Cell Reports

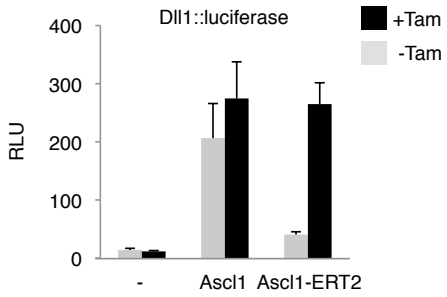
Supplemental Information

Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis

Alexandre A.S.F. Raposo, Francisca F. Vasconcelos, Daniela Drechsel, Corentine Marie, Caroline Johnston, Dirk Dolle, Angela Bithell, Sébastien Gillotin, Debbie L.C. van den Berg, Laurence Ettwiller, Paul Flicek, Gregory E. Crawford, Carlos M. Parras, Benedikt Berninger, Noel J. Buckley, François Guillemot, and Diogo S. Castro

Supplemental Data

A



B

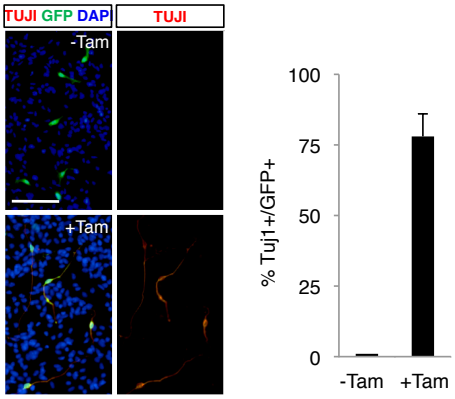


Figure S1. Generation of an inducible version of Ascl1 (related to Figure 1).
A. Induction of a *Dll1* enhancer by Ascl1 and Ascl1-ERT2 in a reporter gene assay in transfected NS5 cells, in the presence and absence of tamoxifen.
B. Differentiation of NS5 cells transduced with an adenovirus vector expressing Ascl1-ERT2, assessed by expression of the neuronal marker TUJ1. Infected cells are labeled with GFP (green). Values: mean±SD. Scale bar: 200μm.

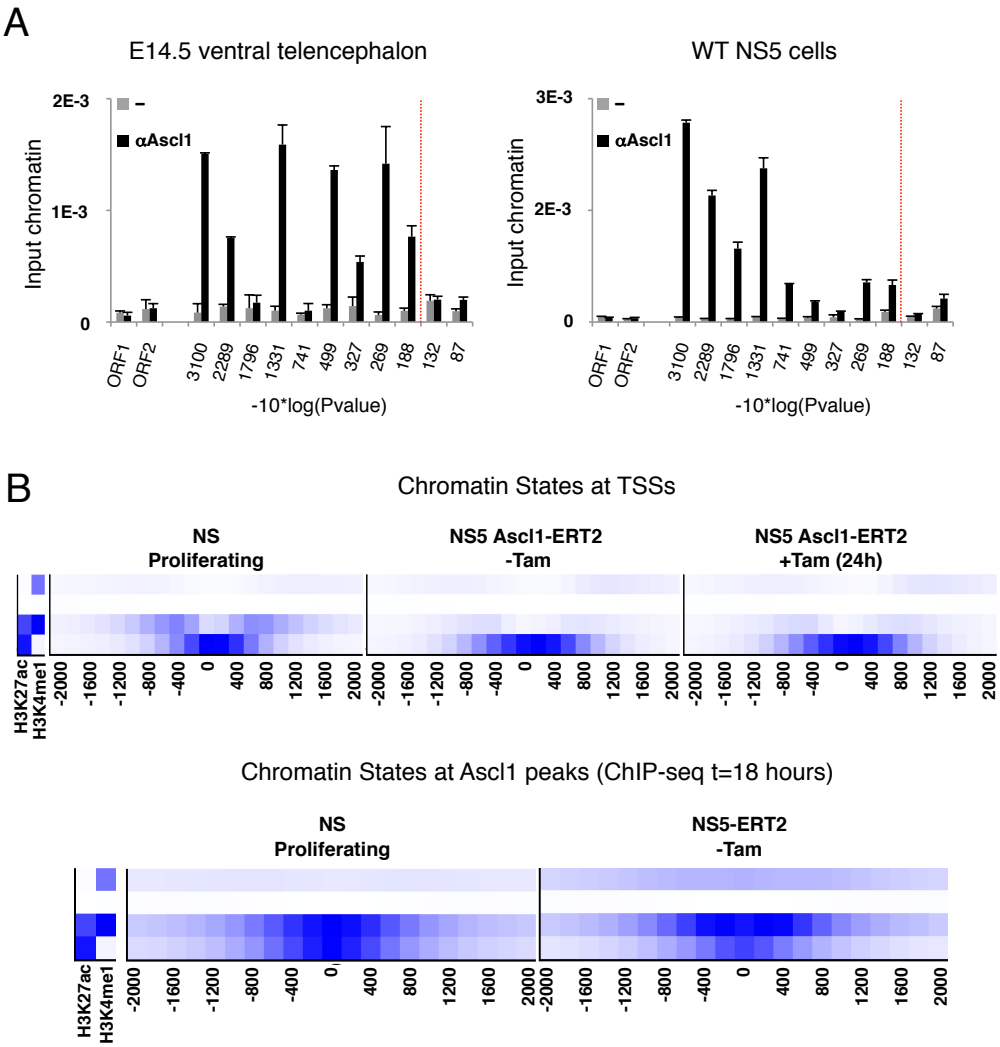


Figure S2. ChIP-PCR validation of genome-wide Ascl1 binding and characterization of chromatin states by ChromHMM in NSCs (related to Figure 2).

A. Validation of Ascl1-ERT2 binding events at various P values by ChIP-PCR against wild type Ascl, using chromatin extracted from embryonic E14.5 ventral telencephalic progenitors (left), and wild type proliferating NS5 cells (right). Red line marks threshold limit for validation ($P<10^{-18}$). Values: mean±SD.

B. *Top*: Heat maps of chromatin states for H3K27ac and H3K4me1 within ±2kb of TSSs in proliferating NS cells (*NS*, data sets from Creyghton, 2010; and Stadler, 2011) and non-induced NS5 Ascl1-ERT2 cells (*-Tam*), or in differentiating NS5 Ascl1-ERT2 cells 24 hours after induction (*+Tam*). *Bottom*: Heat maps of chromatin states for H3K27ac and H3K4me1 within ±2kb of Ascl1-ERT2 (peaks at t=18h), determined by ChromHMM in proliferating NS cells (*NS*, data sets from Creyghton, 2010; and Stadler, 2011) and in non-induced NS5 Ascl1-ERT2 cells (*-Tam*).

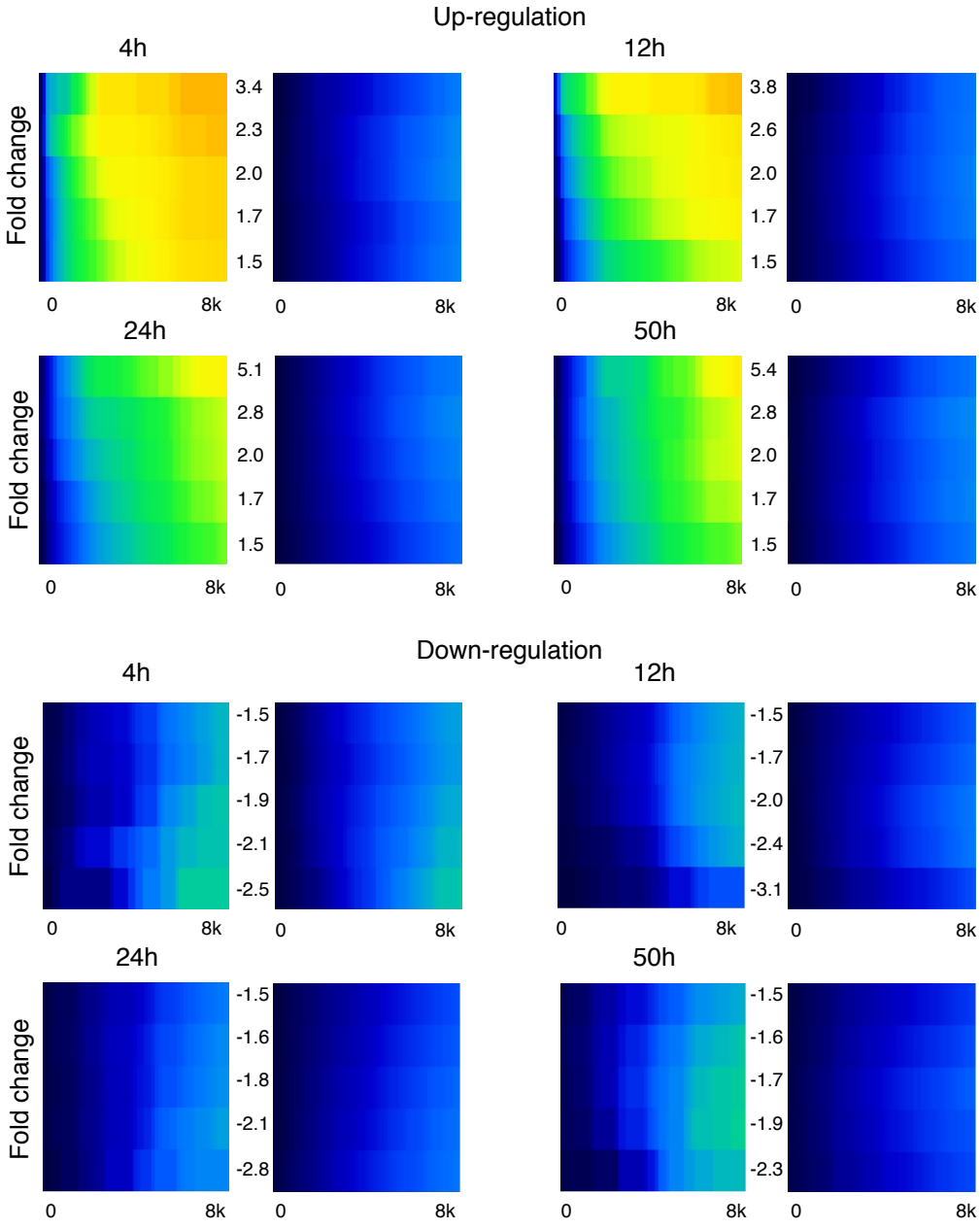


Figure S3. Ascl1 functions as a transcriptional activator at a genome-wide level (related to Figure 3). Heat maps displaying the cumulative fraction of deregulated genes at 4, 12, 24, and 50 hours after induction of differentiation. Bins defined as in main figure 3: deregulated transcripts (up – top panels; down – bottom panels) are divided in bins of equal size, with indicated fold change cutoffs. Ascl1 BEs are divided in bins of 166 peaks, with increasing P value. As control, comparisons were made against 100 randomized sets of binding events (right panels, where the average fraction is shown).

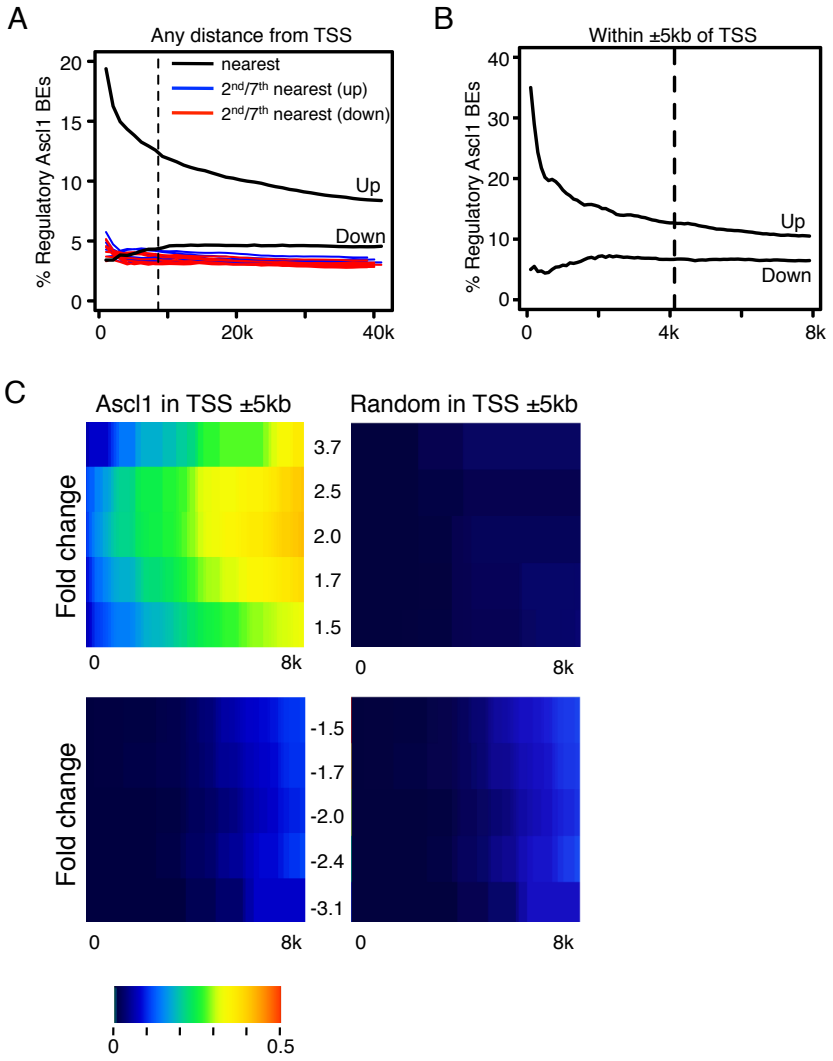


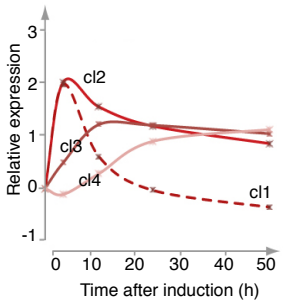
Figure S4. Integration of binding and regulation data using alternative annotation methods (related to Figure 3).

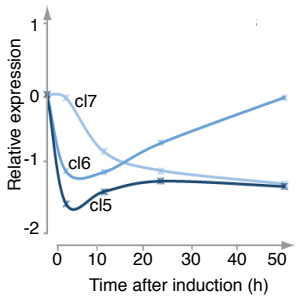
A. Fraction of Ascl1 BEs considered regulatory when peaks are annotated to the nearest TSSs (black), or 2nd to 7th nearest TSSs (red/blue).

B. Fraction of Ascl1 BEs considered regulatory when peaks are annotated to the nearest TSSs (maximum 5kb distance). Bin=100 BEs, dashed line: $P< 10^{-25}$.

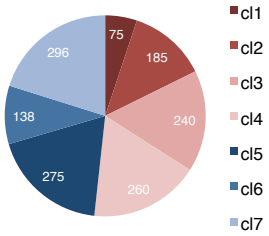
C. Heat maps displaying the cumulative fraction of deregulated genes that are directly bound by Ascl1 at 12 hours after induction of differentiation (nearest gene annotation with maximum 5kb distance). Bins defined as in main figure 3: deregulated transcripts (up – top panels; down – bottom panels) are divided in bins of equal size, with indicated fold change cutoffs. Ascl1 BEs are divided in bins of 166 peaks, with increasing P value. As control, comparisons were made against 100 randomized sets of binding events (right panels, where the average fraction is shown).

A





B



C

Cluster 2	P value
Cell communication	1.37E-3
Signal transduction	2.10E-3
Phospholipid metabolic process	2.72E-3
Small GTPase regulator activity	9.30E-4

Cluster 3	
Transcription	2.09E-5
Transcription factor activity	1.56E-4
Establishment or maintenance of chromatin architecture	7.55E-3
Transcription cofactor activity	7.02E-3
Notch signaling pathway	1.90E-3

Cluster 4	
Structural constituent of cytoskeleton	3.05E-5
Notch signaling pathway	1.29E-4
Nervous system development	3.92E-4
Developmental process	7.88E-4
Transmembrane receptor protein kinase activity	2.10E-3
Axon guidance mediated by Slit/Robo	4.62E-3

Figure S5. Temporal patterning of the Ascl1 transcriptional program (related to Figure 3).

A. Average profile of temporal clusters of genes associated with gene activation (red, left) or repression (blue, right).

B. Composition of clusters according to number of genes.

C. Enrichment of Gene Ontology biological process terms, amongst direct Ascl1 target genes that belong to clusters 2-4, and associated P value.

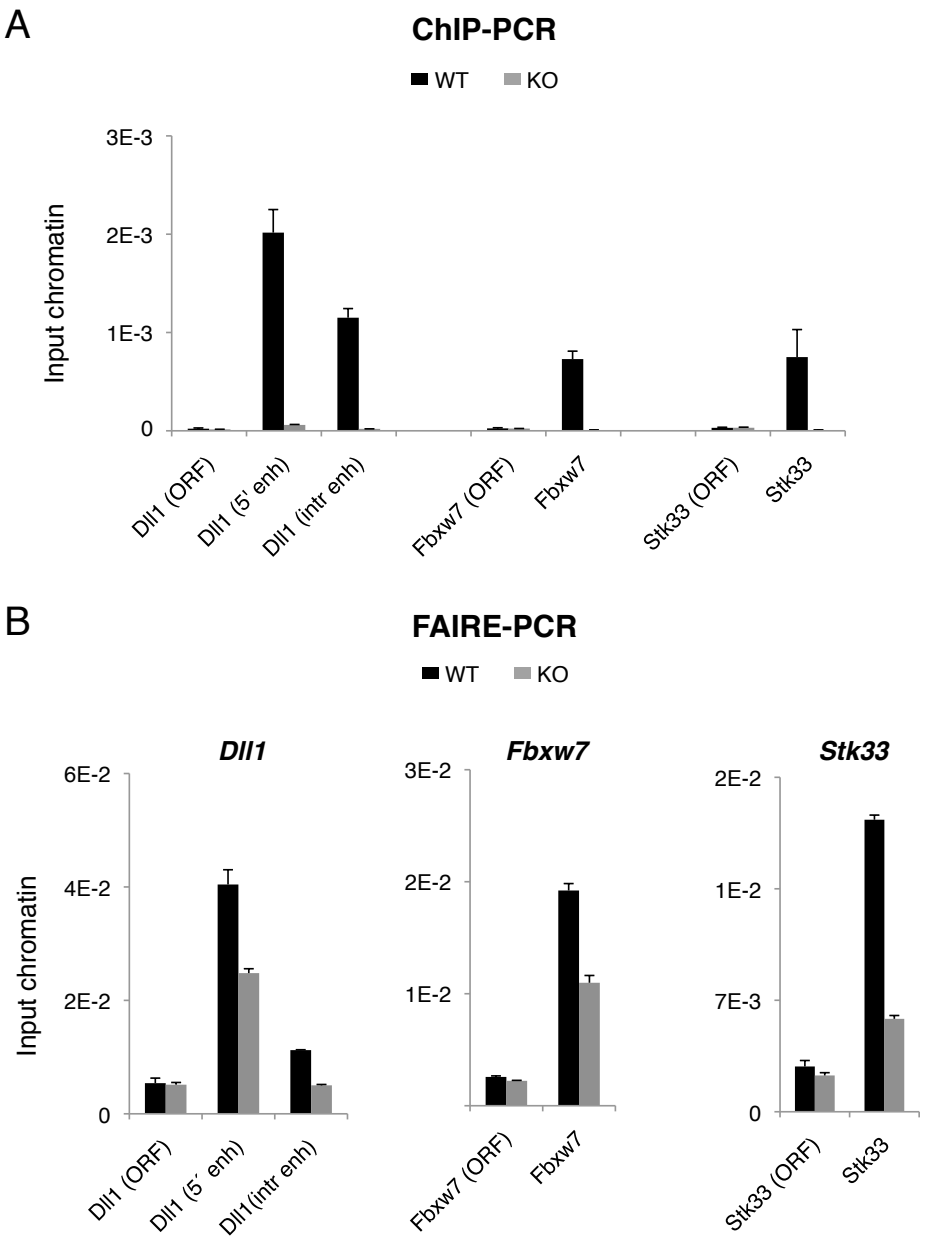


Figure S6. Analysis of Ascl1 bound regions by ChIP- and FAIRE-PCR in chromatin from Ascl1 null embryos (related to Figure 6).

A. ChIP-PCR with αAscl1 for Ascl1-bound regions associated with *Dll1*, *Fbxw7*, and *Stk33* genes, using chromatin extracted from ventral telencephalon of E14.5 wild type (WT) or Ascl1 null embryos (KO). Values: mean±SD.

B. Quantification of nucleosome-depleted chromatin by FAIRE-PCR at Ascl1-bound regions, using chromatin extracted from ventral telencephalon of E14.5 wild type (WT) or Ascl1 null embryos (KO).

In parallel with Ascl1-bound sites, one negative control region within the gene open reading frame (ORF) was tested for each locus. Values: mean±SD.

Supplemental Tables

S1: Expression profiling data set, related to Figure 1

S2: List of Ascl1 binding events at t=18h, related to Figure 2

S3: Segregation of deregulated genes in clusters, related to Figure 3

S4: List of Ascl1 direct targets, related to Figure 3

S5: DHSs in proliferating NS, related to Figure 5

S6: DHSs in differentiated NS, related to Figure 5

S7: DHS specific to differentiated NS, related to Figure 5

S8: Occupied sites identified by DGF, related to Figure 5

S9: List of Ascl1 binding events at t=30', related to Figure 4

S10: Oligonucleotides used in this study, related to Experimental Procedures

Datasets generated in this study

The dataset ChIP-seq Ascl1-ERT2 (t=30') was submitted to the ArrayExpress database (www.ebi.ac.uk/arrayexpress) and are publicly available under the accession number E-MTAB-2384.

The DNase-seq datasets were submitted to the ArrayExpress database (www.ebi.ac.uk/arrayexpress) and are publicly available under the accession number E-MTAB-2270.

The datasets ChIP-seq for H3K4me1 and H3K27ac were submitted to the ArrayExpress database (www.ebi.ac.uk/arrayexpress) and are publicly available under the accession number E-MTAB-3104.

Publicly available datasets used

ChIP-seq Ascl1-ERT2, t=18h: Walpinski OL et al, SRX323564, GSM1187228, GSE43916, PMID:24243019

ChIP-seq Ascl1-ERT2, input: Walpinski OL et al, SRX323563, GSM1187227, GSE43916, PMID:24243019

ChIP-seq H3K4me1: Stadler, M.B. et al. SRX095620, PMID: 22170606.

ChIP-seq H3K27ac: Creyghton, M.P. GSE24164, GSM94585, PMID: 21106759.

Supplemental Experimental Procedures

Reporter gene assays

Reporter gene assays were performed as previously described (Castro et al., 2006), after transfection with Lipofectamine Plus (Invitrogen). DeltaM-Luc encodes the Ascl1 specific enhancer of the mDll1 gene in frame with luciferase and beta-Glob minimal promoter (Castro et al., 2006).

List of primary antibodies used

Primary antibodies used were against Tuj1 (MMS-435P, Babco), Gadd65/67 (G5163, Sigma), GFP (ab16901, Chemicon), Ascl1 (556604, BD Pharmingen), HA-tag (ab1424, Abcam), H3K27ac (ab4729, Abcam) and H3K4me1 (ab8895, Abcam).

ChIP-PCR

Chromatin extractions and ChIP-PCR for Ascl1 in NS5 or embryonic progenitors was performed as previously described (Castro et al., 2011)

Electrophysiology studies

Electrophysiological properties of neurons derived from NS5-Ascl1-ERT2 cells were analyzed 14 d following onset of differentiation. Single perforated patch-clamp recordings were performed at room temperature with amphotericin-B (Calbiochem) for perforation (Heinrich et al., 2010). Micropipettes were made from borosilicate glass capillaries (Garner, Claremont, CA, USA). Pipettes were tip-filled with internal solution and back-filled with internal solution containing 200 μg/mL amphotericin-B. The electrodes had resistances of 2–2.5 MΩ. The internal solution contained 136.5 mM K-gluconate, 17.5 mM KCl, 9 mM NaCl, 1 mM $MgCl_2$, 10 mM HEPES, and 0.2 mM EGTA (pH 7.4) at an osmolarity of 300 mOsm. The external solution contained 150 mM NaCl, 3 mM KCl, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 10 mM HEPES, and 5 mM glucose (pH 7.4) at an osmolarity of 310 mOsm. The recording chamber was continuously perfused with external solution at a rate of 0.5 mL/min. GFP-positive cells were visualized with an epifluorescence microscope (Axioskop2, Carl Zeiss) equipped with the appropriate filter sets. Digital pictures of the recorded cells were acquired using a digital camera (AxioCam, Carl Zeiss). Signals were sampled at 10 kHz with an Axopatch 200B patch-clamp amplifier (Axon Instruments, Foster City, CA, USA), filtered at 5 kHz and analyzed with Clampfit 9.2 software (Axon Instruments). For assessing a cell's excitability, cells received hyper- and depolarizing step-current injections in current clamp mode.

Expression profiling data analysis

Data analysis was performed using GeneSpring X software (Agilent). Normalization was done as follows: (i) Threshold raw signals were set to 1.0. (ii) 75th percentile normalization was chosen as normalized algorithm. (iii) Baseline was transformed to median of all samples. For significance analysis, one-way ANOVA with subsequent P value correction by Benjamini-Hochberg post-hoc correction were used. Significantly de-regulated probes were defined as probes with a corrected P value ≤0.05 and a fold change ≥1.5 relative to non-induced cells (t=0) for at least one time point in a volcano plot analysis. To assign probe sets to genes, probe coordinates previously published (Barbosa-Morais et al., 2010) were used, after annotation to the corresponding gene using ENSEMBL mm9/NCBI37 genome assembly annotation data.

Clustering of regulated genes

In order to separate distinct sets of genes with different kinetics, fuzzy c-means clustering was performed, an algorithm that relies on the same basic principles of k-means clustering. Different to K-means clustering however, fuzzy c-means does not separate entities into different clusters but instead assigns membership values for each cluster. Before clustering of expression data, variance stabilization transformation on quantile normalized “per probe” BeadStudio-output data was performed. Data for significantly deregulated probes were extracted and Z-normalized using the R/Bioconductor packages *beadarray*, *mfuzz* and *clValid*. Optimal number of clusters was pre-defined by calculating inter-cluster as well as intra-cluster connectivity of entities for a total

number of clusters ranging from 1 to 20 and chose the cluster number with the smallest intra-cluster/inter-cluster-connectivity ratio. Per-probe data was then clustered using the fuzzy c-means algorithm implemented in the cfuzz-package (parameters: number of clusters=7, fuzziness coefficient=1-25). Per-probe data was then collapsed into per-gene data using the mean membership-coefficient per cluster and gene.

Motif search and Gene Ontology analyses

We have used CisFinder (Sharov and Ko, 2009) in order to identify motifs enriched in the vicinity of Ascl1 peak summits and in DHSs. Searches were run against a control background of 100 bp genomic regions located 3kb upstream input regions, using default parameters. Gene ontology analysis of gene clusters was performed using the Panther classification system (http://www.pantherdb.org), using genes represented in Mouse Ref-8 v2 array as reference and default parameters. In addition, cluster 4 was further analyzed with GOToolBox, (http://genome.crg.es/GOToolBox/), using Hypergeometric test and the same reference control.

ChIP-seq data analysis and integration

Sequenced reads were processed after mapping with SAMTools for format conversion and removal of PCR duplicates (Li et al., 2009). Ascl1-ERT2 data sets were subsampled where necessary to balance each other for better comparison and peak-calling accuracy (Picard tools, http://picard.sourceforge.net/). Peaks for each sample were called against the input using MACS 1.4.1 (MACS 2.1.0 for histone datasets), with P value cutoff

at 10^{-10} (Zhang et al., 2008). Subsampling of the data sets confirmed that peak calling saturation was achieved with approximately 90% of sequenced reads. Peaks were then annotated to the nearest TSS using PeakAnalyzer 1.4 (Salmon-Divon et al., 2010), and annotated from 2^{nd} to 7^{th} nearest TSS with GREAT (McLean et al., 2010). Peak overlap with expression or DHS data calculated and plotted as heat maps with R/Bioconductor packages "genomeIntervals", "gplots", and in-house developed scripts [5,6,7]. Calculation of P values for the association between binding events and each cluster of deregulation was performed by sampling the total number of genes represented in the microarray 1000 times and assuming a normal distribution. Analysis of regions within DHSs protected from cleavage for high-resolution prediction of specific binding sites occupied by TFs at DHSs was performed with the Digital Genomic Footprinting algorithm Wellington (Piper et al., 2013).

DNase-seq data analysis and integration

Sequenced reads of 20bp for both conditions, proliferating and differentiating cells, were processed as for ChIP-seq, except subsampling. DHSs for each sample were defined with MACS 1.4.1 (P value cutoff at 10^{-5}) by extending mapped reads in 60bp as an estimation for the maximum distance between two nucleosomes (linker DNA). DHS annotation and overlap with expression data and clusters of deregulated genes were performed as described for ChIP-seq.

Density plots

Overlapping (minimum 1 bp) and non-overlapping genomic regions between datasets were determined using BEDTools (Quinlan and Hall, 2010). ChIP-seq

and DNase-seq normalized tag signals were calculated using a 10bp sliding window over the ± 2kb region around each peak summit to generate the occupancy profiles (in-house developed algorithm). These were plotted as heat maps of signal density using R/Bioconductor packages (http://www.R-project.org/ and http://CRAN.R-project.org/package=gplots) or used to determine the median of occupancy around peak summits.

Supplemental References

Barbosa-Morais, N.L., Dunning, M.J., Samarajiwa, S. a, Darot, J.F.J., Ritchie, M.E., Lynch, A.G., and Tavaré, S. (2010). A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. Nucleic Acids Res. *38*, e17.

Castro, D.S., Skowronska-Krawczyk, D., Armant, O., Donaldson, I.J., Parras, C., Hunt, C., Critchley, J. a, Nguyen, L., Gossler, A., Göttgens, B., et al. (2006). Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. Dev. Cell *11*, 831–844.

Castro, D.S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., Drechsel, D., Lebel-Potter, M., Garcia, L.G., Hunt, C., et al. (2011). A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. Genes Dev. *25*, 930–945.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

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

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res. *41*, e201.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics *11*, 415.

Sharov, A. a, and Ko, M.S.H. (2009). Exhaustive search for over-represented DNA sequence motifs with CisFinder. DNA Res. *16*, 261–273.

Zhang, Y., Liu, T., Meyer, C. a, Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.