Genomics Data 10 (2016) 51–53

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in brief

Genome wide binding (ChIP-Seq) of murine Bapx1 and Sox9 proteins in vivo and in vitro

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

[a] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, Baltimore, MD 21205, USA
[b] Department of Biology, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699, USA
[c] The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06030, USA
[d] Apta Biosciences, 31 Biopolis Way, Nanos, Singapore 138669, Singapore
[e] Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore

ARTICLE INFO

Article history:
Received 31 August 2016
Accepted 7 September 2016
Available online 8 September 2016

Keywords:
Genome binding
Bapx1
Mouse
ChIP-Seq
Occupancy profiling by NGS

ABSTRACT

This work pertains to GEO submission GSE36672, in vivo and in vitro genome wide binding (ChIP-Seq) of Bapx1/Nkx3.2 and Sox9 proteins. We have previously shown that data from a genome wide binding assay combined with transcriptional profiling is an insightful means to divulge the mechanisms directing cell type specification and the generation of tissues and subsequent organs [1]. Our earlier work identified the role of the DNA-binding homeodomain containing protein Bapx1/Nkx3.2 in midgestation murine embryos. Microarray analysis of EGFP-tagged cells (both wildtype and null) was integrated using ChIP-Seq analysis of Bapx1/Nkx3.2 and Sox9 DNA-binding proteins in living tissue.

Specifications [*standardized info for the reader*] Where applicable, please follow the Ontology for Biomedical Investigations: http://obi-ontology.org/page/Main_Page	
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	*E12.5 Wildtype and gene-targeted Bapx1/Nkx3.2 and Sox9 mouse embryos*
Experimental features	*E12.5 embryos were isolated. We performed ChIP-Seq on vertebral columns isolated from wildtype and gene-targeted embryos 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=GSE36672

* Corresponding author.
E-mail address: tlufkin@clarkson.edu (T. Lufkin).

2. Materials, Methods and Experimental approach

2.1. Experimental approach

The information presented here has been previously described in Chatterjee et al. (2014) "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]. An overview of the experimental approach is shown in Fig. 1. The Bapx1/Nkx3.2 protein was tagged *in vivo* with the S-peptide epitope to give specificity to the isolation of the Bapx1/Nkx3.2 S-peptide fusion protein for immunoprecipitation. In a similar fashion, the *Bapx1/Nkx3.2* cDNA was engineered so as to have an in-frame fusion with the V5 epitope tag that has a well-established high degree of specificity for the corresponding anti-V5 antibody and is commonly used in *in vitro* cell culture systems following the overexpression of the V5-tagged protein. In this case the V5-Bapx1/Nkx3.2 fusion protein was employed in electrophoretic mobility shift assays (EMSA). The vertebral columns of mid-gestation (E12.5) mouse embryos carrying the *Bapx1/Nkx3.2* S-peptide allele were carefully dissected away from surrounding embryonic tissues, disaggregated into micro pieces and chromatin immunoprecipitation was performed to identify the DNA sequences bound by Bapx1/Nkx3.2 protein in the native *in vivo* murine vertebral column.

http://dx.doi.org/10.1016/j.gdata.2016.09.002

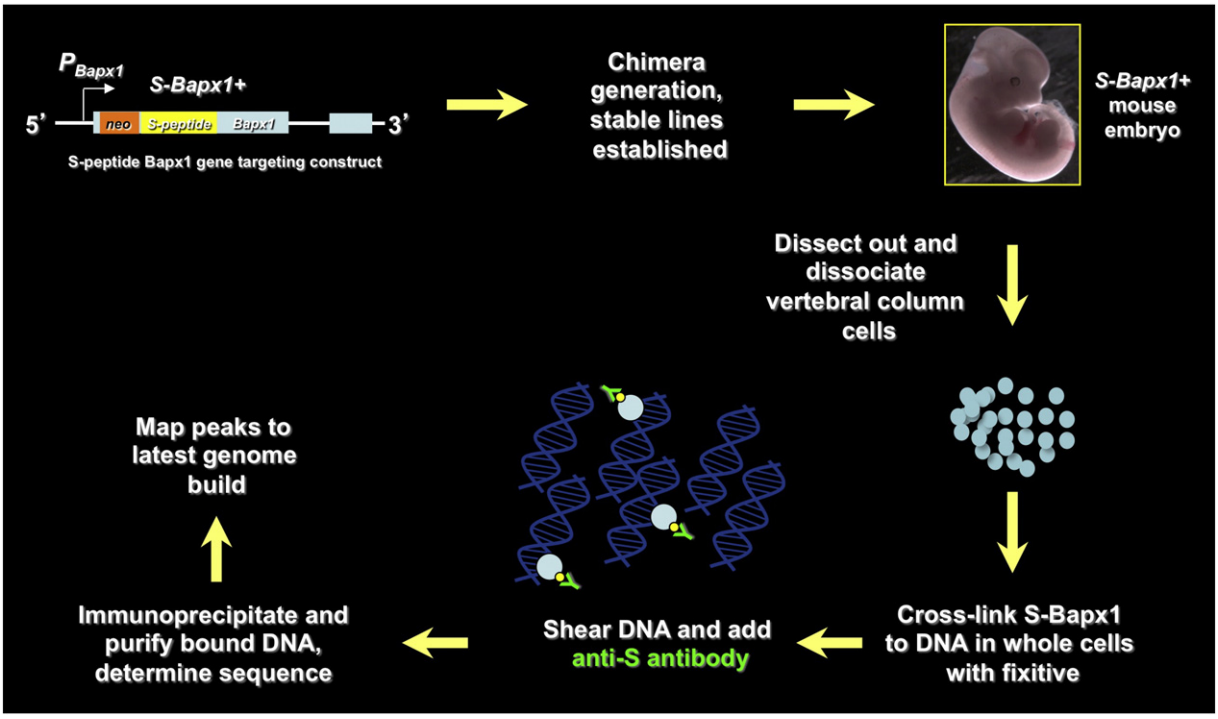

Fig. 1. Overview of the strategy for performing ChIP-seq on genetically modified Bapx1/Nkx3.2 by the addition of an S-protein epitope tag to the N-terminus of the endogenous Bapx1/Nkx3.2 protein-encoding sequence. Once correctly targeted ES clones are identified they are used to generate germline transmitting chimeras and subsequent stable lines of mice expressing the S-protein- Bapx1/Nkx3.2 fusion protein. S-protein tagged Bapx1/Nkx3.2 expressing embryos are isolated mid-gestation and the vertebral columns are removed. The isolated cells are cross-linked with fixitive, DNA is isolated and sheared and anti-S-protein antibody is employed for immunoprecipitation. The isolated DNA which is enriched for Bapx1/Nkx3.2 genomic binding sites is sequenced and mapped to the lastest build of the mouse genome.

2.2. Genetically modified mouse line generation

The "Quick and Easy" BAC modification kit (Gene Bridges) was used to generate *Bapx1/Nkx3.2* gene targeting constructs as we have previously described [2, 3] using C57BL/6J strain murine BAC clones containing the *Bapx1/Nkx3.2* gene obtained from the BACPAC Resources Centre at Children's Hospital Oakland Research Institute (CHORI). Fragments of each modified region flanked by short and long homology arms with a total average length of >10 kb were subcloned into a minimal vector with the Gene Bridges BAC Subcloning kit. The resulting plasmids were linearized and electroporated into ES cells previously we described [2,3]. Normal karyotype ES clones positive by DNA blotting with probes external to the gene targeting constructs were used to generate germline transmitting chimeras as we have previously described [4]. The resulting highly chimeric mice were crossed to C57BL/6J mice and stable mouse lines were established expressing the S-peptide tagged Bapx1/Nkx3.2 protein. The *FRT*-flanked neomycin G418 antibiotic selection cassette in the targeted *Bapx1/Nkx3.2* allele was deleted by crossing to *FLPe*–deleter mice (129S4/SvJaeSor-*Gt(ROSA)26Sor tm1(FLP1)Dym*/J (Stock #3946) from Jackson Laboratories) [5,6]. Genotyping of *Bapx1/Nkx3.2* lines was performed by PCR essentially as described [7]. RNA in situ analysis was performed as previously described [8] on the resulting modified allele carrying embryos to verify that there were no detectable changes in expression of the gene-targeted locus. For analysis of the vertebral column the E12.5 embryos were sectioned through the sagittal plane [9].

2.3. Chromatin immunoprecipitation and DNA-binding motif analysis

Wildtype or *Baxp1/Nkx3.2* gene-targeted mouse lines were mated and noon of the plug date was designated as E0.5. At E12.5 embryos were isolated and the vertebral columns (tagged with S-peptide for Baxp1/Nkx3.2 or wildtype mouse embryos for Sox9) were carefully dissected away from other organs and pooled. An overview of the experimental approach is shown in Fig. 1. Chromatin (~2 mg) was used for chromatin immunoprecipitation as previously described for Bapx1/Nkx3.2 and Sox9 [10]. With 15 ng of purified ChIP DNA from each sample, the sequencing library was generated according to the manufacturer's instructions (Illumina) [1,11]. Antibodies used for Bapx1/Nkx3.2 and Sox were anti S-Peptide antibody (Bethyl laboratories, A190-134A) and anti-Sox9 antibody (R&D Systems, AF3075) respectively. The Illumina Genome Analyzer Pipeline was used to map reads that had passed signal filtering to the murine mm9 genome. The Illumina Genome Analyzer II/lix produced sequence reads. The peak calling algorithm MACS was used to match unique reads with 2 or less mismatches, with default settings as previously described [12]. ChIP-Seq peak annotation is as previously described [1,11]. All data was deposited in Gene Expression Omnibus (GEO) under accession number GSE36672. 2815 Bapx1/Nkx3.2 peaks were detected at the TSS (540). Many binding sites, however, were located far from a TSS; some located at over 25 kb distally in the intergenic (1052) and less located over 25 kb in the intragenic regions (722). Many peaks were identified for Sox9 (3722). DNA-binding motif analysis used the top several hundred peaks and the masked repeat genome sequence spanned on either side from the summit of these peaks by 50 bp. MACS determined peaks were ranked as described [1]. The DNA binding motif for Sox9 was determined mostly by in vitro studies [13]. We identified the Bapx1/Nkx3.2 binding motif [14] enriched in about 40% of the Bapx1/Nkx3.2 binding sites [1,11]. Gene enrichment test for regions bound by Sox9 and Bapx1, we submitted the analyzed ChIP-Seq data as bed format to Genomic Regions Enrichment of Annotations Tool (GREAT) [15] version 1.8 with the following parameters, species assembly-mm9 and Association rule-Single nearest gene: 100,000 bp maximum extension, curated regulatory domains included.

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] 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), e28885, http://dx.doi.org/10.1371/journal.pone.0028885.

[3] W. Wang, P. Lo, M. Frasch, T. Lufkin, Hmx: an evolutionary conserved homeobox gene family expressed in the developing nervous system in mice and *Drosophila*. Mech. Dev. 99 (2000) 123–137.

[4] 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.

[5] 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.

[6] 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 Cre-loxP. Nat. Genet. 25 (2000) 139–140, http://dx.doi.org/10.1038/75973.

[7] 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.

[8] X. Chen, X. Li, W. Wang, T. Lufkin, Dlx5 and Dlx6: an evolutionary conserved pair of murine homeobox genes expressed in the embryonic skeleton. Ann. N. Y. Acad. Sci. 785 (1996) 38–47.

[9] P. Kraus, T. Lufkin, Mammalian Dlx homeobox gene control of craniofacial and inner ear morphogenesis. J. Cell. Biochem. (Suppl 32-33) (1999) 133–140.

[10] 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.

[11] S. Chatterjee, V. Sivakamasundari, P. Kraus, S.P. Yap, V. Kumar, S. Prabhakar, T. Lufkin, Gene expression profiles of Bapx1 expressing FACS sorted cells from wildtype and Bapx1-EGFP null mouse embryos. Genom Data 5 (2015) 103–105, http://dx.doi.org/10.1016/j.gdata.2015.05.031.

[12] 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.

[13] S. Mertin, 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.

[14] 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.

[15] C.Y. McLean, D. Bristor, M. Hiller, S.L. Clarke, B.T. Schaar, C.B. Lowe, A.M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28 (2010) 495–501. 10.1038/nbt.1630.