Genomics Data 5 (2015) 103–105

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Gene expression profiles of *Bapx1* expressing FACS sorted cells from wildtype and *Bapx1-EGFP* null mouse embryos

Sumantra Chatterjee [a], V. Sivakamasundari [b], Petra Kraus [c], Sook Peng Yap [b], Vibhor Kumar [b], Shyam Prabhakar [b], Thomas Lufkin [c]

[a] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, Baltimore, MD 21205 USA
[b] Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore
[c] Department of Biology, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699, USA

ARTICLE INFO

Article history:
Received 13 May 2015
Received in revised form 23 May 2015
Accepted 24 May 2015
Available online 3 June 2015

Keywords:
Bapx1
Mouse
FACS
ChIP-Seq

ABSTRACT

The data described in this article refers to Chatterjee et al. (2015) "In vivo genome-wide analysis of multiple tissues identifies gene regulatory networks, novel functions and downstream regulatory genes for *Bapx1* and its co-regulation with *Sox9* in the mammalian vertebral column" (GEO GSE35649) [1]. Transcriptional profiling combined with genome wide binding data is a powerful tool to elucidate the molecular mechanism behind vertebrate organogenesis. It also helps to uncover multiple roles of a single gene in different organs. In the above mentioned report we reveal the function of the homeobox gene *Bapx1* during the embryogenesis of five distinct organs (vertebral column, spleen, gut, forelimb and hindlimb) at a relevant developmental stage (E12.5), microarray analysis of isolated wildtype and mutant cells in is compared in conjunction with ChIP-Seq analysis. We also analyzed the development of the vertebral column by comparing microarray and ChIP-Seq data for *Bapx1* with similarly generated data sets for *Sox9* to generate a gene regulatory network controlling various facets of the organogenesis.

Specifications	
Organism/cell line/tissue	Mus musculus (C57BL/6J)
Sex	*Pooled male and female embryos*
Sequencer or array type	*Microarray* - MouseWG-6 v2.0 Expression BeadChip microarrays (Illumina). *ChIP Sequencing* -Genome Analyzer II/IIx (Illumina)
Data format	*Analyzed*
Experimental factors	*Wildtype* versus *Bapx1 and Sox9 null embryos at E12.5*
Experimental features	*Microarray was conducted on cells sorted from 5 distinct tissues expressing Bapx1 in both wildtype and Bapx1$^{-/-}$ embryos at E12.5. ChIP-Seq was conducted on ~ 100 vertebral columns dissected out of wildtype embryos at E12.5 to determine genome wide binding.*
Consent	*Level of consent allowed for reuse if applicable (typically for human samples) NA*
Sample source location	*NA*

1. Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35649.

2. Data, experimental design, materials and methods

2.1. Overall experimental strategy

The data presented here pertains to Chatterjee et al. (2015) "In vivo genome-wide analysis of multiple tissues identifies gene regulatory networks, novel functions and downstream regulatory genes for Bapx1 and its co-regulation with Sox9 in the mammalian vertebral column." [1]. We have used stable transgenic mouse lines expressing EGFP under the endogenous promoter of various transcription factors expressed in various organs in the developing embryo to sort out cells to do further genomic experiments like transcriptional profiling and genome wide binding assays (Fig. 1). The study shows the utility of such a method for studying gene regulatory networks in a specific cell population of particular organ as marked by specific transcription factors.

2.2. Transgenic mouse generation

The murine BAC clone RP24-148P5 containing genomic DNA from C57BL/6 J flanking the *Bapx1* gene was obtained from the BACPAC Resources Centre at Children's Hospital Oakland Research Institute (CHORI) and targeting constructs were generated via the Quick and Easy BAC modification kit (Gene Bridges) according to the manufacturer's protocols. The *Bapx1$^{tm2.Tlu}$* allele was generated as previously reported [2] to create wildtype mice expressing EGFP

http://dx.doi.org/10.1016/j.gdata.2015.05.031

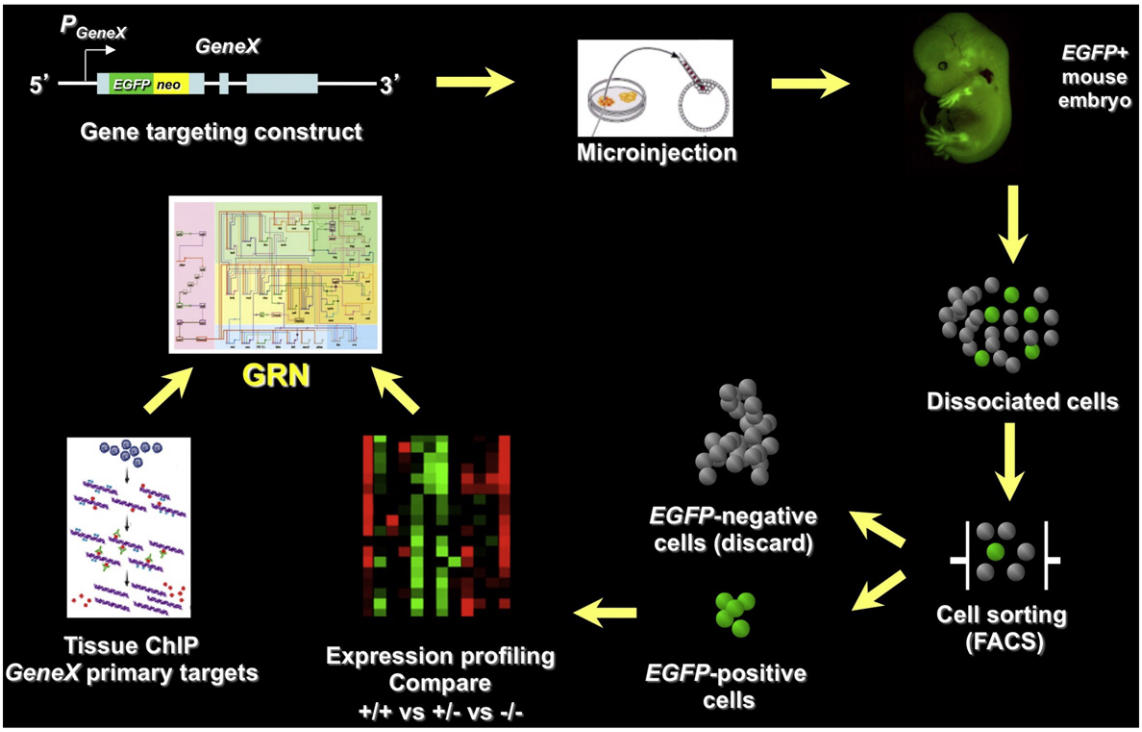

Fig. 1. Workflow for cell-type specific transcriptional profiling via combined transgenic and genomics approaches. *GeneX* represent any gene of interest. In the case described here it is *Bapx1*. *GeneX* is inactivated by insertion of *EGFP* into the coding sequence by homologous recombination in embryonic stem (ES) cells. Chimeric animals which are able to transmit the *EGFP* modified gene are generated by blastocyst injection. Embryos from germ-line transmitting chimeras or from the resulting heterozygous founder animals are used to collect embryos for dissociation into single cells. FACS allows for the separation of *EGFP* positive and negative cells which are subsequently used for expession profiling analysis. The results of the expression profiling combined with the ChIP data allows for the building of the gene regulatory network (GRN). Blue boxes represent exonic sequences. *P-GeneX* represents the promoter of *GeneX*.

and *Bapx1* concurrently in the relevant cells and tissues. The *Bapx1*$^{tm4.1Tlu}$ allele was generated by deleting 90 bp of the endogenous sequence by inserting the EGFP-FRT-PGK-gb2-Neo-FRT cassette immediately upstream of the translational start codon of *Bapx1*, to create an EGFP expressing *Bapx1* null. Two southern blot confirmed gene targeted ES cell clones with normal karyotype were subsequently microinjected into 2–8-cell stage embryos isolated from C57BL/6 mice to generate chimeric mice as previously described [3]. The chimeric mice were bred to wildtype mice to generate stable lines. The *FRT*-flanking neomycin cassette in the targeted *Bapx1* allele was deleted out by breeding to the *FLPe*–deleter mice 129S4/SvJaeSor-*Gt*(*ROSA*)*26Sor tm1*(*FLP1*)*Dym*/J (Stock # 3946) from Jackson Laboratories [4,5]. Routine PCR genotyping of *Bapx1* lines was performed as essentially described in [6]. Constructs of a similar nature were generated for the *Sox9* locus to create *Sox9*^{tm11Tlu} [7] and *Sox9*$^{-/-(EGFP)}$ mice that are to be detailed later.

2.3. Sorting EGFP expressing cells from mouse embryos

Embryos expressing EGFP in the *Bapx1 or Sox9* expression domains were identified under a fluorescent dissection microscope (Leica). Vertebral columns, spleens, guts, hindlimbs and forelimb from Bapx1 tagged line and only vertebral column from Sox9 tagged line were separately dissociated into single cells in a solution made of 100 U/ml Collagenase I & II, 50 U/ml DNAse and 0.05% Trypsin (Invitrogen) by serially filtering them through a 100uM and 40uM cell strainer. The cell pellet was resuspended in 5% FBS, 4 mM EDTA in Leibovitz L-15 medium for cell sorting using FACSAria (BD Biosciences).

2.4. Gene expression microarrays and analyses

For each sorted cell population in both wildtype as well as null embryos, 4 biological replicates were used to extract total RNA using TRIzol (Invitrogen) followed by the RNeasy Micro Kit (Qiagen). Integrity of the RNA was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). TargetAmp™-Nano Labeling Kit for Illumina Expression BeadChip (Epicenter Biotechnologies) was used to label 25 ng total RNA from each sample. MouseWG-6 v2.0 Expression BeadChip microarrays (Illumina) were used to hybridize the samples according to Illumina guidelines. GenomeStudio software (Illumina) was used to prepare background-subtracted data. The background-subtracted data was then imported into Partek Genomics Suite (Agilent) for data normalization. The probe intensities from various biological replicates were normalized using scale normalization using median absolute deviation as a spread measure. This was further filtered on percentile (lower 20 and upper 100) by expression of probe sets without averaging over replicates. These probe sets, which passed this filter, were further used to calculate gene expression changes between different conditions. ANOVA with nominal alpha value set to 0.05 was then used to determine the probe sets significantly different between the different genotypes compared. To reduce the false positive rate, Benjamini and Hochberg Multiple testing correction was applied, and probe sets with expression fold change >2.0 or >1.5 between the different genotypes were selected for further validation and analyses. Data has been deposited in Gene Expression Omnibus (GEO) as described above. Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) analyses were performed using the web-based tools.

2.5. Functional annotation of genes controlled by Bapx1 and Sox9

Our microarray analysis revealed that while *Bapx1* affected a large number of genes (with >2 fold enrichment) in the spleen ($n = 980$), gut ($n = 143$) and vertebral column ($n = 125$), it controlled only several genes in the hind- ($n = 30$) and forelimb ($n = 2$). Sox9 controlled 3593 genes in the vertebral column with >1.5 fold enrichment between the wildtype and the Sox9 null. Analysis of genes controlled by Bapx1 in the Vertebral column showed enrichment of

Table 1
Signaling pathways affected by Bapx1 in the vertebral column.

Bapx1 in Vertebral column	
Pathways	**-log (*p*-value)**
Wnt/β-catenin Signaling	1.26
Inhibition of Matrix Metalloproteases	1.41
VEGF Signaling	1.45
Bapx1 in Spleen	
Pathways	**-log (*p*-value)**
FGF Signaling	1.82
ERK/MAPK Signaling	2.06
Epithelial Adherens Junction Signaling	4.2
Bapx1 in Gut	
Pathways	**-log (*p*-value)**
ERK/MAPK Signaling	1.61
Regulation of the Epithelial-Mesenchymal Transition Pathway	1.66
VEGF Signaling	1.79
Sox9 in Vertebral column	
Pathways	**-log (*p*-value)**
Differentiation of chondrocytes	5.46
Apoptosis	13.41
Proliferation of cells	27.42

genes in the VEGF and Wnt signaling pathways, as well as the inhibition of matrix metalloproteases (Table 1). Similar analysis for genes regulated by Bapx1 in the spleen showed enrichment of genes of the FGF signaling pathway, which is critical for cell survival, differentiation, morphogenesis and angiogenesis. Other enriched genes included members of the epithelial adherens junction signaling and ERK/MAPK signaling. Signaling pathways such as VEGF signaling, ERK/MAPK signaling and the epithelial-mesenchymal transition pathway were enriched in the gut similar to our observations in the spleen. (Table 1). We also looked at the genes which are controlled by *Sox9* in the vertebral column which revealed that most had known roles in chondrocyte differentiation, proliferation and apoptosis (Table 1).

2.6. ChIP-assay and peak calling

For Bapx1 and Sox9 ChIP, vertebral columns from ~100 E12.5 (both S-peptide tagged for Baxp1 and wildtype for Sox9) embryos were dissected and 2 mg of chromatin was used for ChIP as previously described [8] with anti S-Peptide antibody (Bethyl laboratories, A190-134A). For Sox9 ChIP anti-Sox9 antibody (R&D Systems, AF3075) was used for immunoprecipitation. 10-15 ng of purified ChIP DNA from each sample was used to synthesize the sequencing library as per manufacturer's instructions (Illumina). Sequence reads produced by Illumina Genome Analyzer II/IIx that passed the signal purity filtering were mapped to the mouse genome mm9, using the Illumina Genome Analyzer Pipeline. All uniquely mapped reads that are with two or fewer mismatches were retained. Genomic binding sites were identified using the peak calling algorithm MACS (version 1.4.0 beta) with default settings (band width = 300, model fold = 10, 30, *p*-value cutoff = 1.00e-05, range for calculating regional lambda = 1000 and 10,000 bps) [9]. For Bapx1 ChIP, untagged wildtype embryos were used as control and for Sox9 the input DNA was used as control. For annotation of the ChIP-Seq peaks, first the transcription start site (TSS) of the Refseq gene nearest to peaks were found and reported. If the peaks resided in intragenic region of nearest gene it was called "intragenic", otherwise "intergenic". If the peaks resided within 5000 bp of TSS they were called as "promoter" peaks. All data has been deposited in Gene Expression Omnibus (GEO) under accession number GSE35877.

We detected 2815 Bapx1 peaks with 540 of them located at the TSS of various genes. A large proportion of the binding sites were distal from the TSS, 1052 located at >25 kb distally in the intergenic and 722 located >25 kb in the intragenic regions. Similarly for Sox9, we identified 3722 peaks.

2.7. Motif analysis

Peaks called by MACS were ranked as described [1]. The top 200 peaks were used for motif analysis and the repeat masked genome sequence +/− 50 bp from the summit of these 200 peaks was downloaded from the UCSC genome browser (http://genome.ucsc.edu/). We identified the Bapx1 binding motif T(G/A)AGTG [10], enriched in about 40% of the Bapx1 binding sites in the genome specifically near genes controlling cartilage and bone morphogenesis. We also identified the DNA binding motif for Sox9 (A/T)(A/T)CAA(A/T)G identified mostly by in vitro studies [11].

2.8. Bapx1-Sox9 interplay in vertebral column

Of the group of genes repressed by Bapx1 and activated by Sox9, over 85% had both the Sox9 and Bapx1 binding sites within 200 kb of their TSS. These genes represent potential primary targets of both Bapx1 and Sox9, and under opposing regulatory control in the vertebral column to maintain a distinct neuronal and osteo-chondrogenic identity in the developing embryo. This group includes genes like Nrn1, Ctnna2, and Gata3 that are known neuronal genes but have so far not been implicated to be regulated by either Sox9 or Bapx1. Over 75% of the genes repressed by both Bapx1 and Sox9 had binding sites for both Bapx1 and Sox9 within 200 kb of their TSS. These genes are also potential primary targets of both Bapx1 and Sox9, but repressed by both. This group includes the osteogenic, metalloproteinase and angiogenesis genes *Adamts4*, *Col1a1*, *Mmp11*, *Emcn*, *Nrp1* and *Ctsk*.

References

[1] S. Chatterjee, V. Sivakamasundari, S.P. Yap, P. Kraus, V. Kumar, X. Xing, S.L. Lim, J. Sng, S. Prabhakar, T. Lufkin, In vivo genome-wide analysis of multiple tissues identifies gene regulatory networks, novel functions and downstream regulatory genes for Bapx1 and its co-regulation with Sox9 in the mammalian vertebral column. BMC Genomics 15 (2014) 1072, http://dx.doi.org/10.1186/1471-2164-15-1072.

[2] V. Sivakamasundari, H.Y. Chan, S.P. Yap, X. Xing, P. Kraus, T. Lufkin, New Bapx1(Cre-EGFP) Mouse Lines for Lineage Tracing and Conditional Knockout Studies. Genesis 50 (2011) 375–383, http://dx.doi.org/10.1002/dvg.20802.

[3] P. Kraus, G. Leong, V. Tan, X. Xing, J.W. Goh, S.P. Yap, T. Lufkin, A more cost effective and rapid high percentage germ-line transmitting chimeric mouse generation procedure via microinjection of 2-cell, 4-cell, and 8-cell embryos with ES and iPS cells. Genesis 48 (2010) 394–399, http://dx.doi.org/10.1002/dvg.20627.

[4] F.W. Farley, P. Soriano, L.S. Steffen, S.M. Dymecki, Widespread recombinase expression using FLPeR (flipper) mice. Genesis 28 (2000) 106–110, http://dx.doi.org/10.1002/1526-968X(200011/12)28:3/4.

[5] C.I. Rodriguez, F. Buchholz, J. Galloway, R. Sequerra, J. Kasper, R. Ayala, A.F. Stewart, S.M. Dymecki, High-efficiency deleter mice show that FLPe is an alternative to CreloxP. Nat. Genet. 25 (2000) 139–140, http://dx.doi.org/10.1038/75973.

[6] V. Sivakamasundari, H.Y. Chan, S.P. Yap, X. Xing, P. Kraus, T. Lufkin, New Bapx1Cre-EGFP mouse lines for lineage tracing and conditional knockout studies. Genesis 50 (2012) 375–383, http://dx.doi.org/10.1002/dvg.20802.

[7] H.Y. Chan, V. Sivakamasundari, X. Xing, P. Kraus, S.P. Yap, P. Ng, S.L. Lim, T. Lufkin, Comparison of IRES and F2A-based locus-specific multicistronic expression in stable mouse lines. PLoS One 6 (2011)http://dx.doi.org/10.1371/journal.pone.0028885 e28885.

[8] D. Schmidt, M.D. Wilson, C. Spyrou, G.D. Brown, J. Hadfield, D.T. Odom, ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods 48 (2009) 240–248, http://dx.doi.org/10.1016/j.ymeth.2009.03.001.

[9] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu, Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9 (2008) R137, http://dx.doi.org/10.1186/gb-2008-9-9-r137.

[10] D.W. Kim, H. Kempf, R.E. Chen, A.B. Lassar, Characterization of Nkx3.2 DNA binding specificity and its requirement for somitic chondrogenesis. J. Biol. Chem. 278 (2003) 27532–27539, http://dx.doi.org/10.1074/jbc.M301461200.

[11] Mertin, S., S.G. McDowall, V.R. Harley, The DNA-binding specificity of SOX9 and other SOX proteins, Nucleic Acids Res., 27 (1999) 1359–1364. http://dx.doi.org/10.1093/nar/27.5.1359