Figure S1

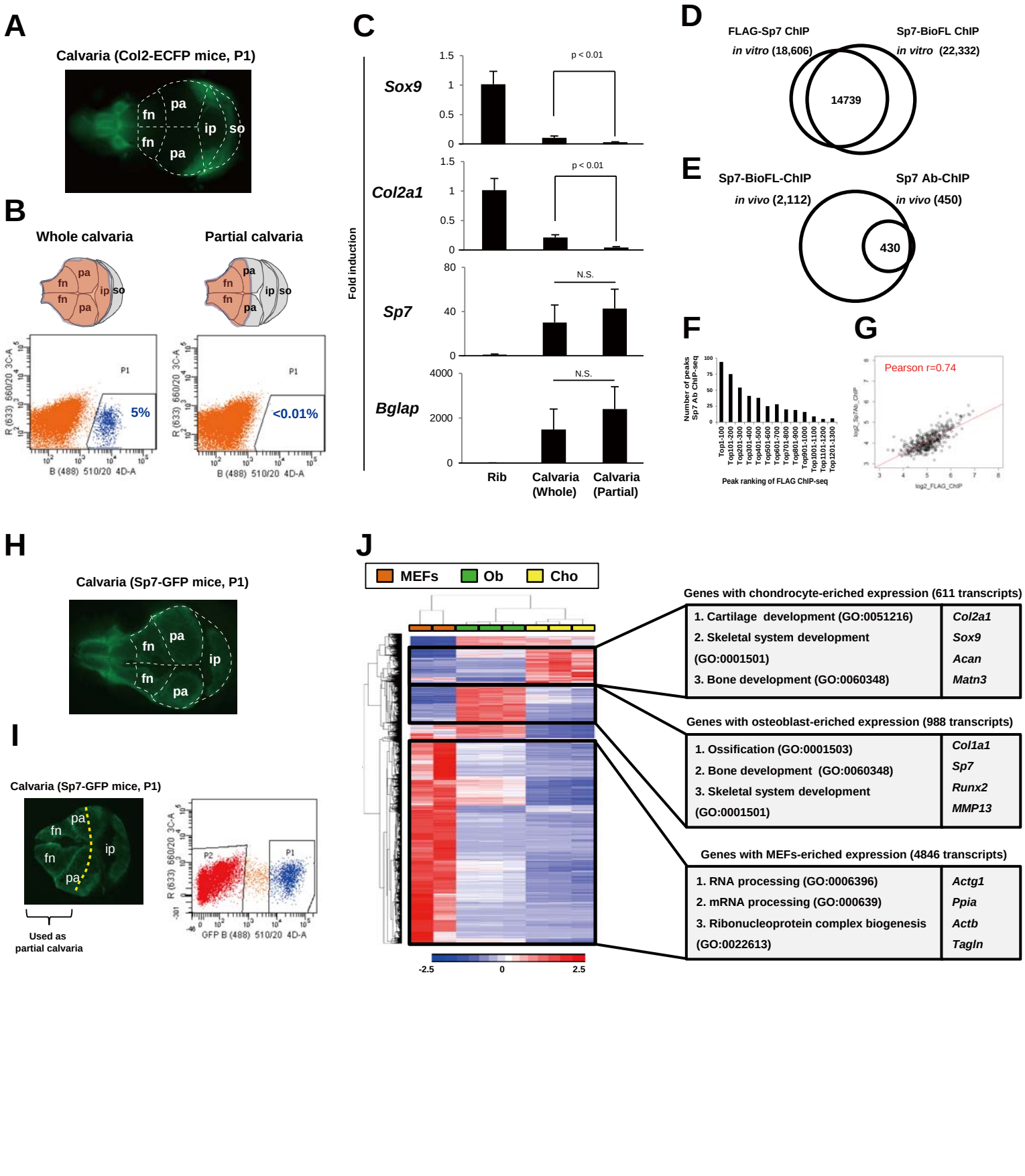

A
Calvaria (Col2-ECFP mice, P1)
pa
fn
ip
so
fn
pa
B
Whole calvaria
Partial calvaria
P1
5%
<0.01%
R (633) 660/20 3C-A
B (488) 510/20 4D-A
C
Sox9
Col2a1
Sp7
Bglap
p < 0.01
N.S.
Fold induction
Rib
Calvaria (Whole)
Calvaria (Partial)
D
FLAG-Sp7 ChIP
in vitro (18,606)
Sp7-BioFL ChIP
in vitro (22,332)
14739
E
Sp7-BioFL-ChIP
in vivo (2,112)
Sp7 Ab-ChIP
in vivo (450)
430
F
Number of peaks Sp7 Ab ChIP-seq
Peak ranking of FLAG ChIP-seq
G
Pearson r=0.74
log2_FLAG_ChIP
H
Calvaria (Sp7-GFP mice, P1)
I
Calvaria (Sp7-GFP mice, P1)
P2
P1
GFP B (488) 510/20 4D-A
Used as partial calvaria
J
MEFs
Ob
Cho
Genes with chondrocyte-eriched expression (611 transcripts)
1. Cartilage development (GO:0051216)
2. Skeletal system development (GO:0001501)
3. Bone development (GO:0060348)
Col2a1
Sox9
Acan
Matn3
Genes with osteoblast-eriched expression (988 transcripts)
1. Ossification (GO:0001503)
2. Bone development (GO:0060348)
3. Skeletal system development (GO:0001501)
Col1a1
Sp7
Runx2
MMP13
Genes with MEFs-eriched expression (4846 transcripts)
1. RNA processing (GO:0006396)
2. mRNA processing (GO:000639)
3. Ribonucleoprotein complex biogenesis (GO:0022613)
Actg1
Ppia
Actb
Tagln
-2.5
0
2.5

Figure S1 (related to Figure 2). Optimization of osteoblast isolation from mouse calvaria, ChIP-seq and RNA-seq analysis

(**A**) GFP fluorescence of neonatal Col2-ECFP mouse calvaria: fn, frontal bone; pa, parietal bone; ip, interparietal bone, so, supraoccipital bone.

(**B**) FACS sorting of the ECFP positive-cells from whole and partial calvaria preparations of neonatal Col2-ECFP mice

(**C**) mRNA expression of chondrocyte and osteoblast marker genes determined by qPCR in rib chondrocytes (Rib), whole calvaria and partial calvaria.

(**D**) Peak overlap between FLAG-Sp7 ChIP-seq and Sp7-BioFL ChIP-seq. ChIP-seq was performed with FLAG antibody on MC3T3-E1 cells expressing Sp7 with either an N-terminal FLAG tag (FLAG-Sp7) or a C-terminal BioFL tag (Sp7-BioFL).

(**E**) Peak overlap in ChIP-seq analysis of primary osteoblasts comparing Sp7 association with Sp7 antibody (Sp7-Ab ChIP) and Sp7-FLAG association with FGAG antibody (FLAG ChIP) (Figure 2).

(**F**) Number of Sp7-Ab ChIP-seq peaks shared by FLAG ChIP-seq in bins of FLAG ChIP-seq peak where each bin comprises 100 Sp7-Ab ChIP-seq peaks with peak ranking in bins decreasing from left to right.

(**G**) Scatter plot showing Sp7-Ab ChIP-seq and FLAG ChIP-seq peak signals. Red line represents Poisson regression line.

(**H**) Visualization of *Sp7*-driven GFP fluorescence in neonatal mouse calvaria

(**I**) FACS sorted GFP positive-cells from partial calvaria preparations of neonatal Sp7-GFP mouse subsequently examined for osteoblast-specific gene activity by RNA-seq. About 15-20% of the isolated single cell population from the partial calvaria was GFP positive.

(**J**) Hierarchical clustering of selected transcripts among MEFs, rib chondrocytes (Cho) and sorted Sp7-GFP positive osteoblasts (Ob). Enriched GO terms from a Database for Annotation, Visualization and Integrated Discovery (DAVID) analysis of mRNA populations. Each grouping shows the top three most enriched GO terms and highlights examples of specific genes within that GO category. Color scale indicates relative gene expression values.

Figure S2

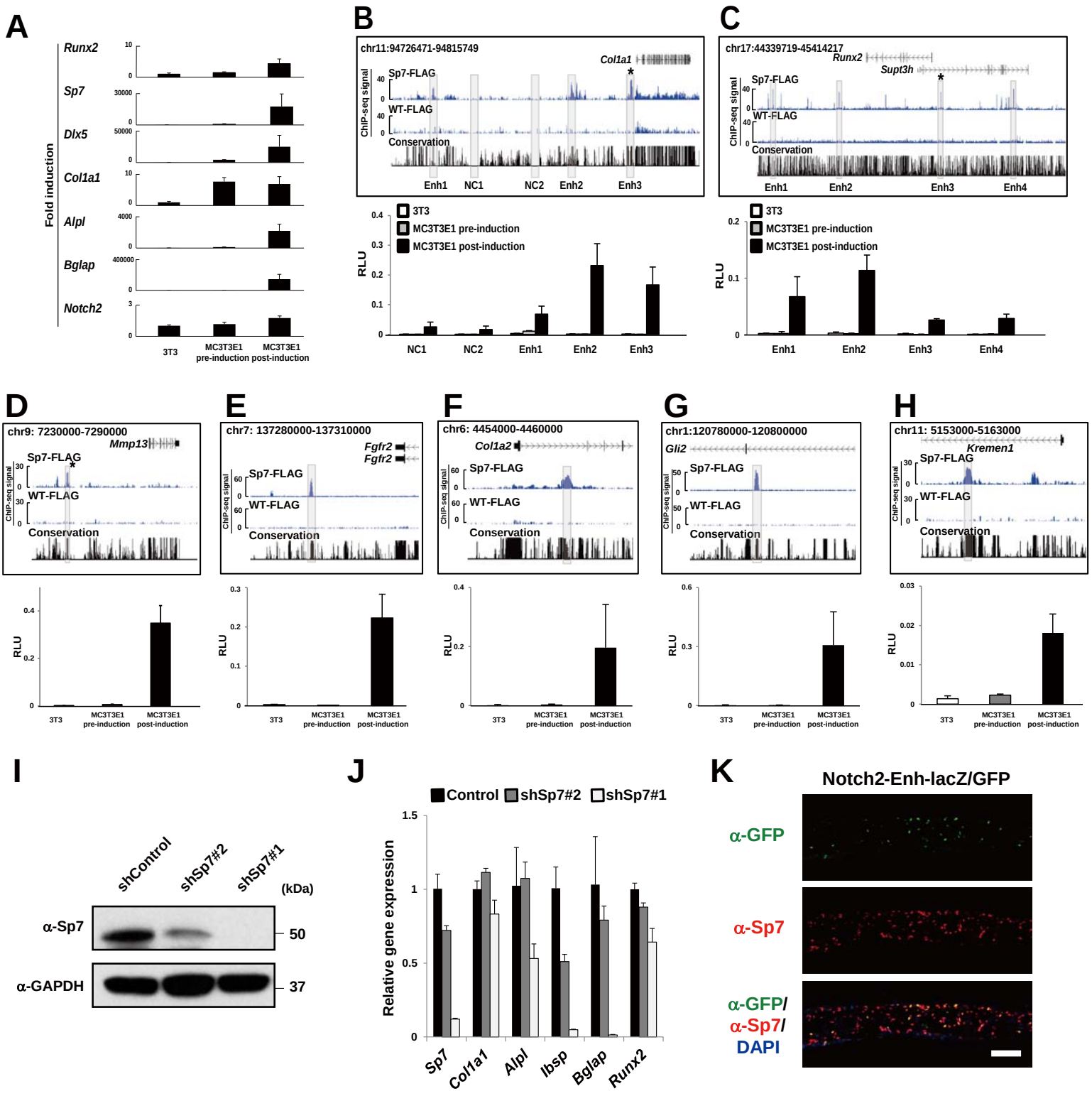

A
Runx2
Sp7
Dlx5
Col1a1
Alpl
Bglap
Notch2
Fold induction
3T3
MC3T3E1 pre-induction
MC3T3E1 post-induction
B
chr11:94726471-94815749
Col1a1
Sp7-FLAG
WT-FLAG
Conservation
ChIP-seq signal
Enh1
NC1
NC2
Enh2
Enh3
3T3
MC3T3E1 pre-induction
MC3T3E1 post-induction
RLU
C
chr17:44339719-45414217
Runx2
Supt3h
Enh1
Enh2
Enh3
Enh4
D
chr9: 7230000-7290000
Mmp13
E
chr7: 137280000-137310000
Fgfr2
F
chr6: 4454000-4460000
Col1a2
G
chr1:120780000-120800000
Gli2
H
chr11: 5153000-5163000
Kremen1
I
shControl
shSp7#2
shSp7#1
(kDa)
α-Sp7
α-GAPDH
50
37
J
Control
shSp7#2
shSp7#1
Relative gene expression
Sp7
Col1a1
Alpl
Ibsp
Bglap
Runx2
K
Notch2-Enh-lacZ/GFP
α-GFP
α-Sp7
α-GFP/
α-Sp7/
DAPI

Figure S2 (related to Figure 3). Functional analysis of putative Sp7-mediated osteoblast enhancers identified through *in vivo* Sp7 osteoblast ChIP-seq

(**A**) RT-qPCR analysis of mRNA expression for osteoblast marker genes in 3T3 cells and in MC3T3E1 cells pre and post osteoblast-induction.

(**B**-**H**) Luciferase enhancer assays for highlighted genomic regions following transfection and osteoblast induction of MC3T3E1 cells. Putative enhancer regions identified by Sp7 binding and defined by species conservation (Enh) and adjacent negative control regions that did not associate with Sp7 (NC) where fused to a luciferase reporter then co-transfected with pGL3-Renilla into 3T3 and MC3T3E1 cell lines. 3T3 cells and MC3T3E1 cells were cultured for 2 days for pre-osteogenic induction (pre-induction). Osteoblast induction (post-induction) requited 7 days of culture of MC3T3E1 cells in osteogenic medium.

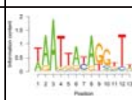
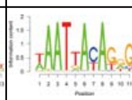
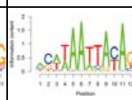
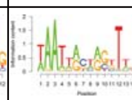
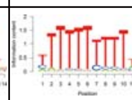
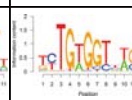
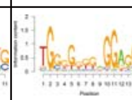
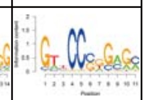
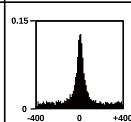
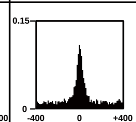
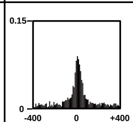
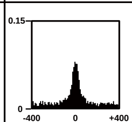
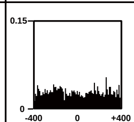
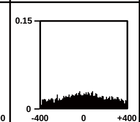
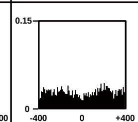
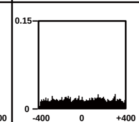
(**I**) Western blot analysis of shRNA knock-down of Sp7 in the MC3T3E1 cell line cultured for 7 days in osteogenic medium. Sp7 levels were compared to GAPDH levels following lentiviral infection with viruses encoding control and two different Sp7 targeted shSp7 RNAs.

(**J**) Quantitative PCR analysis of the requirement for Sp7 for expression of key osteoblast-enriched gene targets as indicated. MC3T3E1 cells were cultured for 7 days in osteogenic medium following lentivirus infection with viruses encoding control and two different Sp7 targeted shSp7 RNAs.

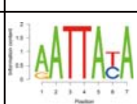
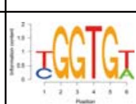
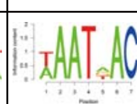
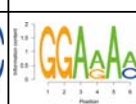
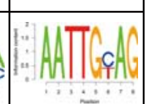
(**K**) P1 calvaria sections immunostained with antibodies specific to GFP and Sp7 in Notch2 enhancer transgenic reporter mice. Nuclei were visualized with DAPI. Bar, 100 μm

Figure S3

A

Ranking	1	2	3	4	5	6	7	8
Motif logo								
Motif score	4.32	4.12	3.59	3.39	3.34	3.30	2.34	2.26
Predicted motif	Homeodomain response element	Homeodomain response element	Homeodomain response element	Homeodomain response element	T-rich	Runx	G-rich	-
% of peaks with motif within 100 bp from the center	67%	59%	50%	43%	10%	24%	14%	13%
Motif distribution from peak center								

B

Ranking	1	2	3	4	5
Logo					
E-value	2.3e-167	8.4e-33	3.5e-29	1.6e-17	4.9e-14
Predicted motif	Homeodomain response element	Runx	Homeodomain response element	NFAT	-

C

Notch2 enhancer

Wild-type	...tgcttgtttcctcTAATTAtatagtacatatccaaacccccAATTAtccag...
Mutant	...ttgcttgtttcctcTcgcTAtatagtacatatccaaacccccAcgcAtccag...

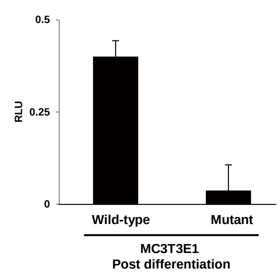
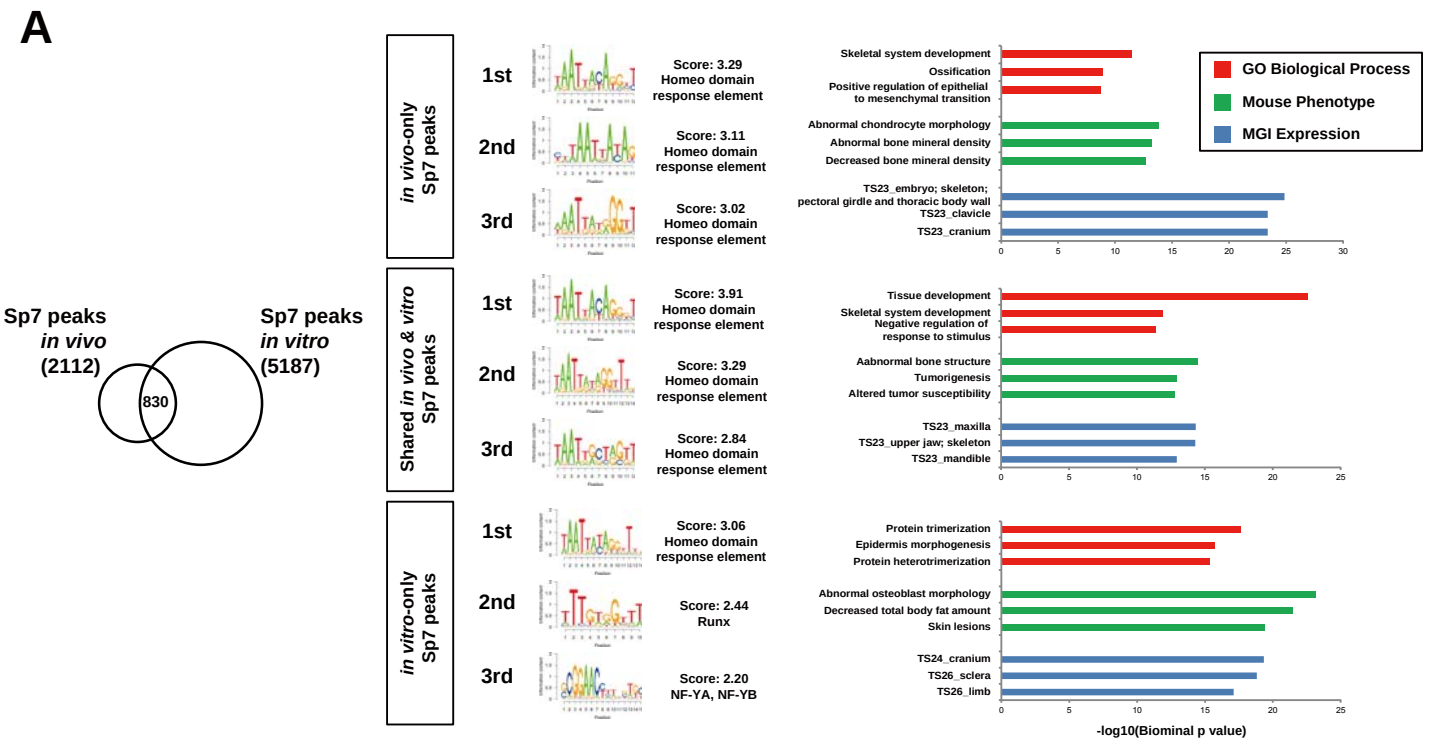


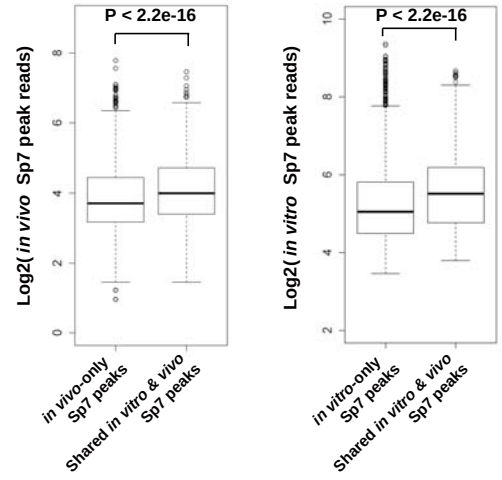
Figure S3 (related to Figure 4). Sp7 motif analysis

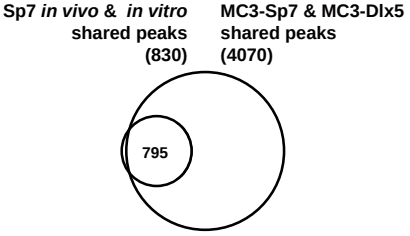
(**A, B**) Summary of *de novo* motif analysis on Sp7 osteoblast ChIP-seq data using Cisgenome (A) and DREME (B) analysis tools. In A, panel displays motif ranking, motif logo, Cisgenome motif score, predicted motif, the percentage of peaks containing the motif and the distribution of the motif within a defined interval around each peak center (+/- 400 bp). The top eight statistically significant motifs are shown. In B, panel displays motif ranking, motif logo, E-value and predicted motif.

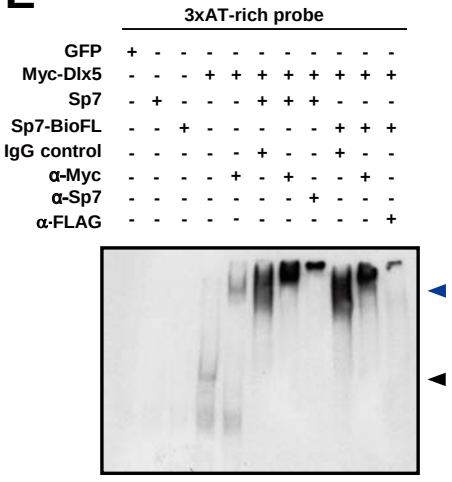
(**C**) Luciferase enhancer assay in MC3T3E1 cells examining the wild-type Notch2 intronic enhancer and the Notch2 enhancer with mutations within the two AT-rich motifs. The reporter construct was co-transfected together with pGL3-Renilla into 3T3 or MC3T3E1 cell lines. MC3T3E cells were cultured in osteogenic media for 7 days.

Figure S4


A
Sp7 peaks in vivo (2112)
Sp7 peaks in vitro (5187)
830
in vivo-only Sp7 peaks
1st
Score: 3.29
Homeo domain response element
2nd
Score: 3.11
Homeo domain response element
3rd
Score: 3.02
Homeo domain response element
Skeletal system development
Ossification
Positive regulation of epithelial to mesenchymal transition
Abnormal chondrocyte morphology
Abnormal bone mineral density
Decreased bone mineral density
TS23_embryo; skeleton; pectoral girdle and thoracic body wall
TS23_clavicle
TS23_cranium
GO Biological Process
Mouse Phenotype
MGI Expression
Shared in vivo & vitro Sp7 peaks
1st
Score: 3.91
Homeo domain response element
2nd
Score: 3.29
Homeo domain response element
3rd
Score: 2.84
Homeo domain response element
Tissue development
Skeletal system development
Negative regulation of response to stimulus
Aabnormal bone structure
Tumorigenesis
Altered tumor susceptibility
TS23_maxilla
TS23_upper jaw; skeleton
TS23_mandible
in vitro-only Sp7 peaks
1st
Score: 3.06
Homeo domain response element
2nd
Score: 2.44
Runx
3rd
Score: 2.20
NF-YA, NF-YB
Protein trimerization
Epidermis morphogenesis
Protein heterotrimerization
Abnormal osteoblast morphology
Decreased total body fat amount
Skin lesions
TS24_cranium
TS26_sclera
TS26_limb
-log10(Biominal p value)


B
P < 2.2e-16
Log2(in vivo Sp7 peak reads)
in vivo-only Sp7 peaks
Shared in vitro & vivo Sp7 peaks
P < 2.2e-16
Log2(in vitro Sp7 peak reads)
in vitro-only Sp7 peaks
Shared in vitro & vivo Sp7 peaks


C
Sp7 in vivo & in vitro shared peaks (830)
MC3-Sp7 & MC3-Dlx5 shared peaks (4070)
795


E
3xAT-rich probe
GFP + - - - - - - - - - -
Myc-Dlx5 - - - + + + + + + + +
Sp7 - + - - - + + + - - -
Sp7-BioFL - - + - - - - - + + +
IgG control - - - - + - - + - -
α-Myc - - - - + - + - - + -
α-Sp7 - - - - - - - + - - -
α-FLAG - - - - - - - - - - +

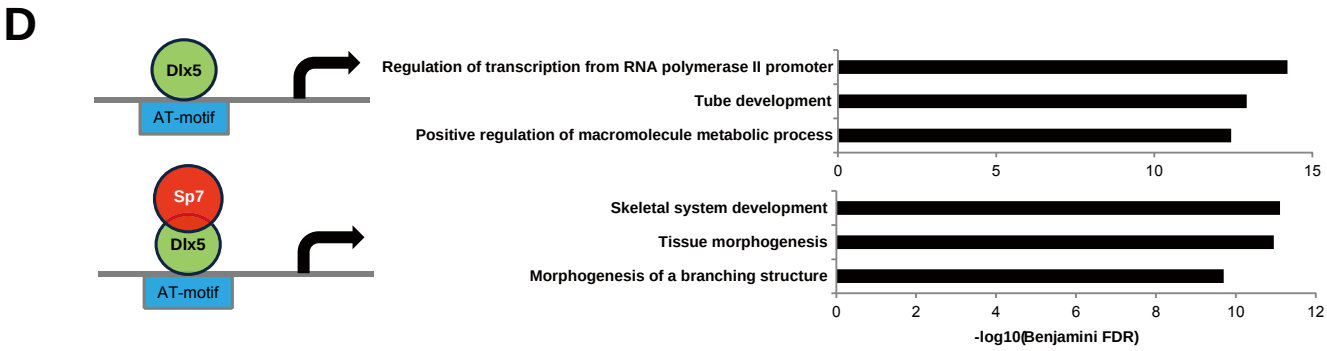

D
Dlx5
AT-motif
Sp7
Dlx5
AT-motif
Regulation of transcription from RNA polymerase II promoter
Tube development
Positive regulation of macromolecule metabolic process
Skeletal system development
Tissue morphogenesis
Morphogenesis of a branching structure
-log10(Benjamini FDR)

Figure S4 (related to Figure 4). Similarities and differences in Sp7-DNA binding signatures from *in vitro* and *in vivo* studies and the genomic architecture of Sp7-Dlx5 shared regions *in vitro*
(**A**) Venn diagram, *de novo* motif analysis and a GREAT GO analysis intersecting Sp7 peaks identified *in vivo* and *in vitro*. For the motif analysis, the top 500 intersected peaks were used in the Cisgenome program. The top three most enriched motifs, motif scores and each predicted motif is shown. All intersected peaks were used in GREAT GO analysis; the top three most enriched GO terms are shown.
(**B**) Intersection of Sp7 peak signals *in vivo* (left panel) and *in vitro* (right panel).
(**C**) Venn diagram showing the overlap of Sp7 associated peaks identified *in vitro* and *in vivo* with common peaks shared by Dlx5 and Sp7 *in vitro*.
(**D**) DAVID GO analysis of genes associated with Dlx5 peaks containing an AT-rich motif. Data shows DAVID GO terms for genes displaying a co-association of Sp7 at identified Dlx5 peaks, or genes with no co-association. The top three most enriched terms are shown.
(**E**) EMSA analysis of the association of Sp7 and Sp7BioFL with Dlx5 on the AT-rich target sequence. Myc-Dlx5 interacted with the target sequence and anti-Myc antibody addition led to a supershift. Neither Sp7 nor Sp7-BioFL interacted with target DNA in this experiment. Co-expression of Myc-Dlx5 with either Sp7 or Sp7-BioFL generated a new slower migrating EMSA product than observed with Myc-Dlx5 alone. Addition of anti-Myc, anti-Sp7 and anti-FLAG antibodies led to an additional super-shift product consistent with ternary complex formation of Sp7/Sp7-BioFL and Myc-Dlx5 on the AT-rich target sequence. Black arrowheads indicate a Myc-Dlx5 shift band and the blue arrow indicate complies of Myc-Dlx5 with either Sp7 or Sp7BioFL.

Figure S5

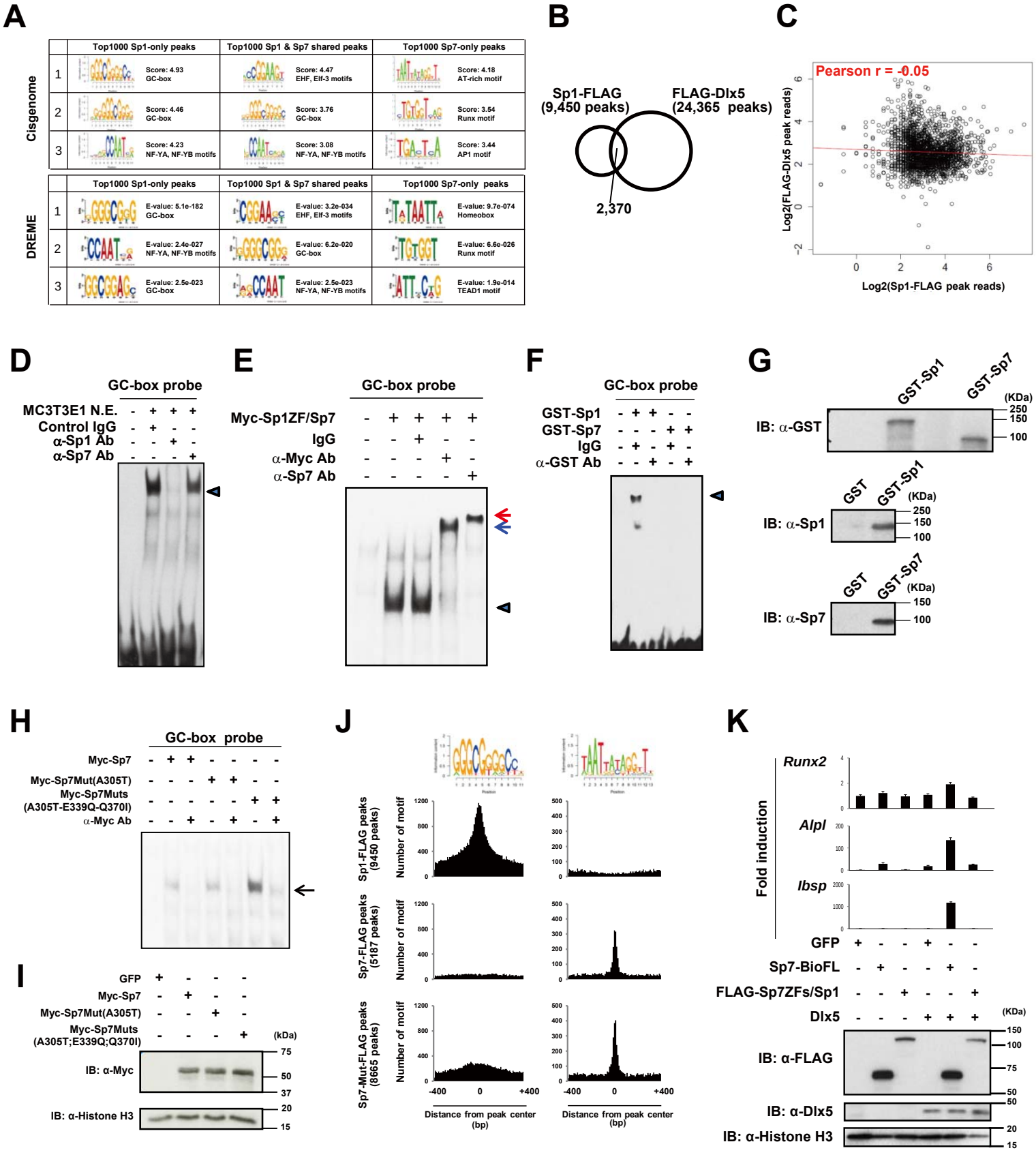

A
Cisgenome
Top1000 Sp1-only peaks
Top1000 Sp1 & Sp7 shared peaks
Top1000 Sp7-only peaks
1
Score: 4.93
GC-box
Score: 4.47
EHF, Elf-3 motifs
Score: 4.18
AT-rich motif
2
Score: 4.46
GC-box
Score: 3.76
GC-box
Score: 3.54
Runx motif
3
Score: 4.23
NF-YA, NF-YB motifs
Score: 3.08
NF-YA, NF-YB motifs
Score: 3.44
AP1 motif
DREME
Top1000 Sp1-only peaks
Top1000 Sp1 & Sp7 shared peaks
Top1000 Sp7-only peaks
1
E-value: 5.1e-182
GC-box
E-value: 3.2e-034
EHF, Elf-3 motifs
E-value: 9.7e-074
Homeobox
2
E-value: 2.4e-027
NF-YA, NF-YB motifs
E-value: 6.2e-020
GC-box
E-value: 6.6e-026
Runx motif
3
E-value: 2.5e-023
GC-box
E-value: 2.5e-023
NF-YA, NF-YB motifs
E-value: 1.9e-014
TEAD1 motif
B
Sp1-FLAG
(9,450 peaks)
FLAG-Dlx5
(24,365 peaks)
2,370
C
Pearson r = -0.05
Log2(FLAG-Dlx5 peak reads)
Log2(Sp1-FLAG peak reads)
D
GC-box probe
MC3T3E1 N.E.
Control IgG
α-Sp1 Ab
α-Sp7 Ab
E
GC-box probe
Myc-Sp1ZF/Sp7
IgG
α-Myc Ab
α-Sp7 Ab
F
GC-box probe
GST-Sp1
GST-Sp7
IgG
α-GST Ab
G
GST-Sp1
GST-Sp7
(KDa)
250
150
100
IB: α-GST
GST
GST-Sp1
IB: α-Sp1
GST
GST-Sp7
150
100
IB: α-Sp7
H
GC-box probe
Myc-Sp7
Myc-Sp7Mut(A305T)
Myc-Sp7Muts
(A305T-E339Q-Q370I)
α-Myc Ab
I
GFP
Myc-Sp7
Myc-Sp7Mut(A305T)
Myc-Sp7Muts
(A305T;E339Q;Q370I)
(kDa)
75
50
37
IB: α-Myc
20
15
IB: α-Histone H3
J
Sp1-FLAG peaks
(9450 peaks)
Sp7-FLAG peaks
(5187 peaks)
Sp7-Mut-FLAG peaks
(8665 peaks)
Number of motif
-400
0
+400
Distance from peak center
(bp)
K
Fold induction
Runx2
Alpl
Ibsp
GFP
Sp7-BioFL
FLAG-Sp7ZFs/Sp1
Dlx5
(KDa)
150
100
75
50
IB: α-FLAG
50
IB: α-Dlx5
20
15
IB: α-Histone H3

Figure S5 (related to Figure 6). Different mode of action between Sp1 and Sp7 in osteoblasts
(**A**) *de novo* motif analysis on Sp1 and Sp7 intersected MC3T3E1 FLAG ChIP-seq data using Cisgenome software (upper) and DREME software (lower). Panel displays top three most enriched motif logos, motif score (upper) or E-value (lower) and the predicted motif sequence.
(**B**) Peak overlap between Sp1-FLAG ChIP-seq and FLAG-Dlx5 ChIP-seq.
(**C**) Scatter plot showing Sp1-FLAG ChIP-seq and FLAG-Dlx5 ChIP-seq peak signals. Red line represents Poisson regression line.
(**D-F**) EMSA with a GC box containing probe examining nuclear extracts of cultured MC3T3E1 cells (D), nuclear extracts of COS7 cells transfected with indicated plasmids (E) and *in vitro* translated proteins (F). Supershift experiments to examine DNA-protein complexes on incubation with anti-Sp1, anti-Sp7, anti-Myc, and anti-GST antibodies. Arrowheads indicate primary EMSA shift bands and the arrows indicate supershift products. In Figure S5E, the specific shift band indicating an interaction of Myc-Sp1-ZFs/Sp7 protein and GC-box probe was supershifted on addition of either anti-Myc antibody or anti-Sp7 antibody. The anti-Sp7 antibody gives as expected EMSA supershift, indicating that the weak band below the Sp1/probe complex band in Figure S5D was unlikely to correspond to an Sp7/GC-box probe complex.
(**G**) Western blotting to analyze GST-Sp1 and GST-Sp7 proteins used in experiments in Figure S5F.
(**H**) EMSA analysis of GC box probe interactions with nuclear extracts from COS7 cells transfected with the indicated plasmids. Arrow indicates primary EMSA shift bands.
(**I**) Western blotting to examine Myc-Sp7, Myc-Sp7mut (A305T) and Myc-Sp7muts (A305T-E339Q-Q370I) proteins used in Figure S5H.
(**J**) Distribution of the GC-box and the AT-rich motif within 400 bp window from the peak center in Sp1-FLAG (upper), Sp7-FLAG (middle), and Sp7-Mut-FLAG (lower) peaks. The number of existing motifs is shown in the Y-axis.
(**K**) RT-qPCR and western blotting analysis of osteoblast differentiation in C3H10T1/2 cells following retrovirus delivery of the indicated gene expression vector. Data displayed are the means ± SDs from triplicate experiments.

Supplemental Tables

Table S1 (related to Figure 2). Summary of ChIP-seq and RNA-seq data sets

Table S2 (related to Figure 2; *see accompanying Excel file*). Cell type-enriched genes
Sheet 1-3: Gene lists of osteoblast-, chondrocyte- and MEFs-enriched expression
Sheet 4-6: Gene ontology analysis of these enriched genes

Table S3 (related to Figure 3; *see accompanying Excel file*). Putative Sp7 target regions

Table S4 (related to Figure 4). Gene list of homeobox containing transcription factors enriched in osteoblast, chondrocyte and MEFs

Table S5 (related to Experimental Procedures). Oligo DNA sequences that were used for RT-qPCR, ChIP-qPCR and EMSA

Table S6 (related to Experimental Procedures). Genomic information used for enhancer assays.

Table S1 (related to Figure 2). Summary of ChIP-seq and RNA-seq data sets

Experiment	Sample Name	Raw sequence reads	Aligned reads	Non-redundant reads	% of Non-redundant reads
ChIP-seq	POB_Sp7-FLAG1	36,103,840	24,774,393	21,941,351	88.56
	POB_Sp7-FLAG2	33,194,333	22,733,681	20,579,301	90.52
	POB_Sp7Ab-ChIP	32,924,692	22,967,394	21,748,764	96.46
	POB_WT-FLAG	35,628,029	22,419,080	13,795,684	61.54
	POB_input	34,301,293	23,297,399	22,472,464	96.46
	MC3_Sp7-FLAG1	38,263,445	27,541,114	23,946,342	86.95
	MC3_Sp7-FLAG2	37,253,827	26,625,495	23,677,714	88.93
	MC3_Dlx5-FLAG1	31,770,764	22,797,664	20,019,290	87.81
	MC3_Dlx5-FLAG2	36,008,576	26,924,794	24,186,120	89.83
	MC3_GFP-FLAG	35,698,722	24,557,367	18,581,940	75.67
	MC3_input1	29,688,793	20,380,712	19,662,939	96.48
	MC3_Sp1-FLAG1	41,280,462	29,896,985	27,635,446	92.44
	MC3_Sp1-FLAG2	32,291,372	23,474,334	21,535,887	91.74
	MC3_Sp7mut-FLAG1	44,169,652	31,588,558	29,213,444	92.48
	MC3_Sp7mut-FLAG2	40,425,136	28,365,768	26,578,478	93.70
	MC3_FLAG-Sp7	37,920,298	26,857,925	24,946,376	92.88
	MC3_Sp7-BioFLAG	36,881,605	26,293,116	24,727,336	94.04
	MC3_input2	37,295,149	26,267,965	24,852,188	94.61
RNA-seq	Ob_Sp7-GFP1	35,890,670	34,707,203	32,274,633	91.18
	Ob_Sp7-GFP2	35,078,539	33,748,214	31,430,653	90.95
	Ob_Sp7-GFP3	34,657,237	27,034,478	23,432,195	69.23
	Cho_WT1	34,580,596	33,653,409	31,320,546	91.63
	Cho_WT2	35,027,305	33,704,131	31,126,885	90.07
	Cho_WT3	37,429,083	35,948,509	33,372,795	90.33

Table S4 (related to Figure 4). Gene list of homeobox containing transcription factors enriched in Sp7-GFP positive osteoblasts, chondrocytes and MEFs

Ref seq	Gene	Sp7-GFP Ob (rpkm)	Rib Cho (rpkm)	MEFs (rpkm)
NM_001081300	Tshz1	5.08	6.60	4.46
NM_008900	Pou3f3	1.30	14.97	0.43
NM_172553	Alx1	7.17	33.97	0.21
NM_010455	Hoxa7	0.01	8.39	0.39
NM_010466	Hoxc8	0.02	25.71	6.26
NM_010454	Hoxa6	0.00	11.98	1.05
NM_010453	Hoxa5	0.02	25.63	2.48
NM_010452	Hoxa3	0.02	9.19	0.90
NM_010789	Meis1	0.71	7.45	1.52
NM_080455	Tshz2	3.41	11.34	6.33
NM_010465	Hoxc6	0.02	9.37	7.63
NM_001193271	Meis1	1.28	11.33	5.89
NM_001042438	Zhx1	1.72	1.67	3.85
NM_173396	Tgif2	5.81	8.90	9.98
NM_183355	Pbx1	5.31	11.52	17.74
NM_016768	Pbx3	5.83	7.36	11.35
NM_009372	Tgif1	6.12	10.07	14.98
NM_175028	Adnp2	4.44	4.38	8.09
NM_008276	Hoxd8	0.13	0.13	11.33
NM_001164076	Tgif1	0.83	0.55	9.55
NM_172856	Cers6	2.27	0.48	19.55
NM_010573	Irx1	0.18	0.01	18.04
NM_010574	Irx2	0.17	0.06	12.27
NM_026058	Cers4	3.06	1.80	6.78
NM_011546	Zeb1	8.04	6.20	7.32
NM_015753	Zeb2	13.59	0.16	20.14
NM_028015	Cers5	24.18	12.17	33.64
NM_029789	Cers2	46.80	24.87	48.65
NM_009572	Zhx1	4.34	1.04	4.45
NM_030708	Zfhx4	8.96	5.25	5.49
NM_009189	Six1	11.92	6.46	5.13
NM_010133	En1	9.81	0.03	5.48
NM_008627	Meis3	15.57	0.70	8.31
NM_009116	Prrx2	140.74	3.82	15.07
NM_013601	Msx2	9.14	0.36	0.63
NM_007442	Alx4	46.41	0.57	2.33
NM_010056	Dlx5	40.29	0.91	0.16
NM_010057	Dlx6	15.73	0.04	0.22
NM_139146	Satb2	22.12	0.15	1.60
NM_010055	Dlx3	9.29	0.10	0.10
NM_018826	Irx5	24.53	0.18	2.36
NM_011380	Six2	33.03	2.09	0.53
NM_010835	Msx1	28.55	1.41	0.84
NM_008393	Irx3	46.86	0.40	3.99
NM_017463	Pbx2	10.50	5.56	5.46
NM_177595	Mkx	9.63	5.87	0.36
NM_011383	Six5	12.94	13.48	0.69

Table S5 (related to Experimental Procedures). Oligo DNA sequences used for RT-qPCR, ChIP-qPCR and EMSA

Application	Name	Sequence
RT-qPCR	Alpl_F	GCTGATCATTCCCACGTTTT
	Alpl_R	CTGGGCCTGGTAGTTGTTGT
	Bglap_F	AAGCAGGAGGGCAATAAGGT
	Bglap_R	TTTGTAGGCGGTCTTCAAGC
	Col1a1_F	ACGTCCTGGTGAAGTTGGTC
	Col1a1_R	CAGGGAAGCCTCTTTCTCCT
	Col1a2_F	AACACCCCAGCGAAGAACTCATACA
	Col1a2_R	ATGTTCTGAGAAGCACGGTTGGCTA
	Col2a1_F	CAACTGGAAAGCAGGGAGAC
	Col2a1_R	ACCTCTAGGGCCAGAAGGAC
	Dlx3_F	CCTGATGTGGTAAAACATCGAAGAAAGGAG
	Dlx3_R	GGAACAGGGTCACATAAATAGGAAACCACA
	Dlx5_F	CAGAAGAGTCCCAAGCATCC
	Dlx5_R	GAGCGCTTTGCCATAAGAAG
	Dlx6_F	CCTTAGGACTGACACAAACACAGGTGAAGA
	Dlx6_R	TGGTGTGGTGAGGAATACCAGTGTGAATAC
	Gapdh_F	AGGTCGGTGTGAACGGATTTG
	Gapdh_R	TGTAGACCATGTAGTTGAGGTCA
	Ibsp_F	CAGAGGAGGCAAGCGTCACT
	Ibsp_R	CTGTCTGGGTGCCAACACTG
	Notch2_F	CCGGTCTCCGTGTAAAAACAAAGGA
	Notch2_R	TGCTGGCACAAGTGTTCAACAGGTA
	Runx2_F	CCCAGCCACCTTTACCTACA
	Runx2_R	TATGGAGTGCTGCTGGTCTG
	Sox9_F	CGACTACGCTGACCATCAGA
	Sox9_R	AGACTGGTTGTTCCCAGTGC
	Sp7_F	ACTCATCCCTATGGCTCGTG
	Sp7_R	GGTAGGGAGCTGGGTTAAGG
ChIP-qPCR	Col1a2_F	TCAACAGATGTTCTCATTCCAAA
	Col1a2_R	TGGCATATAATTGCGGTCAG
	Gapdh_F	AAGCTCATGAGGCACAGAATGGTC
	GapdhR	TGGGTACATGGTGACTTTCCTAGGC
	Notch2_F	ATCCAAACCCCCAATTATCC
	Notch2_R	GAAGGCACTGCCAAGGTCTA
	Runx2_F	CAAAATGGGAGTATGGAAGTGGGGGTA
	Runx2_R	CAAGGCGCTCATTAAGCTATTGGGTTT
EMSA	Notch2 enhancer	TGCTTGTTTCCTCTAATTATATAGTACATATCCAAACCCCCAATTATCCAG
	3xAT-rich motif	ATGGTAATTATAGGTTGGCATAATTATAGGTTTCGATAATTATAGGTTTCAC
	GC-box probe	GTTGCCAGAAGCCCCGCCCCTAGGAGTGATCGGAAAG

Table S6 (related to Experimental Procedures). Genomic information used for enhancer assays

Target	Location	Size (bp)
Notch2 Enhancer	chr3: 97860644-97860985	342
Col1a1 Negative Control 1	chr11:94748641-94749342	702
Col1a1 Negative Control 2	chr11:94767760-94768464	705
Col1a1 Enhancer 1	chr11:94738061-94738710	650
Col1a1 Enhancer 2	chr11:94778427-94779046	620
Col1a1 Enhancer 3*	chr11:94795274-94797693	2420
Runx2 Enhancer 1	chr17:44446734-44447635	902
Runx2 Enhancer 2	chr17:44657550-44658302	753
Runx2 Enhancer 3**	chr17:44978881-44980065	1185
Runx2 Enhancer 4	chr17:45209647-45210552	906
Fgfr2 Enhancer	chr7:137290546-137291249	704
Col1a2 Enhancer	chr6:4457240-4457942	703
Gli2 Enhancer	chr1:120787968-120788378	411
Kremen1 Enhancer	chr11:5155597-5156322	726
Mmp13 Enhancer ***	chr9:7242653-7243273	621

The targets with asterisk were reported as osteoblast enhancers (*, Bedalov et al., 1995; **, Kawane et al., 2014; ***, Meyer et al., 2015)

Supplemental Experimental Procedures

Generation of *Sp7-Biotin-3xFLAG* knock-in mouse line

The targeting strategy to generate the *Sp7-Biotin-3xFLAG* (*Sp7-BioFL*) allele is shown in Figure 1A. Homologous arms were obtained by polymerase chain reaction (PCR) from a BAC clone RP23-263-M4 using Platinum *pfx* DNA Polymerase (11708-013, Invitrogen). 5' homologous arm (chr15:102188571+102191595) carrying *Aat II* and *Nde I* sites at its C-terminus was cloned into *Sac II* and *Not I* sites of the pPGKneobpAlox2PGKDTA (generously provided by P. Soriano, Mount Sinai School of Medicine, NY, U.S.A), upstream of the PGKneo cassette flanked by two loxP sites. *Biotin-3xFLAG* (*BioFL*) sequence (5'-GGGTCCGGCCTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAAGGCGCGCCGAGC TCGAGG-
GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGAC AAG-3') was cloned into the *Aat II* and *Nde I* sites of the 5' homologous arm with a stop codon. 3' homologous arm (chr15:102184004+102188567) was cloned into the *Sal I* site of the pPGKneobpAlox2PGKDTA, downstream of the PGKneo (*Neo*) cassette. The targeting vector was linearized with *Apa I* and electroporated into the 129/Sv x C57BL/6J F1 hybrid mouse embryonic stem cell line (v6.5) (Eggan et al., 2001). After the expansion of G418-resistant clones, recombinants were screened by PCR using 5' and 3' external primers on genomic DNA extracts (Figure 1B). Primer sequences used in the screening are as follows: P1, 5'-ttttattttatgtatgtgggtgtttagc-3' (in 5' external region of the homologous arm); P2, 5'-atccttgtagtcgatgtcatgatctttat-3' (within *BioFL* sequence) ; P3, 5'-gaggattgggaagacaatagcag-3' (within *Neo* sequence); P4, 5'-agtctcactgaagaatgtgggttt -3' (in 3' external region of the homologous arm); and P5, 5'-gctgattggcttcttcttcc-3' (in the 5' homologous arm). Two clones for which homologous recombination was confirmed by the PCR (1A9 and 2E10) were used for blastocyst injection. Male chimeras with 100% chimerism derived from the 2E10 clone were crossed with C57BL/6J to obtain the F1 generation. Heterozygous *Sp7-BioFL-Neo* mice ($Sp7^{BioFL\text{-}neo/+}$) in the F1 generation carrying the *BioFL* and the *Neo* cassette were crossed with *Sox2-Cre* mice (Hayashi et al., 2002) to excise the *Neo* selection cassette generating $Sp7^{BioFL/+}$ mice. The *Sox2-Cre* transgene was segregated from *Sox2-Cre*; $Sp7^{BioFL/+}$ mice, and $Sp7^{BioFL/+}$ mice were crossed to generate $Sp7^{BioFL/BioFL}$ mice. The following primers were used in PCR genotyping of $Sp7^{BioFL\text{-}neo}$ or $Sp7^{BioFL}$ alleles (Figure 1C): P6, 5'-gggaagaagaagccaatcagc-3'; P7, 5'-gcttgtcatcgtcatccttgtagtc-3'; and P8, 5'-tggctaagactcttgacaaagagagaa-3'. The *Sp7-BioFL* mouse line was maintained in $Sp7^{BioFL/BioFL}$ genotype. Skeletons were stained as described previously (Rodda and McMahon, 2006)

Mouse strains

Sp7/Osx-GFP BAC transgenic mice (Rodda and McMahon, 2006) and Col2-ECFP mice (Chokalingam et al., 2009) were used for cell sorting. G0 transgenic analysis for the Notch2 enhancer was performed using Hsp68-lacZ::nGFP reporter construct as previously described (Peterson et al., 2012).

Optimization of osteoblast isolation

Although osteoblast isolation from neonatal calvaria is an established method, chondrogenic subpopulation is suggested to be contaminated (Aberg et al., 2005). To investigate the possibility of chondrocyte contamination, we visualized the chondrocyte distribution in calvaria utilizing a Col2-ECFP reporter mouse (Chokalingam et al., 2009). ECFP fluorescence (Figure S1A) revealed chondrocytes in developing interparietal, supraoccipital and caudal parietal bone forming regions. Fluorescence-activated cell sorting (FACS) analysis showed that about 5% of Col2-ECFP positive cells were isolated from the neonatal whole calvaria (Figure S1B). FACS analysis of Col2-ECFP partial calvaria consisting of the frontal bone and a rostral part of the parietal bone demonstrated less than 0.01% ECFP positive cells. Consistent with the removal of chondrocytes, qPCR analysis using cells isolated from wild-type partial calvaria showed a marked reduction of chondrocyte marker genes such as *Col2a1* and *Sox9* (Figure S1C). Expression of osteoblast marker genes such as *Alpl*, *Ibsp* and *Bglap* were comparable between the cells isolated from whole calvaria and the partial one. These results indicate that isolating osteoblasts from partial calvaria enriched for

osteoblasts and selected against chondrocytes.

Osteoblasts were isolated from neonatal calvaria. After removal of partial interparietal bone, and supraoccipital bone as described in Figure S1I, partial calvaria were digested in three changes of 1 unit/ml Liberase TM (05401119001, Roche) for 10 minutes each at 37 °C. For fluorescence-activated cell sorting (FACS), cells were resuspended in PBS containing 2% FBS, then filtered through a 40 μm nylon cell strainer (352340, BD Falcon) before FACS.

Chondrocytes isolation

For transcription profiles of chondrocytes, rib chondrocytes were manually dissected from ribs of neonatal wild-type mice to exclude hypertrophic regions. The ribs were initially digested in 2 mg/mL PRONASE Protease (53702, EMD Millipore Corporation) for 30 minutes at 37 °C, followed by washed three time with PBS, then digested in 0.2% Collagenase D (1088858, Roche Applied Science) for 3 hours at 37 °C.

Chromatin Immunoprecipitation and DNA sequencing

Protein-DNA complexes were crosslinked with 1% formaldehyde; the crosslinking was quenched by addition of 1 M Glycine. After washing with PBS three times, cells were lysed in lysis buffer (50mM HEPES-KOH pH7.5, 140 mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton-X 100) at 4°C with agitation for 10 minutes then washed with Buffer 2 (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH8.0) for 10 minutes at room-temperature. Chromatin was fragmented by sonication (Branson digital sonifier with microtip) in Buffer 3 (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH8.0) in order to obtain 100 – 600 bp of DNA fragments. Lysate was cleared by centrifugation and supplemented with NaCl, NaDeoxycholate, TritonX 100 and Sarkosyl (L9150, Sigma-Aldrech) to the final concentrations of 130 mM, 0.13%, 1.3%, and 0.5% respectively. Dynabeads (sheep anti-mouse IgG, 11201D or sheep anti-rabbit IgG, 11203D, Life Technologies) were blocked with a blocking buffer (5 mg/ml BSA in PBS) then incubated with Anti-FLAG M2 antibody (F1804, Sigma-Aldrich) or anti-Sp7 antibody (ab22552, Abcam) in the blocking buffer overnight. Unbound antibody was removed by extensive washes in blocking buffer. The beads/antibody complex was incubated with the lysate overnight at 4°C. Beads were washed 5 times with RIPA buffer (1% NP40, 0.7% NaDeoxycholate, 1 mM EDTA, 50 mM HEPES-KOH pH7.5 and 0.5 M LiCl). DNA was eluted with elution buffer (50 mM Tris-HCl pH8.0, 10 mM EDTA, and 1% SDS) at 65°C for 15 min, and the eluate was incubated overnight at 65°C for reverse crosslinking. DNA was purified by PCR purification kit (28106, Qiagen). All buffers were supplemented with protease inhibitor cocktail mix (04693159001, Roche). ChIP-seq libraries were constructed with ThruPLEX®-FD Prep Kit (R40012, Rubicon Genomics) according to manufacturer's instruction. Library quality was validated by Bioanalyzer (Agilent) before sequenced Hiseq2000 and NextSeq500 (Illumina) platforms. We performed 50-nucleotide (nt) single-read (SR) sequencing with Hiseq2000 and 75-nt SR sequencing with NextSeq500.

ChIP-seq data analysis

DNA-sequence information was aligned to the unmasked mouse genome reference sequence mm9 by bowtie aligner (Langmead et al., 2009). Peak calling was performed by two-sample analysis on CisGenome software (Ji et al., 2008) with a P-value cutoff of 10^{-5} comparing with the input control without antibody immunoprecipitation or an IgG control. Peaks were incorporated into further analysis displaying an FDR<0.01. Peak intersection was performed by BEDTools-Version-2.16.2.

To determine evolutionary constraint on peaks, we obtained a "phastCons" conservation score from the UCSC genome browser (Rosenbloom et al., 2014) and scaled the scores to a range from 0 (least conserved) to 255 (most conserved). We then calculated the average over all nucleotide positions in the site to determine a final conservation score for each motif site. For generating heatmaps from ChIP-seq peak reads, we used the Homer program (Heinz et al., 2010).

Genomic Regions Enrichment of Annotations Tool (GREAT) Analysis

For peak distribution in genome and gene ontology analysis, GREAT GO analysis was performed utilizing the online GREAT GO program, version 2.0.2 (McLean et al., 2010). Each peak category was run against whole genome background with assembly mm9. Gene regulatory domains utilized for region annotation were

defined as minimum 5.0 kb upstream and 1.0 kb downstream of the TSS, and extended up to 1000.0 kb to the nearest gene's minimal regulatory domain. MGI expression, GO Biological Processes and Mouse Phenotype annotations were assessed for each peak category.

Motif analysis

de novo motif analysis were performed using the Gibbs motif sampler provided in the CisGenome package (Ji et al., 2008), and DREME (Bailey, 2011). A 100 bp region surrounding peak center was extracted from mm9 and used for these analyses. To search the potential binding protein to *de novo* motif, we compared the motif position weight matrices (PWMs) to all known human and mouse motifs in the TRANSFAC database (Matys et al., 2003), UniProbe database (Hume et al., 2014), TFClass (Wingender et al., 2013) and information in literature (Badis et al., 2009; Berger et al., 2008). To examine distribution of the identified motifs in the peaks, we mapped each PWM back to the whole mouse genome (mm9) using a likelihood ratio (LR) cutoff of 500 comparing to a pre-calculated third-order Markov chain background model; peaks are normalized to 800 bp window at the peak center. Peaks with identified motif were obtained by intersection between the peaks and motif-mapped genomic regions by BEDTools-Version-2.16.2.

Assigning peaks to nearest neighboring genes

We assigned a detected peak region to each member of its nearest gene set by cisGenome (Ji et al., 2008) package, which include: (a) the nearest gene, which is the gene with the smallest distance from the center of the gene body to the center of the peak region; (b) the nearest upstream gene, which is the gene that located at the 5' upstream side of the peak region and has the smallest distance from its TSS to the center of the peak region; (c) the nearest downstream gene, which is the gene that located at the 3' downstream side of the peak region and has the smallest distance from its TSS to the center of the peak region. The annotation of genes and gene isoforms in the mouse mm9 genome assembly was obtained from the UCSC Genome Browser (Rosenbloom et al., 2014).

ChIP-qPCR analysis

ChIP-qPCR was performed with SYBR Premix Ex Taq II (RR820, Takara). Fold enrichment was calculated by normalizing ChIP sample against input, and target region against control region as follows. ΔCt=Ct(ChIP)-Ct(input); $\Delta\Delta Ct$=ΔCt(target region)-ΔCt(control region); Fold enrichment=$2^{-\Delta\Delta Ct}$. Primer sequences for ChIP-qPCR are shown in Table S5

RT-qPCR and RNA-seq analysis

Total RNA was extracted and purified by RNeasy mini kit (74106, Qiagen) or NucleoSpin RNA (740955, Clontech). For cDNA synthesis, 1 μg of total RNA was used with the qScript™ cDNA SuperMix (95048, Quanta Bioscience) according to the manufacturer's instructions. mRNA expression was measured by qPCR using SYBR Premix Ex Taq II (RR820, Clontech) with the ViiA 7 (Applied Biosystems). qPCR reactions were run in triplicate and mRNA expression values normalized to GAPDH expression levels. Heat map in Figure 5I was generated by pheatmap package in R. Primer sequences are shown in Table S5.

For RNA-seq, 200 ng of total RNA was subjected to analysis. RNA quality was confirmed by Experion (700-7105, Bio-lad) analysis and the RNA libraries were constructed by Tru-Seq Illumina kit (RS-122-2001, illumina) with some modifications. Sequence was performed with Hiseq2000 (Illumina). The outputted sequence reads were aligned and mapped using Partek Flow software v2.2 (Partek). The raw reads were first subjected to pre-alignment Quality Assurance and Quality Control (QA/QC). Any base below a Phred value of 20 was trimmed from either side of the read and reads shorter than 25 nt length were removed. The processed reads were aligned by Tophat2-2.0.8 (Kim et al., 2013) to mm9 reference genome. The mapping quality and coverage were checked by post-alignment QA/QC. Aligned reads were quantified by an expectation maximization algorithm (Xing et al., 2006), followed by a normalization of mapped reads per kilo base length of transcript per million reads (RPKM). As a control, mouse embryo fibroblast (MEFs; GSM1173355 and GSM1173356) RNA-seq data was used from a published resource (Hou et al., 2013).

Hierarchical clustering with RNA-seq data in Figure S1J was performed with selected genes based on the following three criteria: fold change >2, false discovery rates (FDR) < 0.05, and reads per kilobase per

million (rpkm) >2 by one-to-one comparisons (i.e., osteoblasts to chondrocytes, osteoblasts to MEFs and chondrocytes to MEFs), followed by pooling of these three sets together for analysis by Partek Genomic Suite 6.6. The gene-lists obtained are presented in Table S2. Hierarchical clustering in Figure 4C was performed with selected genes based on rpkm value more than 5. Gene lists we obtained are shown in Table S4. Each value in these heatmaps is standardized using the following formula: Xis = [Xi - Mean(X)] / std.dev(X), where Xi = expression of gene X for sample i., Xis = standardized expression of gene X for sample i, Mean (X) = average expression of gene X across all samples, and std.dev(X) = standard deviation of gene X across all samples. Gene Ontology analysis was performed by Database for Annotation, Visualization and Integrated Discovery (DAVID Bioinformatics Resources 6.7) (Huang da et al., 2009).

Correlation analysis between ChIP-seq peak reads and gene expression
Peak reads were obtained with the Homer program (Heinz et al., 2010), assigned to the nearest neighboring genes and all peak reads summed for each genes then data was intersected by Cistrome (Liu et al., 2011). Box plots were generated by R program (Team, 2011). Gene expression levels for nearest gene neighbors were compared to Sp7 engagement around those genes through a Cistrome intersectional analysis (Liu et al., 2011), generating box plots for Sp7 associated and non-associated sets by R. The p-value was calculated by Wilcoxon rank sum test with continuity correction in R.

In vitro cell culture
The following cell lines were used in the study: NIH 3T3 fibroblast (CRL-1658, ATCC), MC3T3E1 subclone 4 (CRL-2593, ATCC), 293T (CRL-3216, ATCC), COS-7 (CRL-1651, ATCC) and C3H10T1/2 clone 8 (CRL-226, ATCC). 3T3, 293T, COS-7 and C3H10T1/2 cells were cultured in high-glucose DMEM containing 10% FBS and 1% penicillin/streptomycin. MC3T3E1 cells were cultured in alpha-MEM (A10490-01, Life technologies) containing 10% FBS and 1% penicillin/streptomycin (10% FBS/alpha-MEM). For osteoblast differentiation, MC3T3E1 cells were cultured in 10% FBS/alpha-MEM supplemented with 50 μg/mL ascorbic acid phosphate, 10 mM β-glycerophosphate and 10 ng/ml recombinant human BMP‐2 (rhBMP-2, 355-BM, R&D) for 7 days.

Vectors
pCMV-Myc-Sp1, pCMV-Myc-Sp7, pCMV-Myc-Dlx5, domain deletion of pCMV-Myc-Sp7 and that of pCMV-Myc-Dlx5, pCMV-Myc-Sp7mutant(A305T), pCMV-Myc-Sp7mutants (A305T-E339Q-Q370I), pCMV-Myc-Sp1ZF/Sp7, pCMV-Myc-Sp7ZF/Sp1, pcDNA3.1-Sp7, pcDNA3.1-Sp7-BioFL, pcDNA3.1-Sp7mutant, pcDNA3.1-Sp7mutant-BioFL were constructed by PCR strategy using pCMV-Myc (635689, Clontech) and pcDNA3.1 (V79520, Life technologies) vectors. Domain deletion and point mutation constructs were constructed by PCR strategy. For retrovirus vectors, pMx-GFP, pMx-FLAG-Dlx5, pMx-Sp1-BioFL, pMx-Sp7-BioFL, pMx-Sp7mutant-BioFL, pMx-FLAG-Sp7 and pMx-FLAG-Sp7ZF/Sp1 were constructed by PCR strategy using pMx-IRES-bsr vector (Kitamura, 1998). For the retrovirus infections, these were packaged into Plat-E cells (Morita et al., 2000). For shRNA lentivirus infection, pLKO-scramble shRNA, shSp7 (#1: TRCN0000423959, #2: TRCN0000082144), shDlx3 (TRCN0000430532), shDlx5 (TRCN0000428940) and shDlx6 (TRCN0000238872) from MISSION shRNA (Sigma-Aldrich), psPAX2 (Addgene Plasmid #12260, Cambridge, MA) and pMD2.G (Addgene Plasmid #12259) were used.

In vitro reporter analysis
Cells were plated onto 24-well plates and 0.4 μg of a DNA mixture containing the test reporter plasmids, control reporter plasmids encoding Renilla luciferase and effector plasmids introduced by Fugene HD transfection (E2311, Promega). A dual luciferase assay was performed, as described previously (Ohba et al., 2008). Retrovirus infection was performed as described previously (Kan et al., 2009) and shRNA lentivirus infection was performed according to the manufacturer's instruction (Sigma-Aldrich, MISSION shRNA). The genomic regions used for the reporter analyses are shown in Table S6.

Immunostaining and LacZ staining
Tissues were fixed in 4% paraformaldehyde (PFA)/PBS for 1 hour at 4 °C and soaked in 30% sucrose/PBS

overnight at 4 °C. After embedding in OCT compound, cryosections were cut at 12 μm intervals. Sections were blocked with 3% bovine serum albumin (A7960, Sigma-Aldrich) and 1% heat inactivated sheep serum (S2263, Sigma-Aldrich) in PBST. The following antibodies were used in the study: anti-Sp7 (1:5,000, ab22552, Abcam), anti-FLAG M2 (1:500, F3165, Sigma-Aldrich), anti-GFP (1:500, GFP-1020, Aves labs), Alexa Fluor 488 (1:500, secondary anti-chicken; anti-rabbit, Life Technologies), and Alexa Fluor 568 (1:500, secondary anti-rabbit; anti-mouse, Life Technologies). Primary antibodies were incubated overnight at 4 °C, followed by incubation with secondary antibodies for 1 hour at room temperature. Sections were mounted with Immu-Mount (99-904-02, Fisher Scientific). For Immunostaining with anti-FLAG M2 antibody, we used a Mouse on Mouse (M.O.M.) Basic Kit (BMK-2202, Vector Laboratories). For whole mount lacZ staining, P1 reporter mice were fixed in 1% formaldehyde/ 0.2% glutaraldehyde in PBS for 90 min, then stained in X-gal for 90 min at 37°C, and cleared in 80% glycerol/PBS

Co-immunoprecipitation, SDS-PAGE and western blotting

Nuclear extracts were obtained with a Nuclear Complex Co-IP Kit (54001, Active motif) according to the manufacturer's instruction. Co-IP was performed using 500 ug of nuclear extract, fifty microliters of Dynabeads M-280 Sheep Anti-Mouse IgG (11201D, Life technologies) and five microgram of Anti-FLAG M2 antibody (F1804, Sigma-Aldrich), followed by SDS-PAGE analysis of protein products. Immunoblotting was performed using the following primary antibodies: anti-Myc tag antibody (1:1000, ab9106, Abcam), Anti-FLAG M2 antibody (1:1000, F1804, Sigma-Aldrich), anti-Dlx5 antibody (1:1000, C10748, Assay Biotechnology), anti-Sp7 antibodies (1:3000, ab22552 or 1;250, ab94744, Abcam), anti-GAPDH antibody (1:3000, ab9385, Abcam) and anti-Histone H3 histone antibody (1:3000, ab1791, Abcam). The following secondary antibodies were used for primary antibody detection: HRP-conjugated goat anti-rabbit IgG and HRP-conjugated goat anti-mouse IgG (1:10,000, W4011 and W4021, Promega).

EMSA

For the binding reactions for the AT-rich motif in Figure 4D, we used in vitro translated proteins generated with a PURExpress In Vitro Protein Synthesis Kit (E6800, New England Biolabs) according to manufacturer's instruction and 1 μl of the protein was used for the binding reaction; in Figure 5F, 5G and Figure S4E, nuclear extracts were generated, as previously described (Dignam et al., 1983), from COS7 cells transfected with appropriate expression vectors and 1 μg of the nuclear protein extract was used in a binding reaction. The binding mixture contained 50 mM KCl, 25 mM HEPES-KOH (pH 7.9), 5 mM MgCl2, 50 μM Zn(OAc)2, 1 mM DTT, 12% glycerol, 0.1% NP-40, 1.5 μg of Poly d(I-C), 1.0 μg of Poly d(G-C) and 100 fmol of the Biotin-labeled probe. The binding reaction was placed on ice for 30min. In Figure 4D unlabeled probes added at 100-fold molar excess for competition analysis; in Figure 5G, unlabeled probes added at 5-,and 50-fold molar excess for competition analysis. For supershift assay, we used anti-Glutathione-S-Transferase (GST) antibody (G7781, Sigma-Aldrich), anti-FLAG M2 antibody (Sigma-Aldrich, F1804), anti-Myc antibody (ab9132), anti-DLX5 antibody (C10748, Assay Biotechnology), anti-SP1 antibody (07-645, EMD Millipore), anti-SP7/osterix antibody (ab22552, Abcam) and non-immune IgG (ab46540, Abcam) incubating on ice for 30 min prior to binding with probes.

For binding reactions for GC-box in Figures 6F, 6G, 6H, S5D, S5E, S5F and S5H, nuclear extracts were prepared from COS7 cells transfected with appropriate expression vectors (Figures 6F, 6G, 6H, S5E and S5H), MC3T3E1 cells cultured for 7 days in osteogenic medium (Figure S5D), and from *in vitro* translated proteins prepared by PURExpress In Vitro Protein Synthesis Kit (Figure S5F). The binding mixture contained 50 mM KCl, 25 mM HEPES-KOH (pH 7.9), 5 mM MgCl2, 50 μM Zn(OAc)2, 1 mM DTT, 12% glycerol, 0.1% NP-40, 0.5 μg of Poly d(I-C) and 100 fmol of the Biotin-labeled probe. The binding reaction was performed for 30 min at room temperature. For supershift assays, we used 1 μl of anti-SP1 antibody (07-645; EMD Millipore), anti-SP7/osterix antibodies (ab22552, Abcam), anti-Glutathione-S-Transferase (GST) antibody (G7781, Sigma-Aldrich), anti-Myc tag antibody (ab9106 or ab9132, Abcam) and non-immune IgG (ab46540, Abcam) incubating for 30 min at room temperature prior to binding with DNA probes. The sequence of each DNA probe is shown in Table S5.

Alignment of zinc finger domains in Sp7 and the most closely related Sp7 counterparts amongst

different chordate species

The alignment was performed with sequences from Amphioxus (*Branchiostoma floridae*: XP_002598426), tunicate (*Ciona intestinalis*: XP_002120430), lamprey (*Lethenteron camtschaticum*: BAJ78789), elephant shark (*Callorhinchus milii*: XP_007909749.1), zebrafish (*Danio rerio*: NP_998028.1), coelacanth (*Latimeria chalumnae*: XP_005986366.1), frog (*Xenopus tropicalis*; NP_001128590.1), lizard (*Gekko japonicas*: XP_015277714), alligator (*Alligator mississippiensis*: KQL63681.1) bird (*Cuculus canorus*: XP_009555105.1), mouse (*Mus musculus*: NP_569725.1) and human (*Homo sapiens*: NP_690599.1).

Supplemental References

Aberg, T., Rice, R., Rice, D., Thesleff, I., and Waltimo-Siren, J. (2005). Chondrogenic potential of mouse calvarial mesenchyme. J Histochem Cytochem 53, 653-663.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., *et al.* (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720-1723.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. In Nature (England), pp. 53-59.

Chokalingam, K., Juncosa-Melvin, N., Hunter, S.A., Gooch, C., Frede, C., Florert, J., Bradica, G., Wenstrup, R., and Butler, D.L. (2009). Tensile stimulation of murine stem cell-collagen sponge constructs increases collagen type I gene expression and linear stiffness. Tissue Eng Part A *15*, 2561-2570.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W.M., 3rd, Yanagimachi, R., and Jaenisch, R. (2001). Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. Proc Natl Acad Sci U S A *98*, 6209-6214.

Hayashi, S., Lewis, P., Pevny, L., and McMahon, A.P. (2002). Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mech Dev *119 Suppl 1*, S97-s101.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44-57.

Kan, A., Ikeda, T., Saito, T., Yano, F., Fukai, A., Hojo, H., Ogasawara, T., Ogata, N., Nakamura, K., Chung, U.I., *et al.* (2009). Screening of chondrogenic factors with a real-time fluorescence-monitoring cell line ATDC5-C2ER: identification of sorting nexin 19 as a novel factor. Arthritis Rheum *60*, 3314-3323.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kitamura, T. (1998). New experimental approaches in retrovirus-mediated expression screening. Int J Hematol *67*, 351-359.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., *et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., *et al.* (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res *31*, 374-378.

Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther 7, 1063-1066.

Ohba, S., Kawaguchi, H., Kugimiya, F., Ogasawara, T., Kawamura, N., Saito, T., Ikeda, T., Fujii, K., Miyajima, T., Kuramochi, A., *et al.* (2008). Patched1 haploinsufficiency increases adult bone mass and modulates Gli3 repressor activity. Dev Cell *14*, 689-699.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., *et al.* (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res *43*, D670-681.

Team, R.D.C. (2011). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Xing, Y., Yu, T., Wu, Y.N., Roy, M., Kim, J., and Lee, C. (2006). An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs. Nucleic Acids Res 34, 3150-3160.