 units/ml of penicillin-streptomycin (Gibco-Invitrogen, Carlsbad, CA, USA).

For SILAC experiments, cells were kept in DMEM deficient in arginine and lysine supplemented with dialyzed serum (Gibco-Invitrogen), 2 mM L-Glutamine and 100 units/ml penicillin-streptomycin. Cells were labelled with L-arginine (Arg0) and L-lysine (Lys0) for the unstimulated cells, L-arginine-$^{13}C_6$ (Arg6) and L-

lysine-^{2}H$_4$ (Lys4) for the 30 min and 1 hour conditions, or L-arginine-^{13}C$_6$^{15}N$_4$ (Arg10) and L-lysine-^{13}C$_6$^{15}N$_2$ (Lys8) for the 6 and 24 hours conditions (Cambridge Isotope Laboratories, Inc., Tewksbury, MA, USA) at concentrations of 28 mg/L and 42 mg/L, respectively.

The differentiation of hMSCs cells into the osteoblastic lineage was performed as described previously (Kratchmarova et al. 2005). Briefly, cells were serum starved 24 hours prior to induction of differentiation with 10 mM β-glycerophosphate (Sigma-Aldrich, St. Louis, MO, USA), 80 μg/ml ascorbic acid phosphate (Wako Pure Chemical Industries, Osaka, Japan), 100 nM dexamethasone (Sigma-Aldrich), 10 nM 1,25-dihydroxyvitamin D3 (Biomol GmbH, Hamburg, Germany) and 50 ng/ml EGF (PeproTech, Rocky Hill, NJ, USA). For PRKD1 inhibition, CRT0066101 (Abcam, Cambridge, UK) was added 30 minutes prior to the induction of OB differentiation at a final concentration of 10 μM. For PRKD1 activation, phorbol myristate acetate (PMA) (Sigma-Aldrich) was used at a final concentration of 150 ng/ml.

LC-MS/MS data analysis (protein and site identification and quantification)

All MS raw files from the three biological replicas were processed together using the MaxQuant 1.3.0.5 suite (Cox and Mann 2008) with Andromeda algorithm integrated for database search (Cox et al. 2011) against UniProt release from 05-09-12, including both reviewed and unreviewed sequences. A maximum of two missed cleavages was allowed. Cysteine carbamidomethylation was set as fixed modification and methionine oxidation, asparagine/glutamine deamidation and N-term acetylation as well as serine/threonine/tyrosine phosphorylation were set as variable modifications. The minimum sequence length for the peptides was set to 7. FDR of 1% was set for both protein and site identification after common contaminants were removed. Protein with shared peptides were combined in protein groups, and both unique and razor peptides were used for quantitation.

Kinase substrates

Class I sites (localization probability higher than 0.75) belonging to serine and threonine were subjected to kinase prediction using NetworKIN 2.0 (Linding et al. 2008). Results corresponding to serine phosphorylation were retrieved and crossed with the cluster information to look for kinases with

overrepresented substrates as described (Bennetzen et al. 2012). Class I sites belonging to serine and threonine phosphorylation were also subjected to motif prediction using MotifX (Schwartz and Gygi 2005) with a p value threshold of 10^{-6} and occurrences of 1% of number of sites submitted. Results were then compared with the NetworKIN predictions as well as the quantification values and treated as described elsewhere (Rigbolt et al. 2011a). Briefly, we plotted in a heat map the percentage of sites assigned to a certain motif predicted to each kinase. To evaluate the regulation of sites following the same motif, we used Student´s t-test in regulated sites for each time point. Motifs with no significant regulation ($p > 0.01$) were assigned a value of 0 and the remaining were normalized between 1 (increase) and -1 (decrease).

Alignment of kinases and transcription factors

Protein kinase domains were extracted using Uniprot database as a reference. ClustalW2 (Larkin et al. 2007) was used for alignment and iTOL online software (Letunic and Bork 2011) for visualization and arrangement of the data. A similar approach was used for transcription factors. In this case, domains defined as "DNA binding", "zinc-finger" and others according to PRO Rules were used for filtering. The resulting list was aligned and visualized.

Western blotting

For Western blot analysis, cells were harvested in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% NP-40, 0.25 % Na deoxycholate mixed with 6×loading buffer (4 % SDS, 10% β-Mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Trizma-HCl). Lysates were cleared by centrifugation after brief sonication, boiled at 95 ºC for 5 minutes, loaded on a 4-12% SDS-PAGE precast gel (NuPAGE Novex bis-Tris; Invitrogen) and transferred to 0.45 µm nitrocellulose membranes (GE Healthcare, Little Chalfont, UK) with a semidry blotter (Hoefer, Holliston, MA, USA). Membranes were blocked for 1 hour at room temperature in either 5% BSA or 5% nonfat dry milk PBS with 0.1% Tween 20 (PBS/T) and incubated with primary antibodies overnight at 4 ºC with gentle agitation. The following day, membranes were rinsed in PBS with 0.1% Tween 20 (PBS/T), followed by incubation with species-specific horseradish-coupled secondary antibody (GE Healthcare, #NA931 and #NA934) and rinsing in PBS/T. For detection, membranes were incubated in chemiluminescent HRP substrates (Millipore, Merck, Darmstadt, Germany) and signal detected

on X-ray films. Primary antibodies were directed against PRKD1 (# 2052), phospho-Ser916 PRKD1 (# 2051), phospho-Ser-632/661/486 HDAC4/5/7 (# 3424) and phospho-Thr180/Tyr182 p38/MAPK14 (# 4511) (all Cell Signaling, Danvers, MA, USA), as well as HDAC7 (# sc-74563) and p38/MAPK14 (# sc-7972) (both from Santa Cruz Biotechnology, Dallas, TX, USA).

Quantitative real time polymerase chain reaction (qRT-PCR) and siRNA transfections
RNA was purified using Isol-RNA lysis reagent (5 PRIME, Hilden, Germany) according to the manufacturer's instructions. For generation of cDNA, 1 μg of RNA was reverse transcribed using MMLV (Thermo Fisher Scientific, MA, USA) according to the manufacturer's instructions. Quantitative PCR was performed in 10 μl reactions containing SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich) or FastStart Essential DNA Green Master (Roche, Basel, Switzerland), 3 μl of diluted cDNA and 300 nM of each primer. Reaction mixtures were preheated at 95 °C for 2 min followed by 40 cycles of melting at 95 °C for 15 s, annealing at 60 °C for 30 s, elongation at 72 °C for 45 s on a Mx3000p QPCR instrument (Stratagene, CA, USA) or the LightCycler platform (Roche).

A reverse transfection protocol using Lipofectamine 2000 (Invitrogen, USA) was employed for the siRNA transfections targeting the kinases listed below. hMSC cultures were trypsinized and 18,000 cells/cm^2 were reverse-transfected with two independent siRNAs, in two biological replicates for each siRNA. Below is the list of kinases and the corresponding siRNA sequences:

RIPK1 (#1: GGCUUUGGGAAGGUGUCUCtt, #2: GGCGAAGAUGAUGAACAGAtt)

TAOK1 (#1: CCGUGUUAAUAGAUCUCAUtt, #2: GGAGGCACAUAAUGGACCAtt)

GRK5 (#1: GGGUGUAUAUUUUCUGUGCtt, #2: GGCCUGGGCUGGAGUGUUAtt)

BMP2K (#1: GGUAGUGAAUCAAAUGAAUtt, #2: GGUUGCAUCAGUGUAAGACtt)

MAP3K3 (#1: GGGUAUGAAGAGUGUUAUUtt, #2: GGUCACUCAAAUAGGCAGAtt)

RPS6KA4 (#1: GGAGUUCCUGACGGAGGAGtt, #2: GGAUCCUAAGAAGCGAUUGtt)

GSK3 (#1: GGAUGAAAGAAGUUCUCAGtt, #2: GGAAACAGACCUUAAAAUUtt)

MAP3K5 (#1: GGAAAGCUCGUAAUUUAUAtt, #2: GGUAUACAUGAGUGGAAUUtt)

ERBB2 (#1: GGGAAACCUGGAACUCACCtt, #2: GGACAUCUUCCACAAGAACtt)

AAK1 (#1: GGUUGAAAACAUCCUCUUGtt, #2: GGACAAGCAAUGGGAUGAAtt).

For PRKD1, the experiment was performed in triplicates for each of its siRNAs (si*PRKD1* #1: GGUCAAAUUCUCAAUCAUAtt, si*PRKD1* #2: GGUCUUAAAUGUGAAGGGUtt). The non-targeting control siRNA#1 from Ambion was used as negative control. All siRNAs were used at 25 nM. The transfection media was replaced with normal culture media 8 hours after transfection. Osteogenic induction media was added to the cultures 48 hours after transfection.

Supplemental References

Abdallah BM, Haack-Sorensen M, Burns JS, Elsnab B, Jakob F, Hokland P, Kassem M. 2005. Maintenance of differentiation potential of human bone marrow mesenchymal stem cells immortalized by human telomerase reverse transcriptase gene despite [corrected] extensive proliferation. *Biochem Biophys Res Commun* **326**(3): 527-538.

Al-Nbaheen M, Vishnubalaji R, Ali D, Bouslimi A, Al-Jassir F, Megges M, Prigione A, Adjaye J, Kassem M, Aldahmash A. 2013. Human stromal (mesenchymal) stem cells from bone marrow, adipose tissue and skin exhibit differences in molecular phenotype and differentiation potential. *Stem Cell Rev* **9**(1): 32-43.

Bennetzen MV, Cox J, Mann M, Andersen JS. 2012. PhosphoSiteAnalyzer: a bioinformatic platform for deciphering phospho proteomes using kinase predictions retrieved from NetworKIN. *J Proteome Res* **11**(6): 3480-3486.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**(12): 1367-1372.

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**(4): 1794-1805.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948.

Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**(Web Server issue): W475-478.

Linding R, Jensen LJ, Pasculescu A, Olhovsky M, Colwill K, Bork P, Yaffe MB, Pawson T. 2008. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* **36**(Database issue): D695-699.

Schwartz D, Gygi SP. 2005. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* **23**(11): 1391-1398.