Embryonic stem cell–based mapping of 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[4,5], Richard A Young[4,5], Michael Kyba[3], David K Gifford[2] & Hynek Wichterle[1]

The study of developmentally regulated transcription factors by chromatin immunoprecipitation and deep sequencing (ChIP-seq) faces two major obstacles: availability of ChIP-grade antibodies and access to sufficient number of cells. We describe versatile 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 DNA-binding sites of 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. Whereas many 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 directed differentiation of embryonic stem cells (ESCs) along defined cellular lineages with a versatile system for generating mouse ESC lines containing inducible genes encoding epitope-tagged transcription factors. This system has several advantages: (i) the use of tagged transcription factors or DNA binding proteins obviates the need for validated factor-specific antibodies, (ii) the use of pluripotent cells allows analysis of any developmental cell lineage, and (iii) 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 inefficiency of classical transgenic ESC line production, we relied on a recently developed inducible cassette exchange (ICE) system[2]. The resulting lines contain a single copy of the transgene recombined into a defined expression-competent locus. To 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 N terminus (Flag-Bio) or C 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 states in mouse[3]. We first investigated genome-wide binding of the basic helix-loop-helix (bHLH) transcription factor Olig2 in motor neuron progenitors (pMNs)[4], a rare population of cells (<1% of spinal cells on embryonic day 9.5) found in the embryonic ventral spinal cord[5].

We generated an inducible Olig2 ESC line in which the encoded Olig2 protein is C-terminally tagged with the V5 epitope (iOlig2-V5). To mimic the normal Olig2 pattern of expression, we administered doxycycline late on day 3 of differentiation and analyzed transgene expression 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 its expression was about fourfold greater than that of native Olig2 (**Supplementary 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[4] (**Fig. 1**) and in the repression of Pax6 and Irx3 in dorsal interneuron progenitors[6] (**Fig. 1**). 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 antibody to the V5 epitope. We observed that Olig2-V5 bound in the proximity of the downregulated genes *Irx3*, *Nkx2-2* and *Pax6* (**Fig. 2a** and **Supplementary Fig. 1c**), indicating that Olig2 specifies pMN identity by direct repression of interneuron transcriptional programs.

[1]Departments of Pathology, Neurology and Neuroscience, Center for Motor Neuron Biology and Disease and Columbia Stem Cell Initiative, Columbia University Medical Center, New York, New York, USA. [2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Lillehei Heart Institute and Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. Correspondence should be addressed to H.W. (hw350@columbia.edu).

RECEIVED 8 JUNE; ACCEPTED 19 OCTOBER; PUBLISHED ONLINE 13 NOVEMBER 2011; DOI:10.1038/NMETH.1775

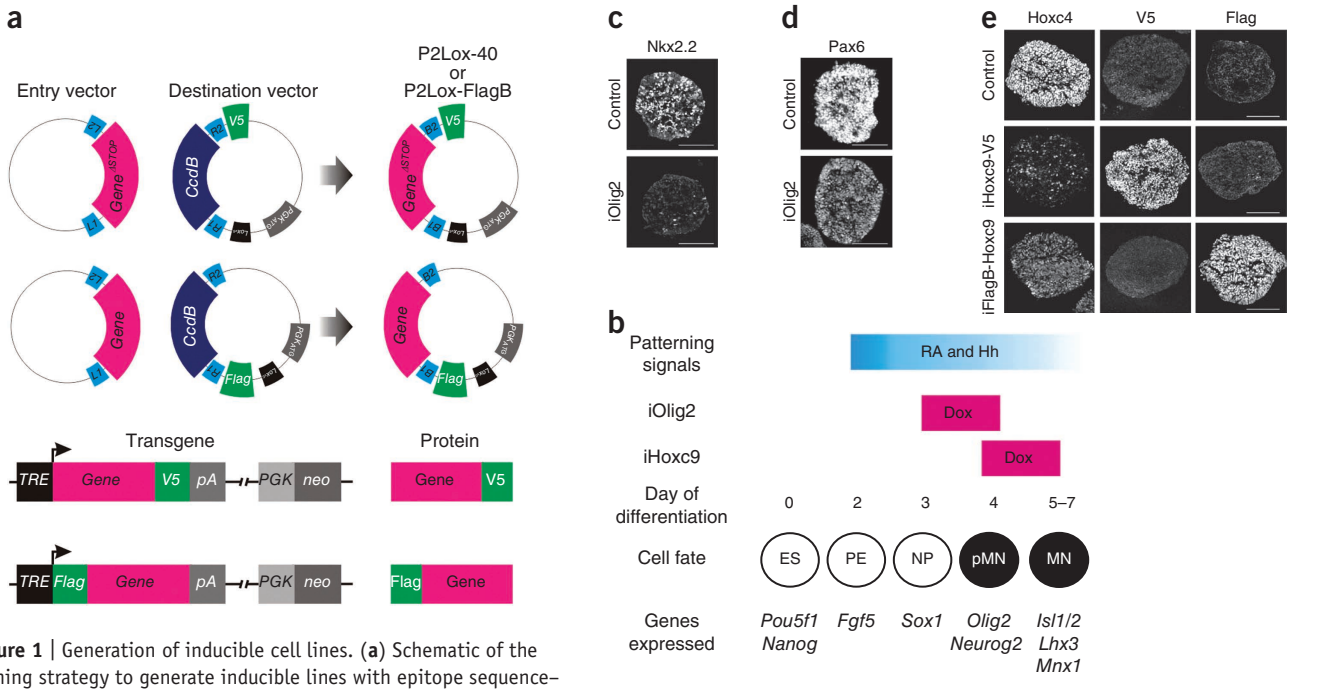

Figure 1 | Generation of inducible cell lines. (**a**) Schematic of the cloning strategy to generate inducible lines with epitope sequence–tagged genes. Coding sequences lacking the stop codon were tagged to encode V5 at the C terminus. R1, R2, L1 and L2, recombination sites for LR-Clonase enzyme (Invitrogen); TRE, tetracycline response element, and pA, polyadenylation signal. (**b**) Overview of ESC-directed differentiation. Differentiating cells become motor neuron progenitors (pMNs) on day 4 and produce motor neurons (MN) on days 5–7. Doxycyline (Dox) was added late on day 3 or 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. ES, embryonic stem cell; PE, primitive ectoderm; NP, neural precursor; RA, retinoic acid; and Hh, hedgehog. (**c**) Nkx2.2 staining in control or Olig2-V5 expressing cells on 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–expressing day 5 embryoid bodies stained with antibodies to Hoxc4, V5 and Flag as indicated. Scale bar, 100 µm.

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 to the native Olig2 protein in the absence of doxycycline. The endogenous Olig2 and inducible Olig2-V5 ChIP-seq experiments were in agreement. The proteins bound to the same regulatory sequences of *Irx3*, *Nkx2-2* and *Pax6* (**Fig. 2a** and **Supplementary Fig. 1c**). As expected for a bHLH transcription factor, motif discovery in 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 motif scoring threshold[8]). To determine whether enriched sequences lacking the E-box motif represent real binding events, we used an *in vitro* ELISA-based DNA-protein interaction assay. We found that Olig2 transcription factor can be recruited to all tested ChIP-seq–identified sequences regardless of whether they contain E-box motif or not (**Supplementary Fig. 1d**), supporting the notion that ChIP-seq data reflect Olig2 binding events.

The binding-site distribution found in both experiments was also highly coincident (**Fig. 2c**). Comparing the read counts at enriched peaks showed that only 0.2% and 1.1% were differently enriched in the native Olig2 and Olig2-V5 ChIP experiments, respectively (**Fig. 2c**). Globally, ChIP-seq enrichment was highly correlated between experiments with a Pearson's correlation coefficient of 0.83, indicating that neither the overexpression of Olig2-V5 in Olig2-expressing 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 (inducible HoxC9-V5, iHoxc9) (**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 pulldowns either with Flag antibodies or streptavidin-based purification in combination with the biotinylation enzyme BirA[10]. We determined that FlagB-Hoxc9 retained its ability to repress cervical *Hoxc4* and *Hoxa5* genes (**Fig. 1** and data not shown). The analysis of FlagB-Hoxc9 binding sites by ChIP-seq with 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 have an identical sequence preference, depicted by a typical Hox binding primary motif (**Fig. 2e**). Moreover, 47.1% of the peaks in experiments with V5-tagged protein and 39.1% in those with Flag-tagged proteins contain the primary motif at a 10% false discovery rate scoring threshold (**Supplementary 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**). Of 22,458 peaks, only 156 peaks (0.7%) were differently enriched in V5-tagged-protein ChIP experiment, and 799 peaks (3.6%) were differently enriched in the

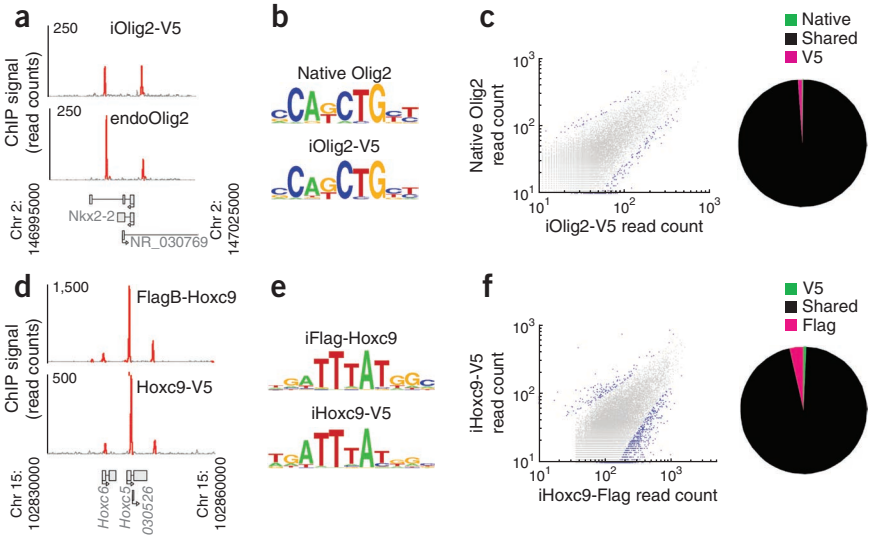

Figure 2 | Native- and tagged-protein ChIP comparisons. (**a**) ChIP signal tracks over *Nkx2-2* genomic loci for endogenous and V5-tagged Olig2. Red peaks represent significant ($P < 0.01$) enrichment over control. Genomic loci with coordinates are shown at the bottom. (**b**) The most overrepresented motifs discovered under ChIP-seq peaks for native Olig2 and Olig2-V5 ChIP experiments. (**c**) A comparison of read enrichment from native and V5-tagged Olig2 ChIP-seq experiments at all detected peaks (left). Blue dots in the scatterplot represent peaks significantly ($P < 0.01$) differently enriched in one experiment over the other. The pie chart shows the distribution of sites differently enriched between native Olig2 and V5-tagged protein ChIP-seq. (**d**–**f**) Same analysis as in **a**–**c** over the *Hoxc5* genomic locus for V5- and Flag-tagged Hoxc9 experiments.

Flag-tagged-protein experiment (**Fig. 2f**). We conclude that genomic regions we identified in both C- and N-terminally tagged ChIP-seq experiments are likely to represent native Hoxc9 binding events.

A comparison of the Olig2-V5 and Hoxc9-V5 ChIP experiments revealed a large fraction of non-overlapping peaks, which contrasts with biological replicates of Hoxc9-V5 ChIP-seq experiments that were virtually indistinguishable (**Supplementary Fig. 3**). Detailed analysis of the overlapping Olig2 and Hoxc9 ChIP-seq–enriched regions by the genome positioning system algorithm[8] in peaks revealed that Olig2 and Hoxc9 occupy proximal but distinct sites in these regions (**Supplementary Fig. 2**). Because a typical ChIP-seq peak covers ~200 base pairs (bp), these experiments might be revealing enhancers that are active in both motor neuron progenitors and postmitotic motor neurons.

The system we presented here is robust and allows the generation of multiple inducible cell lines in parallel. Among the 24 generated lines, we observed for only three lines problems with inducible protein expression, likely owing to the inherent toxicity of introduced transgenes (data not shown). Although the system is versatile and can be used to study both progenitors and differentiated cells, we observed that the efficiency and homogeneity of transgene induction declined in postmitotic neurons. Inducing the transgene at late progenitor stage resulted 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 doxycycline 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 described tools for rapid generation of ESC lines and production of unlimited quantities of isogenic differentiated cells for identification of developmentally relevant transcription-factor binding sites genome-wide. The cell lines can also be used for other studies, including the isolation and identification of transcription-factor binding partners by co-immunoprecipitation followed by mass spectrometry[10]. 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

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Accession codes. Gene Expression Omniubus (GEO): GSE30882.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

E.O.M. is funded as the David and Sylvia Lieb Fellow of the Damon Runyon Cancer Research Foundation (DRG-1937-07). We thank R. Sherwood (Harvard University) for sharing unpublished observations and T Jessell (Columbia University) for anti-Hoxc4 antibody. Personnel and work were supported by US National Institutes of Health grant P01 NS055923 (D.K.G., R.A.Y. and H.W.), R01 NS058502 (H.W.), R01HL081186 (M.K.) and Helmsley Stem Cell Starter grant (H.W.).

AUTHOR CONTRIBUTIONS

E.O.M. and G.M. generated the transcription factor inducible lines; E.O.M. performed phenotypic analysis of the derived lines. E.O.M., W.A.W. and C.A.M. performed ChIP experiments. M.I. and M.K. developed the ICE cell lines and vectors. E.O.M. performed expression analysis. E.O.M. and M.C. performed the western blot and protein-binding to immobilized DNA. S.M. analyzed ChIP-seq data. E.O.M., R.A.Y., D.K.G .and H.W. designed the experiments. E.O.M., S.M. and H.W. wrote the manuscript; D.K.G. revised the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Park, P.J. *Nat. Rev. Genet.* **10**, 669–680 (2009).
2. Iacovino, M. *et al. Stem Cells* **29**, 1580–1588 (2011).
3. Wichterle, H., Lieberam, I., Porter, J.A. & Jessell, T.M. *Cell* **110**, 385–397 (2002).
4. Novitch, B.G., Chen, A.I. & Jessell, T.M. *Neuron* **31**, 773–789 (2001).
5. Mukouyama, Y.S. *et al. Proc. Natl. Acad. Sci. USA* **103**, 1551–1556 (2006).
6. Chen, J.A. *et al. Neuron* **69**, 721–735 (2011).
7. Jessell, T.M. *Nat. Rev. Genet.* **1**, 20–29 (2000).
8. Guo, Y. *et al. Bioinformatics* **26**, 3028–3034 (2010).
9. Jung, H. *et al. Neuron* **67**, 781–796 (2010).
10. Kim, J., Cantor, A.B., Orkin, S.H. & Wang, J. *Nat. Protoc.* **4**, 506–517 (2009).

ONLINE METHODS

Cell culture. Mouse embryonic stem (ES) cells were cultured over a layer of mitomycin-C–treated fibroblast resistant to neomycin (Fisher) in EmbryoMax Dulbecco's modified Eagle medium (DMEM) (Fisher) supplemented with 10% ES tested fetal bovine serum (Invitrogen), L-glutamine (Gibco), 0.1 mM β-mercaptoethanol and 100 U ml^{-1} leukemia inhibitory factor (LIF).

Motor neuron differentiation of ES cells was performed as previously described[3]. Briefly, ES cells were trypsinized (Invitrogen) and seeded at 5×10^5 cells ml^{-1} in ADFNK medium (Advanced DMEM/F-12 medium (Invitrogen): neurobasal medium (Invitrogen) (1:1), 10% knockout serum replacement (Invitrogen), 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, 2 mM L-glutamine and 0.1 mM 2-mercaptoethanol) to initiate formation of embryoid bodies (day 0). Medium was exchanged on days 1, 2 and 5 of differentiation. Patterning of embryoid bodies was induced by supplementing medium on day 2 with 1 μM all-*trans* retinoic acid (RA, Sigma) and 0.5 μM agonist of hedgehog signaling (Smoothened agonist (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 1 μg ml^{-1} 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 sequence encoding 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 FlagB sequence[11] in frame and upstream of the *L1* recombination site.

Open reading frames of genes were cloned by 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 the manufacturer's instructions. The 5′ primer always contained the added CACC sequence to ensure directional integration. For each coding sequence, two alternative 3′ primers were used: with and without the stop codon, generating two pENTR plasmids for each gene (**Supplementary Table 1**).

The LR recombination scheme was as follows. (i) When constructing sequence encoding a V5-His C-terminal fusion protein, the pENTR plasmid with no stop codon was recombined with the p2Lox-V5. (ii) Nontagged protein sequences were generated by recombining the pENTR plasmid with STOP codon with the p2Lox-V5 plasmid. (iii) To generate N-terminal tagged protein sequences, the pENTR plasmid with stop codon was recombined with p2Lox-FlagB.

Inducible lines were generated by treating the recipient ESCs for 16 h with doxycycline to induce Cre recombinase expression followed by electroporation of either p2Lox-V5 and p2Lox-FlagB plasmids containing the desired construct. After selection with 250 ng ml^{-1} of G418 (Cellgro) selection, on average three resistant clones were picked, characterized and expanded.

Immunocytochemistry. Embryoid bodies were fixed with 4% paraformaldehyde in PBS (pH 7.4), embedded in optimal cutting temperature (OCT, Tissue-Tek) and sectioned for staining: 24 h at 4 °C for primary antibodies and 4 h at room temperature (23–26 °C) 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 (1:10,000, AB9610, Millipore); mouse anti-V5 (1:800, R960-25, Invitrogen); mouse anti-Flag M2 (1:400, F1804, Sigma); rabbit anti-Hoxc4 (1:2,000). Alexa Fluor 488–, 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 min at room temperature. Pellets containing ~40×10^6 cells were flash frozen and stored at −80 °C. Cells were thawed on ice, resuspended in 5 ml of lysis buffer A (50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Igepal and 0.25% Triton X-100) and incubated for 10 min at 4 °C in a rotating platform. Samples were spun down for 5 min at 1,350*g*, resuspended in 5 ml lysis buffer B (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, pH 8.0 and 0.5 mM EGTA, pH 8.0) and incubated for 10 min at 4 °C on a rotating platform. Samples were spun down for 5 min at 1,350*g*, resuspended in 3 ml of sonication buffer (50 mM Hepes pH 7.5, 40 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Trition X-100, 0.1% Na-deoxycholate and 0.1% SDS).

Nuclear extracts were sonicated using a Misonix 3000 model sonicator to sheer cross-linked DNA to an average fragment size of ~500 bp. Sonicated chromatin was incubated for 16 h at 4 °C with protein-G beads (Invitrogen) conjugated with 5 μg of 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 containing 500 nM NaCl, once with IgG LiCl wash buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40 and 0.5% Na-deoxycholate) and 1 ml TE. Then, beads were centrifugated at 950*g* for 3 min and residual TE removed with a pipette. Then 210 μl of elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0 and 1% SDS) was added to the beads followed by incubation at 65 °C for 45 min with a brief pulse of vortex every 10 min. Next 200 μl of supernatant was removed after a 1-min centrifugation at 16,000*g*. The cross-link was reversed by 16 h incubation at 65 °C.

RNA was digested by the addition of 200 μl of TE and RNase A (Sigma) at a final concentration of 0.2 mg ml^{-1} and incubated for 2 h at 37 °C. Protein was digested by the addition of proteinase K (0.2 mg ml^{-1} final, Invitrogen) supplemented with $CaCl_2$ followed by a 30-min 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 μl of water. Purified DNA fragments were processed according to the Illumina Solexa sequencing protocol using a Genome Analyzer II (Illumina).

ChIP-seq analysis. Sequence reads were aligned to the mouse genome (version mm9) using Bowtie[12] 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[8]. In GPS, the scaling ratio between ChIP-seq and control channels was estimated using the median ratio of all 10-Kbp windows along the genome. The GPS binding model was initialized to the default and iteratively updated over up to three training rounds. In this study, we required 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 between two ChIP-seq experiments, we first scaled the read counts assigned to each peak using the median ratio of observed read counts across all peaks. The read counts of one experiment were always scaled down to match the scale of the other experiment. We defined 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 as described previously[13]).

DNA motif analysis. *De novo* motif-finding was performed in 200-bp windows centered on the 2,000 top-ranked peaks for each examined ChIP-seq experiment. GimmeMotifs[14] 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[15] 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 1,000,000 200-bp sequences using a third-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 200-bp sequences centered on the Olig2 and Hoxc9 GPS-predicted peak positions.

ELISA DNA binding. PCR amplified and biotin-labeled genomic fragments were gel purified. The fragments were 500–600 bp with the 5′ primers containing a single biotin molecule at the 5′ end. Wash buffer was 0.05% Tween-20 and 0.1% BSA in PBS.

Streptavidin-coated 96-well plates (Fisher) were washed three times with 200 µl of wash buffer. The biotin-labeled PCR reaction was loaded into each well in blocking buffer (0.5% BSA in PBS) up to 125 µl final volume to saturate the binding capacity of each well (plates can bind ~125 pmol per 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 h 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 h with shaking at room temperature. After washing each well three times with 200 µl of wash buffer, horseradish peroxidase (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 min with shaking at room temperature. The wells were washed three times with 200 µl of wash buffer. Then 100 µl of room temperature TMB substrate solution (Fisher) was added to each well. After a 10-min 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.

Primers to amplify the ChIP-seq identified genomic regions are listed in **Supplementary Table 1**.

11. Wang, J., Cantor, A.B. & Orkin, S.H. *Curr. Protoc. Stem Cell Biol.* published online 1 March 2009 (doi: 10.1002/9780470151808.sc01b05s8). (2009).
12. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. *Genome Biol.* **10**, R25 (2009).
13. Benjamini, Y. & Hochberg, Y. *J. Royal Stat. Soc. B* **57**, 289–300 (1995).
14. van Heeringen, S.J. & Veenstra, G.J. *Bioinformatics* **27**, 270–271 (2011).
15. Mahony, S. & Benos, P.V. *Nucleic. Acids. Res.* **35**, W253–W258 (2007).

doi:10.1038/nmeth.1775