nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-019-0399-y

Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes

Begüm Aydin[1,2], Akshay Kakumanu[3], Mary Rossillo[4], Mireia Moreno-Estellés[5], Görkem Garipler[1,2], Niels Ringstad[4], Nuria Flames[5], Shaun Mahony[3*] and Esteban O. Mazzoni[1,2*]

Developmental programs that generate the astonishing neuronal diversity of the nervous system are not completely understood and thus present a major challenge for clinical applications of guided cell differentiation strategies. Using direct neuronal programming of embryonic stem cells, we found that two main vertebrate proneural factors, Ascl1 and neurogenin 2 (Neurog2), induce different neuronal fates by binding to largely different sets of genomic sites. Their divergent binding patterns are not determined by the previous chromatin state, but are distinguished by enrichment of specific E-box sequences that reflect the binding preferences of the DNA-binding domains. The divergent Ascl1 and Neurog2 binding patterns result in distinct chromatin accessibility and enhancer activity profiles that differentially shape the binding of downstream transcription factors during neuronal differentiation. This study provides a mechanistic understanding of how transcription factors constrain terminal cell fates, and it delineates the importance of choosing the right proneural factor in neuronal reprogramming strategies.

Nervous systems are composed of a diverse array of neuronal cell types that form functional circuits. This cellular complexity is generated by the combinatorial activity of transcription factors (TFs). Decades of developmental biology studies have identified a handful of basic helix-loop-helix (bHLH) TFs, called proneural factors, that are necessary and sufficient to initiate neurogenesis[1]. In addition to conferring neuronal fate, proneural factors contribute to the specification of neuronal subtype identity[2]. The molecular mechanisms by which different proneural factors control and coordinate neurogenesis and neuronal-subtype specification have begun to be elucidated[2]. However, remaining gaps in our knowledge make it difficult to generate the vast array of clinically relevant neurons for research and clinical applications.

Ascl1 (also known as Mash1) and neurogenin 2 (Neurog2), which are the mammalian homologs of *Drosophila achaete-scute complex* and *atonal* genes, respectively, are the two main proneural factors that initiate and regulate neurogenesis in vertebrate nervous systems[1–4]. Apart from a few regions in the nervous system where they are co-expressed, these two proneural factors are expressed in a complementary manner and are not interchangeable for neuronal-subtype specification[5–7]. Proneural factors promote neurogenesis and induce distinct subtype identities, and these functions are conserved across phyla. In *Drosophila* ectoderm, *atonal* controls chordotonal organ identity, while *achaete-scute* genes control external sensory organs[8]. In mice, Ascl1 and Neurog2 are respectively required to specify GABAergic (Ascl1) and glutamatergic (Neurog2) neurons in the forebrain and sympathetic (Ascl1) and sensory neurons (Neurog2) of the peripheral nervous system[5,9–14]. Thus, the functional divergence of Ascl1 and Neurog2 is an ancestral trait responsible for the generation of neuronal diversity required in the nervous system that predates the split of vertebrates and invertebrates[15].

The transcriptional programs that establish terminal neuronal identity consist of generic (pan-) neuronal features that are shared by all neurons and subtype-specific features that are shared by specific classes of neurons[1,16,17]. These features are considered to be controlled by the activities of neurogenesis-inducing TFs (including proneural TFs) and TF combinations specific to a particular neuronal subtype[18–20]. While Ascl1 and Neurog2 can induce neurogenesis in neural-lineage and pluripotent cells[21–23], reprogramming of differentiated cells usually couple Ascl1 and/or Neurog2 with additional TFs to promote subtype identity and/or to downregulate the resident transcriptional program[24–26]. However, this model contrasts with the observation that Ascl- and Neurog-proneural families are the dominant force in controlling neuronal subtype identities when expressed in fibroblasts in combination with other TFs[17]. Thus, to better understand the rules that govern neuronal subtype reprogramming, we must understand the differences in Ascl1- and Neurog2-induced neurogenesis.

Direct programming is an advantageous platform to study how proneural TFs, alone or in combination with other TFs, control neuronal gene regulatory networks. An analysis of astrocyte-to-neuronal conversion by Ascl1 and Neurog2 has shown that they initially activate largely non-overlapping genes[27]. Additionally, Ascl1 and Neurog2 can act as 'pioneer factors' in fibroblasts by binding to previously inaccessible regulatory regions and increasing chromatin accessibility following binding[25,28,29]. However, it is unclear whether Ascl1 and Neurog2 would have a similar non-overlapping differentiation trajectory when expressed in pluripotent stem cells compared to differentiated cells and whether their proposed pioneering activity would differentially affect the acquisition of generic and subtype-specific neuronal features. To address these questions, the intrinsic differences between Ascl1 and Neurog2 and their effect

[1]Department of Biology, New York University, New York, NY, USA. [2]Neuroscience Institute, Department of Neuroscience and Physiology, NYU School of Medicine, New York, NY, USA. [3]Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA. [4]Skirball Institute of Biomolecular Medicine, NYU School of Medicine, New York, NY, USA. [5]Developmental Neurobiology Unit, Instituto de Biomedicina de Valencia IBV-CSIC, Valencia, Spain. *e-mail: mahony@psu.edu; eom204@nyu.edu

on the downstream neurogenesis must be studied in a controlled environment that allows for a direct and robust comparison of the induced transcriptional and chromatin dynamics.

Here, we investigated the mechanism by which the two bHLH proneural factors Ascl1 and Neurog2 engage with chromatin and affect the activities of TFs expressed downstream of Ascl1 and Neurog2 during neuronal differentiation. We found that Ascl1 and Neurog2 generate neurons by binding to largely different sets of genomic sites when expressed in similar chromatin and cellular contexts. Their divergent binding is due to distinct DNA sequence specificities of the respective bHLH domains toward preferred E-boxes. The initial divergent binding of Ascl1 and Neurog2 results in distinct regulatory landscapes that influence the binding pattern and the regulatory activity of shared downstream TFs in establishing shared (generic) and neuron-specific (subtype-specific) expression profiles. Thus, we speculate that the intrinsic differences in Ascl1- and Neurog2-induced neurogenesis increase the number of possible neuronal types generated during development by differentially altering the chromatin landscapes upon which the widely expressed downstream TFs operate.

Results

Ascl1 and Neurog2 program neuronal fate with distinct neuronal subtype bias. To investigate the intrinsic activities of Ascl1 and Neurog2, we generated two isogenic mouse embryonic stem cell (mESC) lines that express either Ascl1 (iAscl1 or iA) or Neurog2 (iNeurog2 or iN) following doxycycline treatment but are otherwise identical (Fig. 1a). Induction of Ascl1 and Neurog2 resulted in neuronal differentiation with detectable upregulation of the neuron-specific ßIII-tubulin (Tubb3) within 12 h after induction (Supplementary Fig. 1a). iA and iN neurons adopted typical neuronal morphologies, with projections compatible with axonal and dendritic identity expressing neurofilament (NF) and MAP2 proteins, respectively (Fig. 1b). Both iA and iN neurons responded to KCl-induced depolarization by changing their intracellular Ca^{2+} concentration albeit with different dynamics, with iN neurons exhibiting slower decay (Fig. 1c). In line with previous studies[21–23,30], forced expression of the proneural TFs Ascl1 or Neurog2 triggered a rapid conversion of differentiating mESCs into neurons. Therefore, isogenic iA and iN lines constitute an ideal platform with which to comparatively study the molecular mechanisms of Ascl1- versus Neurog2-induced neurogenesis.

Ascl1 and Neurog2 overexpression transdifferentiates astrocytes into neurons by inducing an early divergent transcriptional profile[27]. To investigate whether Ascl1 and Neurog2 induce neuronal differentiation with similar dynamics during mESC differentiation, we profiled mRNA levels at 12 and 48 h after induction. Around 50% of Ascl1 upregulated genes and 37% of Neurog2 upregulated genes were shared at 12 h (390 genes) (Supplementary Fig. 1b). The percentages of commonly upregulated genes increased to 74 and 80% at 48 h for iA and iN neurons, respectively (2,560 genes). Shared upregulated genes were enriched in gene ontology terms associated with generic neuronal features (Supplementary Fig. 1c). Moreover, by 48 h Ascl1 and Neurog2 had already activated the expression of genes associated with different neuronal subtypes consistent with their requirement during embryonic development such as noradrenergic (*Phox2b* and *Dbh* in iA cells) and sensory neuron markers (*Ret* and *Ntrk1* in iN cells) (Fig. 1d and Supplementary Fig. 1d).

To investigate whether the gene expression differences stem from a subset of neurons in the dish, or the majority of iA and iN neurons differ, we performed a single-cell RNA sequencing (scRNA-seq) experiment at 48 h after induction. The vast majority of cells upregulated the generic neuronal markers *Tubb3* and *Map2* (Fig. 1f). Confirming the hypothesis that Ascl1 and Neurog2 induce neurogenesis through divergent differentiation paths, iA and iN neurons clustered into two distinct groups on the basis of their transgene expression (Fig. 1e). The neuronal subtype markers were not homogeneously distributed across either population or largely co-expressed in the same cells (Fig. 1g; Supplementary Fig. 1d). For example, noradrenergic (*Tfap2b* and *Phox2b*) and cortical interneuron (*Tlx3* and *Arx*) markers were primarily expressed by iA neurons, whereas spinal motor (*Vacht* and *Olig2*) and sensory neuron (*Ret* and *Ntrk1*) markers were expressed by iN neurons (Fig. 1g). Thus, while these results are not indicative of complete neuronal-subtype specification, Ascl1 and Neurog2 expression initiates different neuronal differentiation programs even when expressed under similar chromatin and transcriptional states.

Ascl1 and Neurog2 bind to largely distinct sets of sites in the genome. To understand how Ascl1 and Neurog2 induce neuronal differentiation, we captured their initial binding at 12 h after induction, which is the earliest time point when the doxycycline system induces robust expression of these TFs in most cells (Supplementary Fig. 1a). We identified 20,452 and 28,206 binding sites for Ascl1 and Neurog2, respectively. While analysis of the whole dataset produces similar percentages (Supplementary Fig. 2a), we focused on the top 10,000 binding sites in each dataset for our downstream analysis to eliminate complications that may arise from comparing chromatin immunoprecipitation (ChIP) assays with sequencing (ChIP-seq) signals with different strengths. The initial binding of Ascl1 and Neurog2 was largely non-overlapping, with 90% of all sites confidently called as differentially bound; that is, Ascl1 and Neurog2 each preferentially bound 45% of the sites (Fig. 2a). Only 10% of the sites were bound with similar strength by both TFs. We designated the Ascl1 and Neurog2 differentially bound sites as Ascl1-preferred sites (A > N sites) and Neurog2-preferred sites (N > A sites), respectively. Sites that were bound by both TFs were designated as shared sites (A = N sites). Ascl1 pioneer activity is not enough to allow for its invariable binding across cell types because Ascl1 binding in mESCs does not recapitulate its genomic distribution when expressed in fibroblasts[25] (Supplementary Fig. 2d). Our data recovered some of the few sites previously described as bound by Ascl1 in mESCs, but this comparison was compromised by the radically different ChIP signal strengths[31] (Supplementary Fig. 2e). Thus, in line with *Dll1* activation by distinct Ascl1 and Neurog2 enhancers (Supplementary Fig. 2f), genome-wide comparison of the two proneural bHLH TFs Ascl1 and Neurog2 showed remarkably different binding profiles under similar chromatin and cellular contexts.

Distinct E-box sequences are enriched at Ascl1- and Neurog2-preferred sites. The extensive lack of overlap between Ascl1 and Neurog2 binding prompted us to investigate the possible mechanisms driving their divergent binding patterns. Chromatin accessibility and DNA sequence are the two main factors that dictate in vivo TF binding to regulatory elements[32]. Ascl1 acts as a pioneer factor; however, pioneering activity for Neurog2 was only proposed indirectly when in combination with small molecules that enhance chromatin accessibility[25,28]. When we compared 12 h Ascl1 and Neurog2 binding to the previous chromatin accessibility state by assay for transposase-accessible chromatin using sequencing (ATAC-seq), we observed that both TFs engage with previously accessible and inaccessible sites in roughly the same proportion, with 57% and 43% of A > N sites that were previously accessible and inaccessible, respectively (Fig. 2b). Similarly, 53% and 47% of N > A sites were previously accessible and inaccessible, respectively (Fig. 2b). Therefore, the divergent Ascl1 and Neurog2 binding patterns are not due to major differences in their intrinsic abilities to bind inaccessible chromatin.

It has been observed that the *Drosophila* orthologs Scute and Atonal have different E-box targets, and that Ascl1 and Neurog2 regulate *Dll1* expression by binding to distinct E-box sequences[33,34]. Thus, we investigated whether DNA sequence features could

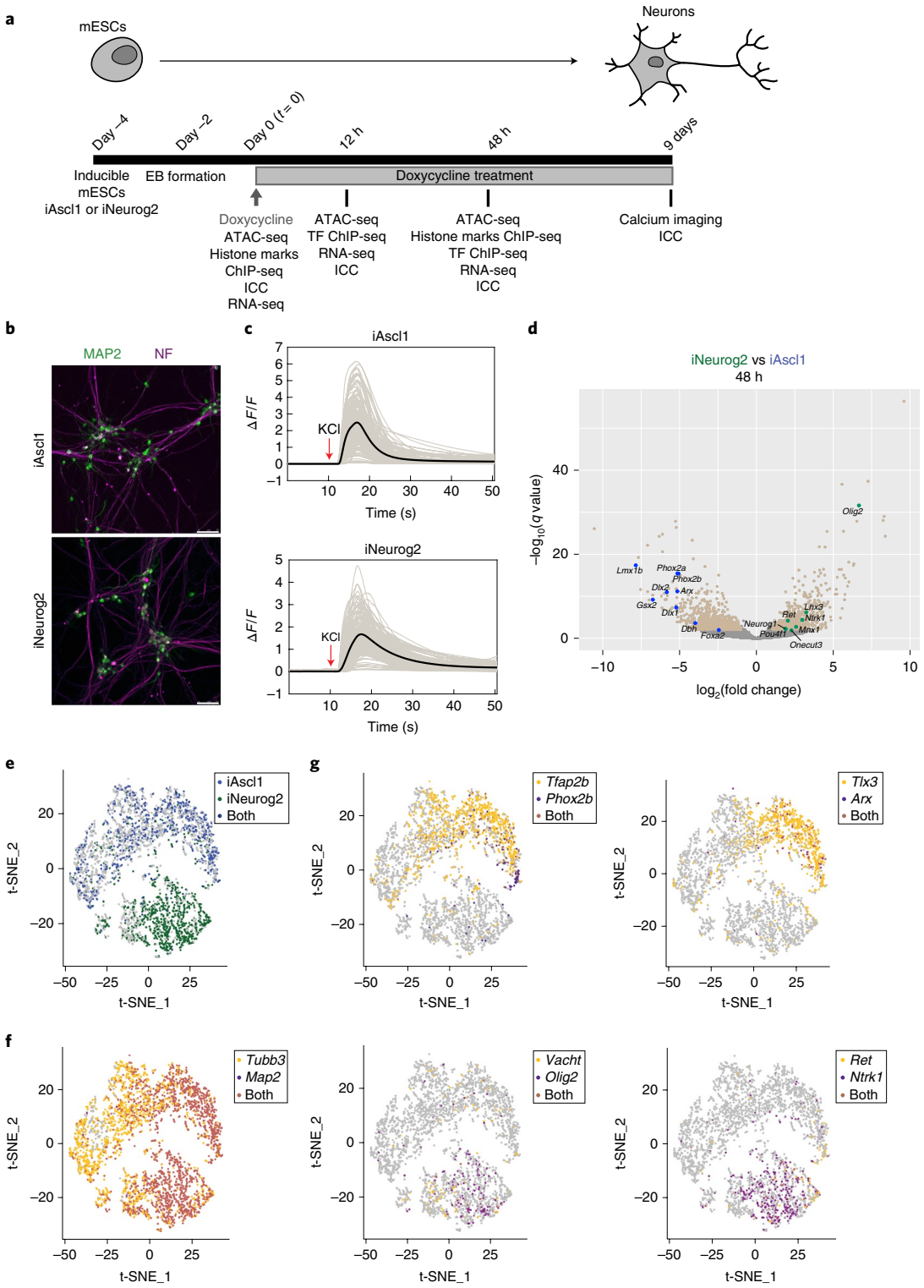


Fig. 1 | Ascl1 and Neurog2 induction in differentiating mESCs generates neurons with distinct neuronal-subtype bias. a, Experimental scheme. ICC, immunocytochemistry. **b**, iA and iN neurons express mature neuronal markers MAP2 and NF 9 d after Ascl1 and Neurog2 induction, respectively. Similar results were obtained from at least $n = 2$ independent cell differentiations. Scale bar, 50 μm. **c**, iA and iN neurons change their Ca^{2+} levels following KCl depolarization (9 d after induction). The thick line shows the average of individual recordings (from $n = 2$ independent cell differentiations). **d**, Volcano plot comparing mRNA levels between iA and iN neurons by RNA-seq at 48 h after induction (iA 48 h, $n = 5$; iN 48 h, $n = 2$). Beige circles represent the differentially expressed genes between iA or iN ($q < 0.01$, Wald test). Green and blue circles represent examples of differentially expressed genes in iN and iA, respectively. **e**, t-SNE plot showing the single-cell clustering of the iA or iN neurons. Circles are colored by the expression of transgenes. iA cells, blue (upper cluster); iN cells, green (lower cluster) ($n = 1$ cell differentiation). **f**, t-SNE plot showing the cells that express generic neuronal markers encoded by *Tubb3* and *Map2*. Note the maturation axis toward the left of the clusters ($n = 1$ cell differentiation). **g**, t-SNE plots showing iA and iN clusters expressing distinct neuronal subtype markers. The circles are colored by expression of Ascl1-specific genes *Tfap2b* and *Phox2b* (noradrenergic) and *Tlx3* and *Arx* (interneuron) (top), or Neurog2-specific genes *Vacht* and *Olig2* (motor neuron) and *Ret* and *Ntrk1* (sensory neuron) (bottom) ($n = 1$ cell differentiation).

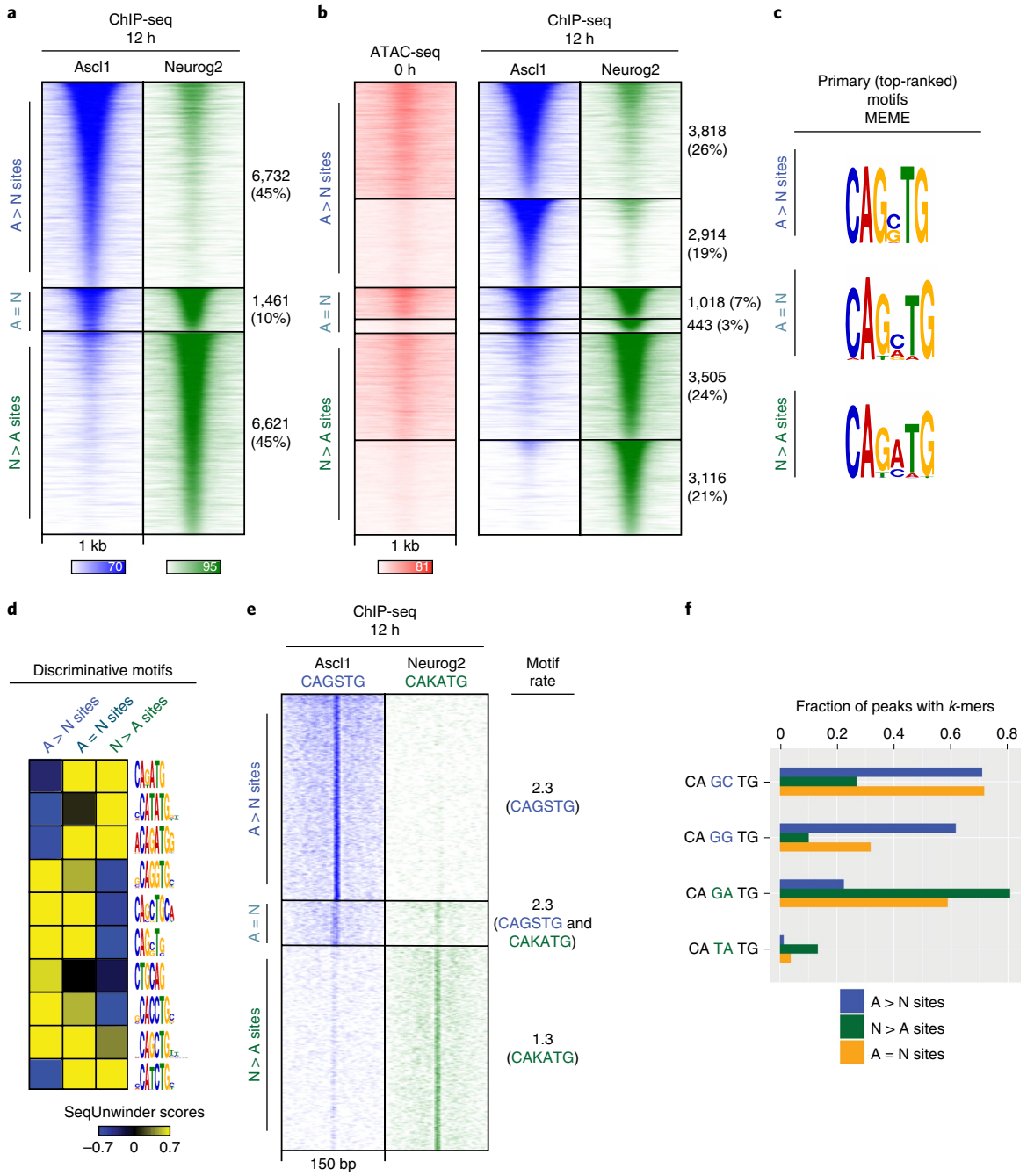


Fig. 2 | Genome-wide characterization of Ascl1 and Neurog2 binding and its determinants. a, ChIP–seq heatmap showing Ascl1 and Neurog2 binding at 12 h after induction. Ascl1- and Neurog2-preferred sites are designated as A > N and N > A, respectively. Sites that are bound by both Ascl1 and Neurog2 are designated as shared sites (A = N). Top 10,000 sites are plotted on the heatmap within a 1-kb window around the peak center ($n = 3$). **b**, Ascl1 and Neurog2 binding does not depend on prior chromatin accessibility. Ascl1 and Neurog2 ChIP–seq heatmap partitioned into previously accessible and inaccessible sites per 0 h ATAC–seq signal at bound sites. **c**, Primary (top-ranked) motifs enriched at the differentially bound A > N, N > A and shared (A = N) sites differ in central nucleotides. **d**, The discriminative motifs enriched at the A > N, A = N and N > A sites corroborate the relative enrichment of the distinct E-box variants. **e**, CAGSTG and CAKATG *k*-mer occurrences plotted at Ascl1 and Neurog2 binding sites within a 150-bp window (S represents G/C; K represents G/T). The rates of motif per *k*-mer occurrence at the binding sites are shown on the right. **f**, Ascl1 and Neurog2 differentially bound sites are distinguished by specific E-box instances. Fraction of differentially bound and shared sites containing various *k*-mer sequences within a 150-bp window around the peak center.

explain the differences in the binding of Ascl1 and Neurog2. The primary (top-ranked) motifs discovered by MEME in each class of Ascl1- and Neurog2-bound sites were variations of canonical E-boxes, differing primarily in the central two nucleotides (Fig. 2c). The primary motif discovered at A > N sites contained the consensus sequence CAGSTG (with S representing G/C nucleotides), encompassing the canonical E-box motif CAGCTG, which had been associated with Ascl1 binding in fibroblasts and neural stem cells[25,30]. Conversely, the primary motif at N > A sites contained the consensus CAKMTG (with K representing G/T nucleotides, and M representing A/C nucleotides). The peaks bound by both TFs (A = N) contained a motif that appeared to be the average between

the motifs found in the other two classes (Fig. 2c). To further identify discriminative motifs between Ascl1 and Neurog2 binding classes, we deployed SeqUnwinder, a tool designed to search for discriminative motifs across ChIP-seq samples. SeqUnwinder identified variations of the canonical E-box motif (CANNTG) that discriminated between A > N, N > A, and A = N shared sites (Fig. 2d). CAGSTG and CAKATG motifs were visibly enriched at the A > N and N > A sites when plotted in a 150-bp window around peaks (Fig. 2e). The CAGSTG motif occurred more than once at the A > N sites, while the CAKATG motif occurred on average once (Fig. 2e). Specifically, CAGCTG and CAGGTG 6-mers were present at 70% and 62% of the A > N sites, with some sites having both 6-mers as opposed to only 27% and 10% of N > A sites (Fig. 2f). Conversely, 81% of the N > A sites contained the CAGATG 6-mer sequence, while this 6-mer was present at only 22% of the A > N sites. Of note, only 13% of the N > A sites contained the CATATG motif described for in vitro Neurog2 binding[35]. Finally, roughly half of the A = N sites contained both Ascl1- and Neurog2-preferred 6-mers, suggesting that Ascl1 and Neurog2 bind to different E-boxes even within shared enhancers (Fig. 2f). Sequences flanking E-boxes have been shown to confer additional specificity to bHLH TFs by affecting the DNA shape[36,37]. Indeed, there were differences in nucleotide preferences flanking the non-discriminative core E-box (CAGNTG), and A > N sites were associated with larger predicted propeller twist and larger predicted minor groove width at alternate sides of the core E-box motif (Supplementary Fig. 3a,b). Thus, Ascl1 and Neurog2 have strong DNA sequence preferences that drive their genomic binding in differentiating mESCs.

bHLH domain controls DNA sequence specificity and neuronal-subtype identity. The basic domain of proneural TFs binds to the major groove of DNA, while the helix-loop-helix (HLH) domain mediates heterodimerization with other HLH proteins[38,39]. To test whether the bHLH (DNA-binding and dimerization) domain is sufficient to induce the divergent Ascl1 and Neurog2 binding patterns, we generated an inducible mESC line expressing a chimeric Ascl1–Neurog2 TF (A[N]^{bHLH} chimera) by swapping the bHLH domain of Ascl1 with that of Neurog2 (Fig. 3a). Similar to Ascl1 and Neurog2, the A[N]^{bHLH} chimera generated neurons that responded to KCl-induced depolarization and expressed mature neuronal cytoskeleton markers (Fig. 3b,c).

A[N]^{bHLH} chimera binding had significantly different ChIP-seq enrichment compared with Ascl1 at 70% of the sites (A[N]^{bHLH} = N sites, and A[N]^{bHLH} > A and N sites) (Fig. 3d). However, only 18% of A[N]^{bHLH} chimera binding sites were significantly different from those of Neurog2 (A[N]^{bHLH} > A and N, and A[N]^{bHLH} = A sites). As expected from its binding pattern, the *k*-mer (6-mers and 8-mers) signatures at the A[N]^{bHLH} chimera binding sites were also similar to that of Neurog2 sites (Supplementary Fig. 4a). For example, the ChIP-seq signal and the *k*-mer signature of the chimera at the shared (*Dll1*) and neuron-specific genes, such as *NeuroD2* (target of Neurog2) and *Dlx2* (target of Ascl1), also resembled that of Neurog2 (Fig. 3e). Thus, the analysis of the A[N]^{bHLH} chimeric TF demonstrates that the differences in Ascl1 and Neurog2 binding patterns are intrinsic and determined by the amino acid sequence of the bHLH domain.

Although the A[N]^{bHLH} chimera bound to Neurog2-preferred sites driven by its Neurog2 bHLH domain, the rest of its amino acid sequence is identical to Ascl1 (Fig. 3a). Specific residues outside the bHLH domain of Ascl1 and Neurog2 were shown to behave as rheostat-like modulators following phosphorylation and dephosphorylation for the context-dependent activity of their proneural functions[2,40–44]. However, the A[N]^{bHLH} chimera induced a gene expression profile similar to that induced by Neurog2 (Fig. 3f,g; Supplementary Fig. 4b,c). A principal component analysis (PCA) of gene expression (RNA-seq) of the A[N]^{bHLH}-, Ascl1-, and Neurog2-induced neurons (iA[N]^{bHLH}, iA, and iN neurons, respectively) revealed that the A[N]^{bHLH} differentiation trajectory is similar to that induced by Neurog2 (Fig. 3f). The first two PCA dimensions of individual replicates explained 83% of the variance, with PC1 reflecting the differentiation time and PC2 reflecting the differences in iA and iN neurons. These results demonstrate that the bHLH domain of Neurog2 is both sufficient to drive sequence-specific DNA binding on chromatin and strongly induces subtype-specific gene expression profiles in differentiating mESCs. Thus, the divergent Ascl1 and Neurog2 binding pattern is the main determinant of the bias in the expression of neuronal-subtype genes.

Ascl1 and Neurog2 binding results in differential chromatin accessibility and enhancer activity. The strong binding preference and the likely importance of the binding pattern in controlling the differentiation trajectory of neurons prompted us to investigate the chromatin landscapes that result from the divergent Ascl1 and Neurog2 binding profiles. We examined genome-wide chromatin accessibility dynamics by ATAC-seq before and after the induction of the proneural TFs. A global accessibility analysis revealed that Ascl1 and Neurog2 induce different accessibility landscapes (Supplementary Fig. 5a). Mirroring the expression dynamics (Supplementary Fig. 1b), the majority of initial accessibility changes were specific to iA or iN neurons (Supplementary Fig. 5a). As differentiation proceeded and the downstream program converged, a larger set of common loci gained accessibility (Supplementary Fig. 5a). We also compared the accessibility landscape in Ascl1-induced neurons from stem cells and fibroblasts[45] (Supplementary Fig. 5b). Following the Ascl1 binding differences, the accessibility landscape between these two neuronal differentiations was markedly dissimilar.

Because of the divergent binding pattern and the resulting accessibility differences following proneural TF induction, we sought to investigate whether Ascl1- and Neurog2-preferred sites gain accessibility during differentiation. Proneural sites gained an ATAC-seq signal after Ascl1 or Neurog2 binding, regardless of their accessibility state before TF induction (Fig. 4a). While Ascl1-preferred sites progressively gained accessibility, Neurog2-preferred sites quickly gained accessibility and remained accessible but lost some ATAC-seq signals at 48 h (Fig. 4b). Interestingly, A[N]^{bHLH} chimera binding also resulted in a rapid gain of accessibility by 12 h, with a pattern similar to that of Neurog2 (Supplementary Fig. 5c). These results demonstrate that, albeit with different dynamics, both bHLH factors induce or maintain regulatory regions in an accessible state. Similarly, independent of the histone 3 lysine 27 acetylation (H3K27ac) status before TF induction, Ascl1 and Neurog2 binding resulted in an increase of H3K27ac at bound sites by 48 h (Fig. 4c,d). Although we observed a gain of accessibility at the previously inaccessible N > A sites in iA neurons by 48 h, these sites did not gain H3K27ac enrichment (Fig. 4b,d). In summary, both Ascl1 and Neurog2 bind to active or inactive regulatory elements, and their binding subsequently increases chromatin accessibility and enhancer activity of bound regulatory regions. Thus, the divergent binding of Ascl1 and Neurog2 results in different chromatin accessibility and activity landscapes during Ascl1- or Neurog2-induced neurogenesis.

Distinct chromatin landscapes induced by Ascl1 and Neurog2 affect the binding of the downstream TFs. We hypothesized that the regulatory activity of the TFs expressed downstream of both proneural factors will be conditioned by the distinct chromatin landscapes induced by Ascl1 and Neurog2. Brn2 (a POU and Homeodomain TF; also known as Pou3f2), Ebf2 (a non-basic HLH and zinc finger TF), and Onecut2 (a CUT and Homeodomain TF) are among the widely expressed neuronal TFs in the nervous system that are induced by both Ascl1 and Neurog2 in differentiating

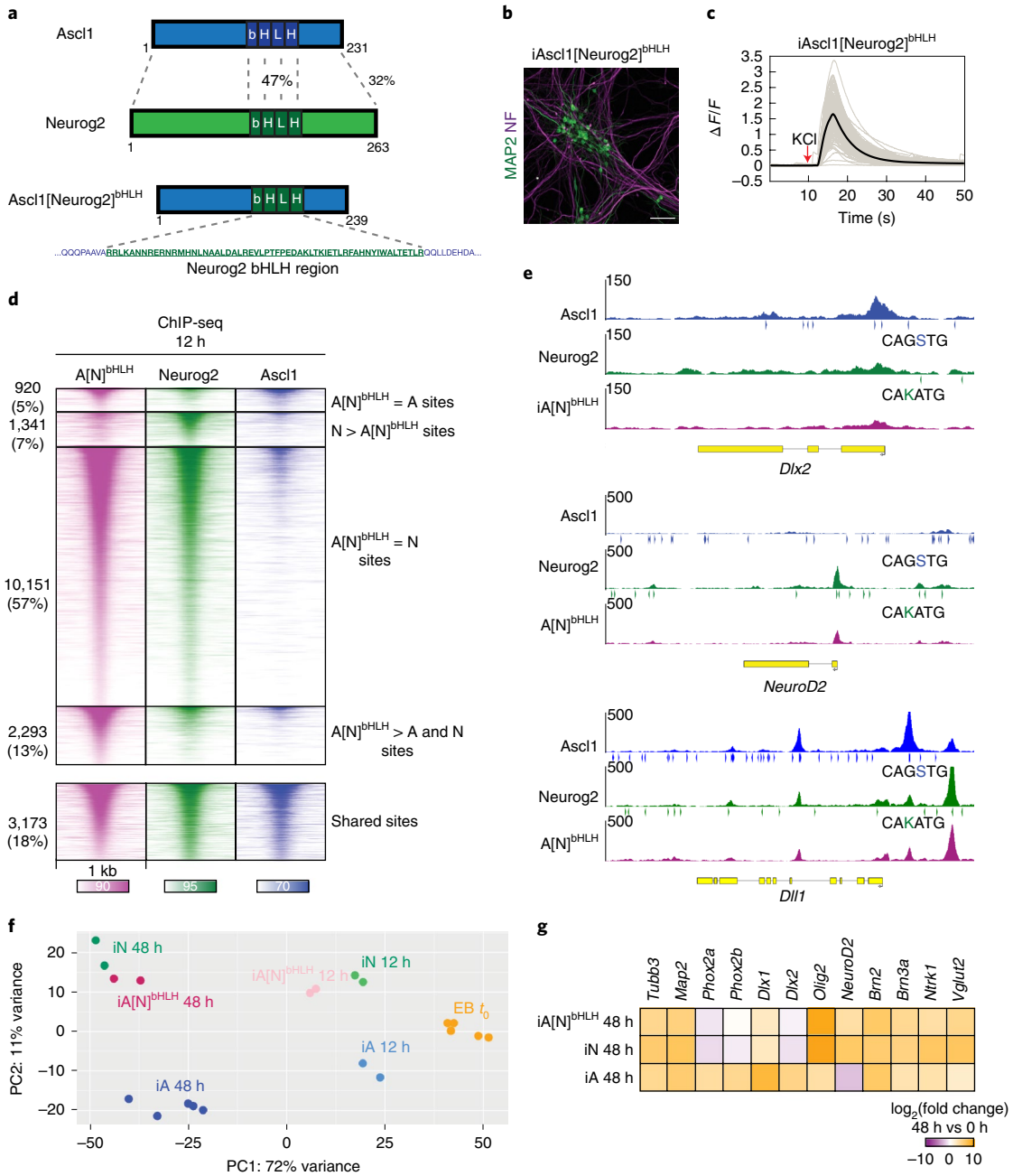


Fig. 3 | bHLH domain of Neurog2 is sufficient to drive both the genomic binding and transcriptional output. a, Schematic of generation of the bHLH chimera by swapping the bHLH domain of Ascl1 with that of Neurog2. The percentages represent the amino acid sequence similarity between the bHLH domains (47%) and the overall protein (32%). The A[N]^{bHLH} chimeric TF construct was used to generate a stable inducible iA[N]^{bHLH} line. **b**, Expression of A[N]^{bHLH} chimeric TF generates neurons that express mature neuronal markers MAP2 (green) and NF (purple) 9 d after induction. Scale bar, 50 µm. Similar results were observed in at least $n = 2$ independent cell differentiations. **c**, iA[N]^{bHLH} neurons respond to KCl-induced depolarization by changing their intracellular Ca^{2+} levels. Thick line represents the average across recordings (from $n = 2$ independent cell differentiations). **d**, The A[N]^{bHLH} chimera binds largely to Neurog2 sites in the genome. ChIP-seq heatmap showing the binding sites of A[N]^{bHLH} chimera in comparison to Neurog2 and Ascl1 (A[N]^{bHLH} $n = 2$). **e**, Genome browser snapshots of Ascl1, Neurog2, and A[N]^{bHLH} chimera binding sites (12 h ChIP-seq) with distribution of Ascl1- or Neurog2-preferred E-boxes (arrowheads) on subtype-specific genes (*Dlx2* and *NeuroD2*) and a shared target *Dll1* (S represents G/C; K represents G/T) (A[N]^{bHLH} $n = 2$). **f**, PCA of the RNA-seq replicates shows A[N]^{bHLH} chimera-induced neurons (iA[N]^{bHLH}) cluster with Neurog2-induced neurons (iN) both at 12 h and 48 h (each circle represents independent cell differentiations). **g**, RNA-seq heatmap showing the expression of representative subtype-specific genes in iA, iN, and iA[N] neurons at 48 h.

mESCs by 48 h (Fig. 5a). Thus, we analyzed Brn2, Ebf2, and Onecut2 genome-wide binding in iA and iN neurons 48 h after induction of the proneural TFs. Around 60% of the Brn2 and Ebf2 binding sites were shared in iA and iN neurons (iA = iN sites), while roughly 40% of Brn2 and Ebf2 sites were differentially enriched in iA or iN neurons (iA > iN and iN > iA sites) (Fig. 5b,c). Binding of Brn2 in ESC differentiation is dissimilar to Brn2 in fibroblasts when expressed alongside Ascl1 and Myt1l (Supplementary Fig. 6a). Among these

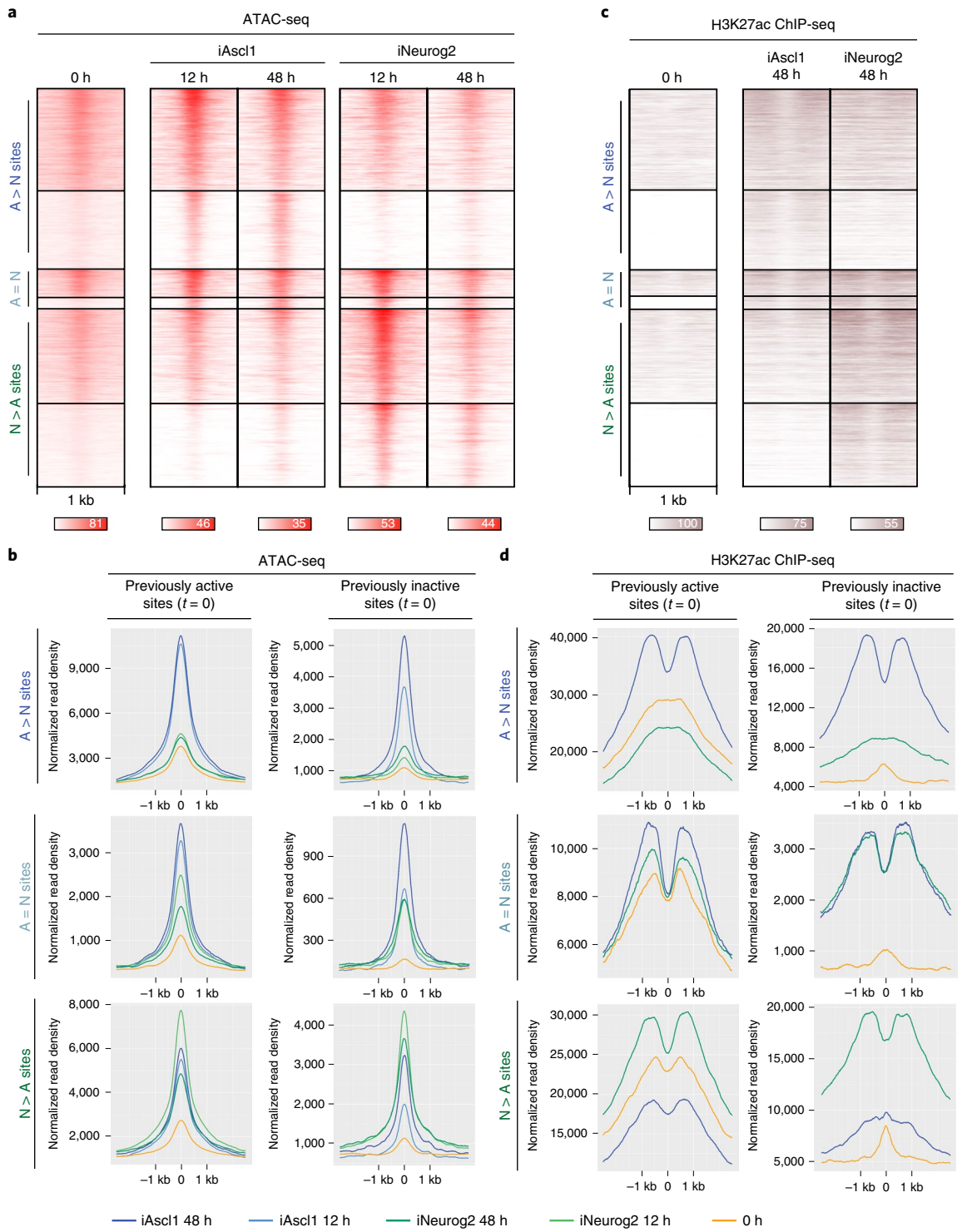


Fig. 4 | Ascl1 and Neurog2 binding results in differential chromatin accessibility and enhancer activity. a, Time-series ATAC–seq heatmaps displaying the gain of accessibility at the Ascl1 and Neurog2 binding sites ($n = 2$). **b**, Metagene plots of accessibility (ATAC–seq reads) at the differentially bound and shared sites of Ascl1 and Neurog2 that were previously active (left) or previously inactive (right) before induction of the TFs (0 h or EB $t = 0$). **c**, H3K27ac ChIP–seq at Ascl1 and Neurog2 binding sites at 48 h shows the gain of enhancer activity at the bound sites in comparison to 0 h ($n = 2$). **d**, Metagene plots of H3K27ac ChIP–seq at the differentially bound and shared sites of Ascl1 and Neurog2 that were previously active (left) or previously inactive (right) before the induction of the TFs (0 h or EB $t = 0$).

TFs, Onecut2 had proportionally fewer differentially bound sites in iA and iN neurons (14%), while the majority of sites bound by Onecut2 were shared in iA and iN neurons (86%) (Fig. 5d).

If the Brn2, Ebf2, and Onecut2 binding differences are shaped by Ascl1- and Neurog2-induced chromatin landscapes, then their differential binding should correlate with the differentially accessible regions established in iA and iN neurons. Indeed, Brn2, Ebf2, and Onecut2 sites that are differentially enriched in iA neurons (iA > iN) occurred in sites that became differentially accessible in iA neurons (Fig. 5e). Similarly, differentially bound sites in iN

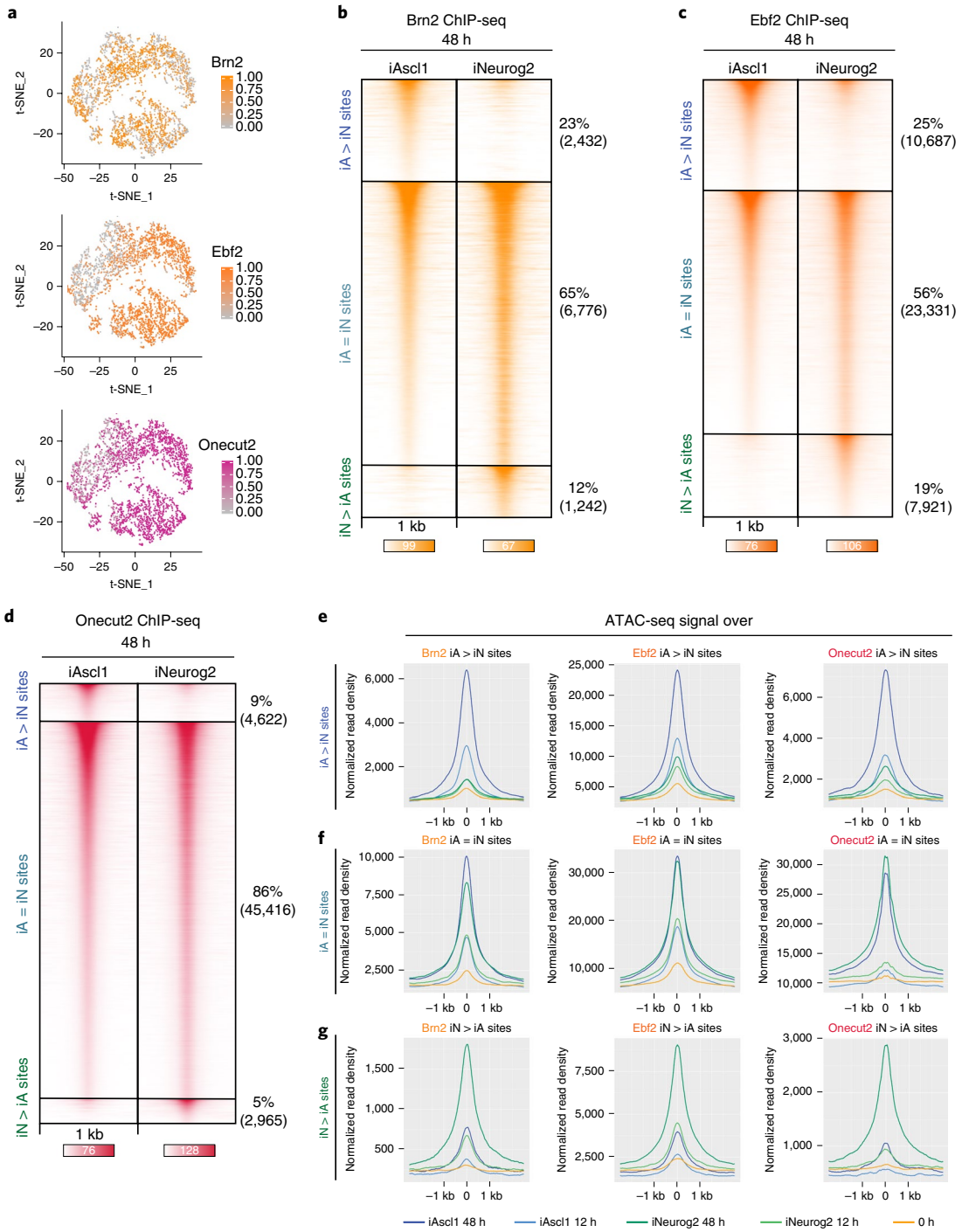


Fig. 5 | Differential chromatin landscapes induced by Ascl1 and Neurog2 shape the binding patterns of the shared downstream TFs. a, t-SNE plots showing the cells that express downstream TFs (upper cluster, iA; lower cluster, iN; see also Fig. 1e). The circles are colored by the expression levels of downstream TF Brn2, Ebf2 and Onecut2 ($n=1$ cell differentiation). **b**–**d**, ChIP-seq heatmaps of endogenous Brn2 (**b**), Ebf2 (**c**) and Onecut2 (**d**) binding in iA and iN neurons at 48 h after induction of Ascl1 and Neurog2. iA > iN designates sites enriched in iA neurons, iN > iA designates sites enriched in iN neurons, and iA = iN designates shared binding in both neurons ($n=2$). **e**, Metagene plots of accessibility (ATAC-seq reads) overlap at the differentially bound sites of Brn2 (left), Ebf2 (middle) and Onecut2 (right) in iA neurons (iA > iN sites). **f**, Metagene plots of accessibility (ATAC-seq reads) overlap at the shared sites of Brn2 (left), Ebf2 (middle) and Onecut2 (right) in iA and iN neurons (iA = iN sites). **g**, Metagene plots of accessibility (ATAC-seq reads) overlap at the differentially bound sites of Brn2 (left), Ebf2 (middle) and Onecut2 (right) in iN neurons (iN > iA sites).

neurons (iN > iA) also occurred in sites that became accessible in iN neurons (Fig. 5g). Meanwhile, the Brn2, Ebf2, and Onecut2 shared binding sites in iA and iN neurons (iA = iN) exhibited high ATAC-seq read counts in both iA and iN neurons, thus were accessible in both neurons (Fig. 5f). Furthermore, the Brn2, Ebf2, and Onecut2 differential binding sites in iA neurons (iA > iN) substantially overlapped with Ascl1-preferred binding (A > N sites) (45%, 35%, 29%, respectively, and only <1% of expected overlap by chance) at 48 h (Fig. 6a–c). Two observations suggest that the differentially bound sites represent direct DNA-binding targets of the downstream TFs. First, a motif-finding

a

48 h Brn2 binding overlap at 48 h Ascl1 and Neurog2 sites

	A > N sites	A = N sites	N > A sites
Brn2 (iA > iN)	45%	5%	3%
Brn2 (iA = iN)	3%	7%	21%
Brn2 (iN > iA)	0%	0%	55%

b

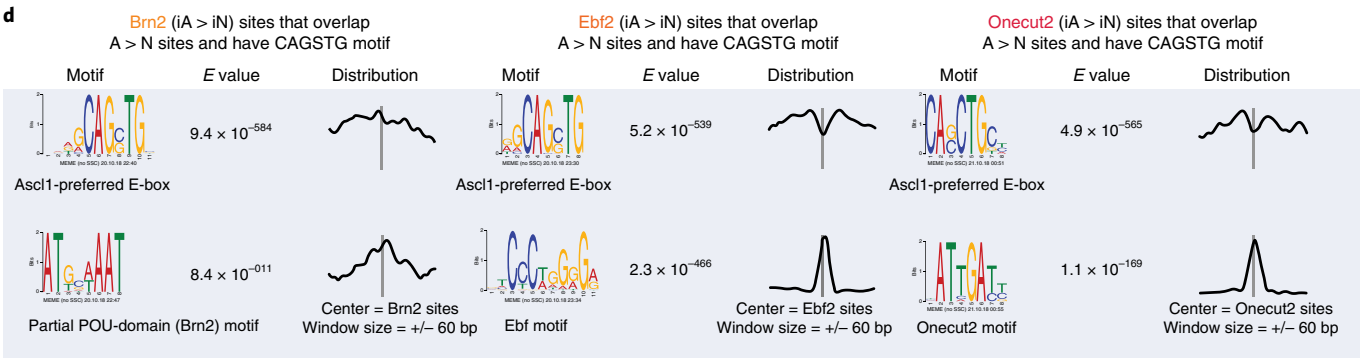
48 h Ebf2 binding overlap at 48 h Ascl1 and Neurog2 sites

	A > N sites	A = N sites	N > A sites
Ebf2 (iA > iN)	35%	8%	4%
Ebf2 (iA = iN)	3%	7%	19%
Ebf2 (iN > iA)	0%	0%	41%

c

48 h Onecut2 binding overlap at 48 h Ascl1 and Neurog2 sites

	A > N sites	A = N sites	N > A sites
OC2 (iA > iN)	29%	5%	3%
OC2 (iA = iN)	1%	2%	8%
OC2 (iN > iA)	0%	0%	36%



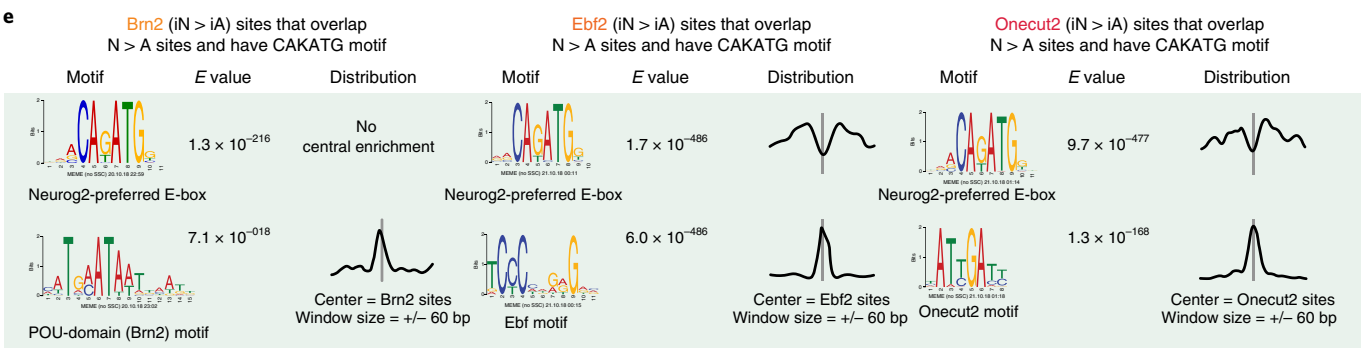


Fig. 6 | Differentially bound sites of downstream TFs in iA or iN neurons overlap with Ascl1 or Neurog2 binding. a–c, A subset of differentially enriched Brn2 (**a**), Ebf2 (**b**) Onecut2 (**c**) binding sites in iA or iN neurons at 48 h overlap with Ascl1 or Neurog2 differential binding at 48 h. **d**, MEME motif search at the 48 h differentially bound Brn2, Ebf2 and Onecut2 sites in iAscl1 neurons (iA > iN) that overlap with differentially bound Ascl1 sites (A > N) with CAGSTG motif. Note that the cognate motif of downstream TFs is present and Ascl1-preferred E-boxes are depleted in the motif distribution graphs centered on downstream TF motifs. The *E* values are reported by MEME and represent an estimate of the expected number of motifs with the same log likelihood ratio that one would find in a similarly sized set of random sequences. **e**. MEME motif search at the 48 h differentially bound Brn2, Ebf2 and Onecut2 sites in iNeurog2 neurons (iN > iA) that overlap with differentially bound Neurog2 sites (N > A) with CAKATG motif.

analysis at the differentially bound Brn2, Ebf2, and Onecut2 sites that harbor an E-box motif revealed Ascl1- versus Neurog2-preferred E-boxes along with appropriate cognate motifs for downstream TFs (Fig. 6d,e). Second, while the downstream TF cognate motifs were enriched at the center of the ChIP-seq peaks, the E-box was depleted at the central peak location (Fig. 6d,e). However, we noted that the downstream TF cognate motif instances were weaker at differentially bound sites compared with other downstream TF binding sites in iA and iN cells (Supplementary Fig. 6b). These results support a model in which the differential chromatin accessibility induced by Ascl1 and Neurog2 binding exposes weaker cognate motifs that can then be bound by downstream TFs. Consequently, the activity of widely expressed TFs is not functionally equivalent in all neurons.

Ascl1 and Neurog2 control initial transcriptional changes and bias the regulatory activity of downstream TFs in the acquisition of neuron-specific identity. To understand how the non-overlapping binding of Ascl1 and Neurog2 induces expression of subtype-specific (neuron-specific) and generic (shared) neuronal genes, we explored the association between binding sites of Ascl1, Neurog2, and downstream TFs with induced gene expression using GREAT. We first investigated the association between differential binding and gene expression at 12 h. This analysis revealed that early (12 h) differential binding of Ascl1 or Neurog2 correlated well with early differentially expressed genes at 12 h (Fig. 7a). Around 65% and 78% of the Ascl1 or Neurog2 differentially expressed genes at 12 h have at least one A > N or N > A peak within GREAT-defined regulatory domains, respectively (Fig. 7a; Supplementary Table 1). Dividing proneural TF binding into previously accessible and inaccessible regions did not dramatically modify the association with transcription (Fig. 7b). Thus, the initial divergent Ascl1 and Neurog2 binding correlates with differential gene expression regardless of the previous accessibility state.

The next challenge was to understand how the 10% overlap in Ascl1 and Neurog2 binding results in ~80% overlap in gene expression by 48 h after induction. Sites that were bound by Ascl1

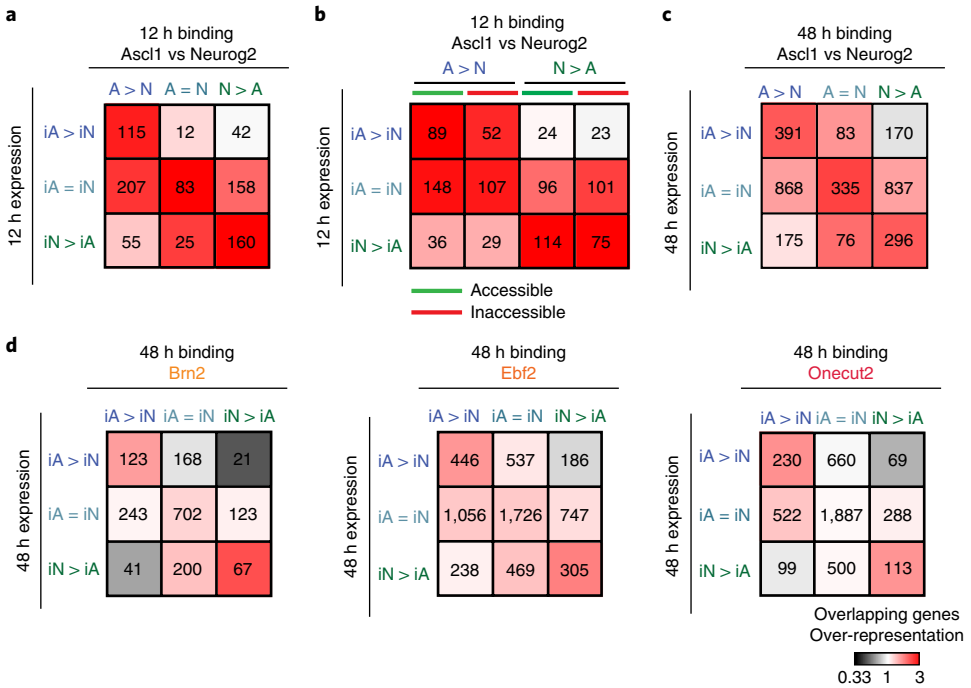


Fig. 7 | Associations between genomic binding sites and gene expression. a, Heatmap representing the associations (ratio between the genes overlapped by Ascl1 or Neurog2 peaks versus random peaks: overlapping genes over-representation) between 12 h Ascl1 and Neurog2 binding with 12 h gene expression. **b**, Heatmap representing the associations between 12 h Ascl1 and Neurog2 sites that were previously accessible (green) or inaccessible (red) with 12 h gene expression. **c**, Heatmap representing the associations between 48 h Ascl1 and Neurog2 binding with 48 h gene expression. **d**, Heatmap representing the associations between 48 h Brn2 (left), Ebf2 (middle) Onecut2 (right) binding in iA and iN neurons with 48 h gene expression in iA and iN neurons. For all panels, the number of genes that overlap with specific binding classes are listed inside the squares.

and Neurog2 (A = N sites) associated with genes that were upregulated in both types of neurons (Fig. 7a–c). Additionally, differentially bound Ascl1 and Neurog2 sites (A > N and N > A) were also associated with genes that were upregulated in both neurons (Fig. 7a–c). Expanding on *Dll1* regulation by differential Ascl1 and Neurog2 binding[34], our results suggest that distinct Ascl1 and Neurog2 regulatory elements are spatially peppered around similar sets of genes, and Ascl1 and Neurog2 induce shared and neuron-specific (subtype-specific) gene expression through different regulatory regions (Supplementary Fig. 6c).

Ascl1 and Neurog2 drive the majority of the expression differences at early time points. We tested whether the downstream factors contribute to gene expression differences. Shared binding sites of Brn2, Ebf2, and Onecut2 in iA and iN neurons were associated with shared upregulated genes in iA and iN neurons at 48 h (Fig. 7d). Similarly, differentially bound Brn2, Ebf2, and Onecut2 sites significantly associated with Ascl1- or Neurog2-specific gene expression at 48 h (Fig. 7d). These results suggest that the initial divergent binding of the proneural TFs biases both binding (Fig. 6) and activity of shared downstream TFs, thus contributing to neuron-specific expression profiles.

Discussion

Here, we probed the molecular mechanisms governing the divergent roles played by the proneural factors Ascl1 and Neurog2 during neuronal differentiation. Using direct neuronal programming of isogenic mESCs, we found that these proneural factors influence cell fate in two ways. First, Ascl1 and Neurog2 bind to and regulate distinct sets of regions in the genome, determined by the intrinsic activity of their bHLH domains. Second, because of this initial divergent binding, Ascl1 and Neurog2 induce differential chromatin landscapes that shape the binding and function of the shared downstream TFs during neuronal fate specification. Hence, we speculate that the regulatory activity of the widely expressed shared TFs will not be identical when expressed downstream of Ascl1 or Neurog2 during neurogenesis and reprogramming experiments.

The question of bHLH TF binding specificity is of importance not only for proneural factors but also for bHLH TFs that regulate various developmental events such as myogenesis, hematopoiesis, and pancreatic development[38]. While extensive binding differences are intuitive for TFs that belong to different bHLH families and induce different cell types such as MyoD versus Ascl1 or NeuroD2[31,46], it was striking to observe the substantial difference in the genomic binding of the proneural bHLHs Ascl1 and Neurog2 even when expressed in similar chromatin contexts. bHLH dimers acquire specificity by recognizing distinct E-box half sites (CANNTG) in DNA[47]. Thus, the non-palindromic Ascl1- and Neurog2-preferred E-boxes (CAGGTG and CAGATG, respectively) enriched at the differentially bound sites could reflect the sites that are bound with their heterodimerization partners.

Several experiments suggest the importance of the bHLH domain for the subtype-specific activity of neural bHLHs[46,48–50]. Using an equivalent chromatin and cellular context for a comprehensive analysis of Ascl1-, Neurog2-, and Ascl1[Neurog2]^{bHLH}-induced neurogenesis led us to an interesting observation. That is, the genomic binding, transcriptional output, and even the chromatin accessibility dynamics induced by the Ascl1[Neurog2]^{bHLH} chimera were similar to that induced by Neurog2. The DNA specificity of the bHLH domain can be further divided by amino acids in mostly the basic domain and helix 1 contacting DNA and helix 2 mediating dimerization[39]. Additional experiments are required to resolve if, in this differentiation system, DNA binding preferences of the amino acids in the basic and helix 1 region or the dimerization surface guides

Ascl1 and Neurog2 to different sites. Phosphorylation of certain residues outside the bHLH domain has been shown to alter the proneural activity[42] and the interactions with putative partners of Ascl1 and Neurog2 homologs in *Xenopus* and in mouse models[40,41,43,44]. Although the controlled mESC differentiation system is ideal for studying the intrinsic differences between these proneural TFs, it might lack the complexity of the extracellular signaling occurring in developing embryos. Alternatively, posttranslational modifications can fine-tune the binding preferences that might have been overshadowed by the high expression levels required to differentiate mESCs into neurons.

Divergent Ascl1 and Neurog2 binding induces initially divergent accessibility and expression patterns that later converge on a generic neuronal fate while maintaining subtype-specific differences (Fig. 1d,g; Supplementary Fig. 1b–d). We found that shared binding of the proneural TFs and the downstream TFs correlated with upregulation of a generic neuronal program. This divergent-to-convergent neuronal differentiation trajectory is in line with previous studies that described NeuroD4 among the common targets regulating the shared genes during astrocyte-to-neuron reprogramming by Ascl1 or Neurog2[27]. The complete cascade of events that lead to this convergence while maintaining some expression differences in astrocyte and pluripotent cell differentiations are yet to be uncovered. Brn2 was proposed to be recruited to its genome-wide sites by Ascl1 in neuronal reprogramming of fibroblasts[25]. We report here that both Ascl1 and Neurog2 influence the binding pattern of several downstream TFs. Our findings propose a novel mechanism that links these previous findings: the widely expressed shared TFs contribute not only to generic neuronal program but also to neuron-specific programs by retaining the memory of the initial neurogenesis triggered by divergent binding of Ascl1 and Neurog2 (Supplementary Fig. 7). Thus, in addition to the differentially expressed TFs and/or terminal selectors, the role of widely expressed TFs should be considered in determining the aspects of neuronal subtype identity.

The ability of Ascl1 and Neurog2 to substitute for each other varies in different regions of the nervous system[5]. We propose that the intrinsic Ascl1 and Neurog2 differences will have a smaller impact on instructing the neuronal-subtype identity in neuronal progenitors in which the chromatin is strongly pre-patterned for a specific neuronal type. However, when expressed in a permissive chromatin and cellular state, Ascl1 and Neurog2 differentially force the specification of distinct neuronal-subtype identities. These findings provide a mechanistic explanation for the importance of choosing the right proneural factor in neuronal differentiation strategies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0399-y.

Received: 12 July 2018; Accepted: 28 March 2019;
Published online: 13 May 2019

References

1. Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3**, 517–530 (2002).
2. Guillemot, F. & Hassan, B. A. Beyond proneural: emerging functions and regulations of proneural proteins. *Curr. Opin. Neurobiol.* **42**, 93–101 (2017).
3. Urbán, N. & Guillemot, F. Neurogenesis in the embryonic and adult brain: same regulators, different roles. *Front. Cell. Neurosci.* **8**, 396 (2014).
4. Schuurmans, C. & Guillemot, F. Molecular mechanisms underlying cell fate specification in the developing telencephalon. *Curr. Opin. Neurobiol.* **12**, 26–34 (2002).
5. Parras, C. M. et al. Divergent functions of the proneural genes *Mash1* and *Ngn2* in the specification of neuronal subtype identity. *Genes Dev.* **16**, 324–338 (2002).
6. Osório, J., Mueller, T., Rétaux, S., Vernier, P. & Wullimann, M. F. Phylotypic expression of the bHLH genes *Neurogenin2*, *Neurod*, and *Mash1* in the mouse embryonic forebrain. *J. Comp. Neurol.* **518**, 851–871 (2010).
7. Simionato, E. et al. atonal- and achaete-scute-related genes in the annelid *Platynereis dumerilii*: insights into the evolution of neural basic-helix-loop-helix genes. *B Evo.l Biol.* **8**, 1–13 (2008).
8. Jarman, A. P. & Ahmed, I. The specificity of proneural genes in determining *Drosophila* sense organ identity. *Mech. Dev.* **76**, 117–125 (1998).
9. Fode, C. et al. A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes Dev.* **14**, 67–80 (2000).
10. Jarman, A. P., Grau, Y., Jan, L. Y. & Jan, Y. N. atonal is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. *Cell* **73**, 1307–1321 (1993).
11. Hirsch, M. R., Tiveron, M. C., Guillemot, F., Brunet, J. F. & Goridis, C. Control of noradrenergic differentiation and Phox2a expression by MASH1 in the central and peripheral nervous system. *Development* **125**, 599–608 (1998).
12. Lo, L., Dormand, E., Greenwood, A. & Anderson, D. J. Comparison of the generic neuronal differentiation and neuron subtype specification functions of mammalian achaete-scute and atonal homologs in cultured neural progenitor cells. *Development* **129**, 1553–1567 (2002).
13. Ma, Q., Fode, C., Guillemot, F. & Anderson, D. J. Neurogenin1 and neurogenin2 control two distinct waves of neurogenesis in developing dorsal root ganglia. *Genes Dev.* **13**, 1717–1728 (1999).
14. Schuurmans, C. et al. Sequential phases of cortical specification involve neurogenin-dependent and -independent pathways. *EMBO J.* **23**, 2892–2902 (2004).
15. Baker, N. E. & Brown, N. L. All in the family: proneural bHLH genes and neuronal diversity. *Development* **145**, 1–9 (2018).
16. Flames, N. & Hobert, O. Transcriptional control of the terminal fate of monoaminergic neurons. *Annu. Rev. Neurosci.* **34**, 153–184 (2011).
17. Tsunemoto, R. et al. Diverse reprogramming codes for neuronal identity. *Nature* **557**, 380 (2018).
18. Wichterle, H., Gifford, D. & Mazzoni, E. Mapping neuronal diversity one cell at a time. *Science* **341**, 726–727 (2013).
19. Hobert, O. Regulation of terminal differentiation programs in the nervous system. *Annu. Rev. Cell Dev. Biol.* **27**, 681–696 (2011).
20. Stefanakis, N., Carrera, I. & Hobert, O. Regulatory logic of pan-neuronal gene expression in *C. elegans*. *Neuron* **87**, 733–750 (2015).
21. Heinrich, C. et al. Generation of subtype-specific neurons from postnatal astroglia of the mouse cerebral cortex. *Nat. Protoc.* **6**, 214–228 (2011).
22. Chanda, S. et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem Cell Rep.* **3**, 282–296 (2014).
23. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
24. Mall, M. et al. Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature* **544**, 245–249 (2017).
25. Wapinski, O. L. et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621–635 (2013).
26. Vadodaria, K. C. et al. Generation of functional human serotonergic neurons from fibroblasts. *Mol. Psychiatry* **21**, 49–61 (2016).
27. Masserdotti, G. et al. Transcriptional mechanisms of proneural factors and REST in regulating neuronal reprogramming of astrocytes. *Cell Stem Cell* **17**, 74–88 (2015).
28. Smith, D. K., Yang, J., Liu, M.-L. L. & Zhang, C.-L. L. Small molecules modulate chromatin accessibility to promote NEUROG2-mediated fibroblast-to-neuron reprogramming. *Stem Cell Rep.* **7**, 955–969 (2016).
29. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
30. Raposo, A. A. et al. Ascl1 coordinately regulates gene expression and the chromatin landscape during neurogenesis. *Cell Rep.* **10**, 1544–1556 (2015).
31. Casey, B. H., Kollipara, R. K., Pozo, K. & Johnson, J. E. Intrinsic DNA binding properties demonstrated for lineage-specifying basic helix-loop-helix transcription factors. *Genome Biol.* **28**, 484–496 (2018).
32. Slattery, M. et al. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).
33. Powell, L. M., Zur Lage, P. I., Prentice, D. R., Senthinathan, B. & Jarman, A. P. The proneural proteins Atonal and Scute regulate neural target genes through different E-box binding sites. *Mol. Cell. Biol.* **24**, 9517–9526 (2004).
34. Castro, D. S. et al. Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Dev. Cell* **11**, 831–844 (2006).
35. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
36. Gordân, R. et al. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* **3**, 1093–1104 (2013).
37. Rohs, R. et al. Origins of specificity in protein–DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).

38. Massari, M. E. & Murre, C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* **20**, 429–440 (2000).
39. Ma, P. C., Rould, M. A., Weintraub, H. & Pabo, C. O. Crystal structure of MyoD bHLH domain–DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77**, 451–459 (1994).
40. Ali, F. R. et al. The phosphorylation status of Ascl1 is a key determinant of neuronal differentiation and maturation in vivo and in vitro. *Development* **141**, 2216–2224 (2014).
41. Hindley, C. et al. Post-translational modification of Ngn2 differentially affects transcription of distinct targets to regulate the balance between progenitor maintenance and differentiation. *Development* **139**, 1718–1723 (2012).
42. Quan, X.-J. et al. Post-translational control of the temporal dynamics of transcription factor activity regulates neurogenesis. *Cell* **164**, 460–475 (2016).
43. Li, S. et al. RAS/ERK signaling controls proneural genetic programs in cortical development and gliomagenesis. *J. Neurosci.* **34**, 2169–2190 (2014).
44. Li, S. et al. GSK3 temporally regulates neurogenin 2 proneural activity in the neocortex. *J. Neurosci.* **32**, 7791–7805 (2012).
45. Wapinski, O. L. et al. Rapid chromatin switch in the direct reprogramming of fibroblasts to neurons. *Cell Rep.* **20**, 3236–3247 (2017).
46. Fong, A. P. et al. Conversion of MyoD to a neurogenic factor: binding site specificity determines lineage. *Cell Rep.* **10**, 1937–1946 (2015).
47. De Masi, F. et al. Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res.* **39**, 4553–4563 (2011).
48. Chien, C. T., Hsiao, C. D., Jan, L. Y. & Jan, Y. N. Neuronal type information encoded in the basic-helix-loop-helix domain of proneural genes. *Proc. Natl Acad. Sci. USA* **93**, 13239–13244 (1996).
49. Nakada, Y., Hunsaker, T. L., Henke, M. R. & Johnson, J. E. Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification. *Development* **131**, 1319–1330 (2004).
50. Quan, X.-J. et al. Evolution of neural precursor selection: functional divergence of proneural proteins. *Development* **131**, 1679–1689 (2004).

Acknowledgements

This work is supported by the NICHD (R01HD079682) and Project ALS (A13-0416) to E.O.M. and by a NYSTEM pre-doctoral training grant (C026880) to B.A. S.M. is supported by the NIGMS (R01GM125722) and the National Science Foundation ABI Innovation grant no. DBI1564466. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. M.R. is supported by NYU MSTP (T32GM007308) and Developmental Genetics T32 (T32HD007520) grants. N.F. and M.M.-E. are supported by an ERC Starting Grant (ERC STG 2011–281920). The authors would like to thank L. Tejavibulya and A. Ashokkumar for their help with molecular biology, M. Khalfan for his help with scRNA-seq analysis, M. Cammer from the NYU Medical Center Microscopy Core for the ImageJ script used in calcium imaging analysis, and the NYU Genomics Core facility. Finally, the authors would like to thank S. Small, N. Konstantinidis, P. Onal, O. Wapinski, S. Ercan, C. Rushlow, C. Desplan, and Mazzoni Lab members for their helpful suggestions on the manuscript.

Author contributions

B.A. performed cell differentiations, RNA-seq, ChIP-seq, ATAC-seq, scRNA-seq, immunohistochemistry experiments, and generated the inducible chimera line. Two replicates of iAscl1 48 h RNA-seq experiments were performed by M.M.-E. with guidance from N.F. M.R. performed calcium imaging of neurons with guidance from N.R. G.G provided the Tubb3::GFP line. A.K., B.A., and S.M. performed analysis of all sequencing data. B.A., S.M., and E.O.M. conceived the experiments and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0399-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.M. or E.O.M.

Journal peer review information: *Nature Neuroscience* thanks Diogo Castro, Carol Schuurmans, and other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Experimental procedures. *Cell line generation and cell differentiation.* Inducible cell lines were generated using a previously described inducible cassette exchange (ICE) method[51]. Resulting transgenic lines contain a single-copy insertion of the transgene into an expression-competent (*HPRT*) locus. The p2Lox-Neurog2 (iNeurog2) plasmid was generated by cloning *Neurog2* complementary DNA into p2Lox-Flag plasmid[52]. Likewise, the p2Lox-Ascl1 (iAscl1) plasmid was generated by cloning mouse *Ascl1* cDNA into p2Lox-V5 plasmid[53]. To generate the p2Lox-iAscl1[Neurog2]^{bHLH} chimera, 396 bp of oligonucleotide gBlocks (IDT) fragment encompassing the Neurog2 bHLH domain fused to the carboxyl terminus of Ascl1 with 1× hemagglutinin (HA) tag sequence was synthesized. The Ascl1 amino-terminal fragment was amplified from mouse *Ascl1* cDNA. In-fusion cloning (Clontech) was used to clone/fuse Ascl1 N-terminal and gBlocks Neurog2 bHLH-Ascl1 C-terminal-HA in a p2Lox plasmid backbone. The inducible cell lines (iA, iN, and iA[N]^{bHLH}) were generated by treating the recipient mESCs for 16 h with 1 μg ml^{-1} doxycycline (Sigma, D9891) to induce Cre recombinase expression to mediate recombination following electroporation of the p2Lox-Ascl1, p2Lox-Neurog2, or p2Lox-iAscl1[Neurog2]^{bHLH} plasmids. After G418 selection (250 ng ml^{-1}, Cellgro), cell lines were characterized by performing antibody staining against the tagged transgenic proteins Ascl1-V5 (anti-V5; R960-25), FLAG-Neurog2 (anti-FLAG; F1804), and A[N]^{bHLH}-HA (anti-HA; ab9110).

The Tubb3::T2A-GFPnls line was generated by designing two single guide RNAs (5′-GCTGCGAGCAACTTCACTT and 5′-GAAGATGATGACGAGGAAT) to target Cas9 to the stop codon on exon 4 of *Tubb3*. A donor vector containing T2A peptide and green fluorescent protein (GFP) with a C-terminal nuclear localization signal was cloned in frame between ~800 bp *Tubb3* homologous arms flanking the stop codon. The coding sequence upstream of the *Tubb3* stop codon was amplified with 5′-CCCTACAACGCCACCCTGTCCAT (forward) and 5′-CTTGGGCCCCTGGGCTTCTGATTCTTC (reverse) primers. The 3′5′ untranslated region sequence downstream of the *Tubb3* stop codon was amplified with 5′-AGTTGCTCGCAGCTGG (forward) and 5′-CCAGCCTTCCCTGCGTTTTTTTC (reverse) primers. Knock-in clones were selected for GFP expression after neuronal differentiation. The p2Lox-Neurog2 plasmid was nucleofected to Tubb3::T2A-GFPnls ESC line to generate iNeurog2 Tubb3::GFP stable line.

The inducible mESCs were grown in 2i (2-inhibitors) based medium (Advanced DMEM/F12: Neurobasal (1:1) Medium (Gibco), supplemented with 2.5% mESC-grade fetal bovine serum (vol/vol, Corning), N2 (Gibco), B27 (Gibco), 2 mM L-glutamine (Gibco), 0.1 mM ß-mercaptoethanol (Gibco), 1,000 U ml^{-1} leukemia inhibitory factor (Millipore), 3 mM CHIR (BioVision) and 1 mM PD0325901 (Sigma) on 0.1% gelatin (Milipore) coated plates at 37 °C, 8% CO_2. To obtain embryoid bodies (EBs), 60–70% confluent mESCs were dissociated using TrpLE (Gibco) and plated in AK medium (Advanced DMEM/F12: Neurobasal (1:1) Medium, 10% Knockout SR (vol/vol) (Gibco), penicillin–streptomycin (Gibco), 2 mM L-glutamine and 0.1 mM ß-mercaptoethanol) on untreated plates for 2 days (day −2) at 37 °C, 8% CO_2. After 2 days, the EBs were passaged 1:2, and expression of the transgenes was induced by adding 3 μg ml^{-1} doxycycline (Sigma, D9891) to the AK medium. For differentiating mESC (EB) antibody stainings, RNA-seq, sc-RNAseq, and ATAC-seq experiments, 2–3 × 10^5 cells were plated in each 100-mm untreated dishes (Corning). For ChIP-seq experiments, the same conditions were used, but the seeded cell number was scaled up to 3–3.5 × 10^6 cells in 245 mm × 245 mm square dishes (Corning).

For day-9 attached neuron antibody stainings and calcium recording experiments, EBs induced with doxycycline for 2 days (48 h + doxycycline) were dissociated with 0.05% Trypsin-EDTA (Gibco) and plated on poly-D-lysine (Sigma, P0899) coated 4-well plates. The dissociated neurons were grown in neuronal medium with supplements (Neurobasal Medium supplemented with 2% fetal bovine serum, B27, 0.5 mM L-glutamine, 0.01 mM β-mercaptoethanol, 3 μg ml^{-1} doxycycline, 10 ng ml^{-1} GDNF (PeproTech, 450–10), 10 ng ml^{-1} BDNF (PeproTech, 450–02), 10 ng ml^{-1} CNTF (PeproTech 450–13), 10 μM Forskolin (Fisher, BP2520–5), and 100 μM IBMX (Tocris, 2845)) at 37 C°, 5% CO_2. Antimitotic reagents (4 μM 5-fluoro-2′-deoxyuridine (Sigma, F0503) and 4 μM uridine (Sigma, U3003)) were used to kill any residual proliferating cells that might have failed neuronal differentiation.

Immunocytochemistry. EBs were collected and fixed in 4% paraformaldehyde (vol/vol) in PBS. Fixed EBs were cryoprotected in 30% sucrose and were embedded in OCT (Tissue-Tek) and sectioned for staining. Primary antibody stainings were done by incubating overnight at 4 °C, and secondary antibody stainings were done by incubating for 1 h at room temperature. Day-9 attached neuron stainings were done on coverslips coated with poly-D-lysine with the same incubation times. After staining, samples were mounted with Fluoroshield with 4,6-diamidino-2-phenylindole (DAPI; Sigma). Images were acquired using a SP5 Leica confocal microscope. The following primary and secondary antibodies were used: anti-Tubb3 (Sigma, T2200, 1:2,000); anti-V5 (Thermo Fisher Scientific, R960-25, 1:5,000); anti-Flag (Sigma, F1804; 1:500); anti-Map2 (Abcam, ab5392, 1:1,000); anti-neurofilament (DSHB, 2H3, 1:1,000); anti-HA (Abcam, ab9110, 1:5,000); goat anti-chicken Alexa 488 (Invitrogen, A-11039, 1:1,000); goat anti-rabbit Alexa 568 (Invitrogen, A-11036, 1:1,000); and goat anti-mouse Alexa 647 (Invitrogen, A-21236, 1:1,000).

Calcium imaging. Dissociated iA, iN and iA[N]^{bHLH} EBs (750,000) were plated on 0.001% poly-D-lysine coated 35-mm glass bottom plates (MatTek, P35GC-1.5-10-C) and incubated for 9 days in neuronal medium (see above). To load neurons with calcium indicator, the cells were incubated for 30–60 min with 2 μM Fluo-4 AM (Thermo Fisher) and 0.02% Pluronic F-127 (Invitrogen) in Ringer's solution (150 mM NaCl, 4 mM KCl, 10 mM HEPES buffer, 10 mM glucose, 2 mM $MgCl_2$, 2 mM $CaCl_2$)[54] at room temperature[55]. Fluo-4 fluorescence was excited with 488 nm light from a monochromatic Polychrome light source (Till Photonics) and emissions were filtered through a 500–550 nm band-pass filter (Chroma). Fluorescence images were acquired at 10 Hz with a cooled EM-CCD camera (Andor). Fluo-4 fluorescence was measured in regions of interest around the cell body of a given neuron. Bath solution exchanges were performed via a computer-controlled gravity-fed perfusion system (Automate Scientific). Excitation light, image acquisition, and hardware control were executed by the Live Acquisition software package (Till Photonics). Post-acquisition analysis was performed using custom Matlab scripts, which normalized changes in fluorescence to the pre-stimulus baseline fluorescence, which was computed as the mean of the 20 lowest fluorescence measurements taken before stimulus application.

RNA-seq. Cells were collected 0, 12, and 48 h after doxycycline induction, and RNA was isolated by resuspending in TRIzol reagent (Invitrogen, 15596026) followed by purification using a RNeasy mini kit (Qiagen, 74106). RNA integrity was measured using Agilent High Sensitivity RNA Screentape (Agilent Tech, 5067–5080). RNA (500 ng) was spiked-in (1:100) with ERCC Exfold Spike-in mixes (Thermo Fisher, 4456739) for accurate comparison across samples. Illumina TruSeq LS kit v2 (RS-122–2001; RS-122–2002) was used to prepare RNA-seq libraries. The final quantification of the library before pooling was done using a KAPA library amplification kit (Roche Lightcycler 480). The libraries were sequenced on an Illumina NextSeq 500 using V2 and V2.5 chemistry for 50 cycles (single-end) at the Genomics Core Facility at NYU.

scRNA-seq. Cells (iAscl1-v5 and iNeurog2 Tubb3::GFP) were collected 48 h after doxycycline induction, and washes were done in 1× PBS with 0.04 mg ml^{-1} BSA (Thermo Fisher, AM2616). Cells were strained through a 30 μM CellTrics filter (cat. no. 04-004-2326) to remove cell clumps. An equal number of iA and iN Tubb3::GFP cells were pooled to have 1,000 cells per μl. A 10× Genomics Chromium Single Cell 3′ library kit was used to generate a single-cell library for a targeted cell recovery rate of 10,000 cells (120262 Chromium i7 Multiplex Kit, 120236 Chromium Single Cell 3′ Chip Kit v2, 120237 Chromium Single Cell 3′ Library and Gel Bead Kit v2). The fragment length distribution of the library was determined using an Agilent High Sensitivity DNA D1000 Screentape (5067–5585) system, and the final quantification of the library before pooling was done using a KAPA library amplification kit (Roche Lightcycler 480). The libraries were sequenced on an Illumina NextSeq 500 High Output using V2.5 chemistry with 26 × 98 bp – 150 cycles run confirmation at the genomics core facility at NYU.

ChIP-seq. Cells were collected at 12 h and 48 h after TF induction and fixed with 1 mM DSG (ProtoChem) followed by 1% FA (vol/vol) each for 15 min at room temperature. Pellets containing 25–30 × 10^6 cells were aliquoted and flash-frozen at −80 °C. Cells were lysed in 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) with 1× protease inhibitors (Roche, 11697498001) at 4 °C. After 10 min, the cells were resuspended in 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 at 4 °C. Nuclear extracts were resuspended in cold 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)). Sonication was performed on ice using a Branson 450 digital sonifier (Marshall Scientific, B450CC) at 20% amplitude, and 18 cycles of 30 s on/60 s off into average size of approximately 300 bp. Immunoprecipitation was done overnight at 4 °C on a rotator with Dynabeads protein-G (Thermo Fisher) conjugated antibodies. A sample of 5 μg of the following antibodies were used for immunoprecipitation: anti-Ascl1 (Abcam, ab74065); anti-Neurog2 (Santa Cruz, SC-19233); anti-HA (Abcam, ab9110); anti-Brn2 (Santa Cruz, SC-6029); anti-Ebf2 (R&D, AF7006); anti-Onecut2 (R&D, AF6294); and anti-H3K27ac (Abcam, ab4729). Washes were done subsequently with 1× with sonication buffer (cold), sonication buffer with 500 nM NaCl (cold), LiCl wash buffer (20 mM Tris-HCl (pH 8.0)) (cold), 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate) (cold), and TE buffer (10 mM Tris, 1 mM EDTA, (pH 8)) (cold). Elution was done by adding elution buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 1% SDS) and incubating for 45 min at 65 °C. Eluted sample and input (sonicated, not chromatin from ChIP assays) were incubated overnight at 65 °C to reverse the crosslink. RNA was digested by the addition of 0.2 mg ml^{-1} RNase A (Sigma) and incubating for 2 h at 37 °C. Protein digestion was performed by adding 0.2 mg ml^{-1} Proteinase K (Invitrogen) for 30 min at 55 °C. Phenol:chloroform:isoamyl alcohol (25:24:1; vol/vol) (Invitrogen) followed by ethanol precipitation were used for DNA extraction. The pellets were suspended in water and one-third of ChIP DNA (1:100 dilution of input DNA) was used to prepare Illumina DNA sequencing libraries. Bioo Scientific multiplexed adapters were ligated after end repair and

A-tailing, and unligated adapters were removed by purification using Agencourt AmpureXP beads (Beckman Coulter). Adapter-ligated DNA was amplified by PCR using TruSeq primers (Sigma). DNA libraries between 300 and 500 bp in size were purified from agarose gel using a Qiagen minElute column, and the final quantification of the library before pooling was done using a KAPA library amplification kit (Roche Lightcycler 480). The libraries were sequenced on an Illumina NextSeq 500 using V2 chemistry for 50 cycles (single-end) and 75 cycles (single-end) at the genomics core facility at NYU.

ATAC-seq. A total of 50,000 cells were harvested and washed twice in cold 1× PBS. Cells were resuspended in 10 mM Tris (pH 7.4), 10 mM NaCl, 3 mM MgCl2, and 0.1% NP-40, and centrifuged immediately at 4 °C. The pellet was resuspended in 25 µl of 2× TD buffer, 2.5 µl TDE1 (Nextera DNA sample preparation kit, FC-121–1030) followed by incubation for 30 min at 37 °C. The reaction was then cleaned using a Min-elute PCR purification kit (Qiagen, 28004). The optimal number of PCR cycles were determined to be the one-third of the maximum fluorescence measured by quantitative PCR reaction with 1× SYBR Green (Invitrogen), custom-designed primers[56] and 2× NEB MasterMix (New England Labs, M0541). Following PCR enrichment, the library was cleaned using the Min-elute PCR kit and quantified using Qubit (Life Technologies, Q32854). The fragment length distribution of the library was determined using an Agilent High Sensitivity DNA D1000 Screentape (5067–5585) system, and the final quantification of the library before pooling was done using a KAPA library amplification kit (Roche Lightcycler 480). The libraries were sequenced on an Illumina NextSeq 500 using V2 chemistry for 150 cycles (paired-end 75 bp) at the genomics core facility at NYU.

Quantification and statistical analysis. *RNA-seq data analysis.* All RNA-seq fastq files were aligned to the mouse genome (version mm10) using Tophat (v.2.1.1)[57] with options "-r 100–no-coverage-search". Rsubread[58], an R package, was used to assign reads to genes defined using Refseq[59] mm10 gene annotations. The Wald test in the DESeq2 package[60] was used for differential gene expression analysis. A *q*-value cut-off of less than 0.01 was used for calling differentially expressed genes. PANTHER (v.13.1) (http://pantherdb.org) was used to perform gene ontology term enrichment analysis.

scRNA-seq data processing. Fastq files were generated by using CellRanger (v.2.1.0) from 10× Genomics with default settings (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger). We added the transgene sequences to the reference genome manually to distinguish the two pooled cell lines: V5 (iAscl1) and GFP (iNeurog2 Tubb3::GFP) exogenous sequences were added to the end of chromosome 1 in FastA and GTF files of the mouse reference genome (mm10). A custom reference genome was generated using the CellRanger mkref function by passing the modified FastA and GTF files. CellRanger count function was used to generate single cell feature counts for the library. Downstream analyses and graph visualizations were performed in Seurat R package[61] (v.2.3.4). Briefly, we removed the cells that have unique gene counts greater than 6,800 (potential doublets) and less than 200. After removing the unwanted cells, we normalized the data by a global-scaling normalization method (logNormalize) with the default scale factor (10,000). Linear dimensional reduction was performed by PCA, and the clustering was performed by using the statistically significant principal components (identified using the jackStraw method and by the standard deviation of principle components). The results were visualized using t-distributed stochastic neighbor embedding (t-SNE) plots.

ChIP-seq data processing. All ChIP-seq fastq files were aligned to the mouse genome (version mm10) using Bowtie (1.0.1)[62] with options "-q–best–strata -m 1–chunkmbs 1024". Only uniquely mapped reads were considered for further analysis. MultiGPS (v.0.74) was used to define TF DNA-binding events[63]. Cut-offs of fold enrichment values ≥1.5 and $q < 0.01$ (assessed using binomial tests and Benjamini–Hochberg multiple hypothesis test correction) were used to call statistically significant binding events. Differential binding analysis between proneural TFs (Ascl1 versus Neurog2), between time points (12 h versus 48 h), or between factor inductions for the downstream TFs (iAscl1 versus iNeurog2) was also performed using MultiGPS, which calls EdgeR[64] internally. Differentially bound sites were defined as those that displayed significantly greater read enrichment levels ($q < 0.01$) as determined by EdgeR's negative binomial generalized linear models applied to MultiGPS' per-replicate count data (TMM normalized). Shared binding events were defined as those that were called in both conditions, and not displaying significant differences in read enrichment level. To account for differences in the numbers of peaks called for Neurog2 and Ascl1, some analyses of differential and shared binding were restricted to the top 10,000 most ChIP-enriched binding events for each of those TFs. When comparing binding site locations across distinct TF classes (for example, Fig. 6a–c), we used a window size of 200 bp to define overlapping sites.

ATAC-seq data processing. All ATAC-seq data were mapped to the mouse genome (version mm10) using Bowtie2-2.2.2 (ref. [64]) using "-q–very-sensitive" options. Enriched domains were identified using the DomainFinder module in SeqCode (https://github.com/seqcode/seqcode-core/blob/master/src/org/seqcode/projects/seed/DomainFinder.java). Briefly, contiguous 50-bp genomic bins with significantly higher read enrichment compared with normalized input were identified (binomial test, $P < 0.05$). Furthermore, contiguous blocks within 200 bp were joined together to call enriched domains. Differential ATAC-seq analysis was performed by first merging accessible domains across compared conditions (bedtools v.2.26.0: merge function with parameter -d100), counting ATAC-seq reads from each replicate that overlapped the merged domains, and performing differential enrichment analysis with EdgeR[64] (v.3.24, thresholds: twofold, $P < 0.01$ (EdgeR's negative binomial generalized linear models)).

Defining 0 h 'active' and 'inactive' regions. A random forest classifier was trained to classify binding event locations as either being active or inactive at the 0 h time point (EBs). The classifier was trained using H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K27me3, and ATAC-seq windowed read-enrichment as predictors. A union list of ~300,000 500-bp regions comprising the enriched domains (see above) of H3K4me1, H3K4me2, H3K4me3, H3K27ac, and ATAC-seq was used as the positive set for training the classifier. An equal number of unmarked 500-bp regions were randomly selected and used as the negative set for training the classifier. Weka's implementation of Random Forests was used to train the classifier (https://github.com/seqcode/seqcode-core/blob/master/src/org/seqcode/ml/classification/BaggedRandomForest.java). Briefly, the forest contained 10,000 trees. Each tree was trained with ten randomly sampled features on 1% bootstrapped samples of the entire dataset. Every binding event that was predicted to be in active 0 h chromatin with a probability of greater than 0.8 was placed in the 'active' class, while the remaining events were placed in the 'inactive' class.

De novo motif discovery and k-mer analysis. MEME-ChIP (MEME suite v.4.11.3)[65] was run on each of the subsets of Ascl1, Neurog2, Brn2, Ebf2, and Onecut2 binding sites using parameters "-meme-mod zoops -meme-minw 6 -meme-maxw 20", and default parameters otherwise. Primary motif finding analyses (for example, Fig. 2c) were performed on 50-bp windows centered on the MultiGPS-defined binding event locations. Motif-finding analysis that aimed to find both primary and secondary motif signals (for example, Fig. 6d,e) were performed on 150-bp windows centered on the MultiGPS-defined binding event locations. Motif distribution plots (Fig. 6d,e) were produced using the Centrimo function of MEME-ChIP.

SeqUnwinder[66] was used for label-specific de novo motif discovery. Briefly, all *k*-mers with lengths of 4 and 5 were used as predictors. The SeqUnwinder classifier was trained to predict iAscl1-specific, iNeurog2-specific, and shared binding events. The heatmaps associating discovered motifs with each label were produced using SeqUnwinder.

For flanking *k*-mer analysis, we started with all possible 8-mers with the following restrictions: the 8-mers were restricted to contain the CAGNTG 6-mer subsequence and the remaining 2 characters were picked from the following set {A, T, G, C, N}. These restrictions resulted in a total of 150 8-mers. We used these 150 8-mers as predictors for a logistic regression classifier with L1 regularization. The classifier was trained on Ascl1- and Neurog2-specific binding sites. All non-zero weighted 8-mers were used for further analysis.

DNA shape properties around Ascl1 and Neurog2 sites were calculated using the DNAshapeR R package[67] (v.1.10.0).

TF binding site and ATAC-seq heatmaps. The MetaMaker program from the SeqCode project was used to generate heatmaps (https://github.com/seqcode/seqcode-core/blob/master/src/org/seqcode/viz/metaprofile/MetaMaker.java). Briefly, each row in a heatmap represents a 1,000-bp window centered on the midpoint of a TF binding event. Reads were extended to 100 bp, and overlapping read counts were binned into 10-bp bins. Color shading between white and a maximum color were used to represent the depth of read coverage in each heatmap. We used a systematic approach to choose the read depth represented by the maximum color for each track. We first calculated the read counts in 10-bp bins at all identified binding sites for the given TFt and then used the 95th percentile value as the maximum value for the color palette. The following are the read depths represented by the maximum color for different heatmaps: Ascl1:–*linemin 15–linemax 70*; Neurog2:–*linemin 15–linemax 95*; A[N]^{bHLH}:–*linemin 15–linemax 90*; Brn2 (iAscl1):–*linemin 10–linemax 99*; Brn2 (iNeurog2):–*linemin 10–linemax 67*; Ebf2 (iAscl1):–*linemin 5–linemax 76*; Ebf2 (iNeurog2):–*linemin 5 –linemax 106*; Onecut2 (iAscl1):–*linemin 5 –linemax 76;* Onecut2 (iNeurog2):–*linemin 5 –linemax 128*. H3K27ac (EB)–*linemin 15–linemax 100*; H3K27ac (iAscl1)–*linemin 10–linemax 75*; H3K27ac (iNeurog2)–*linemin 10–linemax 55*. ATAC-seq (EB)–*linemin 10–linemax 81*; ATAC-seq (iAscl1 12 h)–*linemin 10–linemax 46*; ATAC-seq (iAscl1 48 h)–*linemin 10–linemax 53*; ATAC-seq (iNeurog2 12 h)–*linemin 10–linemax 35*; ATAC-seq (iNeurog2 48 h)–*linemin 10–linemax 44*; ATAC-seq (iA[N]bHLH 12 h)–*linemin 10–linemax 33*; and ATAC-seq (iA[N] bHLH 48 h)–*linemin 10–linemax 27*.

Browser snapshots. The ChipSeqFigureMaker program from the SeqCode project was used to generate the browser shots. (https://github.com/seqcode/seqcode-core/blob/master/src/org/seqcode/viz/genomicplot/ChipSeqFigureMaker.java). Reads from both strands were merged and extended to 100 bp. The colors of the tracks were matched to the colors of the TF heatmaps.

Ascl1 and Neurog2 binding site comparison. All Ascl1 and Neurog2 binding site comparative analyses were restricted to the top 10,000 binding events. The binding events were sorted based on the *q* value indicating significant enrichment over input ChIP-seq experiments. All top 10,000 Ascl1 binding events that showed significantly differential higher ($q < 0.01$, EdgeR's negative binomial generalized linear models) ChIP enrichment over Neurog2 ChIP were defined as Ascl1-preferred or Ascl1 > Neurog2 binding sites. Similarly, all top 10,000 Neurog2 binding events that showed significantly differential higher ($q < 0.01$, EdgeR's negative binomial generalized linear models) ChIP enrichment over Ascl1 ChIP were defined as Neurog2-preferred or Neurog2 > Ascl1 binding sites. All binding events in the top 10,000 Ascl1 and Neurog2 lists, which were also not significantly enriched in either Ascl1 of Neurog2, were defined as shared or A = N sites.

Associations between differential binding sites and differential expression. The GREAT command-line tools[68] were used to define gene regulatory domains and to assess the associations between sets of binding sites and gene categories defined by the differential expression analyses. Regulatory domains were defined using the GREAT "basal plus extension" model with the following settings: basalUpstream = 5,000, basalDownstream = 1,000, maxExtension = 100,000. The gene sets shown in Fig. 7 represent genes that were significantly upregulated in both iA and iN cells compared with EBs (iA = iN), and genes that were significantly differentially expressed between iA and iN cells (iA > iN and iN > iA) for each relevant time point.

Sample size and statistical analysis. No statistical methods were used to predetermine sample sizes, but our sample sizes were similar to those reported in previous publications[69,70]. Data collection and analysis were not performed blind to the conditions of the experiments. Biologically independent cell differentiations were used as replicates.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data (RNA–seq, ChIP–seq, ATAC–seq, and scRNA–seq) produced for this study are available from the GEO database under accession GSE114176. We performed a re-analysis of data sourced from GEO database entries GSE101397, GSE97715 and GSE43916.

Code availability

Analysis scripts are available at https://github.com/seqcode/Aydin_2019_iAscl1-vs-iNeurog2.

References

51. Iacovino, M. et al. Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* **29**, 1580–1588 (2011).
52. Mazzoni, E. O. et al. Embryonic stem cell-based mapping of developmental transcriptional programs. *Nat. Methods* **8**, 1056–1058 (2011).
53. Zappulo, A. et al. RNA localization is a key determinant of neurite-enriched proteome. *Nat. Commun.* **8**, 583 (2017).
54. Groth, R. D., Lindskog, M., Thiagarajan, T. C., Li, L. & Tsien, R. W. Ca^{2+}/CaM-dependent kinase type II triggers upregulation of GluA1 to coordinate adaptation to synaptic inactivity in hippocampal neurons. *Proc. Natl Acad. Sci. USA* **108**, 828–833 (2011).
55. Bootman, M. D., Rietdorf, K., Collins, T., Walker, S. & Sanderson, M. Loading fluorescent Ca^{2+} indicators into living cells. *Cold Spring Harb. Protoc.* **8**, 122–125 (2013).
56. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
57. Kim, D. et al. TopHat2 : accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
58. Liao, Y. & Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
59. O'Leary, N. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 2832 (2014).
61. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411 (2018).
62. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, 1–10 (2009).
63. Mahony, S. et al. An integrated model of multiple-condition ChIP-Seq data reveals predeterminants of Cdx2 binding. *PLoS Comput. Biol.* **10**, e1003501 (2014).
64. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
65. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
66. Kakumanu, A., Velasco, S., Mazzoni, E. O. & Mahony, S. Deconvolving sequence features that discriminate between overlapping regulatory annotations. *PLoS Comput. Biol.* **13**, 1–22 (2017).
67. Chiu, T. P. et al. DNAshapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics* **32**, 1211–1213 (2016).
68. McLean, C. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, nbt.1630 (2010).
69. Velasco, S. et al. A multi-step transcriptional and chromatin state cascade underlies motor neuron programming from embryonic stem cells. *Cell Stem Cell* **20**, 205–217.e8 (2017).
70. Mazzoni, E. O. et al. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat. Neurosci.* **16**, 1219–1227 (2013).

Corresponding author(s): Esteban O. Mazzoni
Shaun Mahony

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	RNA-seq, ChIP-seq, ATAC-seq, and sc-RNA-seq data was collected by Illumina NextSeq 500 sequencer. Confocal images of embryoid bodies and neurons were taken by Leica SP5 Confocal microscope. Live Acquisition software package (Till Photonics) was used to collect calcium imaging data.

Data analysis	SeqUnwinder Kakumanu et al., 2017 GREAT McLean et al., 2010 MultiGPS Mahony et al., 2014 DNAShapeR Chiu et al., 2016 Bowtie Langmead et al., 2009 DESeq2 Love et al., 2014 Tophat Kim et al., 2013 Rsubread Liao et al., 2013 Matlab- MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States. Panther (version 13.1) - http://pantherdb.org Bedtools - Quinlan et al., 2016 EdgeR - Robinson et al. 2009 RefSeq - O'Leary et al. 2016

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data produced for this study are available under GSE114176.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications (Velasco et al., 2017; Mazzoni et al., 2013). Biologically independent cell differentiations were used as replicates. We used 3 independent biological replicates (n=3) for Ascl1 and Neurog2 ChIP-seq experiments. The rest of the ChIP-seq experiments, all ATAC-seq experiments, antibody stainings, and calcium imaging were performed in 2 independent biological replicates. Most of the RNA-seq experiments were performed in 2 biological replicates except iAscl1 48 (n=5) and EB t0 (n=5) because of the availability of additional replicates performed in the past for other studies.
Data exclusions	None
Replication	All attempts at replication were successful. For all experiments, replication was achieved through multiple independent cell differentiations.
Randomization	Experiments were not randomized and we did not perform any in vivo studies. We compared control and test samples through treatment and genetic information.
Blinding	The blinding of investigators were not performed but different people analyzed the same data.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Mouse monoclonal anti-V5_Thermo Fisher Scientific_Cat# R960-25_RRID:AB_2556564_1:5000 dilution
Rabbit polyclonal anti-β3-TUBULIN_Sigma_Cat# T2200_RRID:AB_262133_1:2000 dilution.
Sheep polyclonal anti-EBF2_R&D systems_Cat# AF7006_RRID:AB_10972102_5 ug for ChIP
Sheep polyclonal anti-ONECUT2_R&D systems_Cat# AF6294_RRID:AB_10640365_5 ug for ChIP
Goat polyclonal anti-NGN2_Santa Cruz_Cat# SC-19233_RRID:AB_2149513_5 ug for ChIP
Rabbit polyclonal anti-MASH1 (Ascl1)_abcam_Cat# ab74065_RRID:AB_1859937_5 ug for ChIP
Rabbit polyclonal anti-HA_abcam_Cat# ab9110_RRID:AB_307019_1:5000 dilution
Mouse monoclonal anti-FLAG_Sigma_Cat# F1804_RRID:AB_262044_1:500 dilution
Chicken polyclonal anti-MAP2_abcam_Cat# ab5392_RRID:AB_2138153_1:1000 dilution
Mouse monoclonal anti-NEUROFILAMENT_DSHB_Cat# 2H3_RRID:AB_531793_1:1000 dilution
Goat polyclonal anti-BRN2_Santa Cruz_Cat# SC-6029_RRID: AB_2167385_5 ug for ChIP
Rabbit polyclonal anti-H3K27ac_abcam_Cat# ab4729_RRID:AB_2118291_5 ug for ChIP
Goat anti-chicken 488 AlexaFluor_Invitrogen_Cat# A-11039_RRID:AB_142924_1:1000 dilution
Goat anti-mouse 647 Alexa Fluor_Invitrogen_Cat# A-21236_RRID:AB_141725_1:1000 dilution
Goat anti-rabbit 568 Alexa Fluor_Invitrogen_Cat# A-11036_RRID:AB_143011_1:1000 dilution

Validation

Antibodies are validated by using uninduced (mouse embryonic stem cells) cells as negative controls and comparison of the antibody staining/ChIP with the ones performed with tagged proteins. Additional validation and reports from the manufacturer and peer-reviewed articles can be found on manufacturer's websites.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Inducible cell lines were generated using the inducible cassette exchange (ICE) method that was previously described in Iacovino, M. et al. Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. Stem Cells 29, 1580–1588 (2011).

Authentication

We genotyped the cell lines.

Mycoplasma contamination

These cell lines used in the study were not tested for mycoplasma. However, the parental cell line that was used to generate these inducible lines are routinely checked in the lab and have not been found positive for mycoplasma.

Commonly misidentified lines
(See ICLAC register)

None of the lines used were misidentified.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

To review GEO accession GSE114176: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114176

Files in database submission

HL7T3BGXY_n01_ba48.fastq.gz
HL7T3BGXY_n01_ba49.fastq.gz
HFW7KBGXY_n01_ba63.fastq.gz
HJY7YBGX2_n01_ba65.fastq.gz
C7KY3ACXX_l08n01_sv106.351000000040810.fastq.gz
C7KY3ACXX_l08n01_sv107.351000000040897.fastq.gz
HFW7KBGXY_n01_ba64.fastq.gz
C7KY3ACXX_l08n01_sv110.351000000409f3.fastq.gz
HJY7YBGX2_n01_ba77.fastq.gz
HJY7YBGX2_n01_ba78.fastq.gz

HJY7YBGX2_n01_ba76.fastq.gz
bgm3_ACAGTG_L002_R1_001.fastq.gz
C5WCJACXX_l06n01_bgm6.3410000000bfd5.fastq.gz
H7KNHBGX3_n01_ba83.fastq.gz
H7KNHBGX3_n01_ba84.fastq.gz
H3FWTBGX5_n01_ba126.fastq.gz
H3FWTBGX5_n01_ba127.fastq.gz
HL7T3BGXY_n01_ba50.fastq.gz
HFW7KBGXY_n01_ba59.fastq.gz
H3FWTBGX5_n01_ba128.fastq.gz
H3FWTBGX5_n01_ba129.fastq.gz
H7KNHBGX3_n01_ba67.fastq.gz
HL2NTBGX3_n01_ba108.fastq.gz
H7KNHBGX3_n01_ba85.fastq.gz
HL2NTBGX3_n01_ba107.fastq.gz
H3FWTBGX5_n01_ba130.fastq.gz
H3FWTBGX5_n01_ba131.fastq.gz
HL7T3BGXY_n01_ba51.fastq.gz
HFW7KBGXY_n01_ba61.fastq.gz
H3FWTBGX5_n01_ba133.fastq.gz
H3FWTBGX5_n01_ba134.fastq.gz

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

12h Ascl1 and Neurog2 ChIP-seq experiments were performed in 3 biological replicates. Rest of the ChIP-seq experiments are performed in 2 biological replicates.

Sequencing depth

File(s) Read length Paired or single-end seq Sequenced reads Uniquely mapped reads
HL7T3BGXY_n01_ba48.fastq.gz 51 single 41214096 30324495
HL7T3BGXY_n01_ba49.fastq.gz 51 single 37113165 27261442
HFW7KBGXY_n01_ba63.fastq.gz 76 single 34172197 26232594
HJY7YBGX2_n01_ba65.fastq.gz 50 single 47389148 26908039
C7KY3ACXX_l08n01_sv106.35100000040810.fastq.gz 51 single 42745488 32304763
C7KY3ACXX_l08n01_sv107.35100000040897.fastq.gz 51 single 60891637 34923424
HFW7KBGXY_n01_ba64.fastq.gz 76 single 32221513 25339660
C7KY3ACXX_l08n01_sv110.351000000409f3.fastq.gz 51 single 21141657 15391085
HJY7YBGX2_n01_ba77.fastq.gz 50 single 47382123 36758251
HJY7YBGX2_n01_ba78.fastq.gz 50 single 41435336 33111788
HJY7YBGX2_n01_ba76.fastq.gz 50 single 33191686 26401326
bgm3_ACAGTG_L002_R1_001.fastq.gz 51 single 29254740 22068673
C5WCJACXX_l06n01_bgm6.3410000000bfd5.fastq.gz 51 single 52494545 39539853
H7KNHBGX3_n01_ba83.fastq.gz 75 single 39374378 30086226
H7KNHBGX3_n01_ba84.fastq.gz 75 single 40159227 30518273
H3FWTBGX5_n01_ba126.fastq.gz 75 single 35918702 29038259
H3FWTBGX5_n01_ba127.fastq.gz 75 single 29446502 22765267
HL7T3BGXY_n01_ba50.fastq.gz 51 single 30042940 25695212
HFW7KBGXY_n01_ba59.fastq.gz 76 single 22333815 18954309
H3FWTBGX5_n01_ba128.fastq.gz 75 single 33047029 26099022
H3FWTBGX5_n01_ba129.fastq.gz 75 single 35249661 26716434
H7KNHBGX3_n01_ba67.fastq.gz 75 single 35428163 28605942
HL2NTBGX3_n01_ba108.fastq.gz 75 single 33487953 26778366
H7KNHBGX3_n01_ba85.fastq.gz 75 single 44875183 34995379
HL2NTBGX3_n01_ba107.fastq.gz 75 single 39157095 29516002
H3FWTBGX5_n01_ba130.fastq.gz 75 single 36018293 28555170
H3FWTBGX5_n01_ba131.fastq.gz 75 single 26957572 23011879
HL7T3BGXY_n01_ba51.fastq.gz 51 single 34287101 29026034
HFW7KBGXY_n01_ba61.fastq.gz 76 single 37968301 32408097
H3FWTBGX5_n01_ba133.fastq.gz 75 single 33712513 25428696
H3FWTBGX5_n01_ba134.fastq.gz 75 single 37727645 30308287

Antibodies

anti-Ascl1(abcam, ab74065), anti-Neurog2 (Santa Cruz, SC-19233), anti-HA (abcam, ab9110) anti-Brn2 (Santa Cruz, SC-6029), anti-Ebf2 (R&D, AF7006), anti-Onecut2 (R&D, AF6294), anti-H3K27ac (abcam, ab4729)

Peak calling parameters

All ChIP-seq fastq files were aligned to the mouse genome (version mm10) using Bowtie (1.0.1) with options "-q --best --strata -m 1 --chunkmbs 1024". Only uniquely mapped reads were considered for further analysis. MultiGPS was used to define transcription factor DNA binding events. A q-value cutoff of 0.01 (assessed using binomial tests and Benjamini-Hochberg multiple hypothesis test correction), was used to call statistically significant binding events. Differential binding analysis between time points (12hr vs 48hr) or factor inductions (iAscl1 vs iNeurog2) was also performed using MultiGPS, which calls EdgeR internally. Differentially bound sites are defined as those that display significantly greater read enrichment levels (minimum 1.5-fold, q-value < 0.05) as determined by EdgeR's negative binomial generalized linear models applied to MultiGPS' per-replicate count data (TMM normalized). Shared binding events are defined as those that are called in both conditions, and not displaying significant differences in read enrichment level. To account for some differences in the

	numbers of peaks called for Neurog2 and Ascl1, some analyses of differential and shared binding restrict analysis to the top 10,000 most ChIP-enriched binding events for each of those transcription factors.
Data quality	We assessed the quality of the sequencing reads using FastQC and assessed library complexity using custom scripts
Software	SeqUnwinder Kakumanu et al., 2017 ChIPEnrich Welch et al., 2014 MultiGPS Mahony et al., 2014