Title

Embryonic stem cell–based mapping of developmental transcriptional programs

Authors

Mazzoni, Esteban O
Mahony, Shaun
Iacovino, Michelina
et al.

Embryonic stem cell-based system for mapping developmental transcriptional programs

Esteban O. Mazzoni[1], **Shaun Mahony**[2], **Michelina Iacovino**[3], **Carolyn A. Morrison**[1], **George Mountoufaris**[1], **Michael Closser**[1], **Warren A. Whyte**[4], **Richard A. Young**[4], **Michael Kyba**[3], **David K. Gifford**[2], and **Hynek Wichterle**[1]

[1]Departments of Pathology, Neurology, and Neuroscience, Center for Motor Neuron Biology and Disease and Columbia Stem Cell Initiative, Columbia University Medical Center, 630 168 street, New York, NY 10032, USA

[2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA

[3]Lillehei Heart Institute and Department of Pediatrics, University of Minnesota, 312 Church St. SE, Minneapolis, MN 55455, USA

[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

Abstract

Study of developmentally regulated transcription factors by chromatin immunoprecipitation and sequencing (ChIP-seq) faces two major obstacles: availability of ChIP grade antibodies and access to sufficient number of cells. We describe a versatile method for genome-wide analysis of transcription factor binding sites by combining directed differentiation of embryonic stem cells and inducible expression of tagged proteins. We demonstrate its utility by mapping transcription factors involved in motor neuron specification.

The study of transcriptional networks provides an opportunity to gain fundamental insight into complex molecular processes that govern cell fate specification and embryonic development. While numerous transcription factors controlling cell differentiation have been functionally characterized, their cell type specific patterns of DNA binding remain largely unknown.

The method of choice for genome-wide mapping of transcription factor binding sites is chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) [1]. Although powerful, current ChIP-seq technology is limited by two critical factors when applied to developmental studies. First, ChIP-seq profiling demands a large number of cells (20–50 million) separated from other cell types expressing the transcription factor of interest, and second, it requires antibodies with high affinity and specificity that recognize transcription factors in their native form bound to DNA. To overcome these two hurdles, we combined a versatile system for generating mouse embryonic stem cell (ESC) lines harboring inducible and epitope-tagged transcription factors with directed differentiation of ESCs along defined cellular lineages. This system presents several advantages: 1) the use of tagged transcription factors or DNA binding proteins obviates the need for validated factor-specific antibodies; 2) the use of pluripotent cells allows analysis of any developmental cell lineage; and 3) the inducible expression makes it possible to examine binding of developmentally regulated transcription factors in their correct developmental context, as well as to study tagged transcription factors by gain-of-function analysis.

To overcome the inconsistency and inefficiency of classical transgenic ESC line production, we relied on a recently developed inducible cassette exchange (ICE) system [2]. The resulting transgenic lines harbor a single copy of the transgene recombined into a defined expression-competent locus. To further streamline the generation of inducible cell lines, we introduced Gateway (Invitrogen) landing sites into the shuttle vector and a short epitope tag either at the amino- (Flag-Bio) or carboxy-terminus (His-V5) of the protein (Fig. 1a). Because of the high efficiency of all steps, parallel production of multiple inducible tagged lines can be accomplished in as little as three weeks.

Differentiation of mouse ESCs to spinal motor neurons yields scalable and largely homogeneous populations of cells mirroring developmentally relevant motor neuron differentiation states in mouse [3]. To test our approach, we first investigated genome-wide binding of the bHLH transcription factor Olig2 in motor neuron progenitors (pMNs) [4], a rare population of cells (<1% of spinal cells on e9.5) found in the embryonic ventral spinal cord [5].

We generated an inducible Olig2 ESC line in which Olig2 protein is carboxy-terminal tagged with the V5 epitope (iOlig2-V5). To mimic the normal Olig2 pattern of expression, doxycycline (Dox) was administered late on Day 3 and the expression of the transgene was analyzed on Day 4 (Fig. 1b) when cells reach pMN stage. The transgenic Olig2-V5 protein was expressed uniformly in pMNs, exhibited correct nuclear localization, and is expressed at levels ~4 fold higher than native Olig2 (Suppl. Fig. 1a–b). The V5 sequence did not perturb the function of the tagged Olig2-V5 protein. As expected, ectopic expression of Olig2-V5 resulted in the repression of Nkx2.2 in ventral interneuron progenitors (Fig. 1c) [4] and in the repression of Pax6 and Irx3 in dorsal interneuron progenitors (Fig. 1d) [6]. Therefore, a tagged version of Olig2 recapitulates in differentiating ESCs the normal function of native Olig2 during spinal cord development [7].

To profile Olig2 binding, we induced Olig2-V5 in pMNs and performed a ChIP-seq experiment with an anti-V5 antibody. We observed that Olig2-V5 binds in the proximity of

the downregulated genes Irx3, Nkx2.2 and Pax6 (Fig 2a and Suppl. Fig. 1c), indicating that Olig2 specifies pMN identity by direct repression of interneuron transcriptional programs.

The overexpression of the Olig2 transgene or the addition of the short tag sequence might affect the genomic binding pattern of the Olig2-V5 protein. For comparison, we therefore performed a ChIP-seq experiment in ESC-derived pMNs with antibodies against the native Olig2 protein in the absence of Dox. The endogenous Olig2 and inducible Olig2-V5 ChIP-seq experiments revealed a remarkable level of agreement. The proteins bind to the same regulatory sequences of Irx3, Nkx2.2 and Pax6 (Fig 2a and Suppl. Fig. 1c). As expected for a bHLH transcription factor, motif discovery within the ChIP-enriched sites revealed an E-box motif consensus (Fig 2b) that is present at 58.8% of Olig2-V5 and 60.4% of native Olig2 binding sites (10% false discovery rate (FDR) motif scoring threshold [8]). To determine whether enriched sequences lacking E-box motif represent real binding events we employed an in vitro ELISA based DNA-protein interaction assay. We demonstrate that Olig2 transcription factor can be recruited to all tested ChIP-seq identified sequences regardless whether they contain E-box motif or not (Suppl. Fig. 1d), supporting the notion that ChIP-seq data reflect Olig2 binding events.

The distribution of binding sites found in both experiments is also highly coincident (Fig. 2c). Comparing the read counts at enriched peaks shows that only 0.2 % and 1.1% are differentially enriched in the native Olig2 and Olig2-V5 ChIP experiments, respectively (Fig. 2c). Globally, the levels of ChIP-seq enrichment are highly correlated between experiments with a Pearson's correlation coefficient of 0.83, indicating that neither the overexpression of Olig2-V5 in Olig2+ pMNs, nor the addition of an epitope tag, affects Olig2 activity or DNA binding preference.

Next we compared the binding site preference of tagged transcription factors in a postmitotic motor neuron stage. We have previously demonstrated that Hoxc9 represses cervical programs and promotes specification of thoracic motor neurons [9]; the study of Hoxc9-V5 (iHoxc9-V5) (Fig. 1e) revealed a direct repression of cervical Hox genes [9]. We compared binding sites of C- and N-terminally epitope tagged Hoxc9, reasoning that overlapping sites are most likely to reflect native Hoxc9 binding events. We modified the inducible system to accommodate a Flag-Bio (FlagB) amino-terminal tag (Fig. 1a) that can be used for ChIP pull-downs either with anti-Flag antibodies or streptavidin-based purification in combination with the biotinylation enzyme BirA [10]. We determined that FlagB tagged Hoxc9 retained its ability to repress cervical Hoxc4 and Hoxa5 genes (Fig. 3e and data not shown). Importantly, the genome-wide binding of the iFlagB-Hoxc9 with anti-Flag antibodies shows a high degree of agreement with the Hoxc9-V5 binding profile. Both Hoxc9 proteins associate with rostral Hox genes regulatory elements, indicating their direct repression (Fig. 2d). At the genomic level, both proteins share an identical sequence preference, depicted by a typical Hox binding primary motif (Fig. 2e). Moreover, 47.1% of the peaks in the V5-tagged and 39.1% in the Flag-tagged experiments contain the primary motif at a 10% FDR scoring threshold (Suppl. Fig. 2). Although we estimate that the proportion of ChIP-seq reads located in enriched regions is approximately three times higher in the FlagB-Hoxc9 experiment than in the Hoxc9-V5 experiment, the detected peaks are highly coincident across experiments (Fig. 2f). Out of 22,458, only 156 peaks (0.7%) are differentially

enriched in V5 ChIP and 799 peaks (3.6%) are differentially enriched in the Flag experiment (Fig. 2f). We conclude that genomic regions shared between C- and N- terminally tagged ChIP-seq experiments are likely to represent native Hoxc9 binding events.

The high degree of overlap between ChIP-seq experiments for one transcription factor contrasts with the binding profiles of two unrelated transcription factors. The comparison of the Olig2-V5 and Hoxc9-V5 ChIP experiments revealed a large fraction of non-overlapping peaks, which is in striking contrast to biological replicates of Hoxc9-V5 ChIP-seq experiments that are virtually indistinguishable (Suppl. Fig 3a–c). Detailed analysis of predicted binding positions by the GPS algorithm [8] in co-bound peaks reveals that Olig2 and Hoxc9 occupy proximal but distinct sites within the peaks (Suppl. Fig 2e and f). Because a typical ChIP-seq peak covers ~200 bp, these experiments might be revealing enhancers that are active in both motor neuron progenitors and postmitotic motor neurons.

The system we present here is robust and allows the generation of multiple inducible cell lines in parallel. Of twenty-four generated lines, only three exhibited problems with inducible protein expression, likely due to the inherent toxicity of introduced transgenes (data not shown). While the system is versatile and can be employed to study both progenitors and differentiated cells, we observed that the efficiency and homogeneity of transgene induction declines in postmitotic neurons. Inducing the transgene at late progenitor stage results in maintained and homogenous expression in postmitotic neurons, offering a reasonable workaround for this problem. Some transcription factors control their targets in a concentration dependent manner. In those instances, it will be important to first establish the Dox concentration and timing of the treatment that result in desired phenotypes, to ensure that the transcription factor binding studies produce biologically relevant information.

In summary, we provide a set of tools for rapid generation of ESC lines and production of unlimited quantities of isogenic differentiated cells that enable identification of developmentally relevant transcription factor binding sites in a genome-wide manner. The cell lines can also be utilized for other biochemical studies, including the isolation and identification of transcription factor binding partners by coimmunoprecipitation followed by mass spectrometry [11]. We believe that the combination of these powerful techniques will pave the way to a detailed mechanistic understanding of transcriptional networks that govern mammalian development.

Methods

Cell culture

ES cells were cultured over a layer of Mitomycin-C treated fibroblast resistant to Neomycin (Fisher) in EmbryoMax D-MEM (Fisher) supplemented with 10% ES-FBS (Invitrogen), L-Glutamine (Gibco), 0.1 mM β-mercaptoethanol and 100 U/ml LIF.

Motor neuron differentiation of ES cells was performed as previously described [1]. Briefly, ES cells were trypsinized (Invitrogen) and seeded at 5×10^5 cells/ml in ANDFK medium (Advanced DMEM/F12:Neurobasal (1:1) Medium, 10% Knockout-SR, Pen/Strep, 2 mM L-

Glutamine, and 0.1 mM 2-mercaptoethanol) to initiate formation of embryoid bodies (Day 0). Medium was exchanged on Day 1, Day 2 and Day 5 of differentiation. Patterning of embryoid bodies was induced by supplementing media on Day 2 with 1 μM all-*trans*-Retinoic acid (RA, Sigma) and 0.5 μM agonist of hedgehog signaling (SAG, Calbiochem). For ChIP experiments, the same conditions were used but scaled to seed 1×10^7 cells on Day 0. Doxycycline (Sigma) was added to the culture medium at 1ug/ml when required.

Generation of inducible lines

The p2Lox-V5 plasmid was generated by replacing GFP with the L1-L2 Gateway cassette from pDEST-40 (Invitrogen) in the p2Lox plasmid. The cassette contains a V5-His double epitope tag in frame downstream of the L2 recombination site. p2Lox-FlagB was generated by replacing GFP in the p2Lox plasmid with the L1-L2 Gateway cassette from pDEST-40 without the V5-His sequence but with the addition of a Flag-Biotin sequence [2] in frame and upstream the L1 recombination site.

Open reading frames of genes are cloned by polymerase chain reaction (PCR). To minimize the introduction of mutations during PCR amplification, Phusion polymerase was used (New England Biolabs). Open reading frames were directionally inserted into pENTR/D-TOPO vector (Invitrogen) following manufacturer instructions. The 5' primer always contains the addition of the CACC sequence to ensure directional integration. For each coding sequence, two alternative 3' primers were used: with and without STOP codon, generating two pENTR plasmids for each gene.

The LR recombination scheme is as follow: 1) When constructing a V5-His C-terminal fusion protein the pENTR plasmid with NO STOP codon is recombined with the p2Lox-V5. 2) Non-tagged proteins are generated by recombining the pENTR plasmid with STOP codon with the p2Lox-V5 plasmid. 3) To generate N-terminal tagged proteins, the pENTR plasmid with STOP codon is recombined with p2Lox-FlagB.

Inducible lines were generated by treating the recipient ESCs for 16 hours with doxycycline to induce Cre followed by electroporation of either p2Lox-V5 and p2Lox-FlagB plasmids harboring the desired construct. After G418 selection, on average three resistant clones were picked, characterized and expanded.

Immunocytochemistry

Embryoid bodies were fixed with 4% paraformaldehyde in PBS, embedded in OCT (Tissue-Tek) and sectioned for staining: 24 hours at 4C for primary antibodies and 4 hours at RT for secondary antibodies. After staining, samples were mounted with Aqua Poly Mount (Polyscience). Images were acquired with a LSM 510 Carl Zeiss confocal microscope. Antibodies used in this study include: Rabbit anti-Olig2 (Millipore); mouse anti-V5 (Invitrogen); mouse anti-Flag M2 (Sigma); rabbit anti-Hoxc4 are gifts from T Jessell. Alexa488-, FITC-, Cy3- and Cy5-conjugated secondary antibodies were obtained from either Invitrogen or Jackson Immunoresearch.

ChIP-seq

Differentiating embryoid bodies were washed with PBS and then dissociated by mild Trypsinization (Invitrogen) followed by mechanical dissociation until single-cell suspension was obtained. Cells were fixed with 1% formaldehyde for 15 minutes at room temperature. Pellets containing ~40×10^6 cells were flash frozen and stored at −80 °C. Cells were thawed on ice, resuspended in 5ml of Lysis Buffer A and incubated for 10 minutes at 4 °C in a rotating platform. Samples were spun down for 5 minutes at 1,350g, resuspended in 5ml Lysis Buffer B and incubated for 10 minutes at 4 °C in a rotating platform. Samples were spun down for 5 minutes at 1,350g, resuspended in 3ml of Sonication Buffer (SB).

Nuclear extracts were sonicated using a Misonix 3000 model sonicator to sheer cross-linked DNA to an average fragment size of approximately 500bp. Sonicated chromatin was incubated for 16 hours at 4C with Protein-G beads (Invitrogen) conjugated with either rabbit anti-V5 (Abcam, ab15828), mouse anti-Flag M2 (Sigma, ab15828) or rabbit anti-Olig2 (Millipore, AB15328). After incubation and with the aid of a magnetic device, beads were washed once with SB+500nM NaCl, once with LiCl Wash Buffer (LiClB) and 1ml of TE. Then, beads were centrifuged at 950g for 3 min and residual TE removed with a pipette. 210 ul of Elution Buffer was added to the beads followed by incubation at 65 °C for 45 minutes with a brief pulse of vortex every 10 minutes. 200 ul of supernatant was removed after a 1 minute centrifugation at 16,000g. The crosslink was reversed by 16 hours incubation at 65 °C.

RNA was digested by the addition of 200 ul of TE and RNAseA (Sigma) at a final concentration of 0.2mg/ml and incubated for 2 hours at 37C. Protein was digested by the addition of Proteinase K (0.2 mg/ml final, Invitrogen) supplemented with CaCl^2 followed by a 30 minutes incubation at 55 °C. DNA was extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and then recovered with an ethanol precipitation with glycogens as carrier. The pellets were suspended in 70 ul of water. Purified DNA fragments were processed according to the Illumina/Solexa sequencing protocol using a Genome Analyzer II (Illumina, http://www.illumina.com/pages.ilmn?ID=252).

Lysis Buffer A (50 mM Hepes-KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA; 10% Glycerol; 0.5% Igepal; 0.25% Triton X-100). **Lysis Buffer B** (10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA, pH 8.0; 0.5 mM EGTA, pH 8.0). **Sonication Buffer** (50mM Hepes pH 7.5; 40mM NaCl; 1mM EDTA; 1mM EGTA; 1% Trition X-100; 0.1% Na-deoxycholate; 0.1% SDS). **Sonication Buffer High Salt** (50mM Hepes pH 7.5; 500mM NaCl; 1mM EDTA; 1mM EGTA; 1% Trition X-100; 0.1% Na-deoxycholate; 0.1% SDS). **IgG LiCl Wash Buffer** (20mM Tris-HCL pH8.0; 1mM EDTA; 250mM LiCl; 0.5% NP-40; 0.5% Na-deoxycholate). **Elution Buffer** (50 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0; 1% SDS).

ChIP-seq analysis

Sequence reads were aligned to the mouse genome (version mm9) using Bowtie [3] version 0.12.5 with options "-q --best --strata -m 1 -p 4 --chunkmbs 1024". Only uniquely mapping reads were analyzed further. Binding events were detected using GPS [4]. In GPS, the scaling ratio between IP and control channels was estimated using the median ratio of all 10Kbp

windows along the genome. The GPS binding model was initialized to the default and iteratively updated over up to 3 training rounds. In this study, we require that reported peaks contain a ChIP-seq enrichment level that is significantly greater than 1.5 times the control level with *p*-value <0.01 as tested using the Binomial distribution. Signal-to-noise ratios are estimated by comparing the ChIP-seq read count occurring at any peak found for a given transcription factor in any condition to the count of remaining reads in that experiment.

When comparing enrichment levels between two ChIP-seq experiments, we first scale the read counts assigned to each peak using the median ratio of observed read counts across all peaks. The read counts of one experiment are always scaled down to match the scale of the other experiment. We define differentially enriched sites as those that have a scaled read count in one experiment that is significantly greater than 1.5 times the scaled read count from the other experiment (*p*<0.01, Binomial test, adjusted for multiple testing using Benjamini & Hochberg's method). All microarray and ChIP-seq data are available from the GEO database under accession number GSE31456

DNA motif analysis

De-novo motif-finding was performed in 200bp windows centered on the 2,000 top-ranked peaks for each examined ChIP-seq experiment. GimmeMotifs [5] was used to discover motifs by running and combining results from the motif-finders MDmodule, MEME, GADEM, MotifSampler, trawler, Improbizer, MoAn, and BioProspector. The settings "-w 200 -a large -g mm9 -f 0.5 -l 500" were used with GimmeMotifs. STAMP [6] was used to determine the similarity of discovered motifs to known DNA-binding preferences. Log-likelihood scoring thresholds for the discovered motifs were calculated by simulating 1,000,000 200bp sequences using a 3rd-order Markov model of the mouse genome. The motif scoring thresholds that yield false discovery rates of 10% in this set of sequences were recorded and used to scan 200bp sequences centered on the Olig2 and Hoxc9 GPS-predicted peak positions.

ELISA DNA Binding

PCR amplified and biotin-labeled genomic fragments are gel purified. The fragments are between 500 and 600 bp with the 5' primers containing a single biotin molecule at the 5'end. Streptavidin coated 96 well plates (Fisher, PI-15500) are washed three times with 200 μl of Wash Buffer. The biotin-labeled PCR reaction is loaded into each well in Blocking Buffer up to 125 μl final volume to saturate the binding capacity of each well (Plates can bind ~125pmol/well). The plates were incubated overnight at 4°C then washed each well three times with 200 μl of Wash Buffer. Serial dilutions of the cell extract in PBS were added to each well and incubated for 3 hrs with shaking at room temperature. After washing each well three times with 200 μl with Wash Buffer the primary antibody in 0.5% BSA in PBS was added to final volume of 100 μl to each well and incubate plate for 2 hours with shaking at room temperature. After washing each well three times with 200μl of Wash Buffer, HRP-conjugated secondary antibody diluted in 0.5% BSA in PBS to final volume of 100 μl to was added to each well for 30–45 minutes with shaking at room temperature. The wells were washed three times with 200 μl of Wash Buffer. 100 μl of equilibrate the TMB (Fisher EN-N301) to room temperature was added to each well. After 10 minutes incubation, the

reaction was stopped by adding 100 μl of 2 M sulfuric acid to each well. The absorbance was measured of each well at 450 nm using a plate reader.

Wash Buffer: 0.05% Tween-20 + .1% BSA in PBS.

Blocking Buffer: 0.5% BSA in PBS.

Priers to amplify the ChIp-seq identified genomic regions in Suppl. Table 1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Park PJ. Nat Rev Genet. 2009; 10(10):669. [PubMed: 19736561]

2. Iacovino M, Bosnakovski D, Fey H, et al. Stem Cells. 2011; 29(10):1580. [PubMed: 22039605]

3. Wichterle H, Lieberam I, Porter JA, et al. Cell. 2002; 110(3):385. [PubMed: 12176325]

4. Novitch BG, Chen AI, Jessell TM. Neuron. 2001; 31(5):773. [PubMed: 11567616]

5. Mukouyama YS, Deneen B, Lukaszewicz A, et al. Proc Natl Acad Sci U S A. 2006; 103(5):1551. [PubMed: 16432183]

6. Chen JA, Huang YP, Mazzoni EO, et al. Neuron. 69(4):721. [PubMed: 21338882]

7. Jessell TM. Nat Rev Genet. 2000; 1(1):20. [PubMed: 11262869]

8. Guo Y, Papachristoudis G, Altshuler RC, et al. Bioinformatics. 2010; 26(24):3028. [PubMed: 20966006]

9. Jung H, Lacombe J, Mazzoni EO, et al. Neuron. 67(5):781. [PubMed: 20826310]

10. Kim J, Cantor AB, Orkin SH, et al. Nat Protoc. 2009; 4(4):506. [PubMed: 19325547]

11. Wang J, Cantor AB, Orkin SH. Curr Protoc Stem Cell Biol. 2009 Chapter 1, Unit1B 5.

REFERENCES

1. Wichterle H, Lieberam I, Porter JA, et al. Cell. 2002; 110(3):385. [PubMed: 12176325]

2. Wang J, Cantor AB, Orkin SH. Curr Protoc Stem Cell Biol. 2009 Chapter 1, Unit1B 5.

3. Langmead B, Trapnell C, Pop M, et al. Genome Biol. 2009; 10(3):R25. [PubMed: 19261174]

4. Guo Y, Papachristoudis G, Altshuler RC, et al. Bioinformatics. 2010; 26(24):3028. [PubMed: 20966006]

5. van Heeringen SJ, Veenstra GJ. Bioinformatics. 27(2):270. [PubMed: 21081511]

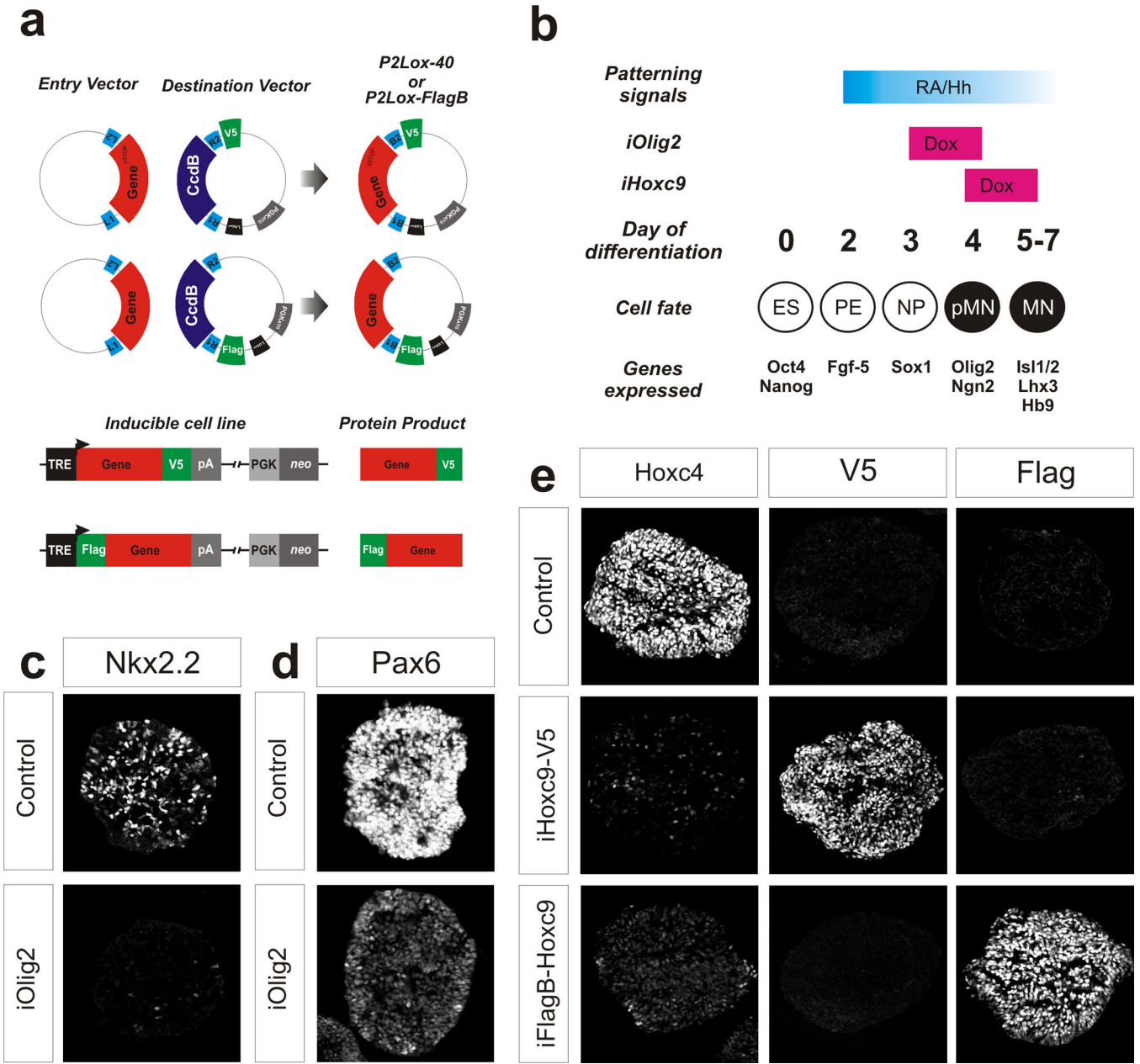6. Mahony S, Benos PV. Nucleic Acids Res. 2007; 35:W253. Web Server issue. [PubMed: 17478497]

Figure 1. Generation of inducible cell lines
(**a**) The schematic depicts the cloning strategy to generate inducible lines with epitope tagged genes. Coding sequences lacking the STOP codon are V5-tagged at the C-terminus. (**b**) Overview of embryonic stem cell (ESC) directed differentiation. Differentiating cells become motor neuron progenitors (pMNs) at Day 4 and produce motor neurons (MNs) from Day 5–7. Doxycyline (Dox) is added late on Day 3 or Day 4 to mimic the expression pattern of the endogenous Olig2 and Hoxc9 respectively. Olig2-V5 is analyzed at Day 4 and Hoxc9-V5 or FlagB-Hoxc9 at Day 5. (**c**) Nkx2.2 staining in control or Olig2-V5 expressing cells at Day 4 of differentiation under high Hh concentration (500 nM). (**d**) Pax6 staining in control or Olig2-V5 expressing cells at Day 4 of differentiation under low Hh (5 nM). (**e**) iHoxc9-

V5 and iFlagB-Hoxc9 day 5 embryoid bodies stained with anti-Hoxc4, V5 and Flag antibodies.
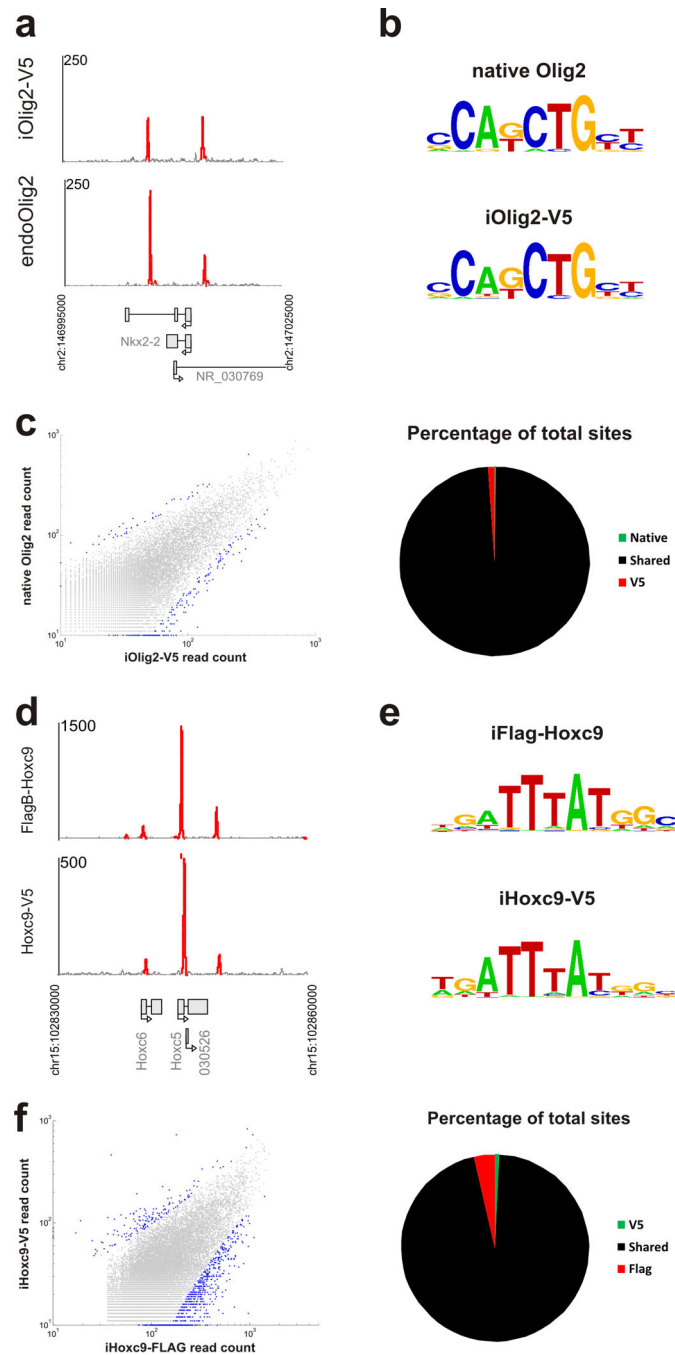
Figure 2. Native and tagged ChIP comparisons

(**a**) ChIP signal tracks over Nkx2.2 genomic loci for endogenous and V5-tagged Olig2. Red peaks represent significant (p<0.01) enrichment over control. (**b**) The most over-represented motifs discovered under ChIP-seq peaks for native Olig2 and Olig2-V5 ChIP experiments. (**c**) The plots show a comparison of read enrichment from native and V5-tagged Olig2 ChIP-seq experiments at all detected peaks. Blue dots in the scatterplot represents peaks significantly differentially enriched in one experiment over the other. Pie chart represents numbers of sites differentially enriched between native Olig2 and V5 tagged ChIP

experiments as a percentage of total sites. (**d**) ChIP signal tracks over Hoxc5 genomic locus for V5- and Flag- tagged Hoxc9 experiments. Red peaks represent significant enrichment over control (p<0.01). (**e**) The most over-represented motifs discovered under ChIP-seq peaks for Hoxc9-V5 and FlagB-Hoxc9 ChIP. (**f**) The plots show a comparison of read enrichment from FlagB and V5-tagged Hoxc9 ChIP-seq experiments at all detected peaks. Blue points in the scatterplot represent peaks significantly differentially enriched in one experiment over the other. The pie chart shows numbers of sites differentially enriched between Hoxc9-V5 and FlagB-Hoxc9 ChIP experiments as a percentage of total sites.