

Published in final edited form as:

*Nat Neurosci.* 2013 September ; 16(9): . doi:10.1038/nn.3467.

## Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity

Esteban O Mazzone<sup>1,2,6</sup>, Shaun Mahony<sup>3,4,6</sup>, Michael Closser<sup>1</sup>, Carolyn A Morrison<sup>1</sup>, Stephane Nedelec<sup>1</sup>, Damian J Williams<sup>5</sup>, Disi An<sup>2</sup>, David K Gifford<sup>3</sup>, and Hynek Wichterle<sup>1</sup>

<sup>1</sup>Departments of Pathology and Cell Biology, Neurology, and Neuroscience, Center for Motor Neuron Biology and Disease, Columbia Stem Cell Initiative, Columbia University Medical Center, New York, New York, USA

<sup>2</sup>Department of Biology, New York University, New York, New York, USA

<sup>3</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

<sup>4</sup>Department of Biochemistry and Molecular Biology, Center for Eukaryotic Gene Regulation, Pennsylvania State University, University Park, Pennsylvania, USA

<sup>5</sup>Department of Physiology and Cellular Biophysics, Columbia University Medical Center, New York, New York, USA

### Abstract

Efficient transcriptional programming promises to open new frontiers in regenerative medicine. However, mechanisms by which programming factors transform cell fate are unknown, preventing more rational selection of factors to generate desirable cell types. Three transcription factors, Ngn2, Isl1 and Lhx3, were sufficient to program rapidly and efficiently spinal motor neuron identity when expressed in differentiating mouse embryonic stem cells. Replacement of Lhx3 by Phox2a led to specification of cranial, rather than spinal, motor neurons. Chromatin immunoprecipitation–sequencing analysis of Isl1, Lhx3 and Phox2a binding sites revealed that the two cell fates were programmed by the recruitment of Isl1-Lhx3 and Isl1-Phox2a complexes to distinct genomic locations characterized by a unique grammar of homeodomain binding motifs. Our findings suggest that synergistic interactions among transcription factors determine the specificity of their recruitment to cell type–specific binding sites and illustrate how a single transcription factor can be repurposed to program different cell types.

© 2013 Nature America, Inc. All rights reserved.

Correspondence should be addressed to E.O.M. (eom204@nyu.edu), H.W. (hw350@columbia.edu) or D.K.G. (gifford@mit.edu).

<sup>6</sup>These authors contributed equally to this work.

**Accession codes.** All microarray and ChIP-seq data have been submitted to the US National Institutes of Health GEO database under accession number GSE31456.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

### AUTHOR CONTRIBUTIONS

E.O.M., S.M., D.K.G. and H.W. conceived the experiments, analyzed the data and wrote the manuscript. E.O.M. generated and validated inducible cell lines and performed the majority of experiments. S.M. performed all of the computational and statistical analyses of genomic, expression and sequencing data. M.C. performed motor neuron quantifications, synapse analysis and co-immunoprecipitation experiments. C.A.M. assisted with ChIP-seq experiments. S.N. performed axon pathfinding analysis. D.J.W. performed electrophysiological recordings. D.A. assisted with the analysis of single gene inducible lines and motor neuron induction.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

Recent progress in programming cell fate by transcription factors has given hope to the goal of producing clinically relevant cell types for disease modeling and direct therapeutic transplantation. Muscle cells, pluripotent stem cells, pancreatic beta cells, hepatocytes and several types of neurons have all been created by the forced expression of combinations of transcription factors known as programming modules<sup>1–7</sup>. However, the process of transcriptional programming remains largely enigmatic. Understanding the mechanism by which programming modules convert one expression profile to another one would not only illuminate the process of cell-fate specification during normal embryonic development, but would also have important implications for the rational design of programming modules for production of cell types that are difficult to generate using available methodologies.

When considering how programming modules associate with *cis*-regulatory elements to bring about a change in cellular identity, two extreme mechanisms can be hypothesized: the independent model, in which individual programming factors bind to distinct genomic locations and control independent subcircuits of the cell specific gene regulatory network<sup>8</sup>, and the synergistic model, in which the factors jointly bind to common regulatory elements to cooperatively activate the global cell type-specific expression program. The independent model predicts that the effect of a programming module could be estimated by additively combining the effects of each individual factor. The synergistic model, on the other hand, predicts that the DNA sequence binding preference of the programming module is encoded by the multimeric transcriptional complex<sup>9</sup> and that regulatory effects of the programming module will therefore be impossible to extrapolate from effects of individual factors studied in isolation.

Cellular complexity in the CNS is established during development by closely related programming modules acting in a cell type-specific manner<sup>10</sup>, providing a unique opportunity to study contributions of individual factors to the specification of cell fate. Motor neurons are cholinergic cells located in the ventral and caudal CNS, whose developmental program is particularly well mapped<sup>10</sup>. Two cardinal types of motor neurons are present in mammals, each expressing different transcriptional programs. Spinal somatic motor neurons (referred to here as spinal motor neurons) innervating skeletal muscles are derived from the ventral spinal progenitor domain and are characterized by coexpression of *Isl1*, *Lhx3* and *Hb9* (*Mnx1*) at the time of their birth<sup>10</sup>. Branchiomotor and visceromotor neurons (referred to here as cranial motor neurons) located in the ventral midbrain, hindbrain and cervical spinal cord are defined by the coexpression of *Isl1*, *Phox2a/2b* and *Tbx20* (refs. 10,11). The combined expression of *Isl1* and *Lhx3*, together with the proneural gene *Ngn2* (NIL factors), is sufficient to bestow spinal motor neuron identity on dorsal spinal progenitors and on spinal progenitors derived from embryonic stem cells (ESCs)<sup>12–14</sup>. Our current insights into the mechanisms through which NIL factors program spinal motor neuron identity are based on analysis of *Isl1* and *Lhx3* mutant phenotypes and on functional mapping of a spinal motor neuron specific *Hb9* (*Mnx1*) enhancer<sup>12,15</sup>. Although NIL factors synergize to control cell type-specific expression of the *Hb9* gene, mutations in *Isl1* and *Lhx3* result in distinct phenotypes, indicating that the two transcription factors may also possess independent functions<sup>16,17</sup>. Thus, the question of whether the NIL factors act primarily synergistically or independently at the genomic level remains unanswered.

To overcome the low efficiency of cell programming that limits biochemical analysis of the process, we established inducible ESC lines that harbor the NIL programming module or a module in which *Lhx3* is replaced by the cranial motor neuron determinant *Phox2a* (the NIP programming module)<sup>18–20</sup>. We found that NIL induction in differentiating ESCs resulted in rapid and highly efficient specification of spinal motor neurons and that NIP induction in an identical cellular context programmed cranial motor neuron identity. Taking advantage of these robust and efficient programming systems, we mapped genome-wide binding sites of

programming factors in both inducible lines. Computational analysis of occupied *cis*-regulatory elements revealed that *Isl1* directly interacts and synergizes with *Lhx3* or *Phox2a* in the relevant cellular contexts. The *Isl1*-*Lhx3* and *Isl1*-*Phox2a* heterodimers exhibited different DNA-sequence preferences, forming the basis of cell-specific programming module activities and indicating that synergistic interactions between programming factors underlie specification of alternate motor neuron fates.

## RESULTS

### **Ngn2, *Isl1* and *Lhx3* program spinal motor neuron fate**

To study programming of spinal and cranial motor neuron identity, we generated two doxycycline (Dox) inducible ESC lines<sup>21</sup>: one line harboring a polycistronic expression construct in which the open reading frames of spinal motor neuron determinants *Ngn2*, *Isl1* and *Lhx3* (refs. 12–14) are separated by 2A peptides (iNIL line), and a second line in which we replaced *Lhx3* with a cranial motor neuron determinant *Phox2a* (iNIP line) (Fig. 1). NIL factors have been shown to activate specification of motor neuron identity in retinoic acid-treated differentiating ESCs<sup>13,14</sup>. We found that NIL factors were sufficient to induce expression of spinal motor neuron markers even in the absence of retinoic acid. Treatment of differentiating ESCs with Dox resulted in robust induction of the tricistronic transgene 24 h later (Supplementary Fig. 1a,b). Notably, despite continuing Dox treatment, *Ngn2* expression was extinguished in most cells by 48 h, consistent with its transient pattern of expression in cells transitioning from progenitors to postmitotic motor neurons<sup>22,23</sup> (Supplementary Fig. 1c).

NIL-expressing cells plated on laminin adopted a typical neuronal morphology, expressed neuronal marker class III  $\beta$ -tubulin (Tub 3, recognized by Tuj1 antibody; Fig. 1b) and the spinal motor neuron marker Hb9, and were negative for cranial motor neuron marker *Phox2b* (Fig. 1b). Quantification revealed that the majority of transgenic cells (labeled by antibody to V5) expressed the postmitotic neuronal marker NeuN (99.72%  $\pm$  0.27 of V5<sup>+</sup> cells express NeuN) and Hb9 (99.82%  $\pm$  0.17 express Hb9), but rarely expressed *Phox2b* (0.24%  $\pm$  0.28 express *Phox2b*) (Fig. 1d and Supplementary Fig. 2a,b). The NIL programming module was capable of activating expression of Hb9 even when expressed in the context of naive ESCs (Supplementary Fig. 2c,d).

### **Ngn2, *Isl1* and *Phox2a* program cranial motor neuron fate**

Cranial motor neurons share many features with spinal motor neurons: they are cholinergic and express the transcription factor *Isl1* (ref. 17). However, unlike ventral spinal motor neurons, they express *Tbx20* instead of Hb9, and their specification depends on the paired-like homeodomain transcription factors *Phox2a* and *Phox2b* instead of LIM homeodomain factor *Lhx3* (refs. 19,24). Misexpression of the *Phox2a* and *Phox2b* transcription factors in the developing spinal cord is sufficient to induce ectopic cranial branchiomotor neurons<sup>18</sup>. However, the expression of *Phox2a* alone in differentiating ESCs resulted in only a small increase in the number of *Isl1*-positive cells, and most of the cells failed to acquire neuronal identity, as shown by the lack of Tub 3 expression (Supplementary Fig. 1d).

We reasoned that joint expression of *Ngn2* and *Isl1* with *Phox2a* (NIP module; Fig. 1a) might lead to a more robust and uniform specification of cranial motor neurons. Indeed, iNIP cells treated for 48 h with Dox, dissociated and plated on laminin acquired uniform neuronal morphology and identity (Tub 3 and NeuN expression) and expressed *Phox2b* in the absence of Hb9 (Fig. 1c,d and Supplementary Fig. 2a,b). The high efficiency of NIP programming was comparable to that of NIL programming: 99.77%  $\pm$  0.22 V5<sup>+</sup> cells expressed NeuN, 99.03%  $\pm$  0.08 expressed *Phox2b* and 0.11%  $\pm$  0.11 expressed Hb9 were

detected (Supplementary Fig. 2a,b). Together, these data indicate that replacement of *Lhx3* in the programming module with *Phox2a* results in efficient specification of neurons that acquired molecular properties of cranial motor neurons.

### Functional characterization of induced NIL and NIP neurons

To determine whether transcriptionally programmed cells acquired key properties of mature motor neurons, we cultured induced NIL and NIP cells alone or on monolayers of primary cortical mouse astrocytes for 7–10 d. Immunostaining of NIL and NIP cells cultured on monolayers of astrocytes revealed dense arrays of synapses marked by the synaptic vesicle marker SV2 (Fig. 1f). Notably, many of the synapses exhibited accumulation of vesicular acetylcholine transporter (Vacht), a marker of mature cholinergic synapse (Fig. 1f). Cholinergic identity of NIL and NIP induced cells was further documented by an approximately 70-fold increase in the levels of choline acetyl transferase (*Chat*) mRNA (Fig. 1e) and by Chat immunostaining (Fig. 1f).

Electrophysiologically mature motor neurons fire trains of action potentials following depolarization<sup>25,26</sup>. Whole-cell patch current-clamp recordings of NIL and NIP induced cells cultured on astrocytes for 7 d revealed that action potentials could be evoked by 20–150-pA, 1-s current injection in all cells tested (12 NIL cells, 12 NIP cells). Furthermore, nearly all patched cells (11 of 12 NIL cells, 11 of 12 NIP cells) fired trains of action potentials, sustained for the duration of the depolarizing current step (Fig. 1g). Together, these observations suggest that inducible expression of NIL and NIP programming modules is sufficient to differentiate ESCs into electrically mature cholinergic neurons.

Motor neurons project axons outside of the CNS to innervate peripheral synaptic targets. To examine whether induced motor neurons acquired this defining characteristic, we implanted control, iNIL and iNIP cells treated with Dox from days 2 to 4 of differentiation into the developing cervical and brachial neural tube of developing chick embryos<sup>27,28</sup>. We detected robust outgrowth of axons (labeled by mouse specific NCAM antibody) exiting spinal cord via the ventral root and extending along all major spinal motor nerves 2 d after implantation of iNIL neurons (four of five successfully transplanted embryos; Fig. 1h). In contrast, axons of control transplants stayed in the spinal cord and failed to project to the periphery (Fig. 1h). To further test the specificity of iNIL cell axonal pathfinding, we examined projections of iNIP cells transplanted into the same region of the chick neural tube. Unlike spinal motor neurons, cranial motor neurons do not exit the CNS through the ventral horn, preferring a more dorsal exit point<sup>19</sup>. iNIP axons accumulated selectively at the lateral region of the developing spinal cord, coalescing with spinal accessory nerve populated by branchiomotor cranial motor axons originating from the spinal accessory nucleus in the lateral cervical spinal cord (four of four successfully transplanted embryos; Fig. 1h)<sup>29</sup>. The same axonal trajectory has been observed for ectopic cranial motor neurons formed in the developing spinal cord following misexpression of *Phox2a* or *Phox2b*<sup>18</sup>. These results indicate that induced expression of the NIL and NIP modules programs cell phenotypes that are, by all examined criteria, consistent with spinal and cranial motor neuron identities (induced cranial and spinal motor neurons).

### Expression profiles of motor neuron programming

Effective programming of ESCs into motor neurons should be accompanied by a repression of the stem cell expression program and induction of the spinal or cranial motor neuron specific transcriptome. Global expression profiling (Affymetrix GeneChip ST arrays) revealed that 48 h of Dox treatment of iNIL and iNIP cells resulted in a marked change in gene expression profile (3,185 and 1,852 genes were more than twofold differentially expressed following NIL and NIP induction, respectively,  $P < 0.001$ ; Fig. 2a). Induction of

NIL and NIP programming modules extinguished the expression of pluripotency genes (*Oct4*, *Nanog*), upregulated generic motor neuron genes (endogenous *Isl1*, *Ebf1/3*, *Onecut1/2*), cholinergic genes (*Vacht* (also known as *Slc18a3*), *Chrn4*) and genes encoding axon guidance molecules (*Nrp1*, *Robo1/2*, *Dcc*) (Fig. 2b and Supplementary Fig. 3a). Comparison of iNIL and iNIP induced cells revealed significant differences between the two samples (2,731 differentially expressed more than twofold,  $P < 0.001$ ; 1,878 genes were upregulated in iNIL cells compared with iNIP, 857 genes were upregulated in iNIP cells compared with iNIL; Fig. 2a). Although spinal motor neuron genes *Hb9*, *Isl2*, endogenous *Lhx3* and *Slit1/2* (ref. 30) were selectively expressed in iNIL cells, iNIP cells upregulated expression of cranial motor neurons markers *Tbx20*, endogenous *Phox2a*, *Phox2b*, *Rgs4* and *Ga $\beta$ <sup>1-33</sup>* (Fig. 2b, Supplementary Fig. 3a). Unsupervised clustering of expression profiles revealed that Dox-treated iNIP cells segregated from Dox-treated iNIL cells (Fig. 2c), indicating that their identities are molecularly distinct.

Spinal motor neurons can be generated from ESCs by directed differentiation controlled by patterning signals retinoic acid and sonic hedgehog (Shh)<sup>27</sup>. Retinoic acid- and Shh-driven differentiation recapitulates the normal process of motor neuron development. ESCs first acquire neural progenitor identity on day 3 (Sox1<sup>+</sup>, Olig2<sup>-</sup>), followed by motor neuron progenitor stage on day 4 (Olig2<sup>+</sup>, Hb9<sup>-</sup>), and differentiate to postmitotic motor neurons (Olig2<sup>-</sup>, Hb9<sup>+</sup>) on days 5–6 (refs. 21,27). To follow spinal motor neuron induction, we introduced a GFP transgene driven by the *Hb9* promoter into the iNIL cell line<sup>27</sup>. Although control cultures of iNIL cells contained few or no GFP-positive cells, ~40% of the cells treated with retinoic acid and Shh for 4 d, and the majority of cells treated with Dox for 48 h, became GFP positive (Supplementary Fig. 3b).

To examine how closely programmed neurons correspond to retinoic acid- and Shh-generated motor neurons, we compared expression profiles of FACS-purified Hb9-GFP<sup>+</sup> retinoic acid- and Shh-generated motor neurons on day 5 of differentiation with Hb9-GFP<sup>+</sup> cells purified from iNIL cultures treated with Dox for 48 h. The induced motor neurons were markedly similar to retinoic acid- and Shh-generated motor neurons (Fig. 2b–d). Most genes (97.4%) were expressed at levels that were not significantly different between the two samples ( $P < 0.001$ ), and only 1.6% of all genes exhibited divergent expression (that is, are induced in one cell type and repressed in the other). The similarity of induced and control motor neurons is further supported by unsupervised clustering of gene expression profiles (Fig. 2c). Although key motor neuron-specific genes were correctly regulated, a set of genes controlling rostro-caudal neural identity and motor neuron subtype identity was differentially expressed in retinoic acid- and Shh-generated cells and induced iNIL cells (Fig. 2e). Induced iNIL motor neurons expressed low levels of *Hox* transcription factors and high levels of rostral neural markers (*Otx1*, *Otx2*). To rectify this difference, we asked whether programmed iNIL motor neurons are responsive to the caudalizing signal retinoic acid<sup>27,34</sup>. Treatment of iNIL cells with retinoic acid during the Dox treatment resulted in correct specification of cervical spinal identity, marked by the expression of *Hox* genes from paralogous groups 4 and 5 and suppression of rostral markers *Otx1/2* (Fig. 2e). Thus, programmed cells acquire generic motor neuron identity following induction of NIL factors, but specification of rostro-caudal subtype identity depends on the treatment of the cells with caudalizing patterning signals.

### Motor neuron programming bypasses neural progenitor stages

Rapid activation of postmitotic motor neuron markers following NIL and NIP induction raised the question of whether transcriptionally programmed cells transit through neural progenitor stages. To capture cells during the transition from ESCs to spinal motor neurons, we profiled induced cells 24 h after Dox treatment. At this time point, NIL factors had

effectively repressed key stem cell genes (for example, *Oct4*, *Nanog*; Fig. 3a) and had already induced expression of markers associated with postmitotic spinal motor neuron identity, such as *Hb9* (*Mnx1*), *Isl2*, *Lhx4*, *VACHT* (*Slc18a3*), *Robo2*, *Slit2* and *Nrp1* (Fig. 3a). Notably, although proliferating cells (Ki67<sup>+</sup>) were intermixed with cells expressing *Hb9*, none of the dividing cells expressed the motor neuron progenitor marker *Olig2* (Fig. 3b).

Neither NIL nor NIP induced expression of genes associated with progenitor stages (*Sox1*, *Olig2* and *Ngn2*) that were highly expressed at 24 and 48 h following retinoic acid and Shh treatment (days 3 and 4; Fig. 3a,c and Supplementary Fig. 3c). These results indicate that programming modules initiate a state transition from ESCs to postmitotic motor neurons that bypasses key steps in the normal motor neuron developmental program.

### Isl genome binding is dependent on programming partners

Efficient and rapid programming of ESC differentiation into phenotypically distinct neurons by modules that differ only in one transcription factor provides an ideal system in which to study whether individual transcription factors act independently or engage in synergistic interactions. If the individual factors are recruited to DNA independently, replacing *Lhx3* with *Phox2a* in the programming module should not affect the DNA binding preference of *Isl1*. To test this independent model, we performed chromatin immunoprecipitation–sequencing (ChIP-seq) analyses of *Isl1* in iNIL and iNIP cells 48 h after Dox induction. Inducible *Isl1* was not epitope tagged and we optimized ChIP using a pool of monoclonal antibodies to *Isl1*. As these antibodies cross-react with both *Isl1* and the closely related *Isl2*, we refer to the data as *Isl* ChIP-seq.

We observed extensive condition-specific *Isl* recruitment to genomic loci in the iNIL and iNIP induced cells (Fig. 4a). We identified 18,187 significant *Isl* binding events (Online Methods) in the two conditions, of which 38% were significantly differentially enriched ( $P < 0.001$ ) between iNIL and iNIP lines (Fig. 4b). In contrast, only 9.6% of the *Isl* binding sites were differentially enriched between iNIL cells and retinoic acid– and Shh-derived motor neurons. To further test the synergistic model, we profiled the binding of *Isl* when ectopically expressed alone in differentiating ESCs. The genomic occupancy of *Isl* was substantially different from that of *Isl* expressed in the context of either iNIL or iNIP cells (Supplementary Fig. 4b): 67.5% of the *Isl* ChIP peaks were differentially enriched between iNIL and i*Isl1* cells, and 48.8% were differentially enriched between iNIP and i*Isl1* cells. These results indicating that recruitment of *Isl1* to DNA binding sites depends on the composition of programming modules. Our findings are consistent with the synergistic model, implicating functional interactions between programming factors.

Next we examined whether identified *Isl* binding sites are distributed randomly across the genome or whether their positions correlate with tissue specific *cis*-regulatory elements. We took advantage of ENCODE project data that identified putative regulatory regions in mouse ESCs, whole brain, heart, kidney, liver and spleen, defined using combinations of DNaseI hypersensitivity and enrichment in H3K4me1 and H3K27ac histone modifications<sup>35</sup>. Of all of the tissues examined, *Isl* binding sites correlated best with whole brain putative regulatory regions (Supplementary Fig. 4a). Notably, the overlap with regulatory regions in ESCs was as low as in unrelated tissues (Supplementary Fig. 4a). These findings indicate that expressed NIL and NIP factors are not passively recruited to existing stem cell regulatory regions, but that these factors actively engage neuronal regulatory regions.

### Cell-specific *Isl* binding correlates with gene expression

The identification of condition-specific *Isl* binding prompted us to examine whether differentially occupied sites in the iNIL and iNIP cells are associated with the establishment

of cell type-specific gene expression profiles. We observed condition-specific Isl binding in the vicinity of developmental genes that were selectively induced by the NIL or NIP programming modules. For example, three sites that were bound by Isl in the iNIL cells, but not in the iNIP cells, were located downstream of the endogenous *Lhx3* gene; conversely, two sites that were bound in the iNIP cells, but not in the iNIL cells, were located near the *Phox2b* transcription start site (TSS) (Fig. 4a). Meanwhile, we observed shared Isl binding sites near a subset of genes that are induced in both cell types, such as *Chat* (Fig. 4a).

We then asked what fraction of genes differentially expressed between the two conditions were proximal to sites that were differentially occupied by Isl. To extend this analysis, we asked what fraction of genes differentially expressed between the two conditions are proximal to sites that are differentially occupied by Isl. We first subdivided all induced genes that have nearby (overlapping the gene or <10 kbp upstream or downstream from the gene TSS) Isl binding sites into three categories: those that were induced in both cell lines, those that were induced selectively in iNIL cells (NIL induced) and those that were selectively induced in iNIP cells (NIP induced). Similarly, Isl binding sites were subdivided into condition-specific sites that were most differentially enriched ( $P < 0.001$ ) in iNIL cells (5,285 sites, NIL>NIP), those most differentially enriched in iNIP cells (1,657 sites, NIP>NIL; Fig. 4b) and condition-independent sites that were similarly enriched in the two cell lines (1,705 sites, NIL = NIP; Online Methods). Of all of the NIL induced genes, 57% had a nearby NIL>NIP site, 26% had a nearby NIL = NIP site and only 13% had a nearby NIP>NIL site (Fig. 4c). Conversely, 70% of genes induced selectively in iNIP cells had a nearby NIP>NIL Isl binding site, 22% had sites similarly occupied in both cell lines and only 14% had a nearby NIL>NIP (Fig. 4c). Based on the correlation between condition-specific Isl binding and condition-specific activation of gene expression, we propose that a subset of Isl binding sites function as context-dependent enhancers contributing to the establishment of the observed cell type-specific pattern of gene expression and to cell fate programming.

### Sequence motifs explain differential Isl binding

Given that Isl genomic binding depends on the programming module context, we reasoned that Isl1 might partner with different transcription factors during NIL- and NIP-mediated cell fate programming, resulting in a global change in its DNA binding preference. To elucidate the mechanisms underlying differential recruitment of Isl1 to genomic sites, we analyzed the DNA motifs enriched in the condition-specific and condition-independent binding sites. Motif analysis identified a monomeric sequence with consensus TAAKKR under the condition-independent (NIL = NIP) sites, which is identical to the *in vitro* binding preference characterized for Isl2 (Supplementary Fig. 4c)<sup>36</sup>.

The analysis of differentially enriched sites revealed more complex dimeric motifs composed of a combination of two homeodomain binding sites (Fig. 5a,b). Notably, the motifs associated with iNIL and iNIP condition-specific sites exhibited different motif grammar. Although the homeodomain half-sites formed an inverted repeat in the motif enriched under NIL-specific sites, the motif enriched under NIP-specific sites contained an everted half-site configuration (Fig. 5b). The motifs were highly enriched under NIL- and NIP-specific Isl ChIP-seq peaks, with 60.1% of NIL and 33.5% of NIP peaks containing the NIL- and NIP-specific motif, respectively ( $2.5 \times 10^{-23}$  false discovery rate). The marked specificity of the ordering of homeodomain binding motifs in selectively occupied sites suggests that Isl1 partners with two different homeodomain transcription factors in iNIL and iNIP cell lines. The differences in the structure of these transcription factor complexes likely underlie their sequence specific recruitment to DNA, providing a physical mechanism by which one transcription factor can regulate different targets to establish alternate cellular identities.

### Lhx3 and Phox2a co-occupy Isl-binding sites

Previous analysis of the spinal motor neuron-specific *Hb9* enhancer revealed that Isl1 forms a multimeric complex with Lhx3, Ldb1 and Ngn2 or Neurod4<sup>12</sup>. We asked whether Lhx3 co-occupies other sites selectively bound by Isl in iNIL cell line. Taking advantage of the V5 epitope tag on the *Lhx3* transgene<sup>21</sup>, we performed ChIP-seq analysis of Lhx3 binding in the iNIL cells 48 h after Dox induction. We identified 47,908 Lhx3 binding sites in the genome and found that these sites were highly coincidental with the sites occupied by Isl in the iNIL cell line. We observed that only 1.7% of all sites were significantly differentially enriched ( $P < 0.001$ ) in one experiment compared with the other (Fig. 5a,c). These findings suggest that Isl1 and Lhx3 bind to DNA as an obligatory heterodimer during spinal motor neuron differentiation.

Although there is no prior evidence that Phox2a heterodimerizes with Isl1, we examined whether V5 epitope tagged Phox2a might pair with Isl1 in the iNIP cell line. Although ChIP-seq analysis revealed only 1,568 significant Phox2a binding events, Phox2a and Isl binding events were highly coincident and the magnitude of ChIP enrichment at the co-bound sites was also highly correlated, mirroring the co-binding of Isl and Lhx3 in iNIL cells. We observed that only 4.6% of all sites were significantly differentially enriched in one experiment compared with the other (Fig. 5a,c). The high degree of co-binding of Isl and Phox2a raised the possibility that the two factors might be parts of the same transcriptional complex. It has been shown that purified Isl1 and Lhx3 transcription factors interact in solution<sup>126</sup>. Co-immunoprecipitation experiments confirmed the Isl1-Lhx3 interaction in induced iNIL cells and revealed that Isl1 and Phox2a are members of the same transcriptional complex in iNIP cells (Fig. 5d). Together, these results indicate that the alternate cellular fates produced by NIL and NIP programming modules are encoded by cooperative recruitment of Isl-Lhx3- and Isl-Phox2a-containing complexes to enhancers with distinct motif grammar.

## DISCUSSION

We exploited the differentiation potential of pluripotent ESCs to study how transcription factor modules control specification of distinct neuronal cell types. Inducible expression of two programming modules differing in one transcription factor led to a rapid and efficient specification of cells expressing key molecular and functional properties of spinal and cranial motor neurons. Isl1 transcription factor changed its genome binding preference when expressed alone or in the context of either the NIL or NIP programming modules. Because the factors were expressed in identical cellular context, the different binding preference of Isl cannot be attributed to differential chromatin accessibility or initial presence of distinct cofactors. Our data support a model in which Isl forms transcriptional complexes with Lhx3 or Phox2a. The complexes are recruited to condition-specific enhancers with differential motif grammar leading to activation of cell type-specific expression programs and to specification of spinal or cranial motor neurons (Supplementary Fig. 5). These findings have broader consequences for the rational design of programming modules, as mapping an individual transcription factor's DNA binding preference is insufficient to predict its binding and its potential for cellular programming when it is coexpressed with other cooperating programming factors. Systematic computational and experimental analysis of interactions among programming factors, along with decoding the grammar of their cooperative binding motifs, will be a fundamental step toward rational design of programming modules for predictable production of diverse cell types of interest.

The synergistic nature of the programming module's activity could explain why collections of factors are typically required to program terminal cell fate<sup>2-7,37</sup>. It is of interest that Oct4, Klf4 and Sox2 (core module) co-occupy regulatory elements in ESCs<sup>38-40</sup>, suggesting that



combinatorial programming modules may be a general developmental strategy. A second set of transcription factors (Myc module) appears to operate in parallel to the core module in pluripotent stem cells<sup>40</sup>. We therefore anticipate that additional transcriptional modules besides NIL and NIP will contribute to the establishment of terminal motor neuron expression profiles. Notably, the NIL programming module does not activate expression of Hox transcription factors that control specification of motor neuron subtype identity<sup>41</sup>. This is consistent with our recent demonstration that rostro-caudal patterning signals specify motor neuron positional identity by remodeling Hox chromatin landscape during early neural progenitor stages that are bypassed during direct programming by NIL factors<sup>42</sup>. Thus, generic motor neuron identity can be experimentally uncoupled from the Hox-driven program controlling subtype-specific motor neuron properties. Evolution of a generic motor neuron program that operates in parallel with transcription factors controlling subtype-specific programs would provide a versatile and efficient system for diversification of generic motor neurons into distinct subtypes necessary for the assembly of a functioning motor system.

Currently, the identification of effective programming modules relies on empirical testing of combinations of transcription factors expressed in the target cell type. In contrast, the most effective programming module for specification of motor neuron identity is composed of transcription factors expressed only transiently during the transition from motor neuron progenitor to postmitotic state. We propose that selection of effective programming modules for other types of nerve cells should focus on transcription factors expressed during similar developmental windows. Without doubt, direct programming of cellular identity will have a substantial effect on human stem cell applications<sup>13</sup>. Differentiation of human pluripotent stem cells to neurons is currently relatively inefficient and slow, taking weeks to months of *in vitro* culture<sup>43–45</sup>. Understanding the logic and function of programming modules might not only inform ways to generate cell types refractory to efficient programming by extrinsic patterning signals, but might also substantially accelerate production of homogenous cell populations necessary for human disease modeling, cell-based drug screening and transplantation therapy.

## ONLINE METHODS

### Cell culture

ESCs were cultured over a layer of Mitomycin-C–treated fibroblast resistant to Neomycin (Fisher) in EmbryoMax D-MEM (Fisher) supplemented with 10% ESC-grade fetal bovine serum (vol/vol, Invitrogen), L-glutamine (Gibco), 0.1 mM  $\beta$ -mercaptoethanol and 100 U ml<sup>-1</sup> leukemia inhibitory factor.

Motor neuron differentiation of ESCs was performed as previously described. Briefly, ESCs were trypsinized (Invitrogen) and seeded at  $5 \times 10^5$  cells per ml in ANDFK medium (Advanced DMEM/F12:Neurobasal (1:1) Medium, 10% Knockout-SR (vol/vol), Pen/Strep, 2 mM L-glutamine, and 0.1 mM 2-mercaptoethanol) to initiate formation of embryoid bodies (day 0). Medium was exchanged on days 2 and 5 of differentiation. Patterning of embryoid bodies was induced by supplementing media on day 2 with 1  $\mu$ M all-trans retinoic acid (Sigma) and 0.5  $\mu$ M Smo agonist of hedgehog signaling (SAG, Calbiochem). For ChIP experiments, the same conditions were used, but scaled to seed  $10^7$  cells on day 0. Dox (Sigma) was added to the culture medium at 3  $\mu$ g ml<sup>-1</sup> 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<sup>21</sup>.

Open reading frames of genes are cloned by PCR with or without overhangs containing the two alanine sequences. 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.

Inducible lines were generated by treating the recipient ESCs for 16 h with Dox to induce Cre followed by electroporation of p2Lox-NIL:V5 or p2Lox-Oc1: V5 plasmid. After G418 selection, resistant clones were picked, characterized and expanded. An *Hb9-GFP* transgene was incorporated by electroporation of the *Hb9-GFP* plasmid with Hygromycin resistance. After 15 d of selection, clones were expanded. The clone where *Hb9-GFP* recapitulates Hb9 expression was kept.

## Immunocytochemistry

Embryoid bodies were fixed with 4% paraformaldehyde (vol/vol) in phosphate-buffered saline, embedded in OCT (Tissue-Tek) and sectioned for staining: 24 h at 4 °C for primary antibodies and 4 h at 20–25 °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. We used antibodies to Olig2 (MABN50, Millipore, 1:2,000), Ngn2 (SC-19233, Santa Cruz, 1:200), V5 (R960, Invitrogen, 1:500), Ki67 (550609, BD Biosciences, 1:100), Tuj1 (ab7751, Abcam, 1:500), SV2 (SV2, Developmental Studies Hybridoma Bank, 1:100), 3-Tubulin (D71G9, Cell Signaling, 1:1,000), Vacht (AB1578, Chemicon, 1:1,000); NeuN (MAB377, Chemicon, 1:500), mNCAM (1:1,000), Hb9 (1:5,000), Isl1 (1:100); Chat (1:500) (gifts from T. Jessell); and Phox2b (1:500; gift from J.-F. Brunet, Institut de Biologie de l'École Normale Supérieure). Alexa 488–(A11001, A11008), FITC–(715-095-150, 706-095-148), Cy3–(715-165-150, 711-165-152, 706-165-148) and Cy5-conjugated (715-175-150, 711-175-152) secondary antibodies were used (Invitrogen or Jackson Immunoresearch, 1:2,000). For quantification, 100 cells in random fields were scored for coexpression of transgene and motor neuron markers by at least two independent authors ( $n = 1$ ). Data distribution was assumed to be normal, but this was not formally tested. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those generally employed in the field.

## FACS

Differentiating embryoid bodies were washed with phosphate-buffered saline and then dissociated by mild trypsinization (Invitrogen) followed by mechanical dissociation into single-cell suspension. GFP intensity of single cells was measured on Beckman Coulter MoFlo XDP FACS sorter.

## Expression analysis

RNA was extracted with Qiagen RNeasy at different time points following manufacturer's instructions. For quantitative PCR analysis, cDNA was synthesized using SuperScript III (Invitrogen) and amplified using SYBR green brilliant PCR amplification kit (Stratagene) and quantified using an Mx3000 thermocycler (Stratagene). Data distribution was assumed to be normal, but this was not formally tested. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those generally employed in the

field. One independent differentiation was considered to be a biological replicate ( $n = 1$ ). We used primers to *Actb* (TGAGAGGGA AATCGTGCCTGACAT and ACCGCTCGTTGCCAATAGTGATGA), *eIsl1* (GTTGGAGAAAGTGGGAAATGAC and TAGAACAGACTTCATGCGCTTC), *eNgn2* (TCGGCTTTAACTGGAGTGCC and GTGTGTTGTCGTTCTCGTGC), *eLhx* (CGTAGCCTCTAAATGCGAGA and TGGCAAAGGTGTCTGTTTAC), *ePhox2a* (CTCCTCCAACGTGTGCGCTT and CTTGTAGGGAACAGTCCG AGT), *Hb9* (GAACACCAGTTCAAGCTCAACA and CTCTTCCGTCTT CTCCTCACTG), *Tbx20* (TGACATTGAGAGGGAGAGTGTG and GAAG ATGACGATACCCAGGAAC), *Chat* (CGGTTTATTCTCTCCACCAG and TAACAGGCTCCATACCCATT) and *Vacht* (AGACTATGCCACGCTCTTCG and AGGCTCCTCGGGATACTTGT). For arrays, RNA was amplified using NuGen Applause bundle amplification and labeling kit according to manufacturer's instructions. Labeled cDNA hybridized to Affymetrix Mouse Gene 1.0 ST microarrays. Arrays were scanned using the GeneChip Scanner 3000. Data analysis was carried out using the oneChannelGUI BioConductor package<sup>46</sup>. RMA-sketch normalization was performed across all arrays. Differentially expressed genes were defined by ranking all probesets by the moderated  $t$ -statistic-derived  $P$  value (adjusted for multiple testing using Benjamini & Hochberg's method and setting a threshold of  $P < 0.001$ ). Expression clustering of normalized data transformed by fold-change with respect to the average of the control replicates was performed using the MeV software (version 4.7). When clustering the expression data, MeV used average linkage clustering with Pearson correlation as the distance metric.

In the text, we use the term divergent expression to refer to a subset of genes that are significantly differentially expressed between two conditions. With this term, we aim to ignore genes that are differentially expressed between conditions, but also similarly differentially expressed between each condition and the control; for example, the case in which a gene is twofold upregulated (w.r.t. control) in condition A and fourfold upregulated in condition B. For a gene to be called divergently upregulated, it must be significantly upregulated between conditions A and B, significantly upregulated between A and control, but not significantly upregulated between B and control (or conversely, significantly downregulated between B and control and not significantly downregulated between A and control).

## ChIP

Cells were fixed with 1% formaldehyde (vol/vol) for 15 min at 20–25 °C. Pellets containing  $\sim 40 \times 10^6$  cells were flash frozen in liquid nitrogen 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 (vol/vol), 0.5% Igepal (vol/vol), 0.25% Triton X-100 (vol/vol)) 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 200 mM NaCl, 1 mM EDTA, pH 8.0, 0.5 mM EGTA, pH 8.0) 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 3 ml of Sonication Buffer (50 mM Hepes (pH 7.5), 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate (wt/vol), 0.1% SDS (wt/vol)). Nuclear extracts were sonicated using a Misonix 3000 model sonicator to shear cross-linked DNA to an average fragment size of approximately 500 bp. Sonicated chromatin was incubated for 16 h at 4 °C with Protein-G beads (Invitrogen) conjugated with either rabbit antibody to V5 (Abcam) or monoclonal antibody to Isl1 (generous gift from T. Jessell and S. Morton, Columbia University). After incubation, and with the aid of a magnetic device, beads were washed once with Sonication Buffer and 500 nM NaCl and once with LiCl Wash Buffer (20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate) and 1 ml of TE (10 mM Tris, 1 mM EDTA, pH 8). Then, beads were centrifuged at 950  $g$

for 3 min and we removed residual TE with a pipette. 210  $\mu$ l of Elution Buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 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. 200  $\mu$ l of supernatant was removed after a 1 min centrifugation at 16,000 *g*. The crosslink was reversed by 16 h incubation at 65 °C. RNA was digested by the addition of 200  $\mu$ l of TE and RNaseA (Sigma) at a final concentration of 0.2 mg ml<sup>-1</sup> and incubated for 2 h at 37 °C. Protein was digested by the addition of Proteinase K (0.2 mg ml<sup>-1</sup> final, Invitrogen) supplemented with CaCl<sub>2</sub> 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  $\mu$ l of water. Purified DNA fragments were processed according to the Illumina/Solexa library protocol and sequenced using a Genome Analyzer II (Illumina, <http://www.illumina.com/pages.ilmn?ID=252>).

### ChIP-seq analysis

Sequence reads were aligned to the mouse genome (version mm9) using Bowtie<sup>47</sup> 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<sup>48</sup>. In GPS, the scaling ratio between immunoprecipitation 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. We required that reported peaks contain a ChIP-seq enrichment level that was significantly greater than 1.5 times the control level with a *P* value <0.01 as tested using the binomial distribution. When comparing enrichment levels between the two ChIP-seq experiments for different factors, 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 had a scaled read count in one experiment that was 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). When comparing enrichment levels between two ChIP-seq experiments profiling the same factor in different conditions, we made the assumption that the protein would likely bind to the same bases if it bound to a given region across multiple conditions. We therefore used the multi-condition mode in GPS to predict consistent binding event locations across all profiled conditions. Reads were assigned to binding events by GPS in a replicate-wise manner. EdgeR was then used to normalize the count data and to call differentially enriched binding events<sup>49</sup>. We defined differentially enriched events between a pair of conditions as those assigned a *P* value less than 0.001 by EdgeR. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those generally employed in the field.

### DNA motif analysis

*De novo* motif-finding was performed in 200-bp windows centered on the 500 top-ranked peaks for each examined subset of ChIP-seq binding events. GimmeMotifs (van Heeringen and Veenstra) 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.

### Comparison of peaks and expression

To compare peak locations with gene expression, we focused only on the genes that were probed on the Affymetrix Mouse Gene 1.0ST array that also had annotations in the Ensembl mouse annotation (release 62). We defined a peak as being located near a gene (or vice

versa) if the peak was located within 10 kbp of the gene's TSS or was overlapping the gene body (up to a maximum of 500 kbp from the TSS). Data distribution was assumed to be normal, but this was not formally tested. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those generally employed in the field.

### Western blot

Day 4 cultures, after 48 h of Dox, were lysed in RIPA buffer and loaded with Gel Loading Buffer II (Amersham) after boiling on NuPAGE Novex 4–12% gradient gels (Invitrogen). Transferred protein on PDVF membrane was incubated with mouse antibody to V5 (Invitrogen, R960, 1:5,000) and revealed with GE Amersham ECL kit according to manufacturer instructions.

### Electrophysiology

For whole-cell patch-clamp recordings, NIP and NIL cells were co-cultured with mouse primary cortical astrocytes. Briefly, astrocytes were prepared as previously described<sup>50</sup> and plated on 25-mm diameter coverslips at a density of 50,000 cells per well in a six-well plate. 4 d following astrocyte plating, freshly dissociated NIP and NIL cells were added to the wells at a density of 100,000 cells per well. Cultures were maintained for 7 d before recording. Current-clamp recordings were performed using an Axopatch 2B amplifier. Data were digitized using a Digidata 1322A digital to analog converter and were recorded at a 10-kHz sample rate using pClamp 10 software (all equipment from Molecular Devices). Patch pipettes were fabricated using a P-97 pipette puller (Sutter Instruments). The external recording solution contained 145 mM NaCl, 5 mM KCl, 10 mM HEPES, 10 mM glucose, 2 mM CaCl<sub>2</sub> and 2 mM MgCl<sub>2</sub>. The pH was adjusted to 7.3 using NaOH and the osmolality adjusted to 325 mOsm with sucrose. The pipette solution contained 130 mM CH<sub>3</sub>KO<sub>3</sub>S, 10 mM CH<sub>3</sub>NaO<sub>3</sub>S, 1 mM CaCl<sub>2</sub>, 10 mM EGTA, 10 mM HEPES, 5 mM MgATP and 0.5 mM Na<sub>2</sub>GTP (pH 7.3, 305 mOsm). Experiments were performed at 21–23 °C. During recordings, current was injected to hold the cells at –60 mV. Action potentials were evoked using incrementally increasing current steps 1 s in duration. The maximum amplitude of the current step (20–50 pA) and the size of the increment were calculated from the input resistance of the cell. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those generally employed in the field.

### Co-immunoprecipitation

Cells were grown, differentiated and crosslinked as described for chromatin immunoprecipitation experiments ( $n = 1$ ). Cells were lysed in RIPA (50 mM Tris (pH 8.0), 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) with protease inhibitors (Roche, 11697498001). Extracts were sonicated four times for 20 s using a Misonix 3000 at 70%. Extracts were then centrifuged at 4 °C for 10 min at 20,000 *g*. Protein complexes were immunoprecipitated overnight at 4 °C using 5 µg of rabbit antibody to V5 (Abcam, ab1229) or rabbit IgG (NB810-56910, Novus Biologicals) bound to 50 µl Dynabeads (Life Technologies, 100-04D). Immunoprecipitates were washed three times with lysis buffer followed by 10 min of boiling in SDS running buffer. Western blots were developed using mouse antibody to Isl1 (39.4D5, 1:1,000, Developmental Studies Hybridoma Bank).

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank I. Lieberam (King's College) for assistance introducing the *Hb9-GFP* transgene, members of the Wichterle laboratory for helpful comments, S. Brenner-Morton and T. Jessell (Columbia University) for sharing clones of Isl monoclonal antibodies, J.-F. Brunet for Phox2a and Phox2b antibodies, and I. Schieren for technical assistance with flow cytometry. E.O.M. receives funding from the Damon Runyon Cancer Research Foundation (DRG-1937-07). Personnel and work were supported by the Project ALS foundation and US National Institutes of Health grants P01 NS055923 (D.K.G. and H.W.) and R01 NS078097 (H.W.).

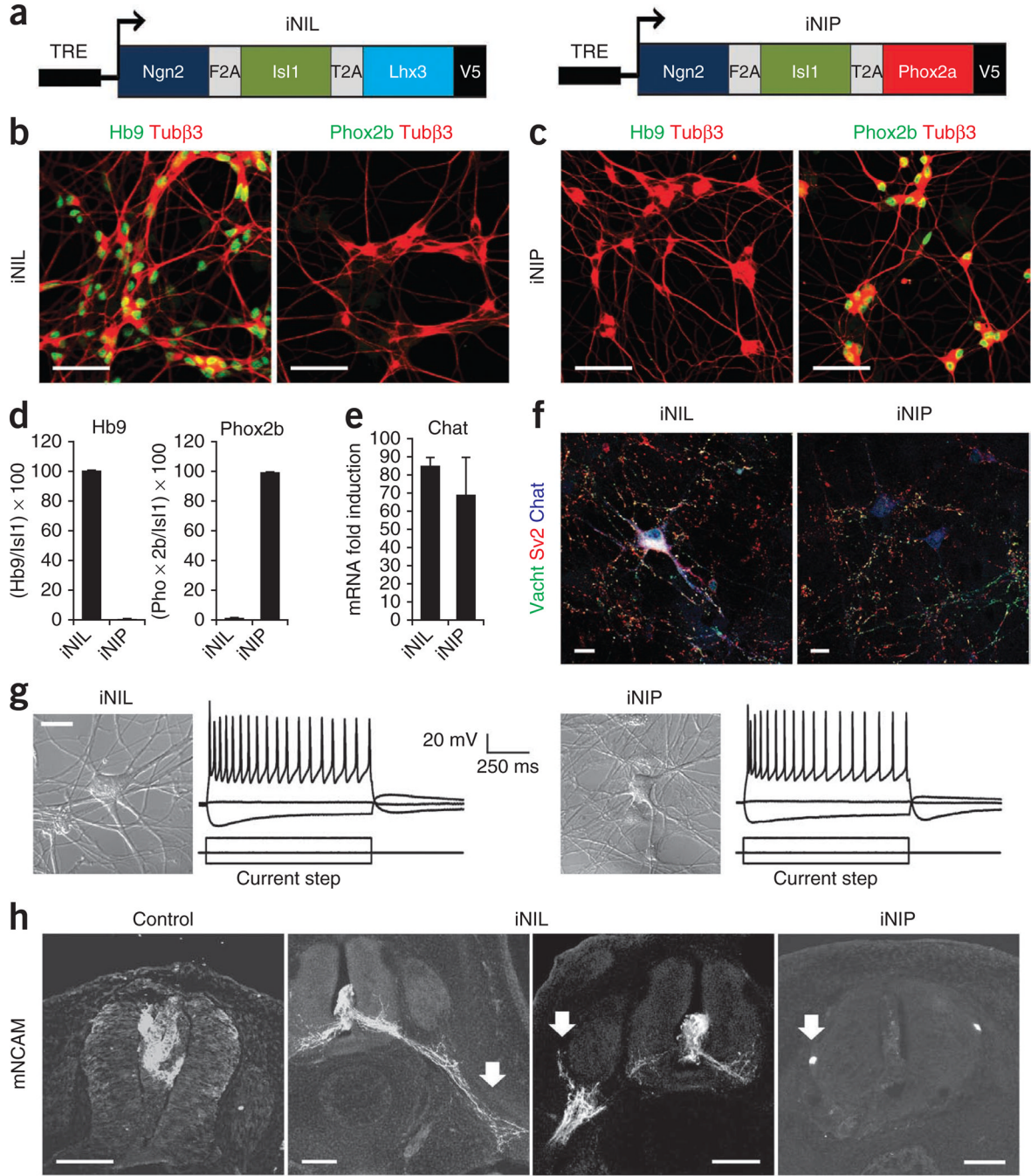
## References

1. Mann RS, Carroll SB. Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.* 2002; 12:592–600.
2. Tapscott SJ, et al. MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science.* 1988; 242:405–411.
3. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006; 126:663–676. [PubMed: 16904174]
4. Pfisterer U, et al. Direct conversion of human fibroblasts to dopaminergic neurons. *Proc. Natl. Acad. Sci. USA.* 2011; 108:10343–10348. [PubMed: 21646515]
5. Sekiya S, Suzuki A. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature.* 2011; 475:390–393.
6. Son EY, et al. Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell.* 2011; 9:205–218.
7. Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. *In vivo* reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature.* 2008; 455:627–632. [PubMed: 18754011]
8. Peter IS, Davidson EH. Evolution of gene regulatory networks controlling body plan development. *Cell.* 2011; 144:970–985.
9. Slattery M, et al. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell.* 2011; 147:1270–1282.
10. Jessell TM. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Rev. Genet.* 2000; 1:20–29.
11. Song MR, et al. T-Box transcription factor Tbx20 regulates a genetic program for cranial motor neuron cell body migration. *Development.* 2006; 133:4945–4955.
12. Lee SK, Pfaff SL. Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors. *Neuron.* 2003; 38:731–745.
13. Hester ME, et al. Rapid and efficient generation of functional motor neurons from human pluripotent stem cells using gene delivered transcription factor codes. *Mol. Ther.* 2011; 19:1905–1912. [PubMed: 21772256]
14. Lee S, et al. Fusion protein Isl1-Lhx3 specifies motor neuron fate by inducing motor neuron genes and concomitantly suppressing the interneuron programs. *Proc. Natl. Acad. Sci. USA.* 2012; 109:3383–3388.
15. Thaler JP, Lee SK, Jurata LW, Gill GN, Pfaff SL. LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell type-specific protein-protein interactions. *Cell.* 2002; 110:237–249.
16. Sharma K, et al. LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. *Cell.* 1998; 95:817–828.
17. Pfaff SL, Mendelsohn M, Stewart CL, Edlund T, Jessell TM. Requirement for LIM homeobox gene *Isl1* in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell.* 1996; 84:309–320.
18. Hirsch MR, Glover JC, Dufour HD, Brunet JF, Goridis C. Forced expression of Phox2 homeodomain transcription factors induces a branchio-visceromotor axonal phenotype. *Dev. Biol.* 2007; 303:687–702.
19. Pattyn A, Morin X, Cremer H, Goridis C, Brunet JF. Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis. *Development.* 1997; 124:4065–4075.

20. Coppola E, Pattyn A, Guthrie SC, Goridis C, Studer M. Reciprocal gene replacements reveal unique functions for *Phox2* genes during neural differentiation. *EMBO J.* 2005; 24:4392–4403. [PubMed: 16319924]
21. Mazzoni EO, et al. Embryonic stem cell-based mapping of developmental transcriptional programs. *Nat. Methods.* 2011; 8:1056–1058. [PubMed: 22081127]
22. Novitsch BG, Chen AI, Jessell TM. Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor *Olig2*. *Neuron.* 2001; 31:773–789. [PubMed: 11567616]
23. Mizuguchi R, et al. Combinatorial roles of *olig2* and *neurogenin2* in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. *Neuron.* 2001; 31:757–771. [PubMed: 11567615]
24. Hasan KB, Agarwala S, Ragsdale CW. *PHOX2A* regulation of oculomotor complex nucleogenesis. *Development.* 2010; 137:1205–1213.
25. Miles GB, et al. Functional properties of motoneurons derived from mouse embryonic stem cells. *J. Neurosci.* 2004; 24:7848–7858. [PubMed: 15356197]
26. Gao BX, Ziskind-Conhaim L. Development of ionic currents underlying changes in action potential waveforms in rat spinal motoneurons. *J. Neurophysiol.* 1998; 80:3047–3061. [PubMed: 9862905]
27. Wichterle H, Lieberam I, Porter JA, Jessell TM. Directed differentiation of embryonic stem cells into motor neurons. *Cell.* 2002; 110:385–397.
28. Wichterle H, Peljto M, Nedelec S. Xenotransplantation of embryonic stem cell-derived motor neurons into the developing chick spinal cord. *Methods Mol. Biol.* 2009; 482:171–183.
29. Dillon AK, et al. Molecular control of spinal accessory motor neuron/axon development in the mouse spinal cord. *J. Neurosci.* 2005; 25:10119–10130.
30. Holmes G, Niswander L. Expression of *slit-2* and *slit-3* during chick development. *Dev. Dyn.* 2001; 222:301–307. [PubMed: 11668607]
31. Brunet JF, Pattyn A. *Phox2* genes: from patterning to connectivity. *Curr. Opin. Genet. Dev.* 2002; 12:435–440. [PubMed: 12100889]
32. Grillet N, Dubreuil V, Dufour HD, Brunet JF. Dynamic expression of *RGS4* in the developing nervous system and regulation by the neural type-specific transcription factor *Phox2b*. *J. Neurosci.* 2003; 23:10613–10621. [PubMed: 14627646]
33. Moore RY. Cranial motor neurons contain either galanin or calcitonin gene-related peptide like immunoreactivity. *J. Comp. Neurol.* 1989; 282:512–522. [PubMed: 2470788]
34. Mahony S, et al. Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol.* 2011; 12:R2. [PubMed: 21232103]
35. Shen Y, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature.* 2012; 488:116–120. [PubMed: 22763441]
36. Berger MF, et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell.* 2008; 133:1266–1276. [PubMed: 18585359]
37. Vierbuchen T, et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature.* 2010; 463:1035–1041. [PubMed: 20107439]
38. Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell.* 2008; 132:1049–1061. [PubMed: 18358816]
39. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell.* 2005; 122:947–956. [PubMed: 16153702]
40. Kim J, et al. A *Myc* network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell.* 2010; 143:313–324. [PubMed: 20946988]
41. Dasen JS, Jessell TM. *Hox* networks and the origins of motor neuron diversity. *Curr. Top. Dev. Biol.* 2009; 88:169–200. [PubMed: 19651305]
42. Mazzoni EO, et al. Saltatory remodeling of *Hox* chromatin in response to rostro-caudal patterning signals. *Nat. Neurosci.* in the press.
43. Li XJ, et al. Specification of motoneurons from human embryonic stem cells. *Nat. Biotechnol.* 2005; 23:215–221. [PubMed: 15685164]

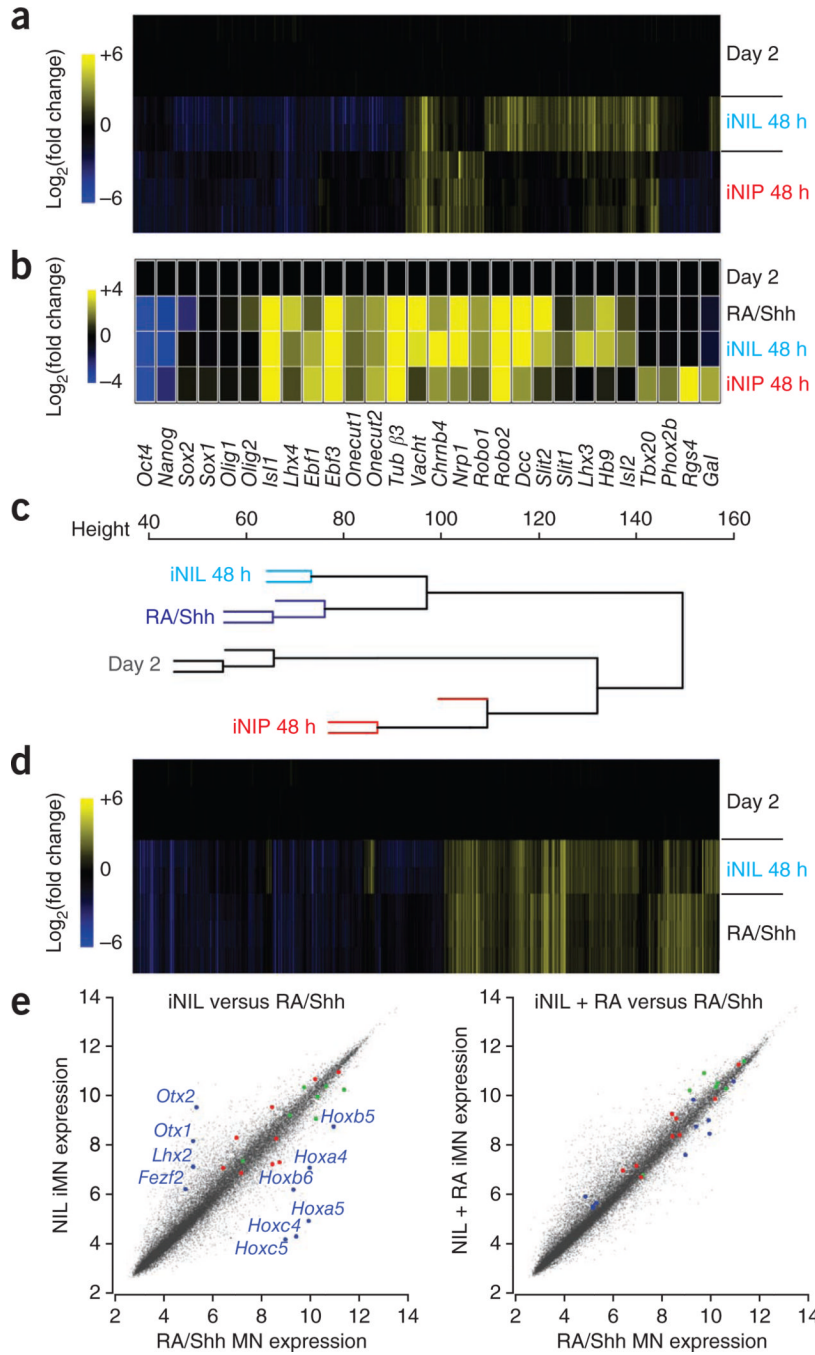
44. Perrier AL, et al. Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc. Natl. Acad. Sci. USA*. 2004; 101:12543–12548. [PubMed: 15310843]
45. Boulting GL, et al. A functionally characterized test set of human induced pluripotent stem cells. *Nat. Biotechnol.* 2011; 29:279–286. [PubMed: 21293464]
46. Sanges R, Cordero F, Calogero RA. oneChannelGUI: a graphical interface to Bioconductor tools, designed for life scientists who are not familiar with R language. *Bioinformatics*. 2007; 23:3406–3408. [PubMed: 17875544]
47. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009; 10:R25. [PubMed: 19261174]
48. Guo Y, et al. Discovering homotypic binding events at high spatial resolution. *Bioinformatics*. 2010; 26:3028–3034. [PubMed: 20966006]
49. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26:139–140. [PubMed: 19910308]
50. Albuquerque C, Joseph DJ, Choudhury P, MacDermott AB. Dissection, plating, and maintenance of cortical astrocyte cultures. *Cold Spring Harbor Protoc.* 2009; 8 pdb prot5273.





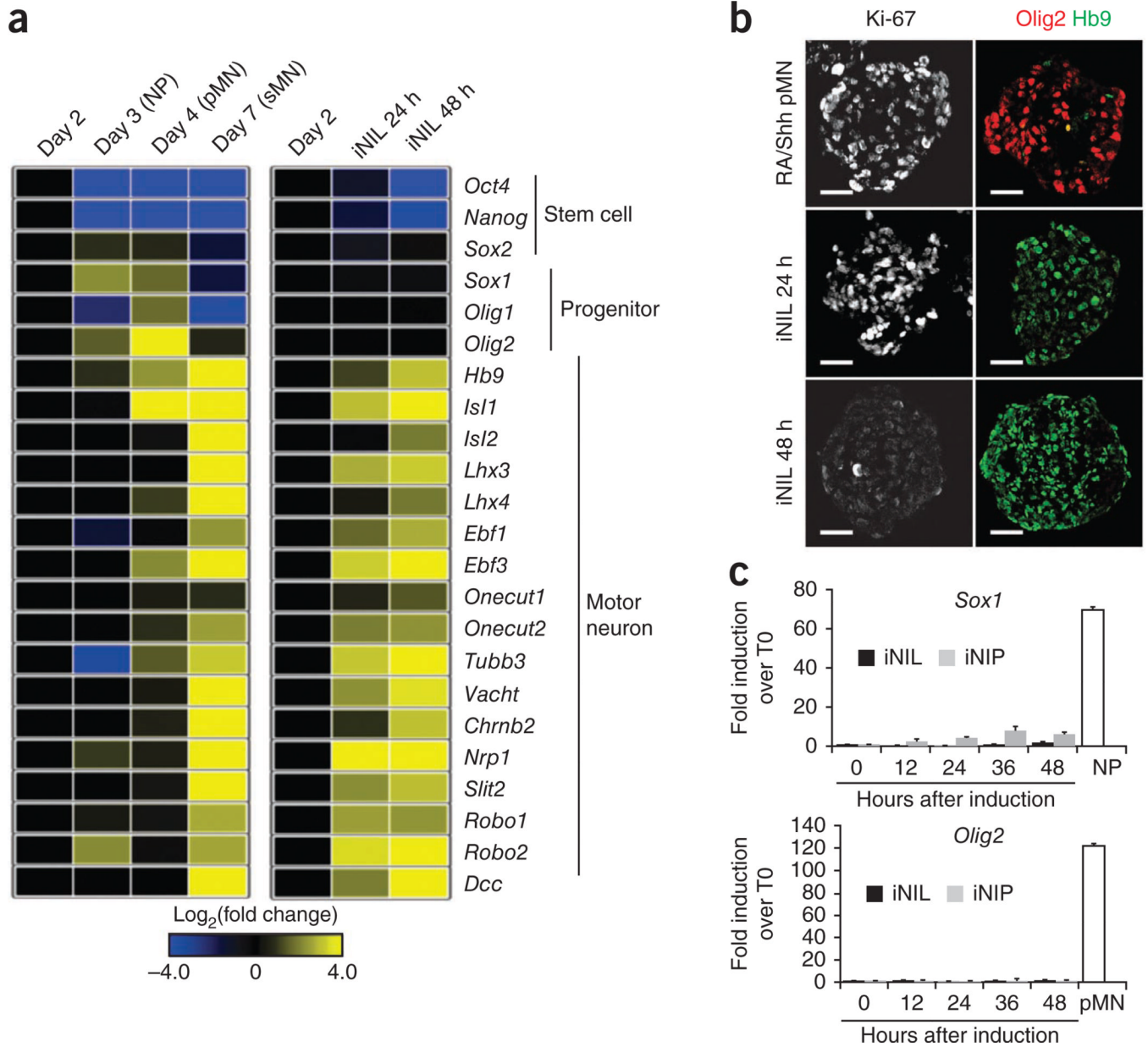
**Figure 1.** NIL and NIP transcription factors program spinal and cranial motor neurons, respectively. **(a)** Schematic representation of Dox-inducible NIL and NIP programming modules. TRE, tetracycline response element; F2A and T2A, 2A peptide sequences from foot-and-mouth disease virus. **(b)** NIL programmed spinal motor neuron in the absence of patterning signals. Induced iNIL cells exhibited neuronal morphology with multiple Tuj1-immunoreactive processes, expressed Hb9 and did not express Phox2b. Day 2 embryoid bodies treated with Dox for 48 h were dissociated, plated on laminin-coated substrate and analyzed 24 h later. Scale bars represent 50 μm. **(c)** NIP-programmed cells exhibited neuronal morphology, were

Tuj1 immunoreactive and did not express Hb9, but were positive for Phox2b. Day 2 embryoid bodies treated with Dox for 48 h were dissociated and plated on laminin-coated substrate and analyzed 24 h later. Scale bars represent 50  $\mu\text{m}$ . **(d)** Efficient induction of Hb9 and Phox2b by NIL and NIP, respectively. Shown is a quantification of the percentage of NIL or NIP induced cells ( $\text{Isl1}^+$ ) expressing Hb9 or Phox2b ( $n = 3$ ). Data are presented as mean  $\pm$  s.e.m. **(e)** NIL and NIP induced the cholinergic fate. Shown is fold induction of *Chat* mRNA in NIL and NIP cells treated with Dox for 48 h ( $n = 3$ ; normalized to control). Data are presented as mean  $\pm$  s.e.m. **(f)** NIL and NIP programmed cells contain cholinergic synaptic vesicles. Dissociated iNIL and iNIP cells induced with Dox were cultured on astrocyte monolayers for 7 d and stained with SV2, Vacht and Chat. Scale bars represent 10  $\mu\text{m}$ . **(g)** NIL and NIP programmed neurons fired repetitive action potentials. Dox-induced iNIL and iNIP cells cultured for 7 d on astrocyte monolayers were analyzed by electrophysiology. Representative (11 of 12 cells for each cell line) current-clamp recordings are shown. The current command step that induced repetitive action potentials is shown under the membrane traces. Scale bar represents 25  $\mu\text{m}$ . **(h)** NIL- and NIP-induced motor neurons projected axons toward different targets. Control and Dox-induced day 4 embryoid bodies were implanted into the Hamburger Hamilton stage 16 developing chick cervical spinal cord *in vivo*. Embryos were fixed 2 d later, sectioned and stained with a mouse-specific NCAM antibody. Dense bundles of axons emanating from NIL-induced transplants were observed in the ventral root and in axial (left arrow) and limb (right arrow) nerve branches (four of five successfully transplanted embryos). In contrast, NIP-induced cells projected axons dorso-laterally toward the spinal accessory nerve (arrows) (four of four successfully transplanted embryos). Axons of control transplants remained in the confines of the neural tube. Scale bars represent 100  $\mu\text{m}$ .



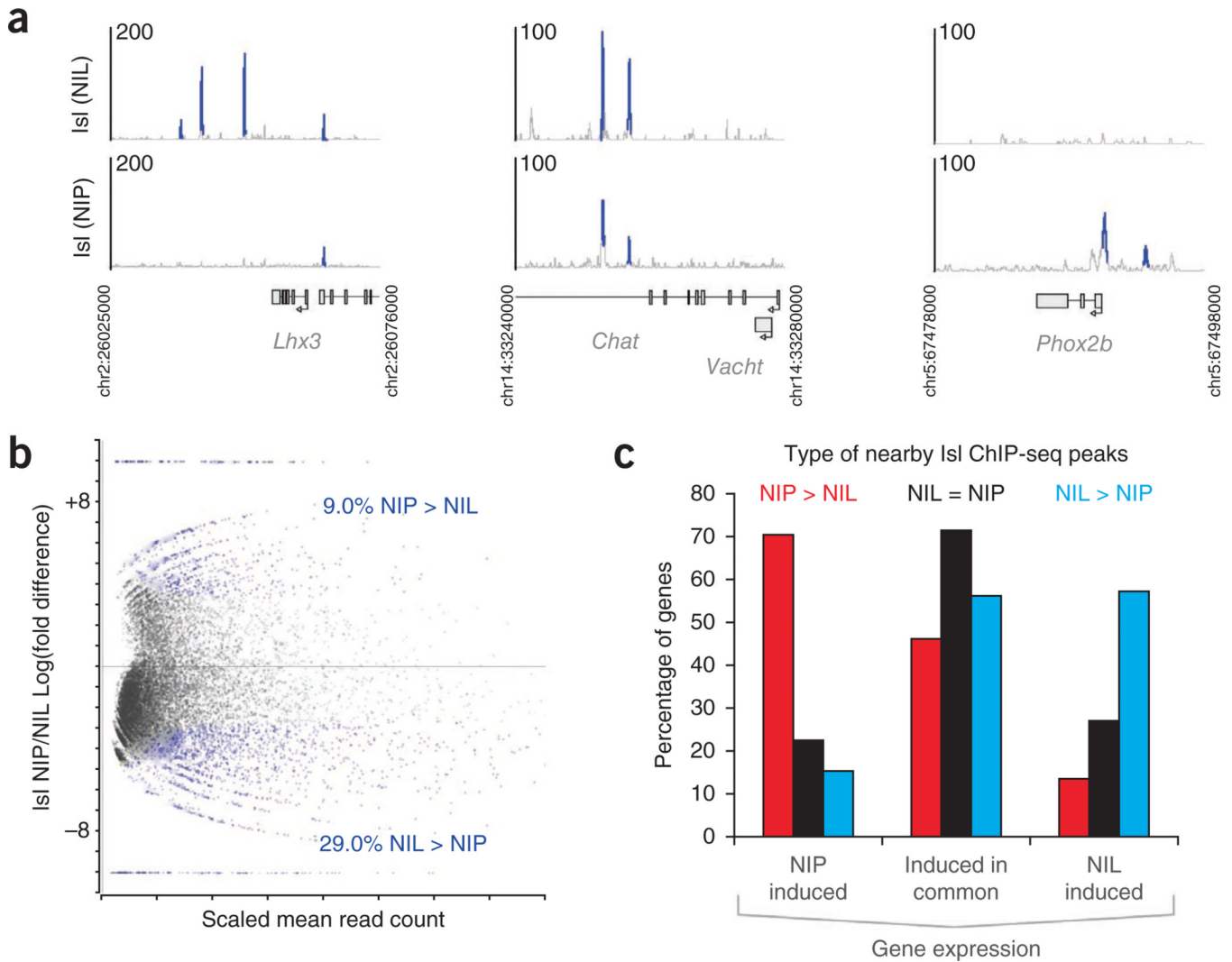
**Figure 2.** NIL and NIP induce different transcriptomes. (a) NIL and NIP factors induced different transcriptomes. Shown is a clustergram of all differentially expressed genes in day 2 embryoid bodies (before Dox induction, three replicates) and in iNIL (two replicates) and iNIP (three replicates) cells induced for 48 h with Dox. (b) The expression of a relevant subset of genes revealed the identity of NIL- and NIP-programmed cells. Shown is a heat map of average expression of genes associated with spinal and cranial motor neuron identity in day 2 embryoid bodies, retinoic acid- and Shh-derived (RA/Shh) spinal motor neurons (day 6 FACS-purified spinal motor neurons following 4 d of differentiation by retinoic acid

and Shh treatment), and NIP- and NIL-programmed neurons induced for 48 h with Dox. **(c)** Unsupervised clustering of individual expression profiles revealed that NIL-induced cells clustered with retinoic acid- and Shh-differentiated spinal motor neurons, whereas the NIP programming module induced a different cell type that clustered apart from spinal motor neurons. **(d)** NIL expression induced a spinal motor neuron-specific transcriptome. Shown is a clustergram of all differentially expressed genes in day 2 embryoid bodies, Dox-treated iNIL cells and retinoic acid- and Shh-differentiated motor neurons. **(e)** Retinoic acid imposed cervical identity onto NIL-programmed spinal motor neurons. Left, scatter plot of mRNA expression intensities in Dox-induced iNIL cells versus retinoic acid- and Shh-differentiated spinal motor neurons. Right, scatter plot of mRNA expression intensities in Dox-induced iNIL cells treated with 1  $\mu$ M retinoic acid for 48 h versus retinoic acid- and Shh-differentiated spinal motor neurons. Rostro-caudal patterning genes are shown in blue, spinal motor neuron associated transcription factors in red, and spinal motor neuron associated receptors and enzymes in green. iMN, induced motor neuron.

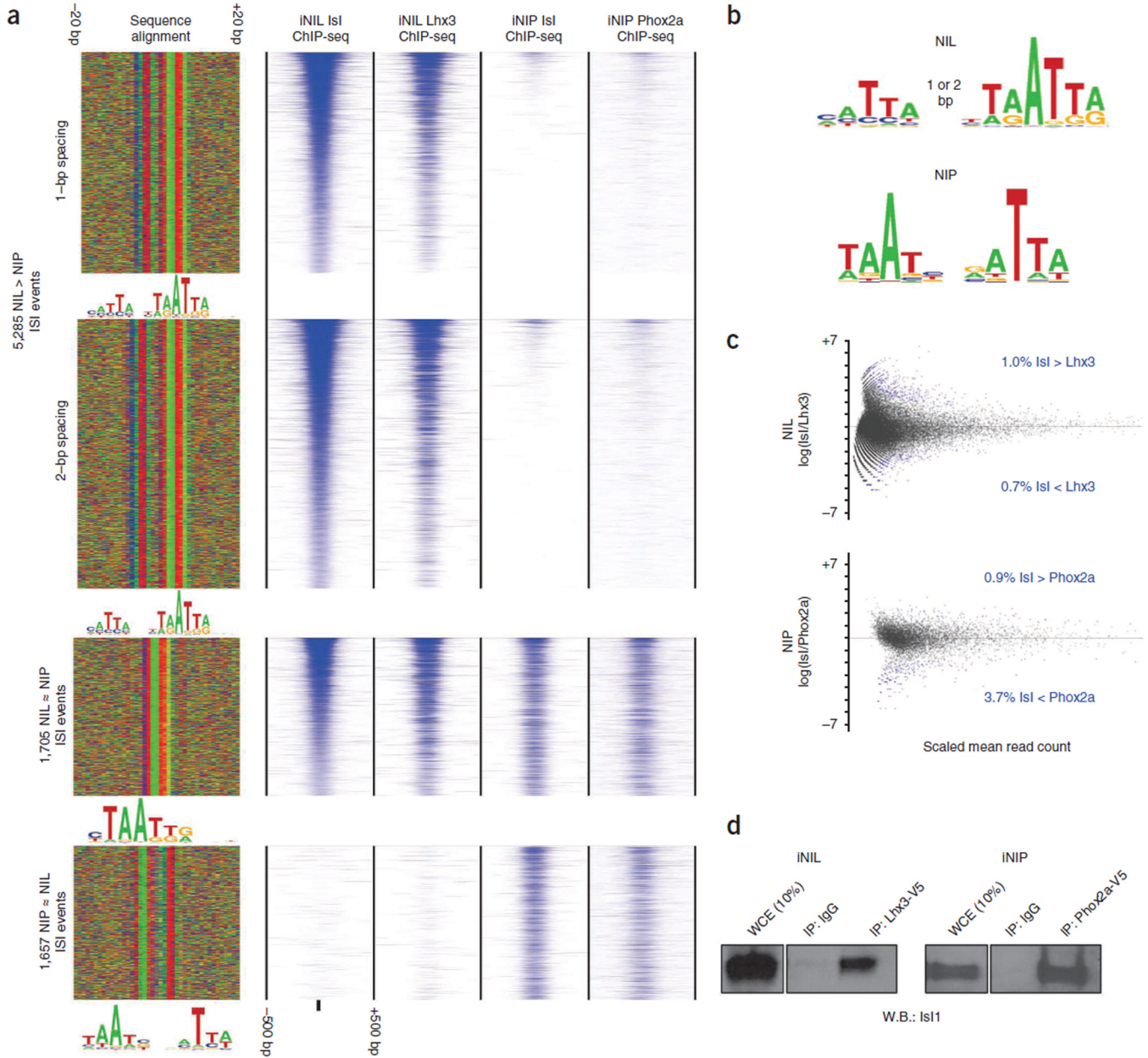


**Figure 3.** Induced iNIL cells bypass progenitor stages. **(a)** NIL-programmed spinal motor neurons directly induced terminal genes. Shown are expression profiles of genes associated with pluripotency (stem cell), neural and motor neuron progenitors (progenitor, NP and pMN), and postmitotic spinal motor neuron (motor neuron, sMN) identities. Clustergram of gene expression changed in a time series of embryoid bodies treated with retinoic acid and Shh from days 2 to 7 (left) and iNIL embryoid bodies treated with Dox from days 2 to 4 (right). **(b)** NIL-programmed spinal motor neurons did not express Olig2. Retinoic acid- and Shh-induced motor neuron progenitors and NIL-programmed spinal motor neurons (24 and 48 h of Dox treatment) were stained for Olig2 and Hb9. By 48 h after NIL induction, cells had exited the cell cycle, as revealed by the lack of Ki-67 staining. Scale bars represent 20  $\mu$ m (top four panels) and 25  $\mu$ m (bottom two panels). **(c)** Neither NIL nor NIP transcription factors induced expression of progenitor markers Sox1 or Olig2. Shown is a time series of

*Sox1* and *Olig2* mRNA levels analyzed by quantitative PCR after the Dox treatment of iNIL and iNIP cells. Neural progenitors (NP; 24 h after retinoic acid and Shh treatment of day 2 embryoid bodies) and motor neuron progenitors (pMN; 48 h after retinoic acid and Shh treatment of day 2 embryoid bodies) served as positive controls for *Sox1* and *Olig2*, respectively. Data are presented as mean  $\pm$  s.e.m. ( $n = 3$ ). T0, before induction.

**Figure 4.**

Context-specific Isl1 genome association in iNIL and iNIP cells correlates with differential gene expression. **(a)** Isl binding at developmentally regulated genes in iNIL and iNIP cells treated with Dox for 48 h. Isl ChIP-seq signals over *Lhx3*, *Chat* and *Phox2b* are shown. Blue peaks represent significant ( $P < 0.01$ ) read enrichment over control. Genomic loci coordinates are shown next to the x axis. **(b)** Isl genome association was NIL and NIP specific. Shown is a comparison of Isl read enrichment from iNIL and iNIP cells at all detected peaks. Blue represents peaks that were significantly differentially enriched in one experiment over the other ( $\log_2$ ; Online Methods). **(c)** Condition-specific Isl binding was associated with condition-specific gene expression. Differentially expressed genes were divided into ones selectively induced in iNIL and iNIP cells and ones induced in common in both cell types. The bar graph represents the percentage of genes containing a proximal Isl peak that was condition specific (Fig. 2b) for each group.



**Figure 5.** Sequence motifs occupied by Isl1 in iNIL and iNIP cells. **(a)** Condition-specific Isl enrichment is explained by condition-specific DNA sequence motifs. All Isl binding sites determined to be significantly differentially enriched in either iNIL or iNIP cells were plotted, alongside a set of sites that were similarly enriched in both conditions. The sites are centered on the closest match to the condition-specific or condition-independent motifs within 25 bp of the predicted binding position. Isl ChIP-seq peaks that were similarly enriched or significantly enriched in either iNIL or iNIP cells were ordered on the basis of Isl read enrichment levels. Each column in the figure plots either a representation of the sequence alignment in a 40-bp window around the motif match or ChIP-seq read enrichment in a 1-kbp window around the predicted binding site. **(b)** Primary DNA motifs over-represented under enriched peaks obtained from Isl ChIP-seq experiments in iNIL and iNIP cells treated for 48 h with Dox. **(c)** Lhx3 and Phox2a colocalized with Isl genomic binding



sites in iNIL and iNIP cells, respectively. Shown is a comparison of read enrichment from Isl1 with either Lhx3 or Phox2a at all detected peaks. Blue represents peaks that were significantly differentially enriched in one experiment over the other. **(d)** Isl1 associated with Lhx3 and Phox2a. V5 epitope-tagged Lhx3 or Phox2a transcription factor containing protein complexes were immunoprecipitated with antibody to V5 from Dox-treated iNIL or iNIP cells. The presence of Isl1 was examined by western blot analysis. WCE, whole cell extract; IP: IgG, negative control to test nonspecific binding; IP, V5 immunoprecipitation of Lhx3 from iNIL cells and Phox2a from iNIP cells. Shown is a representative blot ( $n = 3$ ).