

1 **Regionalization of the nervous system requires axial allocation prior to neural**
2 **lineage commitment**

3 Vicki Metzis¹, Sebastian Steinhauser¹, Edvinas Pakanavicius¹, Mina Gouti^{1,2},
4 Despina Stamataki¹, Robin Lovell-Badge¹ Nicholas M Luscombe^{1,3,4} and James
5 Briscoe^{1,*}

6
7 ¹The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

8 ²Max-Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, Berlin 13092, Germany

9 ³UCL Genetics Institute, Department of Genetics Evolution and Environment, University
10 College London, Gower Street, London WC1E 6BT, UK

11 ⁴Okinawa Institute of Science & Technology Graduate University, 1919-1 Tancha, Onna-son,
12 Kunigami-gun, Okinawa 904-0495, Japan

13

14 *Author for correspondence: james.briscoe@crick.ac.uk

15

16

17

18 **Summary**

19 Neural induction in vertebrates generates a central nervous system that extends the
20 rostral-caudal length of the body. The prevailing view is that neural cells are initially
21 induced with anterior (forebrain) identity, with caudalising signals then converting a
22 proportion to posterior fates (spinal cord). To test this model, we used chromatin
23 accessibility assays to define how cells adopt region-specific neural fates. Together
24 with genetic and biochemical perturbations this identified a developmental time
25 window in which genome-wide chromatin remodeling events preconfigure epiblast
26 cells for neural induction. Contrary to the established model, this revealed that cells
27 commit to a regional identity before acquiring neural identity. This “primary
28 regionalization” allocates cells to anterior or posterior regions of the nervous system,
29 explaining how cranial and spinal neurons are generated at appropriate axial
30 positions. These findings prompt a revision to models of neural induction and support
31 the proposed dual evolutionary origin of the vertebrate central nervous system.

32

33

34 **Introduction**

35 The acquisition of neural identity, known as neural induction (Stern,
36 2006), represents one of the most widely studied events in embryogenesis. In
37 vertebrates, this process begins at gastrulation and continues as the principal axis
38 elongates resulting in a nervous system extending along the anterior-
39 posterior (AP) length of the body. The critical role of the organizer in specifying
40 neural fate from ectoderm was initially established by the pioneering work of
41 Spemann and Mangold (Spemann and Mangold, 1924). Attention then turned to
42 identifying the inducing signals emanating from the organizer and understanding how
43 different rostral-caudal regions of the nervous system are generated (Anderson and
44 Stern, 2016; Stern, 2001; Stern et al., 2006).

45

46 Several models have been proposed to explain rostral-caudal regionalisation.
47 Embryological experiments led Otto Mangold to suggest separate activities are
48 responsible for inducing distinct regions of the nervous system (Mangold, 1933). He
49 proposed that different parts of the organiser, or the organiser at different times,
50 produced these distinct signals. Subsequently, Nieuwkoop, building on the work of
51 Conrad Waddington (Waddington, 1940), proposed a two-step mechanism to explain
52 the formation and regionalization of the nervous system known as “activation-
53 transformation” (Nieuwkoop, 1952). This hypothesis contends that cells first adopt a
54 neural identity equivalent to the anterior nervous system (“activation”).
55 “Transformation”, in a subsequent step, is responsible for converting a proportion of
56 these cranial-like cells to more caudal fates such as the midbrain, hindbrain
57 and eventually, spinal cord (Nieuwkoop and Nigtevecht, 1954; Stern, 2001).

58

59 In this view, anterior neural cells are considered the precursors of the entire nervous
60 system. This implies that cells that form the nervous system are first specified
61 with an anterior identity before they acquire more caudal axial fates such as
62 hindbrain or spinal cord. Whether this mechanism is valid and applicable to all axial
63 levels of the nervous system is unresolved. Nevertheless, it remains the prevailing
64 view of nervous system regionalisation (Stern, 2001, 2005, 2006) and has influenced
65 the development of methods for the directed differentiation of embryonic stem cells
66 to specific classes of neurons, where regionalising signals are assumed to act after

67 neural identity has been established in cells (Davis-Dusenbery et al., 2014; Wichterle
68 et al., 2002).

69

70 The anterior nervous system in vertebrates, comprising fore-, mid- and hindbrain,
71 has an anatomically and molecularly distinct origin from the spinal cord. The anterior
72 nervous system is formed during gastrulation from cells that remain in
73 the anterior epiblast. By contrast, spinal cord cells are produced during axis
74 elongation by axial stem cells, often referred to as neuromesodermal progenitors
75 (NMPs) (Henrique et al., 2015). These bipotent cells arise in the caudal lateral
76 epiblast, adjacent to the node, and contribute progeny to both the paraxial
77 mesodermal tissue and spinal cord (Garriock et al., 2015; Tzouanacou et al., 2009;
78 Wymeersch et al., 2016). NMPs are exposed to WNT and FGF signalling
79 and are marked by the expression of transcription factors *Sox2*, *T/Brachyury* and
80 *Cdx1, 2, 4* (Gouti et al., 2014, 2017; Henrique et al., 2015; Tsakiridis et al., 2014;
81 Wymeersch et al., 2016). Deletion of *T/Bra*, *Cdx* genes or the absence of WNT
82 signalling severely abrogates axis elongation in mouse embryos, resulting in a failure
83 to form spinal cord and somites at post occipital levels (Amin et al., 2016;
84 Chawengsaksophak et al., 1997; Takada et al., 1994; Yamaguchi et al., 1999; Young
85 et al., 2009). Thus, anterior and posterior parts of the nervous system are populated
86 by distinct groups of cells. Similar to the case *in vivo*, timely pulses of WNT and FGF
87 signals to ESCs that have acquired an epiblast-like state results in the generation of
88 cells resembling NMPs found in embryos (Gouti et al., 2017; Koch et al., 2017).
89 These cells have the capacity to differentiate into spinal cord progenitors that
90 express 5' *Hox* genes characteristic of thoracic and lumbar spinal cord (Gouti et al.,
91 2014; Lippmann et al., 2015). Single-cell transcriptome analysis further emphasises
92 the correspondence between *in vitro* and *in vivo* cell populations (Gouti et al., 2017;
93 Koch et al., 2017). By contrast, ESCs that are differentiated to an epiblast state in
94 the absence of a WNT pulse will generate neural progenitors that display a caudal
95 limit at the level of the hindbrain and cervical spinal cord (Gouti et al., 2014;
96 Lippmann et al., 2015). These observations appear to challenge the Activation-
97 Transformation hypothesis, and are reminiscent of Mangold's model of distinct
98 mechanisms specifying different regions of the nervous system (Mangold, 1933).

99

100 To address the sequence of events that lead to the establishment of a regionalised
101 nervous system an unbiased definition of neural cell identity is required. Early
102 embryological experiments relied on morphological criteria to define cell types and
103 thus do not provide sufficient molecular detail to understand nervous system
104 regionalisation. More recently, gene expression has been used as a proxy for identity.
105 However, this has raised further questions. While common gene regulatory networks
106 (GRNs) are used to globally define neural progenitor (NP) populations, it remains a
107 challenge to understand how these networks are established genome wide, leaving
108 open the question of how neural progenitors become refined into functionally distinct
109 neural cell types along the AP axis. For instance, the SoxB1 family of transcription
110 factors play critical roles in neural progenitors along the AP axis and are broadly
111 expressed, yet it remains unclear how they act at different axial levels (Kondoh et al.,
112 2016). By contrast, distinct enhancer usage in cells has been used to define different
113 cell types (Soucie et al., 2016) and has been shown to better resolve cell identity
114 than conventional transcriptome comparisons (Corces et al., 2016). A repertoire of
115 enhancers is known to drive AP-specific expression of genes that are broadly
116 expressed throughout the nervous system, including the major neural
117 regulators SOX2 and SHH (Epstein et al., 1999; Kutejova et al., 2016; Peterson et al.,
118 2012; Uchikawa et al., 2003). Thus, regulatory element usage provides a reliable
119 and objective correlate of cell identity (Buecker and Wysocka, 2012). One way to
120 assay this is to systematically map and compare chromatin accessibility in different
121 cell types. Techniques such as DNase-seq (Song and Crawford, 2010) and ATAC-
122 seq (Buenrostro et al., 2013) provide this opportunity and the regions identified by
123 these approaches show a high degree of correlation with active histone marks and
124 known enhancers (Buenrostro et al., 2013; Lavin et al., 2014; Wu et al., 2016).

125

126 Here, we define the enhancer landscape using ATAC-seq, in cells with anterior
127 (forebrain/midbrain), hindbrain and spinal cord identity. We take advantage of the
128 temporal resolution afforded by the *in vitro* differentiation of ESCs into defined neural
129 fates to determine how chromatin accessibility changes in time and how this relates
130 to the progression of cells to anterior and posterior neural fates. Combined with *in*
131 *vivo* validation, we show that the difference between anterior and posterior neural
132 progenitors is reflected in their respective chromatin accessibility
133 profiles. We provide evidence that AP identity precedes the acquisition of neural fate.

134 Furthermore, we find that the genomic landscape of NMPs is distinct from other cell
135 types and is dependent on the presence of CDX TFs that remodel the
136 chromatin landscape in response to FGF and WNT signalling. This transition is
137 essential not only to elicit induction of posterior *Hox* genes, but also to repress
138 cranial neural fates. The ability to induce an NMP state in cells is transient (Gouti et
139 al., 2014; Turner et al., 2014) and restricted to stages prior to the acquisition of
140 neural identity; continual changes in the genomic accessibility of cells undergoing
141 neural induction are sufficient to change the intrinsic response of a cell to the same
142 extrinsic signals and the resulting cell fate identity. Together with genetic
143 perturbations and alterations in the timing of posteriorizing signals, the data reveal
144 that, contrary to the activation-transformation hypothesis, axial identity is established
145 before neural induction. These findings are consistent with the proposed dual origin
146 of the central nervous system during animal evolution (Arendt et al., 2016) and
147 prompts a revision to models of neural induction and nervous system regionalization.

148

149 **Results**

150 **In vitro generation of anterior, hindbrain or spinal cord neural progenitors**

151 To define the sequence of events that commit neural cells to different anterior-
152 posterior (AP) identities, we took advantage of mouse embryonic stem cells (ESCs),
153 which, as shown previously, can be directed to differentiate into NPs with anterior,
154 hindbrain or spinal cord identities (Gouti et al., 2014, 2017) (see Methods). In each
155 case, ESCs were transferred from pluripotent conditions to serum-free media
156 containing bFGF (FGF) to induce an epiblast identity. For anterior NPs, FGF was
157 removed after three days and the SHH agonist SAG added, promoting ventral neural
158 identity. By Day (D) 5 these cells expressed a mixture of forebrain and midbrain
159 markers (Gouti et al., 2014). For the generation of hindbrain NPs, cells were
160 exposed to retinoic acid (RA), in addition to SHH signals from D3 to D5 (Figure 1A).
161 This produced visceral motor neuron (MN) progenitors expressing PHOX2B and
162 somatic MNs expressing OLIG2, similar to the brainstem (Figure 1B) (Gouti et al.,
163 2014; Pattyn et al., 1999, 2000). For the generation of spinal cord NPs, ESCs were
164 cultured in the same serum-free media containing bFGF for two days (Figure 1A)
165 and then received a 24 hour pulse of both FGF and WNT (FGF/WNT) signals from
166 D2-3 (WNT signalling was induced with the GSK3 β inhibitor, CHIR99021; see

167 Methods). At D3, cells were transferred to medium containing RA and SHH signals,
168 similar to the hindbrain condition (Figure 1A). At D5, this resulted in the generation of
169 *Olig2* positive spinal somatic MN progenitors (Figure 1B) but no visceral MNs (Figure
170 1B) and expressed *Hox* genes characteristic of cervical/brachial and thoracic regions
171 (see Figure 3E-M).

172

173 **Chromatin accessibility defines neural progenitor identity**

174 Genes expressed in neural progenitors can be controlled by different regulatory
175 elements at different AP positions (Brunelli et al., 2003; Epstein et al., 1999;
176 Uchikawa et al., 2003), therefore we hypothesized that differential enhancer usage
177 would provide an objective definition of cell identity, and a means to follow the
178 transition of cells into distinct neural cell types. We therefore used ATAC-seq
179 (Buenrostro et al., 2013) to examine chromatin accessibility in cells across all stages
180 and conditions of ESC differentiation to anterior, hindbrain or spinal cord fate
181 (Figures 1C-D and Fig S1). As anticipated, distinct chromatin accessibility profiles
182 were evident in different cell types. For example, enhancers directing pluripotency
183 genes, such as *Oct3/4/Pou5f1* (Yeom et al., 1996), were accessible in ESCs but not
184 in any of the three neural conditions, D5A, D5H or D5SC (Figure 1C). By contrast,
185 neural enhancers, such as a previously identified enhancer of *Olig2* (Oosterveen et
186 al., 2012; Peterson et al., 2012) showed the opposite behaviour, exhibiting
187 accessibility in neural conditions, particularly in hindbrain and spinal cord NPs where
188 it is highly expressed, but not in ESCs (Figure 1D). Genome wide comparisons
189 between ESCs and neural progenitors further revealed widespread differences in
190 accessibility between D0 ESCs and each of the D5 NPs, in line with previous
191 observations that different tissue types present entirely different chromatin
192 landscapes (Soucie et al., 2016; Visel et al., 2009). Compared to ESCs, D5 spinal
193 cord NPs increased accessibility at 10,880 sites genome wide, while 13,804 sites
194 became inaccessible in these cells. Further comparisons revealed that as cells
195 differentiated to neural progenitors, more differences in chromatin accessibility were
196 established (Figure 1F) and that sites open in all neural conditions (D5A, D5H and
197 D5SC) displayed accessibility at active neural regulatory regions marked by
198 H3K27ac (Peterson et al., 2012) (Figure 1G). Thus, ATAC-seq allows the
199 identification of regulatory regions that define the neural progenitor lineage.

200

201 **Differences in chromatin accessibility define axial identity of neural** 202 **progenitors**

203 We next sought to identify the regulatory signatures that define neural progenitors
204 with different AP identities. To this end, we performed unsupervised clustering using
205 self-organizing maps (SOM) (Haberle et al., 2014; Törönen et al., 1999) of all
206 regulatory regions that became accessible after removal of ESCs from pluripotent
207 conditions (Figure 2A and see Methods). This allowed us to classify chromatin
208 accessible regions that displayed different dynamics and to explore their relatedness
209 to each other (see below). In particular, we recovered a large set of accessible
210 regions that were common to all three neural progenitor subtypes (Figure 2A, black
211 clusters). We mined 2.56 million DNaseI-hypersensitive sites (DHS) from the
212 ENCODE regulatory element database (Consortium, 2012; Sloan et al., 2016), which
213 covers regulatory sites across a range of stages and tissues. Comparison between
214 the sites we identified in neural progenitors (Figure 2A, black clusters) and the
215 ENCODE data demonstrated that our data were enriched with chromosomal
216 locations associated with open chromatin in neural tissues (Figure S2A). We refer to
217 these common sites as “neural sites”.

218
219 In addition to neural sites, clustering revealed that distinct sets of regulatory regions
220 became accessible in neural progenitors depending on their AP identity: 1863 sites
221 were enriched specifically in anterior neural progenitors (Figure 2A, orange cluster),
222 2509 in hindbrain progenitors (Figure 2A, blue clusters) and 1538 in spinal cord
223 progenitors (Figure 2A, red clusters). Examining the position of these “region-specific”
224 regulatory sites indicated that they were also enriched for open chromatin regions
225 enriched in neural tissues (Figure S2B-D) and displayed activity in neural tissues
226 when compared to other tissues documented in the VISTA enhancer database (Visel
227 et al., 2007) (Figure S2E). In contrast to common neural sites, however, which were
228 mainly (~63.8%) located close to the transcriptional start site (TSS) of coding genes
229 (Figure S2F), region-specific sites displayed the opposite behaviour, lying
230 predominantly in distal regions of the genome (Figure S2G-I).

231
232 Gene-to-peak associations revealed that region-specific sites flanked many neural
233 genes and reflected the AP identity of the cells (Figure 2B-D and Table S1). Anterior
234 NPs displayed region-specific sites at *Shh* (Figure 2B), overlapping the previously

235 identified Shh brain enhancer (Epstein et al., 1999). By contrast, hindbrain region-
236 specific sites flanked the cranial MN gene *Phox2b* (Figure 2C), and in the spinal cord,
237 region-specific sites flanked many posterior *Hox* genes expressed in the spinal cord
238 including *Hoxc8* (Figure 2D) in addition to neural genes including *Nkx6-1*, *Lhx1*,
239 *Sox11* and *Sox2* (Table S1 lists all peak coordinates for region-specific sites and
240 associated genes).

241

242 To test whether the different region-specific signatures observed *in vitro* reflect
243 differences present *in vivo*, we performed ATAC-seq on mouse NPs isolated from
244 different AP levels: anterior (forebrain/midbrain/anterior hindbrain levels) or spinal
245 cord levels (cervical/thoracic) of e9.5 mouse embryos (Figure 2E-G). This
246 developmental time point corresponds most closely to the gene expression profile of
247 D4-5 NPs (Gouti et al., 2017). To isolate neural progenitors from the surrounding
248 tissue, we used *Sox2eGFP* reporter mice that express GFP throughout the nervous
249 system (Ellis et al., 2004) to obtain distinct anterior and posterior regions of the
250 neural tube (Figure 2E and see Methods). To ask if AP differences observed *in vitro*
251 predict AP identity *in vivo*, we examined the change in accessibility of *in vitro*-defined
252 spinal cord sites, relative to *in vitro*-defined anterior sites, under *in vivo* conditions
253 (Figure 2F). Strikingly, region-specific signatures showed an enrichment that
254 reflected their AP identity *in vivo* similar to that observed *in vitro*, whereas neural
255 sites common to all *in vitro* derived neural progenitors were equally enriched in both
256 populations *in vivo* (Figure 2F). Anterior *in vivo* NPs demonstrated increased
257 accessibility at sites that define *in vitro* anterior neural identity and not spinal cord
258 (Figure 2G). For example, regulatory regions included sites that flanked neural
259 genes, such as *Shh*, where accessibility overlapped the *Shh* brain enhancer (Epstein
260 et al., 1999), under anterior but not spinal cord *in vivo* conditions (Figure S2J, arrow).
261 Similar results were obtained when examining *in vivo*-derived spinal cord NPs:
262 increased accessibility was observed in regions that define spinal cord NPs *in vitro*
263 (Figure 2G). For example, accessibility was observed at the *Olig2* enhancer
264 (Oosterveen et al., 2012; Peterson et al., 2012), which is repressed in anterior neural
265 conditions (Figure S2K, arrow). In summary, changes in chromatin accessibility
266 reflect the AP identity of neural progenitors of the nervous system and are
267 recapitulated by NPs generated *in vitro*.

268

269 **Context-dependent binding of neural TFs defines axial identity**

270 We next interrogated the *in vitro* region-specific signatures of anterior, hindbrain and
271 spinal cord progenitors to identify TFs that recognise these AP-specific sites. We
272 performed ChIP-seq enrichment analysis using a set of publicly available datasets
273 (Sheffield and Bock, 2015) covering 910,490 regulatory regions from 270 different
274 TFs (Table S2). Our analysis revealed that distinct TF binding was enriched in the
275 three distinct subtypes: anterior NPs were enriched with FOXA2 and NKX2-2 binding
276 sites (Figure 2H); by contrast, hindbrain sites contained OLIG2 and NKX6-1 (Figure
277 2I); and spinal cord sites were enriched for CDX2 and HOXC9 binding (Figure 2J).
278 Motif enrichment with Homer (Heinz et al., 2010) predicted an enrichment of SoxB1
279 TF motifs (SOX1/2/3), in all three neural progenitors subtypes (Figure S3A-C),
280 consistent with their expression throughout the neuraxis and central role in neural
281 development (Avilion et al., 2003; Kamachi and Kondoh, 2013; Wood and Episkopou,
282 1999; Ying et al., 2003). Notably, hindbrain and spinal cord cells, which are both
283 exposed to the same signals (RA/SHH) from D3-D5 are enriched for SOXB1 binding
284 but at distinct genomic sites (Figure S3B-C). The presence of posterior HOX binding
285 events together with SOX in spinal cord progenitors suggested that region-specific
286 expression of *Hox* genes influenced the binding site preference of the core neural
287 SOXB1 TFs. Likewise, posterior *Hox* genes can alter the binding site preference of
288 SOX factors when misexpressed in the cortex (Hagey et al., 2016).

289
290 To validate these in-silico findings, we performed ChIP-seq of the SOXB1 TF, SOX2,
291 in D5 hindbrain and spinal cord NPs (Figure 2I-J). This confirmed hindbrain predicted
292 SOX sites were physically engaged with SOX2 in hindbrain NPs (Figure 2I) and not
293 spinal cord NPs at day 5 of the differentiation (Figure 2J). Conversely, SOX2
294 accessible sites specific to the spinal cord condition showed increased engagement
295 of SOX2 in spinal cord versus hindbrain conditions (Fig.2I-J). Collectively, these data
296 demonstrate the utility of ATAC-seq for predicting factors enriched at region-specific
297 sites in NPs. Furthermore, it demonstrates that NPs develop region-specific
298 transcription factor binding at a time when they are exposed to the same extrinsic
299 signals – RA/SHH. This raises the question of how differences in regional identity
300 and transcription factor engagement are established.

301

302 **Posteriorising signals that promote spinal cord identity depend on**
303 **developmental timing**

304 We took advantage of the temporal resolution afforded by the *in vitro* differentiation
305 to define when AP identity is established in neural progenitors. To address this
306 question, we examined when region-specific regulatory regions became accessible,
307 and under which conditions this occurred (Figure 3A-C). The prevailing view is that
308 to generate neural cells with a spinal identity, anterior neural cells must be gradually
309 posteriorised to acquire a spinal fate (Davis-Dusenbery et al., 2014; Stern, 2001). To
310 test this assumption, we asked if spinal cord progenitors transitioned via an anterior
311 or hindbrain identity, before acquiring spinal cord identity. We assessed region-
312 specific sites and their behaviour over the course of the differentiation. In contrast to
313 previous models, we found that spinal cord cells failed to exhibit transient
314 accessibility at either anterior (Figure 3A) or hindbrain (Figure 3B) region-specific
315 sites, challenging this view. Instead, spinal cord-specific sites became accessible in
316 spinal cord conditions by D4 of the differentiation (Figure 3C).

317

318 Examining the timing of region-specific sites further revealed a synchronicity
319 between neural progenitor identity and the establishment of AP fate in cells.
320 Specifically, anterior, hindbrain and spinal cord progenitors begin to exhibit region-
321 specific accessibility between D3-4, coincident with the emergence of neural sites
322 (Figure 2A, black cluster). Both the hindbrain and spinal cord progenitors are
323 exposed to the same conditions (RA + SHH; Figure 1A) yet adopt region-specific
324 signatures at this same time point (Figure 3B-C). Thus, extrinsic signals present at
325 the time the regional signatures emerge are not sufficient to promote distinct
326 regulatory element usage in cells. The only difference in generating hindbrain vs
327 spinal cord progenitors *in vitro*, is the addition of WNT signals that is added together
328 with FGF between D2-3 of the spinal cord differentiation (Figures 1A and 3D). This
329 suggests that WNT signalling, together with FGF, plays a critical role in posteriorising
330 cells to adopt spinal rather than hindbrain fates, consistent with previous studies
331 (Gouti et al., 2014; Nordström et al., 2006). Our data suggest that at the genomic
332 level, this requires the suppression of hindbrain sites in response to RA/SHH signals,
333 and adoption of a distinct set of SC specific sites (Figures 2A and Figure 3B-C).

334

335 While the specific timing and onset of neural progenitor fates has been difficult to
336 define *in vivo*, we took advantage of the *in vitro* system to alter the timing of extrinsic
337 signals to directly test the temporal requirement for WNT signalling. To this end, we
338 provided the same 24h pulse of FGF/WNT but at later stages of the differentiation, to
339 hindbrain progenitors at D4-5 (Figure 3D'). We then asked whether exposure to
340 these signals at this time was sufficient to promote a spinal cord instead of hindbrain
341 identity (Gouti et al., 2014; Nordström et al., 2006). Altering the timing of FGF/WNT
342 from D2-3 to D4-5 (Figure 3D') resulted in the induction of canonical WNT signalling
343 target genes such as *Notum*, to levels comparable with the induction observed at D3
344 when provided to cells between D2-3 (Figure 3E, compare D3NMP and D5H+).
345 However, shifting the treatment of FGF/WNT to D4-5 was no longer sufficient to
346 induce expression of posterior *Hox* genes characteristic of the spinal cord condition,
347 including *Hoxc8*, *Hoxb9*, which are induced from D3 following D2-3 FGF/WNT
348 treatment (Figure 3F-G). Likewise, we found that the induction of *Brachyury* (*T/Bra*)
349 and *Cdx2*, normally induced at D3 by a D2-3 pulse of FGF/WNT (Gouti et al., 2014)
350 (Figures 3H-I and S4A-C), was no longer observed at D5 following FGF/WNT
351 treatment between D4-5 (Figure 3H-I, compare D3NMP and D5H+).

352

353 The failure of hindbrain progenitors to upregulate spinal cord genes suggests that
354 administration of FGF/WNT signals at this stage is not capable of posteriorising cells.
355 We further investigated the neural identity of the cells resulting from FGF/WNT
356 treatment between D4-5. We found that the expression of the ventral neural markers
357 *Phox2b* and *Olig2*, normally expressed in hindbrain conditions (Figure 1B), was no
358 longer maintained following FGF/WNT treatment in these cells (Figure 3J-K). By
359 contrast, expression of *Pax7* and *Dbx1*, markers of dorsal and intermediate neural
360 tube fates, respectively, was observed (Figure 3L-M). WNT signalling is known to
361 promote dorsal neural fate at the expense of ventral fate in the neural tube (Lei et al.,
362 2006; Wang et al., 2011). Thus, the later treatment of cells with FGF/WNT, a point at
363 which cells have begun expressing neural progenitor markers such as *Sox1* and
364 *Pax6* (Fig. 3N,O) in response to neuralising signals provided from D3 (Figure 1A), is
365 consistent with WNT promoting dorsal (*Pax7*, *Dbx1*) at the expense of ventral
366 (*Phox2b*, *Olig2*) neural cell fates during neural tube development (Alvarez-Medina et
367 al., 2008; Lei et al., 2006; Muroyama et al., 2002; Wang et al., 2011). The
368 posteriorising activity of WNT together with FGF signalling is thus dependent on

369 precise developmental timing: prior to neural progenitor establishment (D2-3), these
370 signals are capable of posteriorising cells, yet after cells commit to the neural lineage
371 (D4-5), the same combination of extrinsic signals promotes dorsal in place of
372 posterior cell fates in the nervous system (Le Dréau and Martí, 2012; Gouti et al.,
373 2015). Taken together these data indicate that the generation of spinal cord fates
374 does not follow a gradual posteriorisation of more anterior neural progenitors, as
375 cells lose the competency to be posteriorised following establishment of neural fate.

376

377 **WNT directs a transient set of chromatin remodelling events**

378 To understand how FGF/WNT signals exert a stage-specific, posteriorising effect in
379 cells prior to neural progenitor establishment, we examined the chromatin landscape
380 in cells following FGF or FGF/WNT treatment at D3. We found that at D3, cells that
381 had been exposed to FGF/WNT for 24h displayed accessibility at 875 unique regions
382 (Figure 2A, NMP/SC and NMP cluster). Strikingly, of these 875 sites, 454 (51.8%)
383 were immediately downregulated as cells committed to spinal cord fates by D5
384 (Figure 2A, NMP cluster). Thus, as cells transition to a spinal cord identity, they
385 transiently adopt a genomic signature, in response to FGF/WNT signals, that is
386 distinct from both the epiblast (Figure 2A, epiblast cluster) and neural regulatory
387 signatures (Figure 2A, black clusters). These cells include the bipotential population
388 of NMPs, which contributes to both the spinal cord and somites (Gouti et al., 2014;
389 Tsakiridis et al., 2014; Turner et al., 2014; Tzouanacou et al., 2009; Wymeersch et
390 al., 2016), that expresses the transcription factors Sox2, *T/Bra* and *Cdx1,2,4* (Martin
391 and Kimelman, 2012; Olivera-Martinez et al., 2012; Tsakiridis et al., 2014; Young et
392 al., 2009).

393

394 To further validate the *in vitro* signature, we asked to what extent these sites overlap
395 with the *in vivo* chromatin accessibility associated with NMPs. We took advantage of
396 existing ATAC-seq data collected from whole mouse epiblasts from E6.0-7.2 and
397 from E7.5 posterior mouse tissue – the tissue which contains NMPs (Neijts et al.,
398 2016). We found that more than 71% of sites induced by FGF/ WNT signalling *in*
399 *vitro* at D3 overlapped with accessible sites found in E7.5 posterior mouse embryos
400 (Figure 4A), while the overlap with either the epiblast (29% at E6.0) or purified NPs
401 from E9.5 embryos is much less (less than 5%, this study; Figure 4A). We also found
402 that the epiblast-specific sites defined in the self-organizing map (Figure 2A, epiblast

403 cluster) showed the greatest overlap with those found *in vivo* in E6.0 epiblast (Figure
404 2A), while very little overlap was evident with *in vivo* neural progenitors (Figure 4A).
405 These data suggest that the epiblast and NMP signatures identified *in vitro*
406 correspond to their respective, tissue-specific, regulatory signatures *in vivo*.

407

408 As NMPs generate paraxial mesoderm and are present in posterior tailbud tissues, it
409 remained a possibility that the NMP sites we had recovered *in vitro* represented a
410 nascent mesodermal population. Both NMP and paraxial mesodermal progenitors
411 express *T/Bra* (Garriock et al., 2015; Tsakiridis et al., 2014; Wymeersch et al., 2016).
412 Therefore, we examined chromatin accessibility in ESCs in which *T/Bra* had been
413 genetically inactivated. ESCs lacking *T/Bra* are able to generate spinal cord
414 progenitors, but are incapable of forming paraxial mesoderm (Gouti et al., 2014).
415 Strikingly, the loss of *T/Bra* had a negligible effect on both the NMP/SC shared and
416 NMP-specific signature induced by WNT exposure (Figure 4B), as NMP sites
417 remained accessible in mutant cells (Fig. 4B,D). Furthermore, *T/Bra*-lacking ESCs
418 differentiated to D5SC maintained spinal cord chromatin accessible sites (Figure 4C)
419 and maintained the expression of posterior *Hox* genes such as *Hoxb9* and *Hoxc6*
420 (Figure 4E-F), while expression of 3' *Hox* genes (like *Hoxb4* and *Hoxc4*) was
421 reduced, similar to WT D5SC cells (Figure 4G-H). Thus, *T/BRA*, which is necessary
422 for paraxial mesoderm specification (Nowotschin et al., 2012; Rashbass et al., 1991),
423 is not responsible for the chromatin remodelling events associated with NMP identity
424 (Figure 4B,D), nor is it required to induce posterior *Hox* genes or spinal cord identity
425 (Figure 4C, E-F) (Gouti et al., 2014).

426

427 **The acquisition of spinal cord fate requires CDX to repress hindbrain identity**

428 To establish which factors are responsible for mediating WNT-dependent chromatin
429 remodelling events, we performed ChIP-seq enrichment analysis on NMP-specific
430 sites (Figure 5A). In this analysis, we found that the NMP regulatory regions are
431 highly enriched with CDX2 TF binding events (Figure 5A). Nucleotide resolution
432 analysis of the frequency of transposon-mediated integration events further verified
433 the presence of a CDX “footprint” (Buenrostro et al., 2013) present in the chromatin
434 landscape of D3NMP cells. This suggested physical engagement of these factors
435 occurs at sites of open chromatin (Figure S4D) and supports the idea that CDX plays
436 an important role downstream of WNT signalling to promote NMP identity. Recent

437 observations support these findings, suggesting that CDX factors promote a niche
438 that sustains growth of the posterior tailbud and *in vitro*-derived NMPs (Amin et al.,
439 2016). However, the genetic removal of *Cdx* TFs or combined absence of *Cdx2* and
440 *T/Bra* results in severe axis elongation *in vivo* (Amin et al., 2016; van Rooijen et al.,
441 2012a; Young et al., 2009). This has precluded a direct analysis of the cellular
442 context required for these factors to function *in vivo*, and hence their role during
443 spinal cord generation has remained unclear.

444

445 We took advantage of the *in vitro* differentiation, where we could directly test the
446 function of all three CDX TFs in both the generation of NMPs and spinal cord cell
447 types. Using an ESC line lacking *Cdx1,2,4* (*Cdx^{1,2,4}-/-*) (Gouti et al., 2017), we asked
448 if WNT treatment was sufficient to remodel the chromatin landscape, as observed in
449 NMPs lacking *Bra* (Figure 4B). In contrast to the loss of *T/Bra*, however, the
450 elimination of all three *Cdx* TFs had a profound effect on the response to WNT
451 signalling (Figure 4D). Only a small fraction of the NMP and NMP/SC shared sites
452 remained accessible in *Cdx^{1,2,4}-/-* cells with most NMP sites displaying similar
453 accessibility to D3A cells that were not treated with WNT signals (Figure 4D).
454 Similarly, the CDX footprint observed in WT cells was no longer observed (Figure
455 S4D). This suggests that CDX factors are essential for the remodelling of chromatin
456 accessibility associated with an NMP state (Amin et al., 2016) as well as the
457 transition from an NMP to spinal cord fate (Figure 4D). Furthermore, the
458 differentiation of mutant *Cdx* cells to neural progenitors no longer resulted in the
459 expression of spinal cord 5' *Hox* genes *Hoxb9*, *Hoxc6* (Figure 4E-F). Instead, *Cdx*
460 mutant cells expressed *Hoxb4*, *Hoxc4*, as observed in WT hindbrain conditions at D5
461 (Fig. 4G-H).

462

463 These findings suggest that in the absence of *Cdx* TFs, the application of
464 posteriorising signals no longer promotes a posterior neural identity in cells, and thus
465 the generation of more anterior neural fate ensues. To test this prediction, we
466 performed immunofluorescence on *Cdx^{1,2,4}-/-* mutant cells at D5 of the spinal cord
467 differentiation, to confirm the complement of MN subtypes present at this time. We
468 found ectopic generation of visceral MNs marked by PHOX2B in *Cdx* mutant cells
469 (Figure 5B), which normally occurs in hindbrain but not spinal cord conditions (Figure
470 1B). In addition, sustained *Olig2* induction was observed suggesting the removal of

471 CDX TFs does not impede neural progenitor establishment, per se (Figure 5B). In
472 agreement with this observation, analysis of the chromatin accessibility landscape
473 present in D5 *Cdx*^{1,2,4/-} cells differentiated under spinal cord conditions revealed that
474 these cells lacked spinal cord identity sites and instead, gained accessibility at
475 hindbrain identity sites (Figure 5C). These data demonstrate that CDX factors are
476 required to suppress the accessibility of hindbrain identity sites in response to
477 RA/SHH signals, and allow the specific acquisition of spinal cord accessible regions.

478

479 To address how CDX factors are capable of both repressing hindbrain accessible
480 sites and allowing the induction of a spinal cord regulatory program, we took
481 advantage of previously published CDX2 ChIP-seq from *in vitro*-generated NMPs
482 (Amin et al., 2016) and motor neuron progenitors (pMNs) (Mazzoni et al., 2013) and
483 mapped the proportion of bound sites in the accessible regions clustered in the self-
484 organising map (Figure 2A) of ATAC-seq accessible regions (Figure 5D). This
485 allowed us to define direct targets of CDX2 (Table S3) and monitor in which cell
486 types these regions were accessible. As expected, CDX2 binding was highly
487 enriched at genomic sites accessible in NMPs, as well as SC specific sites and sites
488 found in both conditions (Figures 5D and S5A). This overlap suggests a critical role
489 for CDX factors in the transition to a spinal neural identity, where it targets many
490 regions of the posterior *Hox* genes (Table S3) (Amin et al., 2016; Mazzoni et al.,
491 2013; Neijts et al., 2017; Young et al., 2009). More strikingly, however, we found that
492 CDX2 bound to additional target sites, outside of the *Hox* locus, which were
493 accessible in hindbrain and spinal cord conditions (Figures 5D, green clusters and
494 S5A), as well as sites exclusive to the hindbrain (Figures 5D, orange cluster and
495 S5A). Examination of these CDX2-bound sites revealed a shared hindbrain/spinal
496 cord site lying upstream of *Phox2b* (Figure 5E and TABLE S3 list of regions). This
497 demonstrates that CDX factors are capable of directly targeting neural genes,
498 including those that are repressed from spinal cord conditions. Similarly, previous
499 studies have suggested that CDX1 directly binds to regulatory elements at *Mafb*, and
500 represses the expression of this hindbrain marker (Kim et al., 2005; Sturgeon et al.,
501 2011). Consistent with this we found that CDX2 directly targets *Mafb* at a region only
502 accessible under hindbrain conditions (Figure 5F). A systematic analysis further
503 revealed that a substantial number of hindbrain genes are repressed by CDX, as
504 demonstrated by mRNA-seq analysis *in vivo* (Figure S4E) (Amin et al., 2016). This

505 includes *Mafb* and the binding of CDX2 correlates with preventing accessibility at a
506 nearby regulatory region, as *Cdx* mutant cells show striking increases in accessibility
507 at this site from D4-5 under spinal cord conditions (Figure S4F, arrows). Gene
508 ontology analysis of the major pathways and processes enriched at CDX2-engaged
509 sites (Figure S5B-E) confirmed that CDX2 occupancy at hindbrain-accessible sites is
510 directly linked with neural genes (Figure S5C), whereas in both NMP/SC and SC the
511 CDX2 bound accessible sites were associated with genes implicated in anterior-
512 posterior patterning (Figure S5D-E).

513

514 In summary, genome-wide analyses indicate that in addition to establishing NMP
515 identity and driving activation of posterior *Hox* genes (Amin et al., 2016; Mazzoni et
516 al., 2013; Neijts et al., 2017), CDX factors play a central role in directly targeting
517 neural genes at regulatory sites associated with hindbrain identity (Figure 5D-F). We
518 propose that this mechanism ensures the repression of hindbrain genes in response
519 to RA/SHH signals, in addition to the priming of posterior *Hox* genes, which drive
520 spinal cord identity. The induction of CDX, prior to neural induction, is therefore
521 critical in establishing the appropriate SC-specific regulatory signature, while further
522 repressing hindbrain fate in response to neuralising signals. We propose that this
523 dual functionality of CDX is essential to restrict the generation of specific neuronal
524 subtypes to discrete AP levels of the neural tube. Such a mechanism ensures the
525 production of, for example, visceral MNs in the posterior hindbrain but not the spinal
526 cord (Figure 6A).

527

528 **DISCUSSION**

529 The prevailing view of nervous system formation, summarised in the “activation-
530 transformation hypothesis”, proposes that nervous system induction occurs in two
531 phases (Eyal–Giladi, 1954; Nieuwkoop, 1952; Nieuwkoop and Nigtevecht, 1954;
532 Toivonen and Saxen, 1968). In the first step, ectoderm is induced to become anterior
533 neural tissue. Following this “activation” step, posteriorising signals convert –
534 “transform” – anterior neurectoderm into the complete range of positional identities
535 that comprise the neuraxis (Stern, 2001). In this view of neural induction, the
536 generation of posterior regions of the nervous system, such as the spinal cord,
537 require that cells first acquire forebrain identity before being caudalised to a posterior
538 fate. Despite the widespread acceptance of this model, previous studies have lacked

539 the cellular and temporal resolution necessary to test this model. To address this, we
540 took advantage of *in vitro* differentiation of ESCs to assay the chromatin landscape
541 as cells transition from pluripotency to different axial levels of the nervous system.
542 This revealed that cells destined to form spinal cord progenitors do not transiently
543 adopt an anterior neural state or acquire their regional identity via a gradual
544 caudalisation of more rostral cell types (Figure 3A-C). Instead, spinal cord cells are
545 regionally restricted prior to their commitment to a neural fate. We provide evidence
546 that CDX remodels the chromatin landscape in cells before neural fate is established.
547 This step is essential for the specification of a spinal cord identity in cells, and the
548 repression of cranial MN fates (Figure 5C). Thus, specification of spinal cord fate
549 involves cells committing to an axial identity (Figure 6A, “*primary regionalisation*”)
550 prior to neural induction, reversing the sequence of events implied by the ‘activation-
551 transformation’ hypothesis and prompting a revision in our understanding of nervous
552 system regionalization (Figure 6A).

553

554 **Regulatory signature dynamics argue against “activation-transformation”**

555 Support for the activation-transformation hypothesis originated in embryological and
556 molecular experiments in chick and frog embryos. For example, explants of posterior
557 axial tissue promote midbrain and hindbrain fates from prospective forebrain tissue
558 (Cox and Hemmati-Brivanlou, 1995) and manipulating WNT, FGF and/or RA
559 signalling in neural plate explants alters rostral caudal identity of neural cells in ways
560 consistent with a graded caudalising activity (Kolm et al., 1997; Lamb and Harland,
561 1995; Muhr et al., 1999; Nordström et al., 2006; Wilson et al., 2001). It is notable that
562 in many of these studies the most caudal markers assayed were representative of
563 the hindbrain or anterior spinal cord and the results were subsequently extrapolated
564 to apply to the entire length of the spinal cord without explicit testing. While RA
565 exposure to neural progenitors is sufficient to posteriorize anterior neural cells to
566 form hindbrain, the most caudal identity generated in these assays corresponds to
567 cervical (anterior) spinal cord (Gouti et al., 2014; Liu et al., 2001; Mahony et al., 2011;
568 Maury et al., 2014; Mazzoni et al., 2013; Niederreither et al., 2000). Furthermore,
569 treatment of anterior neural progenitors with increasing concentrations of WNT fails
570 to caudalise these cells to a spinal cord fate, instead their identity corresponds to the
571 posterior hindbrain (Kirkeby et al., 2012). Thus, the activation-transformation
572 hypothesis seems compatible with the experimental evidence for regionalisation of

573 the fore-, mid- and hindbrain, but extending the model to the spinal cord does not
574 appear to be supported by the data (Lamb and Harland, 1995).

575

576 To systematically define neural cell identity, with the resolution to distinguish AP-
577 specific differences, we took advantage of ATAC-seq (Buenrostro et al., 2013, 2015).
578 We reasoned that enhancer usage, read out by chromatin accessibility, would
579 provide an unbiased means to follow the induction of neural fate and to determine
580 when specific regional identities are established in cells. Using a data-driven
581 approach, we clustered regulatory regions based on their pattern of accessibility over
582 developmental time during the formation of anterior, hindbrain and spinal cord
583 progenitors. This revealed that a common set of regulatory regions becomes
584 accessible across all neural progenitors. In addition, many of these sites overlap with
585 accessible regions present in neural tissues. Thus, a distinct regulatory signature
586 defines the neural lineage. Overlaid with this, we find that neural progenitors display
587 additional regulatory signatures that define their AP identity, even in hindbrain and
588 spinal cord progenitors, where their AP signatures are laid down while exposed to
589 the same neuralising conditions (Figure 2A). Comparisons between *in vitro* and *in*
590 *vivo*-derived neural cells demonstrated that the regulatory signatures observed *in*
591 *vitro* reliably predicted AP position within the neural tube (Figure 2E-G). Taking
592 advantage of the temporal resolution afforded by the *in vitro* differentiation, we traced
593 the emergence of genomic signatures during neural induction (Figure 3A-C). This
594 revealed that AP-specific regulatory regions appear in cells concomitant with the
595 establishment of neural identity (from Day 3-5). Crucially, the pattern of accessibility
596 that defined spinal cord identity (Figure 3C) appeared at the same time as the
597 broader neural signature, as shown by the neural set of regions that open from Day
598 4-5 across all progenitor subtypes (Figure 2A, black clusters show neural sites). This
599 coincides with the addition of RA and SHH – signals that promoted neural induction
600 (Figure 3N,O). In addition, cells differentiating to a spinal cord fate did not display a
601 transient anterior or hindbrain regulatory landscape prior to the induction of the
602 spinal cord signature (Figure 3C). These data argue against the idea that the spinal
603 cord is generated by posteriorising more anterior neural cells and therefore,
604 alternative mechanisms must be involved.

605

606 **AP identity is established before the acquisition of neural identity**

607 *In vitro*, we found that the regulatory signatures that define different regional NPs is
608 established at the same time. Moreover, different AP identities emerge in cells under
609 the same conditions: hindbrain and spinal cord signatures appear in cells during the
610 period in which they are exposed to RA & SHH (Figure 3B-C). Hence, the distinction
611 between regional identities is established before this time, when spinal cord (but not
612 hindbrain) fated cells receive FGF/WNT signalling (Figure 1A). We tested the
613 importance of this timing by altering when the signals were applied during neural
614 differentiation (Figure 3D,D'). This indicated that delaying addition of FGF/WNT
615 signals until after neural identity had been established was unable to convert
616 hindbrain cells to a spinal cord identity (Figure 3F-G). Thus, cells must receive these
617 signals before neural fate is established, as the competence of cells to form spinal
618 cord is lost following neural induction. This suggests that FGF/WNT signalling
619 establishes a posterior genomic programme in cells – a 'primary regionalisation' –
620 before neural induction.

621

622 To determine what factors might determine posterior competency in cells, we
623 examined the chromatin landscape and identified regions that responded to
624 FGF/WNT at D3 (NMP sites; Figure 2A). These NMP sites were enriched in CDX2
625 TF binding sites (Figure 5A) and we identified the presence of a CDX footprint
626 (Figure S4D) suggesting CDX factors physically engage with open chromatin sites
627 following FGF/WNT treatment. Our findings are consistent with the established role
628 of CDX in promoting the formation of posterior embryonic development, downstream
629 of FGF and WNT signalling (Amin et al., 2016; Skromne et al., 2007; Young et al.,
630 2009). The absence of CDX factors *in vivo* severely truncates embryo elongation
631 resulting in a lack of trunk tissue (Amin et al., 2016; Chawengsaksophak et al., 2004;
632 Young et al., 2009). It has therefore proved difficult to define the cellular context
633 required for CDX to function, and to address how CDX may regulate the production
634 of spinal cell types. Taking advantage of the *in vitro* system, we showed by mutating
635 all three *Cdx* genes that these factors are necessary for remodelling chromatin and
636 the presence of a CDX footprint (Figure S4D), making available 875 sites genome
637 wide in response to FGF/WNT signaling (Figures 2A and 4D). Furthermore, we
638 uncovered that the majority of these NMP sites overlap with *in vivo* accessible
639 regions in the posterior mouse tailbud, the tissue in which NMPs reside (Figure 4A).
640 A similar role for CDX was recently proposed in NMPs generated *in vitro* from mouse

641 EpiSCs (Amin et al., 2016; Neijts et al., 2016). Our data agree and extend these
642 findings by demonstrating that cells lacking *Cdx1/2/4* can no longer generate a
643 genomic signature associated with NMP identity (Figure 4D) and ultimately generate
644 hindbrain NPs instead of spinal cord (Figure 5C-D).

645

646 CDX factors are crucial to prevent accessibility at hindbrain specific sites in addition
647 to the maintenance of NMP identity. Loss of *Cdx* function has been shown to
648 promote hindbrain fate at the expense of spinal cord in zebrafish (Skromne et al.,
649 2007) and in mouse (Young et al., 2009). In the latter case, *Hox* genes are mostly
650 capable of recovering the posterior elongation defects observed in these mutants.
651 However, in zebrafish, overexpression of posterior *Hox* genes was incapable of
652 rescuing the defect under these conditions. It has therefore remained difficult to
653 determine the precise function of CDX proteins, the timing and cell type in which they
654 function, and how they exert their effect *in vivo*. We resolve this uncertainty and
655 show that a limited developmental window exists, between pluripotent epiblast cells
656 and neural progenitors, in which FGF/WNT can induce *Cdx* expression (Figures 3I
657 and S4A-C) and thus exert posteriorising effects (Figure 3F,G). We demonstrate that
658 the expression of *Cdx* factors is necessary to establish spinal cord competency, by
659 directly binding to neural regulatory regions that are accessible in hindbrain
660 conditions (Figure 5D-F and TABLE S3) and by directly promoting posterior *Hox*
661 genes (Figure S4E and TABLE S3). Hence, a major anterior-posterior division of the
662 nervous system, separating the spinal cord from more rostral territories, is
663 established prior to neural induction by the chromatin remodeling activities of CDX
664 TFs.

665

666 Newly accessible genomic regions associated with the addition of WNT signals could
667 be divided into those only transiently available and lost by the time cells had adopted
668 a spinal cord identity (Figure 2A, NMP sites) and a set that continued to be
669 accessible in spinal cord progenitors. This suggests in addition to ‘priming’ –
670 identifying and making accessible – regulatory elements that are then sustained in
671 spinal cord progenitors, CDX establishes a transition (“handover”) state between
672 NMP and SC cells. This handover is driven by active remodeling of the chromatin
673 both before (NMP sites; CDX-driven; Figure 5A) and after (SC sites; HOXC9/CDX-
674 enriched; Figure 2J) neural induction. Moreover, CDX2 appears to repress directly

675 genes involved in hindbrain neural identity (Figures 5E-F and S4E) indicating a
676 pivotal function for CDX in securing spinal cord identity.

677

678 The importance of the CDX dependent genomic programme prior to neural induction
679 is reinforced by a direct comparison of chromatin accessibility in hindbrain and spinal
680 cord progenitors. In both conditions, cells have been exposed to RA and SHH
681 signalling and express the neural progenitor transcription factors Sox1-3. Examining
682 accessible genomic sites that distinguish hindbrain and spinal cord revealed an
683 overrepresentation of the SOXB1 DNA binding motif in these sites. ChIP-seq
684 confirmed that the binding site preference of SOX2 depends on AP position;
685 hindbrain accessible regions, predicted to be bound by SOXB1 factors were
686 enriched for SOX2 protein binding in hindbrain progenitors. Conversely, a different
687 set of SOXB1 accessible sites, specific to the spinal cord, shows engagement of
688 SOX2 at these sites in spinal cord but not hindbrain progenitors (Fig. 2). This
689 suggests that deployment of SOX2 in neural progenitors is dependent on the
690 genomic programme established in the epiblast from which the neural cells originate.
691 The activity of CDX before neural induction directs neural TFs such as SOX2 to bind
692 to spinal cord specific locations and abrogates binding to sites occupied only in the
693 hindbrain (Figure 2I-J). CDX proteins regulate posterior *Hox* gene expression and
694 establish the differences in *Hox* complement between spinal cord and hindbrain
695 (Mazzoni et al., 2013; Nordström et al., 2006; van Rooijen et al., 2012b; Young et al.,
696 2009). Thus, the differences in the *Hox* code between spinal cord and hindbrain
697 appear responsible for the region-specific chromatin accessibility and distribution of
698 SOX binding (Hagey et al., 2016). This highlights the importance of primary
699 regionalisation in cells, preceding neural induction, to establish distinct hindbrain and
700 spinal cord identities.

701

702 **Distinct lineages generate the anterior vs posterior nervous system**

703 The divergence in the mechanisms for formation of the anterior and posterior
704 nervous system is reminiscent of older ideas in which separate developmental
705 organizers were proposed to induce different parts of the CNS (Mangold, 1933). This
706 is consistent with experiments in chick embryos in which the identity of neural tissue
707 induced by grafts of the organiser – tissue capable of inducing neural identity –
708 depends on the embryonic age of the organiser: developmentally older organisers

709 induce characteristics of the caudal neural tube but not forebrain (Storey et al., 1992).
710 Thus, transplantation of different epiblast populations may have occurred at these
711 different stages. In addition, it has been suggested that AP patterning events in the
712 zebrafish epiblast is uncoupled from the neural inducing activities provided by the
713 organizer (Koshida et al., 1998). Although in these studies definitive molecular
714 markers were either lacking or specific to the hindbrain (but not spinal cord), these
715 data hinted that axial patterning information might be established in the epiblast
716 before the acquisition of neural identity. Indeed, NMPs, that generate both spinal
717 cord and somites, are specified in the posterior epiblast in response to FGF and
718 WNT signaling (Garriock et al., 2015; Tzouanacou et al., 2009; Wymeersch et al.,
719 2016). Thus, spinal cord cells derive from a different lineage to the rest of the
720 nervous system. These findings may reconcile old observations that grafts of tissue
721 able to generate caudal structures frequently developed into both neural and
722 mesodermal cell types, in contrast to the grafts producing forebrain structures that
723 lacked mesodermal counterparts (Nieuwkoop, 1952).

724

725 The separate lineages generating hindbrain versus spinal cord progenitors
726 prefigures differences in the complement of cell types generated in these regions.
727 Serotonergic neurons and cranial visceral MNs are produced exclusively in the
728 hindbrain (Carcagno et al., 2014; Cordes, 2001). This contrasts with preganglionic
729 neurons of the sympathetic nervous system (Shirasaki and Pfaff, 2002) and the
730 classes of somatic MNs and interneurons that are specific to the spinal cord (Jessell,
731 2000). Moreover, the genetic programme and progeny of neural crest derived from
732 hindbrain and spinal cord levels differs (Simoes-Costa and Bronner, 2016); for
733 example, the skeletogenic potential observed in cranial neural crest cells (NCCs),
734 which form at hindbrain axial levels, is not observed in trunk NCCs generated at
735 more posterior levels of the embryo (Martik and Bronner, 2017; Santagati and Rijli,
736 2003). Primary regionalisation events in the precursors that generate trunk, and not
737 cranial, NCCs, may contribute to their inherent differences along the neuroaxis.
738 Whether a similar strategy also accounts for the generation of posterior tissues
739 contributing to mesodermal or endodermal lineages in the trunk of the embryo
740 remains to be tested.

741

742 The transition from hindbrain to spinal cord specific cell types coincides with the
743 approximate rostral limit of the neural progenitors lineage traced from *T/Bra*
744 expressing NMPs (Garriock et al., 2015). This raises the possibility that the separate
745 origins of the progenitors that populate the hindbrain and spinal cord underpins the
746 regionally restricted generation of different neuronal subtypes. In this view, the
747 differences in the ontogeny of the hindbrain and spinal cord establishes the distinct
748 genomic regulatory environment responsible for generating region-specific cell types
749 along the AP axis. This has implications for regenerative medicine and the design of
750 conditions that produce neurons with authentic molecular identities. The finding that
751 regionalisation is initiated and differences established prior to neural induction
752 highlights the importance of determining the appropriate culture conditions at early
753 stages of the directed differentiation of ESCs (Gouti et al., 2014; Lippmann et al.,
754 2015; Maury et al., 2014). Focusing on these time windows and endeavouring to
755 recapitulate normal developmental processes may contribute to deriving *in vitro*
756 differentiated neuronal subtypes that accurately mimic their *in vivo* counterparts.

757

758 The role of the WNT-CDX genetic network in the specification of caudal tissue has
759 been documented in a range of animals from across the bilaterian clade
760 (Chawengsaksophak et al., 2004; Faas and Isaacs, 2009; Morales et al., 1996). The
761 broad evolutionary conservation suggests that the network had this function in the
762 last common ancestor of all Bilateria (Ryan et al., 2007). Whether the role of WNT-
763 CDX activity in specifying distinct regions of the nervous system extends to non-
764 vertebrates is yet to be firmly established. Nevertheless, the divergent lineage and
765 distinct molecular events of the anterior and posterior nervous system is consistent
766 with the proposed dual evolutionary origins of the central nervous system (Arendt et
767 al., 2016). This hypothesis postulates that the bilaterian nervous system evolved
768 from the merger of nerve centres residing at opposite poles of the ancestral pre-
769 bilaterian animal (Arendt et al., 2016). In this view, the nervous system at the apical
770 pole of the ancestral animal had a primary sensory function and modulated body
771 physiology. Whereas the basally located blastoporal nervous system coordinated
772 feeding movements and locomotion (Figure 6B). The expansion and fusion of these
773 centres is proposed to have led to the bilaterian nerve cord and brain (Arendt et al.,
774 2016; Tosches and Arendt, 2013). Hence, the distinct molecular mechanisms that
775 specify anterior versus posterior vertebrate nervous systems may represent an

776 evolutionary vestige of the processes that once generated neural tissue in pre-
777 bilaterian animals.

778

779 **Author contributions**

780 VM and JB conceived the project, interpreted the data and wrote the manuscript. VM
781 designed and performed the experiments and performed data analysis. SS and EP
782 developed custom code and performed data analysis. DS assisted in the collection
783 of *in vivo* samples and constructed the *T/Bra* and *Cdx* mutant lines. MG performed
784 the initial differentiations with the *T/Bra* and *Cdx* mutant lines and critically revised
785 the manuscript. RLB provided access to animal colonies and performed critical
786 revision of the manuscript. NML interpreted the data and critically revised the
787 manuscript.

788

789 **Acknowledgements**

790 The authors would like to thank the Science Technology Platforms at the Francis
791 Crick Institute (and formerly at the National Institute for Medical Research). In
792 particular, we thank the Advanced Sequencing Facility, Bioinformatics And
793 BioStatistics Facility, the Flow Cytometry Facility, High Performance Computing and
794 the Biological Research Facility for their ongoing support and access to equipment.
795 We are very grateful to Karine Rizzoti for providing *Sox2eGFP* mice, Jean-Francois
796 Brunet for the *Phox2b* antibody and Bennett Novitch for the *Olig2* antibody. The
797 authors also acknowledge and thank the ENCODE Consortium and the production
798 labs of John Stamatoyannopoulos (UW) and Ross Hardison (PennState). We thank
799 Alfonso Martinez-Arias, Francois Guillemot, Jens Kleinjung and Karine Rizzoti for
800 comments on the manuscript and members of the Briscoe lab for useful discussion.
801 We thank Boris Lenhard, Julien Delile and Jens Kleinjung for critical feedback on
802 data analysis. This work was supported by the Francis Crick Institute, which receives
803 its core funding from Cancer Research UK (FC001051, FC0010110), the UK Medical
804 Research Council (FC001051, FC0010110), and the Wellcome Trust (FC001051,
805 FC0010110); JB and NML are funded by the Wellcome Trust (WT098326MA); JB by
806 the BBSRC (BB/J015539/1).

807 **References**

808 Alvarez-Medina, R., Cayuso, J., Okubo, T., Takada, S., and Martí, E. (2008). Wnt
809 canonical pathway restricts graded Shh/Gli patterning activity through the regulation

810 of Gli3 expression. *Development* 135, 237–247.

811 Amin, S., Neijts, R., Simmini, S., van Rooijen, C., Tan, S.C., Kester, L., van
812 Oudenaarden, A., Creighton, M.P., and Deschamps, J. (2016). Cdx and T
813 Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche.
814 *Cell Rep.* 17, 3165–3177.

815 Anderson, C., and Stern, C.D. (2016). Organizers in Development. *Curr. Top. Dev.*
816 *Biol.*, 117, 435-454.

817 Arendt, D., Tosches, M.A., and Marlow, H. (2016). From nerve net to nerve ring,
818 nerve cord and brain - evolution of the nervous system. *Nat Rev Neurosci* 17, 61–72.

819 Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-badge, R.
820 (2003). Multipotent cell lineages in early mouse development depend on SOX2
821 function. 126–140.

822 Brunelli, S., Casey, E.S., Bell, D., Harland, R., and Lovell-Badge, R. (2003).
823 Expression of SOX3 throughout the developing central nervous system is dependent
824 on the combined action of discrete, evolutionarily conserved regulatory elements.
825 *Genesis* 36, 12–24.

826 Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in
827 development: Lessons from genomics. *Trends Genet.* 28, 276–284.

828 Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013).
829 Transposition of native chromatin for fast and sensitive epigenomic profiling of open
830 chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–
831 1218.

832 Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A
833 Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.*
834 1–9.

835 Carcagno, a. L., Di Bella, D.J., Goulding, M., Guillemot, F., and Lanuza, G.M. (2014).
836 Neurogenin3 Restricts Serotonergic Neuron Differentiation to the Hindbrain. *J.*
837 *Neurosci.* 34, 15223–15233.

838 Chawengsaksophak, K., James, R., Hammond, V.E., Köntgen, F., and Beck, F.
839 (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* 386, 84–87.

840 Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J., and Beck, F.
841 (2004). Cdx2 is essential for axial elongation in mouse development. *Proc. Natl.*
842 *Acad. Sci.* 101, 7641–7645.

843 Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human

844 genome. *Nature* 489, 57–74.

845 Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L.,
846 Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-
847 specific and single-cell chromatin accessibility charts human hematopoiesis and
848 leukemia evolution. *Nat. Genet.* 48, 1193–1203.

849 Cordes, S.P. (2001). Molecular genetics of cranial nerve development in mouse. *Nat.*
850 *Rev. Neurosci.* 2, 611–623.

851 Cox, W.G., and Hemmati-Brivanlou, A. (1995). Caudalization of neural fate by tissue
852 recombination and bFGF. *Development* 121, 4349–4358.

853 Davis-Dusenbery, B.N., Williams, L. a, Klim, J.R., and Eggan, K. (2014). How to
854 make spinal motor neurons. *Development* 141, 491–501.

855 Doetschman, T., Gregg, R.G., Maeda, N., Hopper, M.L., Melton, D.W., Thompson,
856 S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse
857 embryonic stem cells. *Nature* 330, 576–578.

858 Le Dréau, G., and Martí, E. (2012). Dorsal-ventral patterning of the neural tube: A
859 tale of three signals. *Dev. Neurobiol.* 72, 1471–1481.

860 Ellis, P., Fagan, B.M., Magness, S.T., Hutton, S., Taranova, O., Hayashi, S.,
861 McMahon, A., Rao, M., and Pevny, L. (2004). SOX2, a persistent marker for
862 multipotential neural stem cells derived from embryonic stem cells, the embryo or the
863 adult. *Dev. Neurosci.* 26, 148–165.

864 Epstein, D.J., McMahon, A.P., and Joyner, A.L. (1999). Regionalization of Sonic
865 hedgehog transcription along the anteroposterior axis of the mouse central nervous
866 system is regulated by Hnf3-dependent and -independent mechanisms.
867 *Development* 126, 281–292.

868 Eyal–Giladi, H. (1954). Dynamic aspects of neural induction in Amphibia. *Arch. Biol.*
869 65, 179–259.

870 Faas, L., and Isaacs, H. V. (2009). Overlapping functions of cdx1, cdx2, and cdx4 in
871 the development of the amphibian *Xenopus tropicalis*. *Dev. Dyn.* 238, 835–852.

872 Garriock, R.J., Chalamalasetty, R.B., Kennedy, M.W., Canizales, L.C., Lewandoski,
873 M., and Yamaguchi, T.P. (2015). Lineage tracing of neuromesodermal progenitors
874 reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and
875 differentiation. *Development* 142, 1628–1638.

876 Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and
877 Briscoe, J. (2014). In Vitro Generation of Neuromesodermal Progenitors Reveals

878 Distinct Roles for Wnt Signalling in the Specification of Spinal Cord and Paraxial
879 Mesoderm Identity. *PLoS Biol.* 12, e1001937.

880 Gouti, M., Metzis, V., and Briscoe, J. (2015). The route to spinal cord cell types : a
881 tale of signals and switches. *Trends Genet.* 1–8.

882 Gouti, M., Delile, J., Stamatakis, D., Wymeersch, F.J., Huang, Y., Kleinjung, J.,
883 Wilson, V., and Briscoe, J. (2017). A Gene Regulatory Network Balances Neural and
884 Mesoderm Specification during Vertebrate Trunk Development. *Dev. Cell* 1–19.

885 Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong,
886 X., Akalin, A., Suzuki, A.M., et al. (2014). Two independent transcription initiation
887 codes overlap on vertebrate core promoters. *Nature* 507, 381–385.

888 Hagey, D.W., Zaouter, C., Combeau, G., Lendahl, M.A., Andersson, O., Huss, M.,
889 and Muhr, J. (2016). Distinct transcription factor complexes act on a permissive
890 chromatin landscape to establish regionalized gene expression in CNS stem cells.
891 *Genome Res.* 26, 908–917.

892 Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X.,
893 Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-
894 Determining Transcription Factors Prime cis-Regulatory Elements Required for
895 Macrophage and B Cell Identities. *Mol. Cell* 38, 576–589.

896 Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal
897 progenitors and the making of the spinal cord. *Development* 142, 2864–2875.

898 Jessell, T.M. (2000). Neuronal specification in the spinal cord: inductive signals and
899 transcriptional codes. *Nat. Rev. Neurosci.* 20, 20–29.

900 Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate
901 specification and differentiation. *Development* 140, 4129–4144.

902 Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A.,
903 van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2017).
904 JASPAR 2018: update of the open-access database of transcription factor binding
905 profiles and its web framework. *Nucleic Acids Res.* gkx1126.

906 Kim, F.A., Sing, A., Kaneko, T., Bieman, M., Stallwood, N., Sadl, V.S., and Cordes,
907 S.P. (2005). The vHNF1 homeodomain protein establishes early rhombomere
908 identity by direct regulation of Kreisler expression. *Mech. Dev.* 122, 1300–1309.

909 Kirkeby, A., Grealish, S., Wolf, D.A., Nelander, J., Wood, J., Lundblad, M., Lindvall,
910 O., and Parmar, M. (2012). Generation of Regionally Specified Neural Progenitors
911 and Functional Neurons from Human Embryonic Stem Cells under Defined

912 Conditions. *Cell Rep.* *1*, 703–714.

913 Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann,
914 B., Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and
915 Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev. Cell.*

916 Kolm, P.J., Apekin, V., and Sive, H. (1997). *Xenopus* Hindbrain Patterning Requires
917 Retinoid Signaling. *Dev. Biol.* *192*, 1–16.

918 Kondoh, H., Takada, S., and Takemoto, T. (2016). Axial level-dependent molecular
919 and cellular mechanisms underlying the genesis of the embryonic neural plate. *Dev.*
920 *Growth Differ.* *58*, 427–436.

921 Koshida, S., Shinya, M., Mizuno, T., Kuroiwa, a, and Takeda, H. (1998). Initial
922 anteroposterior pattern of the zebrafish central nervous system is determined by
923 differential competence of the epiblast. *Development* *125*, 1957–1966.

924 Kutejova, E., Sasai, N., Shah, A., Gouti, M., and Briscoe, J. (2016). Neural
925 Progenitors Adopt Specific Identities by Directly Repressing All Alternative
926 Progenitor Transcriptional Programs. *Dev. Cell* 1–15.

927 Lamb, T.M., and Harland, R.M. (1995). Fibroblast growth factor is a direct neural
928 inducer, which combined with noggin generates anterior-posterior neural pattern.
929 *Development* *121*, 3627–3636.

930 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2.
931 *Nat. Methods* *9*, 357–359.

932 Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M.,
933 Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are
934 shaped by the local microenvironment. *Cell* *159*, 1312–1326.

935 Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R.,
936 Morgan, M.T., and Carey, V.J. (2013). Software for Computing and Annotating
937 Genomic Ranges. *PLoS Comput. Biol.* *9*, e1003118.

938 Lei, Q., Jeong, Y., Misra, K., Li, S., Zelman, A.K., Epstein, D.J., and Matisse, M.P.
939 (2006). Wnt Signaling Inhibitors Regulate the Transcriptional Response to
940 Morphogenetic Shh-Gli Signaling in the Neural Tube. *Dev. Cell* *11*, 325–337.

941 Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring Reproducibility Of
942 High Throughput Experiments. 1–36.

943 Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: Fast, accurate and
944 scalable read mapping by seed-and-vote. *Nucleic Acids Res.* *41*.

945 Lippmann, E.S., E.williams, C., Ruhl, D.A., Estevez-Silva, M.C., Chapman, E.R.,

- 946 Coon, J.J., and Ashton, R.S. (2015). Deterministic HOX patterning in human
947 pluripotent stem cell-derived neuroectoderm. *Stem Cell Reports* 4, 632–644.
- 948 Liu, J.P., Laufer, E., and Jessell, T.M. (2001). Assigning the positional identity of
949 spinal motor neurons: Rostrocaudal patterning of Hox-c expression by FGFs, Gdf11,
950 and retinoids. *Neuron* 32, 997–1012.
- 951 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change
952 and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- 953 Mahony, S., Mazzoni, E.O., McCuine, S., Young, R. a, Wichterle, H., and Gifford,
954 D.K. (2011). Ligand-dependent dynamics of retinoic acid receptor binding during
955 early neurogenesis. *Genome Biol.* 12, R2.
- 956 Mangold, O. (1933). Über die Induktionsfähigkeit der verschiedenen Bezirke der
957 Neurula von Urodelen. *Naturwissenschaften* 21, 761–766.
- 958 Martik, M.L., and Bronner, M.E. (2017). Regulatory Logic Underlying Diversification
959 of the Neural Crest. *Trends Genet.* 33, 715–727.
- 960 Martin, B.L., and Kimelman, D. (2012). Canonical Wnt Signaling Dynamically
961 Controls Multiple Stem Cell Fate Decisions during Vertebrate Body Formation. *Dev.*
962 *Cell* 22, 223–232.
- 963 Maury, Y., Côme, J., Piskorowski, R. a, Salah-Mohellibi, N., Chevaleyre, V.,
964 Peschanski, M., Martinat, C., and Nedelec, S. (2014). Combinatorial analysis of
965 developmental cues efficiently converts human pluripotent stem cells into multiple
966 neuronal subtypes. *Nat. Biotechnol.* 1–10.
- 967 Mazzoni, E.O., Mahony, S., Peljto, M., Patel, T., Thornton, S.R., McCuine, S.,
968 Reeder, C., Boyer, L.A., Young, R.A., Gifford, D.K., et al. (2013). Saltatory
969 remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat.*
970 *Neurosci.* 16, 1191–1198.
- 971 McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger,
972 A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-
973 regulatory regions. *Nat. Biotechnol.* 28, 495–501.
- 974 Morales, A. V., De La Rosa, E.J., and De Pablo, F. (1996). Expression of the cCdx-B
975 homeobox gene in chick embryo suggests its participation in rostrocaudal axial
976 patterning. *Dev. Dyn.* 206, 343–353.
- 977 Mudge, J.M., and Harrow, J. (2015). Creating reference gene annotation for the
978 mouse C57BL6/J genome assembly. *Mamm. Genome* 26, 366–378.
- 979 Muhr, J., Graziano, E., Wilson, S., Jessell, T.M., and Edlund, T. (1999). Convergent

- 980 Inductive Signals Specify Midbrain , Hindbrain , and Spinal Cord Identity in Gastrula
981 Stage Chick Embryos. *Neuron* 23, 689–702.
- 982 Muroyama, Y., Fujihara, M., Ikeya, M., Kondoh, H., and Takada, S. (2002). Wnt
983 signaling plays an essential role in neuronal specification of the dorsal spinal cord.
984 *Genes Dev.* 16, 548–553.
- 985 Neijts, R., Amin, S., Van Rooijen, C., Tan, S., Creyghton, M.P., de Laat, W., and
986 Deschamps, J. (2016). Polarized regulatory landscape and Wnt responsiveness
987 underlie Hox activation in embryos. *Genes Dev.* 30, 1937–1942.
- 988 Neijts, R., Amin, S., van Rooijen, C., and Deschamps, J. (2017). Cdx is crucial for
989 the timing mechanism driving colinear Hox activation and defines a trunk segment in
990 the Hox cluster topology. *Dev. Biol.* 422, 146–154.
- 991 Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P., and Dollé, P. (2000).
992 Retinoic acid synthesis and hindbrain patterning in the mouse embryo. *Development*
993 127, 75–85.
- 994 Nieuwkoop, P.D. (1952). Activation and organization of the central nervous system in
995 amphibians. Part III. Synthesis of a new working hypothesis. *J. Exp. Zool.* 120, 83–
996 108.
- 997 Nieuwkoop, P.D., and Nigtevecht, G. V (1954). Neural activation and transformation
998 in explants of competent ectoderm under the influence of fragments of anterior
999 notochord in urodeles. *J. Embryol. Exp. Morph.* 2, 175–193.
- 1000 Nordström, U., Maier, E., Jessell, T.M., and Edlund, T. (2006). An early role for Wnt
1001 signaling in specifying neural patterns of Cdx and Hox gene expression and motor
1002 neuron subtype identity. *PLoS Biol.* 4, 1438–1452.
- 1003 Novitch, B.G., Chen, A.I., and Jessell, T.M. (2001). Coordinate regulation of motor
1004 neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2.
1005 *Neuron* 31, 773–789.
- 1006 Nowotschin, S., Ferrer-Vaquero, A., Concepcion, D., Papaioannou, V.E., and
1007 Hadjantonakis, A.-K. (2012). Interaction of Wnt3a, Msn1 and Tbx6 in neural versus
1008 paraxial mesoderm lineage commitment and paraxial mesoderm differentiation in the
1009 mouse embryo. *Dev. Biol.* 367, 1–14.
- 1010 Olivera-Martinez, I., Harada, H., Halley, P. a, and Storey, K.G. (2012). Loss of FGF-
1011 dependent mesoderm identity and rise of endogenous retinoid signalling determine
1012 cessation of body axis elongation. *PLoS Biol.* 10, e1001415.
- 1013 Oosterveen, T., Kurdija, S., Alekseenko, Z., Uhde, C.W., Bergsland, M., Sandberg,

- 1014 M., Andersson, E., Dias, J.M., Muhr, J., and Ericson, J. (2012). Mechanistic
1015 Differences in the Transcriptional Interpretation of Local and Long-Range Shh
1016 Morphogen Signaling. *Dev. Cell* 23, 1006–1019.
- 1017 Pattyn, A., Morin, X., Cremer, H., Goridis, C., and Brunet, J.-F. (1997). Expression
1018 and interactions of the two closely related homeobox genes Phox2a and Phox2b
1019 during neurogenesis. *Development* 124, 4065–4075.
- 1020 Pattyn, A., Morin, X., Cremer, H., Goridis, C., and Brunet, J.F. (1999). The
1021 homeobox gene Phox2b is essential for the development of autonomic neural crest
1022 derivatives. *Nature* 399, 366–370.
- 1023 Pattyn, A., Hirsch, M., Goridis, C., and Brunet, J.F. (2000). Control of hindbrain
1024 motor neuron differentiation by the homeobox gene Phox2b. *Development* 127,
1025 1349–1358.
- 1026 Peterson, K. a, Nishi, Y., Ma, W., Vedenko, A., Shokri, L., Zhang, X., McFarlane, M.,
1027 Baizabal, J.-M., Junker, J.P., van Oudenaarden, A., et al. (2012). Neural-specific
1028 Sox2 input and differential Gli-binding affinity provide context and positional
1029 information in Shh-directed neural patterning. *Genes Dev.* 26, 2802–2816.
- 1030 Pique-Regi, R., Degner, J.F., Pai, A. a, Gaffney, D.J., Gilad, Y., and Pritchard, J.K.
1031 (2011). Accurate inference of transcription factor binding from DNA sequence and
1032 chromatin accessibility data. *Genome Res.* 21, 447–455.
- 1033 Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for
1034 comparing genomic features. *Bioinformatics* 26, 841–842.
- 1035 Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne,
1036 S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for
1037 deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165.
- 1038 Rashbass, P., Cooke, L.A., Herrmann, B.G., and Beddington, R.S. (1991). A cell
1039 autonomous function of Brachyury in T/T embryonic stem cell chimaeras. *Nature* 353,
1040 348–351.
- 1041 van Rooijen, C., Simmini, S., Bialecka, M., Neijts, R., van de Ven, C., Beck, F., and
1042 Deschamps, J. (2012a). Evolutionarily conserved requirement of Cdx for post-
1043 occipital tissue emergence. *Development* 139, 2576–2583.
- 1044 van Rooijen, C., Simmini, S., Bialecka, M., Neijts, R., van de Ven, C., Beck, F., and
1045 Deschamps, J. (2012b). Evolutionarily conserved requirement of Cdx for post-
1046 occipital tissue emergence. *Development* 139, 2576–2583.
- 1047 Ryan, J.F., Mazza, M.E., Pang, K., Matus, D.Q., Baxevanis, A.D., Martindale, M.Q.,

1048 and Finnerty, J.R. (2007). Pre-bilaterian origins of the hox cluster and the hox code:
1049 Evidence from the sea anemone, *Nematostella vectensis*. *PLoS One*. 2, e153.

1050 Santagati, F., and Rijli, F.M. (2003). Cranial neural crest and the building of the
1051 vertebrate head. *Nat. Rev. Neurosci.* 4, 806–818.

1052 Sheffield, N.C., and Bock, C. (2015). LOLA: Enrichment analysis for genomic region
1053 sets and regulatory elements in R and Bioconductor. *Bioinformatics* 32, 587–589.

1054 Shirasaki, R., and Pfaff, S.L. (2002). Transcriptional Codes and the Control of
1055 Neuronal Identity. *Annu. Rev. Neurosci.* 25, 251–281.

1056 Simoes-Costa, M., and Bronner, M.E. (2016). Reprogramming of avian neural crest
1057 axial identity and cell fate. *Science* (80-.). 352, 1570–1573.

1058 Skromne, I., Thorsen, D., Hale, M., Prince, V.E., and Ho, R.K. (2007). Repression of
1059 the hindbrain developmental program by Cdx factors is required for the specification
1060 of the vertebrate spinal cord. *Development* 134, 2147–2158.

1061 Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C.,
1062 Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the
1063 ENCODE portal. *Nucleic Acids Res.* 44, D726–D732.

1064 Song, L., and Crawford, G.E. (2010). DNase-seq: A high-resolution technique for
1065 mapping active gene regulatory elements across the genome from mammalian cells.
1066 *Cold Spring Harb. Protoc.* 5.

1067 Soucie, E.L., Weng, Z., Geirsdóttir, L., Molawi, K., Maurizio, J., Fenouil, R.,
1068 Mossadegh-Keller, N., Gimenez, G., VanHille, L., Beniazza, M., et al. (2016).
1069 Lineage-specific enhancers activate self-renewal genes in macrophages and
1070 embryonic stem cells. *Science* 351, aad5510.

1071 Spemann, H., and Mangold, H. (1924). Über Induktion von Embryonalanlagen durch
1072 Implantation artfremder Organisatoren. *Roux's Arch. EntwMech. Org.* 100, 599–638.

1073 Stern, C.D. (2001). Initial patterning of the central nervous system: how many
1074 organizers? *Nat. Rev. Neurosci.* 2, 92–98.

1075 Stern, C.D. (2005). Neural induction: old problem, new findings, yet more questions.
1076 *Development* 132, 2007–2021.

1077 Stern, C.D. (2006). Neural induction: 10 years on since the “default model”. *Curr.*
1078 *Opin. Cell Biol.* 18, 692–697.

1079 Stern, C.D., Charité, J., Deschamps, J., Duboule, D., Durston, A.J., Kmita, M.,
1080 Nicolas, J.F., Palmeirim, I., Smith, J.C., and Wolpert, L. (2006). Head-tail patterning
1081 of the vertebrate embryo: One, two or many unresolved problems? *Int. J. Dev. Biol.*

1082 50, 3–15.

1083 Storey, K.G., Crossley, J.M., DeRobertis, E.M., Norris, W.E., and Stern, C.D. (1992).
1084 Neural induction and regionalization of the chick embryo. *Development* 114, 729–
1085 741.

1086 Sturgeon, K., Kaneko, T., Biemann, M., Gauthier, A., Chawengsaksophak, K., and
1087 Cordes, S.P. (2011). Cdx1 refines positional identity of the vertebrate hindbrain by
1088 directly repressing *Mafb* expression. *Development* 138, 65–74.

1089 Takada, S., Stark, K.L., Shea, M.J., Vassileva, G., McMahon, J.A., and McMahon,
1090 A.P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo.
1091 *Genes Dev.* 8, 174–189.

1092 Toivonen, S., and Saxen, L. (1968). Morphogenetic interaction of presumptive neural
1093 and mesodermal cells mixed in different ratios. *Science* (80-). 159, 539–540.

1094 Törönen, P., Kolehmainen, M., Wong, G., and Castrén, E. (1999). Analysis of gene
1095 expression data using self-organizing maps. *FEBS Lett.* 451, 142–146.

1096 Tosches, M.A., and Arendt, D. (2013). The bilaterian forebrain: An evolutionary
1097 chimaera. *Curr. Opin. Neurobiol.* 23, 1080–1089.

1098 Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R.,
1099 Economou, C., Karagianni, E., Zhao, S., Lowell, S., et al. (2014). Distinct Wnt-driven
1100 primitive streak-like populations reflect in vivo lineage precursors. *Development* 141,
1101 1209–1221.

1102 Turner, D. a, Hayward, P.C., Baillie-Johnson, P., Rué, P., Broome, R., Faunes, F.,
1103 and Martinez Arias, A. (2014). Wnt/ β -catenin and FGF signalling direct the
1104 specification and maintenance of a neuromesodermal axial progenitor in ensembles
1105 of mouse embryonic stem cells. *Development* 141, 4243–4253.

1106 Tzouanacou, E., Wegener, A., Wymeersch, F.J., Wilson, V., and Nicolas, J.-F.
1107 (2009). Redefining the progression of lineage segregations during mammalian
1108 embryogenesis by clonal analysis. *Dev. Cell* 17, 365–376.

1109 Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y., and Kondoh, H. (2003).
1110 Functional analysis of chicken *Sox2* enhancers highlights an array of diverse
1111 regulatory elements that are conserved in mammals. *Dev. Cell* 4, 509–519.

1112 Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer
1113 Browser - A database of tissue-specific human enhancers. *Nucleic Acids Res.* 35.

1114 Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J. a, Holt, A., Plajzer-Frick, I.,
1115 Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-

- 1116 specific activity of enhancers. *Nature* 457, 854–858.
- 1117 Waddington, C.H. (1940). *Organisers and Genes*. Cambridge Univ. Press. 1905–
1118 1975.
- 1119 Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos.
1120 *Bioinformatics* 33, 3645–3647.
- 1121 Wang, H., Lei, Q., Oosterveen, T., Ericson, J., and Matise, M.P. (2011). Tcf/Lef
1122 repressors differentially regulate Shh-Gli target gene activation thresholds to
1123 generate progenitor patterning in the developing CNS. *Development* 138, 3711–
1124 3721.
- 1125 Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed
1126 differentiation of embryonic stem cells into motor neurons. *Cell* 110, 385–397.
- 1127 Wilson, S.I., Rydström, A., Trimborn, T., Willert, K., Nusse, R., Jessell, T.M., and
1128 Edlund, T. (2001). The status of Wnt signalling regulates neural and epidermal fates
1129 in the chick embryo. *Nature* 411, 325–330.
- 1130 Wood, H.B., and Episkopou, V. (1999). Comparative expression of the mouse Sox1 ,
1131 Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mech. Dev.* 86,
1132 197–201.
- 1133 Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q.,
1134 Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian
1135 preimplantation embryos. *Nature* 534, 652–657.
- 1136 Wymeersch, F.J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F.C.K., and
1137 Wilson, V. (2016). Position-dependent Plasticity of Distinct Progenitor Types in the
1138 Primitive Streak. *5*, e10042.
- 1139 Yamaguchi, T.P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A.P. (1999). T
1140 (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification.
1141 *Genes Dev.* 13, 3185–3190.
- 1142 Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K.,
1143 and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the
1144 totipotent cycle of embryonal cells. *Development* 122, 881–894.
- 1145 Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of
1146 embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat.*
1147 *Biotechnol.* 21, 183–186.
- 1148 Young, T., Rowland, J.E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M.,
1149 van Nes, J., de Graaff, W., Duluc, I., Freund, J.N., et al. (2009). Cdx and Hox Genes

1150 Differentially Regulate Posterior Axial Growth in Mammalian Embryos. *Dev. Cell* 17,
1151 516–526.
1152 Zhang, Y., Liu, T., Meyer, C. a, Eeckhoute, J., Johnson, D.S., Bernstein, B.E.,
1153 Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of
1154 ChIP-Seq (MACS). *Genome Biol.* 9, R137.
1155

1156 **Methods**

1157 **Cell lines**

1158 All WT ESC culture was performed using the HM1 line (Doetschman et al., 1987).
1159 *Bra*^{-/-} and *Cdx*^{1,2,4/-} knockout ESC lines were generated in the HM1 line using
1160 CRISPR as previously described (Gouti et al., 2017). Single guide RNAs were used
1161 to target the T-box domain (*T/Bra* mutant), and the caudal-like activation domain of
1162 *Cdx1*, *Cdx2* and *Cdx4* (*Cdx*^{1,2,4/-} triple mutant). Cell lines were validated by DNA
1163 sequencing and western blotting and routinely tested for mycoplasma.

1164

1165 **Cell culture and neural progenitor differentiation**

1166 Mouse ESCs were propagated on mitotically inactivated mouse embryonic
1167 fibroblasts (feeders) in DMEM knockout medium supplemented with 1000U/ml LIF
1168 (Chemicon), 10% cell-culture validated fetal bovine serum, penicillin/streptomycin,
1169 2mM L-glutamine (Gibco). To obtain neural progenitors with anterior, hindbrain or
1170 posterior neural identity, ESCs were differentiated as previously described (Gouti et
1171 al., 2014). Briefly, ESCs were dissociated with 0.05% trypsin, and plated on gelatin-
1172 coated plates for two sequential 20 minute periods in ESC medium to separate them
1173 from their feeder layer cells which adhere to the plastic. To start the differentiation,
1174 cells remaining in the supernatant were pelleted by centrifugation, washed in PBS,
1175 and pelleted again. Cells were counted and resuspended in N2B27 medium
1176 containing 10ng/ml bFGF to a concentration of 10⁶ cells per ml, and 50,000 cells per
1177 35mm CELLBIND dish (Corning) were plated. N2B27 medium contained a 1:1 ratio
1178 of Advanced Dulbecco's Modified Eagle Medium F12:Neurobasal medium (Gibco)
1179 supplemented with 1xN2 (Gibco), 1xB27 (Gibco), 2mM L-glutamine (Gibco), 40
1180 µg/ml BSA (Sigma), penicillin/streptomycin and 0.1mM 2-mercaptoethanol. To
1181 generate anterior neural progenitors, the cells were grown up to day 3 in N2B27 +
1182 10ng/ml bFGF, followed by N2B27 + 500nM smoothed agonist (SAG; Calbiochem)
1183 from day 3-5. To generate hindbrain neural progenitors, cells were cultured under
1184 the same conditions as the anterior, but were additionally exposed to 100nM retinoic
1185 acid (RA; Sigma) from day 3-5. To generate spinal cord neural progenitors, cells
1186 were cultured with N2B27 + 10ng/ml bFGF until day 2, N2B27 + 10ng/ml bFGF +
1187 5µM CHIR99021 (Axon) until day 3, and N2B27 + 100nM RA + 500nM SAG until day
1188 5. For Hindbrain+ treated cells (Fig. 3), cells were differentiated under hindbrain
1189 conditions with one modification between day 4-5, where they were additionally

1190 exposed to 10ng/ml bFGF and 5 μ M CHIR99021 in addition to continued treatment
1191 with RA and SAG as above. For all differentiations, media changes were made every
1192 24 hours from day 2.

1193

1194 **Immunofluorescence and microscopy**

1195 Cells were washed in PBS and fixed in 4% paraformaldehyde in PBS for 15min at 4
1196 degrees, followed by two washes in PBS and one wash in PBST (0.1% Triton X-100
1197 diluted in PBS). Primary antibodies were applied overnight at 4 degrees diluted in
1198 filter-sterilized blocking solution (3% FBS diluted in PBST). Cells were washed 3x in
1199 PBST and secondary antibodies (AlexaFluor conjugated; Invitrogen) were applied at
1200 room temperature, diluted 1:1000 in PBS for 1hr. Cells were washed 3x in PBS,
1201 incubated with DAPI for 5 min in PBS and washed twice before mounting with
1202 Prolong Gold (Invitrogen). Primary antibodies were diluted as follows: Phox2b rabbit,
1203 kindly provided by Jean-Francois Brunet (Pattyn et al., 1997), 1/1000; Olig2 rabbit
1204 (Millipore, AB9610; 1/1000); Olig2 guinea-pig, kindly provided by Ben Novitch
1205 (Novitch et al., 2001), 1/1000, and Sox2 goat (R&D Systems, AF2018; 1/500). Cells
1206 were imaged on a Zeiss Imager.Z2 microscope using the ApoTome.2 structured
1207 illumination platform. Z-stacks were acquired and represented as maximum intensity
1208 projections using ImageJ software. Immunofluorescence was performed on a
1209 minimum of 3 biological replicates, from independent experiments.

1210

1211 **RNA extraction, cDNA synthesis and qPCR analysis**

1212 RNA was extracted from cells using a Qiagen RNeasy kit, following the
1213 manufacturer's instructions. Extracts were digested with DNase I to eliminate
1214 genomic DNA. First strand cDNA synthesis was performed using Superscript III
1215 (Invitrogen) using random hexamers and was amplified using Platinum SYBR-Green
1216 (Invitrogen). qPCR was performed using the Applied Biosystems 7900HT Fast Real
1217 Time PCR. PCR primers were designed using NCBI primer blast or primer3 software,
1218 using exon-spanning junctions where possible. Expression values for each gene
1219 were normalised against β -actin, using the delta-delta CT method. Error bars
1220 represent standard deviation across three biological replicate samples. qPCR was
1221 performed on 3 biological replicates for every primer pair analysed. Primer
1222 sequences are available in Table 4.

1223

1224 **ATAC-seq**

1225 ATAC-seq was performed on ESCs and at each day of the differentiation following
1226 methods previously described (Buenrostro et al., 2013, 2015). Briefly, adherent cells
1227 were treated with StemPro Accutase (A1110501) to obtain a single cell suspension.
1228 Cells were counted and resuspended to obtain 50,000 cells per sample in ice-cold
1229 PBS. Cells were pelleted and resuspended in lysis buffer (10mM Tris-HCl pH 7.4,
1230 10mM NaCl, 3mM MgCl₂, 0.1% IGEPAL). Following a 10 min centrifugation at 4°C,
1231 nucleic extracts were resuspended in transposition buffer for 30 minutes at 37°C and
1232 purified using a Qiagen MinElute PCR Purification kit following manufacturer's
1233 instructions. Transposed DNA was eluted in a 10µl volume and amplified by PCR
1234 with Nextera primers (Buenrostro et al., 2013) to generate single-indexed libraries. A
1235 maximum of 12 cycles of PCR (determined using optimisation experiments) was
1236 used to prevent saturation biases based on optimisation of qPCR cycles as
1237 previously described (Buenrostro et al., 2015). Library quality control was carried out
1238 using the Bioanalyzer High-Sensitivity DNA analysis kit. Libraries were sequenced
1239 as paired-end 50 or 100 bp reads, multiplexing 4 samples per lane on the Illumina
1240 High-Seq 2500 platform at the Francis Crick Institute Advanced Sequencing Facility.
1241 For all conditions, two biological replicate samples were collected from independent
1242 experiments.

1243

1244 ***In vivo* ATAC-seq and mouse lines**

1245 *Sox2eGFP* heterozygous mice (Ellis et al., 2004) were maintained on a C57BL6
1246 background. To obtain embryos, *Sox2eGFP* heterozygous mice were mated to wild
1247 type mice. Embryos for ATAC-seq were harvested at e9.5 in HBSS buffer (GIBCO)
1248 containing 5% FBS. As the ratio of cells to transposase is a critical parameter in
1249 generating ATAC-seq results (Buenrostro et al., 2015), we aimed to use the same
1250 ratio of cells *in vivo* as *in vitro* for maximum comparability. To obtain sufficient
1251 quantities of cells from *in vivo*, embryos from several litters were pooled together and
1252 screened for GFP using a Leica MZFL widefield microscope with a GFP filter set.
1253 Embryos were separated into GFP positive and negative pools. To enrich for anterior
1254 and spinal cord neural progenitors, GFP positive embryos were dissected as follows:
1255 heads were decapitated at the second pharyngeal arch and otocysts removed to

1256 avoid contamination with other *GFP*-expressing cells. To obtain spinal cord NPs, the
1257 neural tube and surrounding somitic tissue was dissected, from the level of caudal
1258 hindbrain to the tailbud posterior neuropore. Both cranial and trunk regions were
1259 minced with forceps, incubated for 5 minutes on ice in enzyme-free dissociation
1260 buffer (Gibco) and then gently passed through a 40 μ m filter using the plunger from a
1261 sterile syringe. Dissociated cells were collected, centrifuged at 4°C for 5 minutes at
1262 1500rpm and resuspended in 500ul HBSS buffer containing 5% FBS. Cells were
1263 passed through a 40 μ m filter and sorted using flow cytometry. Flow analysis and
1264 sorting was performed by the Francis Crick Flow Cytometry facility, using an Aria
1265 Fusion cell sorter with a 488nm laser. GFP negative cells (obtained from negative
1266 littermates collected in parallel) were used as a negative control to set voltage gating.
1267 50,000 GFP positive cells from anterior and spinal cord levels obtained from FACS
1268 were subject to ATAC-seq as described for in vitro-derived cells. Duplicate samples
1269 were collected on independent days to represent biological repeats. All animal
1270 procedures were performed in accordance with the Animal (Scientific Procedures)
1271 Act 1986 under the UK Home Office project licences PPL80/2528 and PD415DD17.

1272

1273 **ChIP-seq**

1274 Sox2 ChIP-seq (Santa cruz antibody SC-17320X) was performed in duplicate as
1275 previously described (Kutejova et al., 2016). Briefly, 10-30 million cells from day 5
1276 hindbrain or day 5 spinal cord neural progenitors were crosslinked in 1%
1277 formaldehyde for 20 min at 4 degrees. Chromatin was sonicated using a Diagenode
1278 Bioruptor (using a cycle of 30sec on, 30 sec off) until fragments were between 200-
1279 400bp. 3ug antibody was incubated together with cell lysate overnight at 4°C on a
1280 rotating wheel. Immunoprecipitation of chromatin fragments were captured using
1281 Protein G-coupled Dynabeads (Life Technologies). Samples were decrosslinked and
1282 purified using the Qiagen MinElute kit. Approximately 10ng ChIP DNA and 10ng
1283 input DNA for each condition was used to prepare ChIP-seq libraries using the KAPA
1284 Hyper Prep Kit (Illumina). Biological duplicates were obtained for both conditions
1285 from separate experiments. Libraries were sequenced as single-end, 50bp reads on
1286 the Illumina High-Seq 2500 platform (Francis Crick Institute).

1287

1288

1289 **Data analysis**

1290

1291 **ATAC-seq pre-processing**

1292 Sequencing adapters and poor quality bases were trimmed from reads using
1293 *trim_galore* with default settings
1294 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were
1295 aligned to the *mm10* reference genome using *bowtie2* with parameters: *-X 2000 ---*
1296 *sensitive-local* (Langmead and Salzberg, 2012). Alignments were filtered for
1297 unmapped, low quality (MAPQ < 30), mapping to chrM or not properly paired
1298 fragments. PCR duplicates were marked using *Piccard* and not used in subsequent
1299 analysis steps.

1300 Signal tracks were computed as *fragments per million per base pair* (FPM) using
1301 *deepTools bamCoverage* with following settings: *--scaleFactor 10⁶/Library size -bs 1*
1302 *-extendReads -samFlagInclude 66 -ignoreDuplicates* (Ramírez et al., 2016).

1303 Read enriched regions were called with MACS2 using the options *-g mm -p 0.01 -*
1304 *nomodel -f BAMPE* (Zhang et al., 2008).

1305 For analysis on insertion level, such as TF footprinting, we adjusted plus-strand
1306 insertions by +4bp and minus-strand insertions by -5bp (Buenrostro et al., 2013).

1307

1308 **ATAC-seq - differential and clustering analysis**

1309 Initially, a robust peak set per condition was obtained by computing the
1310 irreproducible discovery rate (IDR) between duplicates and thresholding for peaks
1311 with an IDR <= 0.1 (Li et al., 2011). These robust peak sets were merged to one
1312 consensus region set using *bedtools merge* (Quinlan and Hall, 2010). We computed
1313 the fragment coverage for all samples of these regions using *featureCounts* with
1314 additional options *-F SAF -p -ignoreDup* (Liao et al., 2013).

1315 The resulting count table was used as input for a differential chromatin accessibility
1316 analysis with DESeq2 (default settings) to compare pairwise all WT in vitro
1317 conditions with D0 (ESC) (Love et al., 2014). We obtained a set of variable regions
1318 by filtering for differential peaks with $\log_2(\text{FC}) > 1$ and adjusted p-value < 0.01.

1319 These variable regions were clustered in order to resolve the complex chromatin
1320 dynamics from our multiple condition time series data. For this, DESeq2 normalized
1321 count data was transformed by computing region-wise z-scores ($z=(x -$
1322 $\text{mean}(x))/\text{sd}(x)$). Variable regions were clustered using the z-score matrix as input

1323 for a self-organizing map (SOM) with a grid of 5x5 cells, a hexagonal topology and a
1324 Gaussian neighborhood function. We found that 5x5 best reflects the complex
1325 chromatin dynamics, though different grid sizes largely reproduced the same
1326 dynamics. SOM clusters were merged by similar trends and manually annotated
1327 using known regulatory regions.

1328

1329 **ATAC-seq - peak annotation**

1330 Consensus regions were annotated with nearby genes using *ChIPseeker*
1331 *annotatePeak* with GENCODE release M14 (Mudge and Harrow, 2015). Regions
1332 were further annotated to nearby genes using default settings (basal rule) in GREAT
1333 (McLean et al., 2010). Each gene is assigned to a regulatory region spanning 5kb
1334 upstream and 1kb downstream of the TSS (irrespective of other genes). This
1335 regulatory domain is extended in both directions to all nearest genes, up to a
1336 maximum of 1000kb (McLean et al., 2010).

1337

1338 **ATAC-seq – *In vitro* vs *in vivo* comparative analysis**

1339 Fragment coverage of *in vitro* consensus regions for *in vivo* ATAC-seq experiments
1340 was computed using *featureCounts* with the same options as above. We normalized
1341 counts from *in vivo* and corresponding *in vitro* samples using DESeq2 with default
1342 settings. The normalized counts were used to compute log₂ fold-changes between
1343 conditions for *in vivo* and *in vitro* respectively. These were filtered for regions falling
1344 into the SOM cluster: Anterior, Neural and Spinal cord. Distribution of log₂ fold-
1345 changes were compared with two-sided Wilcoxon rank sum tests. ATAC-seq meta
1346 profiles were plotted for regions which were enriched in both *in vitro* and the
1347 corresponding *in vivo* sample ($\log_2(\text{FC}) > 0.5$).

1348

1349 **ATAC-seq - ENCODE DHS overlap**

1350 DNase hypersensitive sites (DHS) for various tissues, on the mm10 reference
1351 genome, were obtained from the ENCODE data portal (Consortium, 2012; Sloan et
1352 al., 2016). DHS regions were overlapped with our SOM clustered regions using
1353 *GenomicRanges findOverlaps* (Lawrence et al., 2013).

1354

1355 **ATAC-seq - Vista enhancer enrichment**

1356 Vista enhancer regions were downloaded from the ENCODE data portal (Consortium,
1357 2012; Sloan et al., 2016) (<https://www.encodeproject.org/>). Enhancers were
1358 overlapped with our SOM clustered regions and enrichment was examined with a
1359 one-sided binomial test, similar to methods used for GREAT analysis (McLean et al.,
1360 2010). Resulting p-values were corrected for multiple testing using the Benjamini-
1361 Hochberg procedure.

1362

1363 **ATAC-seq - motif enrichment**

1364 Regions were centered on the summit, rescaled to 300bp using *bedtools slop* and
1365 sequences were queried by *bedtools getfasta* from the mm10 reference genome.
1366 Resulting fasta files were used as input for *HOMER findMotifs.pl* with parameters –
1367 *bits -mset vertebrates* (Heinz et al., 2010). TF motifs were visualized using
1368 ggSeqLogo (Wagih, 2017).

1369

1370 **ATAC-seq - TF footprinting**

1371 TF binding motifs for factors of interest were queried from the JASPAR database
1372 (Khan et al., 2017). We matched these motifs against genomic sequences using
1373 motifmatchr to obtain their genomic positions
1374 (<https://github.com/GreenleafLab/motifmatchr>). Two strategies were used depending
1375 on the question: either the motif was genome-wide matched (for CTCF; Figure S1F)
1376 or +/-5kbp around peak summits was used (for CDX; Figure S4D).
1377 Resulting motif positions were extended +/- 150bp. Adjusted Tn5 insertions from
1378 fragments <= 100bp were counted per base-pair and strand. For the footprinting, we
1379 used PWM scores and corresponding insertion count matrix as input for
1380 CENTIPEDE to compute posterior probabilities that a motif is bound (Pique-Regi et
1381 al., 2011). A different threshold to classify bound/unbound was used depending on
1382 the motif matching strategy (genome-wide: >= 0.99; peak summit: >= 0.9).

1383

1384 **ChIP-seq pre-processing**

1385 Sequencing adapters and poor quality base calls were trimmed from reads using
1386 *trim_galore* with default settings. Trimmed reads were aligned against the mm10
1387 reference genome using *bowtie2* with *--sensitive* as additional option. Alignments
1388 were filtered for unmapped, multi-mapping and duplicated reads.

1389 Signal tracks as log₂ fold-change between ChIP and input were generated using
1390 *deepTools bamCompare* with following parameters *--scaleFactorsMethod SES --ratio*
1391 *log2 --bs 25 --ignoreDuplicates* (Ramírez et al., 2016).

1392 Peak calling was performed using MACS2 with *--g mm --p 0.001* (Zhang et al., 2008).
1393 Re-analysis of publicly available datasets was performed in the same way as for new
1394 samples. All ChIP-seq datasets used in this study are listed in Table S2.

1395

1396 **ChIP-seq enrichment analysis**

1397 Enrichment analysis of TF peaks was performed with LOLA (default settings) using
1398 the set of variable ATAC-seq regions as universe (Sheffield and Bock, 2015). We
1399 considered all TF peak sets with an adjusted p-value < 0.01 as enriched.

1400 To complement the mm10 core database with TFs relevant for neural development,
1401 we added 4 new samples and 39 publicly available TF ChIP-seq datasets (Table S2).
1402 Replicates of ChIP-seq experiments were considered separately where available.

1403

1404 **RNA-seq pre-processing**

1405 RNA-seq experiments in this study were quantified using *Salmon* (quasi-mapping
1406 mode) with the GENCODE release M14 (Mudge and Harrow, 2015). Single-end as
1407 well as paired-end reads were processed using following options: *-l A --seqBias --*
1408 *numBootstraps 50*. Resulting counts and transcripts per million (TPM) were used for
1409 downstream analysis.

1410 Differential analysis of D5H vs D5SC (Gouti et al., 2014) was performed using
1411 DESeq2 with default settings. Resulting adjusted p-values were used for Figure S4E.

1412

1413 **Interaction database**

1414 A dataset of putative gene-region interactions was downloaded from the 4DGenome
1415 database (<https://4dgenome.research.chop.edu/>). The obtained interactions were
1416 mapped to *mm9* and for further downstream analysis re-mapped to *mm10* using
1417 *UCSC-liftOver* with default setting. Interactions for which only one anchor could be
1418 mapped to *mm10* were removed.

1419 Putative chromatin-chromatin interactions were mapped by filtering for anchors
1420 which overlap open chromatin sites from this study.

1421

1422 **Gene ontology enrichment of CDX2 bound open chromatin sites**

1423 MNP CDX2 ChIP-seq peaks from Mazzoni et al., 2013, were overlapped with NMP,
1424 NMP-SC, spinal cord, H/SC and hindbrain regions. Resulting peak sets were used
1425 as input for gene ontology enrichment analysis using GREAT (default settings)
1426 (McLean et al., 2010).

1427

1428 **Data availability**

1429 All data generated in this study is available from the Array Express website.

1430

1431 **Code availability**

1432 Analysis scripts are available at <https://github.com/luslab/NeuralATACseq>

1433 **FIGURE LEGENDS**

1434 **Figure 1. Regulatory element usage distinguishes cell states during neural**
1435 **induction. (A)** Schematic of the five-day directed differentiation of mouse embryonic
1436 stem cells (ESCs) to neural progenitors that yields anterior (top), hindbrain (middle)
1437 or spinal cord (bottom) identities. Note that spinal cord progenitors are generated via
1438 an NMP state induced by the addition of FGF and WNT signals from day 2-3 (light
1439 pink shading). **(B)** Immunofluorescence on day 5 reveals that hindbrain progenitors
1440 generate a mixture of PHOX2B expressing visceral- and OLIG2 expressing somatic
1441 motor neuron progenitors. Spinal cord progenitors lack visceral- but generate OLIG2
1442 expressing somatic motor neuron progenitors. Scale bars = 20 microns. **(C-**
1443 **D)** Genome browser tracks of ATAC-seq accessible regions (mm10 assembly)
1444 present in ESCs (day 0, grey) compared to day 5 anterior (blue), hindbrain (yellow)
1445 and spinal cord (red) progenitors, and associated gene expression levels determined
1446 by mRNA-seq (Gouti et al., 2014) for each stage as indicated on the right (Error bars
1447 = SEM). Cis-interactions indicated below represent known genomic interactions from
1448 published data (see Methods). ESCs show accessibility at *Pou5f1/Oct4* enhancers
1449 (C, arrow) unlike neural progenitors which repress *Oct4*. Instead open regions flank
1450 neural genes such as *Olig2* (D, arrow). **(E)** Genome wide comparison in accessibility
1451 between Day 5 spinal cord (D5SC) and Day 0 ESCs reveals differences in regulatory
1452 element usage (FDR<0.01). **(F)** The proportion of differential sites present in each
1453 condition compared to ESCs demonstrates the gradual change in accessibility during
1454 neural progenitor differentiation. **(G)** Both neural and AP-specific sites, but not ESC
1455 sites, are enriched in H3K27ac marks from neural progenitors (Peterson et al.,
1456 2012). bFGF= basic fibroblast growth factor; ESC=embryonic stem cell; D=day;
1457 FC=fold change; kbp=kilobase pairs; RA=retinoic acid; SHH=sonic hedgehog;
1458 TPM=transcripts per million.

1459 **Figure 2. Differential enhancer usage and transcription factor engagement**
1460 **reveal AP identity of neural progenitors. (A)** Self-organizing map (SOM) of all
1461 regulatory regions from all conditions and time points that show differential
1462 accessibility relative to day 0. Each plot represents the z-score for each region,
1463 across each condition (A'). Regions were classified into 10 clusters on the basis of
1464 their accessibility as outlined in A". Many sites are common ("neural sites") to all

1465 neural progenitors (black cluster, n=5584), these differ from epiblast regulatory
1466 regions that are accessible at early stages of the differentiation (Epi; green cluster,
1467 n=1714). Region-specific sites are also identified in anterior (blue cluster, n=1863),
1468 hindbrain (orange cluster, n=2509) and spinal cord (red cluster, n=1538) progenitors.
1469 A distinct set of regulatory regions was observed specifically opening in D3 NMPs
1470 (pink cluster, n=454 regions). A/H represents accessible regions shared between
1471 anterior and hindbrain (lime cluster, n=1840); H/SC represents shared hindbrain and
1472 spinal cord sites (purple cluster, n=1276); NMP/SC represents shared
1473 neuromesodermal progenitor and spinal cord sites (brown cluster, n=421). Grey
1474 shaded cluster represents unclassified sites. **(B-D)** Examples of ATAC-seq
1475 accessible regions (mm10 assembly) that define anterior (B), hindbrain (C) or spinal
1476 cord (D) day 5 progenitors that were identified using the SOM presented in A and
1477 their gene expression profile determined by mRNA-seq (Error bars = SEM). Anterior
1478 progenitors display region-specific sites opening at *Shh* (B), while hindbrain
1479 progenitors demonstrate a *Phox2b* specific site (C) and spinal cord progenitors open
1480 a *Hoxc8* regulatory region (D). For a complete listing refer to Table S1. **(E-G)** *In*
1481 *vivo* ATAC-seq confirms a correlation between accessibility profiles found in neural
1482 cells *in vitro* and *in vivo*. E9.5 *Sox2eGFP* reporter embryos were dissected to obtain
1483 neural progenitors from anterior (blue shading) and spinal cord (red shading) regions
1484 of the neural tube (E). The fold change in accessibility at anterior (blue), and spinal
1485 cord (red) sites identified *in vitro* in spinal cord progenitors relative to anterior neural
1486 progenitors *in vivo* correlates with AP identity (F). By contrast, common neural sites
1487 identified *in vitro* (black) are similar in both populations *in vivo* and *in vitro*. Anterior
1488 sites identified *in vitro* show preferential accessibility *in vivo* in anterior compared to
1489 spinal cord progenitors (G), while spinal cord *in vitro* sites show more accessibility *in*
1490 *vivo* in spinal cord compared to anterior neural progenitors. (P values determined
1491 using Wilcoxon Rank Sum Test.) **(H-J)** ChIP-seq enrichment analysis of factors
1492 present in region specific sites (see Methods). Anterior regions are enriched in
1493 FOXA2 binding events (H), while hindbrain sites are enriched with OLIG2 and NKX6-
1494 1 (I). Spinal cord sites are instead enriched with CDX2 (J). SOX2 ChIP-seq in day 5
1495 hindbrain (D5H) and spinal cord (D5SC) cells reveals that the binding site preference
1496 of this SOXB1 TF is condition-specific (I-J). Note that CDX2* denotes CDX2 ChIP-
1497 seq performed in the presence of FGF signaling (Mazzoni et al., 2013).
1498 FPM=fragments per million; neural EB=embryoid bodied-derived neural progenitors;

1499 NMP=neuromesodermal progenitors; pMN=motor neuron progenitors; NP= neural
1500 progenitors; NT=neural tube; TPM=transcripts per million.

1501 **Figure 3. Axial identity is established in cells prior to neural identity. (A-C)** The
1502 average accessibility (z-score) of region-specific sites over time in anterior (A),
1503 hindbrain (H) or spinal cord (SC) conditions. In each case (A-C), AP-specific sites
1504 become accessible between D3-4, the time point corresponding to the addition of
1505 neuralising signals to the culture medium. Spinal cord progenitors do not transiently
1506 open sites corresponding to anterior (A) or hindbrain (B) identity before opening
1507 spinal cord-specific sites (C). Error bars=SD. **(D)** Schematic of the differentiation
1508 used to generate hindbrain and spinal cord cells. The only difference between these
1509 two conditions is the addition of WNT signals between day 2-3 in the spinal cord
1510 condition, which is provided together with bFGF (bFGF/WNT). By contrast, hindbrain
1511 cells are only exposed to bFGF at this time (shaded in grey). **(D')** Schematic of the
1512 differentiation used to test the posteriorising effect of bFGF/WNT in hindbrain neural
1513 progenitors. bFGF/WNT signals are provided to cells from day 4-5 (H+ condition,
1514 D'). **(E-M)** RT-qPCR data showing the relative expression of genes at D3 and D5
1515 following the differentiation of cells to hindbrain, spinal cord or "hindbrain+" identity.
1516 The induction of the WNT signaling target gene *Notum* (E) is observed both at day 3
1517 (D3NMP, following day 2-3 treatment with bFGF/WNT) and at day 5 (D5H+,
1518 following treatment with bFGF/WNT between day 4-5). By contrast, induction of
1519 posterior spinal cord *Hox* genes *Hoxb9* and *Hoxc8* is dependent on timing: induction
1520 at day 3 in D3NMP cells is observed following day 2-3 treatment with bFGF/WNT
1521 signals, but not at day 5 in D5H+ cells following day 4-5 treatment with the same
1522 signals (F-G). Similarly, the induction of *T/Bra* and *Cdx2* is dependent on timing,
1523 responding to early (day 2-3) and not late (day 4-5) treatment with bFGF/WNT
1524 signals. The late treatment of bFGF/WNT in the D5H+ condition prevents expression
1525 of ventral neural genes *Phox2b* and *Olig2*, normally induced in the hindbrain at this
1526 timepoint (J-K, compared D5H to D5H+). By contrast, D4-5 treatment with
1527 bFGF/WNT induces dorsal and intermediate neural tube genes, indicated
1528 by *Pax7* (L) and *Dbx1* (M), respectively. (N-O) mRNA-seq expression profile of
1529 neural genes *Sox1* (N) and *Pax6* (O) indicates that neural progenitor genes are
1530 upregulated at D4, which follows treatment with neuralising signals (RA and SHH).
1531 TPM=transcripts per million.

1532 **Figure 4. *T/Bra* is dispensable for the chromatin remodeling events that**
1533 **establish NMP and SC identity. (A)** The proportion of NMP, NMP/SC shared or
1534 epiblast genomic sites identified *in vitro* and their overlap with accessible regions
1535 present *in vivo*. The highest proportion (71%) of NMP site overlap coincides with the
1536 posterior E7.5 embryo (E7.5-P) which harbours NMPs (Neijts et al., 2016) and this
1537 contrasts with neural progenitors which lack this signature in the spinal cord (E9.5-
1538 SC) and anterior nervous system (E9.5-A; this study). **(B)** The average accessibility
1539 profile of NMP/SC shared and NMP specific sites in wild-type and *T/Bra*^{-/-} cells.
1540 Accessibility of these sites remains largely unchanged in cells lacking *T/Bra*^{-/-} cells at
1541 day 3 of the spinal cord differentiation. **(C)** *T/Bra*^{-/-} mutant cells differentiated to D5
1542 under spinal cord conditions retain accessibility at spinal cord (SC) genomic sites.
1543 **(D)** Heatmap showing the accessibility of NMP and NMP/SC (neuromesodermal
1544 progenitors and spinal cord) shared sites is maintained in the absence of *T/Bra* (*Bra*^{-/-}
1545 ^{-/-}) but is dramatically reduced in the absence of all three *Cdx* TFs, *Cdx1,2,4* (*Cdx*^{-/-}).
1546 **(E-H)** RT-qPCR of *Hox* genes at D5 of the *in vitro* differentiation shows the difference
1547 in AP identity between hindbrain (D5H) and spinal cord (D5SC) cells under wildtype
1548 conditions, compared with *T/Bra*^{-/-} mutant (*Bra*^{-/-}) and *Cdx*^{1,2,4/-} mutant (*Cdx*^{-/-}) cells
1549 differentiated under spinal cord conditions. Note that *Bra*^{-/-} mutant cells retain
1550 expression of spinal cord *Hox* genes *Hoxb9* (E) and *Hoxc6* (F) in contrast
1551 to *Cdx* mutant cells which fail to upregulate these genes and instead
1552 express *Hoxb4* (G) and *Hoxc4* (H), which occurs in hindbrain conditions. FPM=
1553 fragments per million; NMP= neuromesodermal progenitors; pMN= motor neuron
1554 progenitors.

1555 **Figure 5. *Cdx* transcription factors remodel the chromatin landscape to**
1556 **posteriorise cells and repress cranial motor neuron identity. (A)** ChIP-seq
1557 enrichment analysis reveals that CDX2 is highly enriched at NMP specific sites. **(B)**
1558 Removal of the three *Cdx* transcription factors *Cdx1/2/4* results in ectopic production
1559 of cranial motor neuron progenitors marked by PHOX2B, that are normally repressed
1560 in spinal cord conditions. Scale bars = 20 microns. **(C)** The average profile of spinal
1561 cord specific sites (left plot) shows that relative to day 5 spinal cord (D5SC, red),
1562 these sites are lost in *Cdx* mutant cells differentiated under the same conditions
1563 (D5SC-*Cdx*^{-/-}, green), to the same extent observed in D5 hindbrain cells (yellow).
1564 Under spinal cord conditions, *Cdx* mutant cells show increased accessibility at

1565 hindbrain specific sites (right plot). **(D)** The proportion of accessible regions bound by
1566 CDX2, as indicated by CDX2 ChIP-seq analysis from neuromesodermal progenitors
1567 (NMP, Amin et al., 2016) and motor neuron progenitors (pMNs, Mazzone et al.,
1568 2016) derived *in vitro*, demonstrates an enrichment for CDX2 at NMP, NMP/SC
1569 (NMP and spinal cord shared) and spinal cord (SC) sites. Furthermore, under pMN
1570 conditions, CDX2 targets accessible regions that are shared between the hindbrain
1571 and spinal cord. **(E)** At the *Phox2b* genomic region, a hindbrain/spinal cord shared
1572 site is bound by CDX2 in pMN conditions (blue shading). **(F)** A hindbrain-accessible
1573 site is bound by CDX2 (blue shading) at the *Mafb* locus in pMN conditions.

1574 **Figure 6. Proposed model of nervous system development.** **(A)** Pluripotent
1575 epiblast cells in the early embryo are first allocated into anterior (blue) or posterior
1576 (red) populations before they have committed to a neural identity. When cells
1577 undergo neural induction to form the progenitors of the nervous system, these two
1578 populations give rise to distinct neural subtypes. Cells which have been posteriorised
1579 form spinal neural cell types (e.g. somatic motor neurons) that make up the posterior
1580 nervous system. By contrast, anterior epiblast cells can generate cranial motor
1581 neurons, and thus, unlike posterior epiblast cells, support the generation of the
1582 anterior nervous system. **(B)** Comparisons between Cnidarian and Bilaterian animals
1583 provides supporting evidence for the dual evolutionary origin of the vertebrate central
1584 nervous system which is proposed to have arisen in a pre-bilaterian animal ancestor
1585 (Arendt et al., 2016). Cnidarians display two distinct nerve centres: apical (blue) and
1586 blastoporal (red). Blastoporal centres show expression of putative CDX orthologues
1587 (Arendt et al., 2016; Ryan et al., 2007). In Bilaterians, these separate nerve centres
1588 have expanded and merged. We speculate that the transition between these two
1589 nervous systems lies at the level of the posterior hindbrain/anterior spinal cord.

1590 **SUPPLEMENTARY FIGURE LEGENDS**

1591

1592 **Figure S1. Quality control for all ATAC-seq samples generated in this study.** **(A)**
1593 The proportion of mitochondrial fragments recovered across each sample. **(B)**
1594 Representative example showing the distribution of fragment lengths recovered from
1595 ATAC-seq, using paired-end sequencing. **(C)** Average level of Tn5 enrichment
1596 (score = maximum(number of insertions)/minimum(number of insertions)) observed

1597 across transcription start sites (TSS) for each sample. **(D)** Summarised Tn5 insertion
1598 profile covering +/-2kb around annotated TSS for sample D5H (replicate 2). Red line
1599 corresponds to a 50bp running average. **(E)** The fractions of fragments that map to *in*
1600 *vitro* consensus peak regions. **(F)** CTCF footprint present in ESC accessible regions
1601 as determined by ATAC-seq.

1602

1603 **Figure S2. Tissue specificity and genomic location of regulatory regions that**
1604 **define neural and region-specific identity. (A-D)** Comparison of ATAC-seq
1605 identified regions with previously published DNase hypersensitivity sites present
1606 across a range of *in vivo* tissues and time points from the ENCODE regulatory
1607 element database (Consortium, 2012; Sloan et al., 2016). Genomic regions
1608 correspond to neural (A), anterior (B), hindbrain (C) and spinal cord (D) specific sites
1609 from Figure 2A. Each set of genomic regions demonstrates an enrichment in
1610 embryonic and neural samples *in vivo*. **(E)** Comparison of ATAC-seq identified
1611 regions with the Vista enhancer database (Visel et al., 2007) shows that ATAC-seq
1612 recovers enhancers that show neural tissue specificity *in vivo*. **(F-I)** Classification of
1613 neural (F), anterior (G), hindbrain (H) and spinal cord (I) sites according to genomic
1614 position. Neural sites are enriched at promoter regions (F), in contrast to the region-
1615 specific sites, which predominantly occupy distal intergenic and intronic regions (G-I).
1616 **(J-K)** Genome browser view (mm10 assembly) showing ATAC-seq from anterior
1617 (blue track) and spinal cord (red track) neural progenitors obtained from e9.5 mouse
1618 embryos at the *Shh* (J) and *Olig2* (K) locus. Arrows indicate known enhancers that
1619 direct *Shh* expression in the midbrain (Epstein et al, 1999; J) and *Olig2* in the spinal
1620 cord (Oosterveen et al, 2012 and Peterson et al, 2012; K). Gene expression levels
1621 determined by mRNA-seq (Gouti et al., 2014) are shown as bar plots from *in vitro*
1622 Day 5 anterior (blue) and spinal cord (red) conditions (error bars = SEM). Chromatin
1623 interactions recovered from indicated tissues are presented below for comparison.
1624 Peak regions are represented with black bars. A=anterior neural progenitor;
1625 DR=dorsal root; NP=neural progenitor; NSC=neural stem cell; SC=spinal cord
1626 progenitor; TPM=transcripts per million.

1627

1628 **Figure S3. Motif analysis of region specific sites that define anterior, hindbrain**
1629 **and spinal cord. (A-C)** Motif analysis performed using Homer (Heinz et al., 2010)
1630 on anterior (A), hindbrain (B) and spinal cord (C) specific sites shows distinct and

1631 common neural factors are enriched at each AP level. For each predicted factor
1632 (shown on the left), their corresponding expression level determined by mRNA-seq
1633 (Gouti et al., 2014) in the same condition at D5 of the *in vitro* differentiation is shown
1634 (central column; error bars = SEM). The top 6 predicted motif logos are presented on
1635 the right. TPM=transcripts per million.

1636

1637 **Figure S4. Expression dynamics of *Cdx* TFs during the spinal cord**
1638 **differentiation. (A-C)** Expression profile determined by mRNA-seq for *Cdx1* (A),
1639 *Cdx2* (B) and *Cdx4* (C) from day 0 to day 5 of the spinal cord differentiation. **(D)**
1640 Nucleotide resolution of the frequency of integration of the Tn5 transposon reveals
1641 increased engagement of CDX2 at NMP accessible sites in WT compared to *Cdx*
1642 mutant D3NMP sites. **(E)** Differential gene expression determined by mRNA-seq in
1643 Day 5 hindbrain (D5HB) versus Day 5 spinal cord (D5SC) *in vitro* conditions (Gouti et
1644 al., 2014) compared with wildtype (WT) and *Cdx* mutant (*Cdx* KO) *in vivo* samples
1645 from microdissected e8.0 posterior tailbud tissue (Amin et al, 2016). CDX positively
1646 regulates *Hoxb9* and other 5' *Hox* genes while it represses *Aldh1a2* in the spinal
1647 cord, in agreement with previous studies (Gouti et al., 2017). CDX negatively
1648 regulates many hindbrain genes including *Mafb*. **(F)** ATAC-seq from wildtype (red)
1649 and *Cdx* mutant (green) cells at indicated stages (grey bars) at the *Mafb* genomic
1650 region. Arrows indicate ectopic accessibility observed in *CDX* mutant cells between
1651 Day 4-5 of the spinal cord differentiation. This region (blue shading) overlaps with a
1652 binding site occupied by CDX in motor neuron progenitor (pMN) conditions
1653 (arrowhead) from previously published studies (Mazzoni et al., 2013).
1654 NMP=neuromesodermal progenitor; SC=spinal cord; TPM=transcripts per million.
1655 Error bars = SEM.

1656

1657 **Figure S5. CDX2 occupancy in open chromatin sites and associated gene**
1658 **ontology enrichment. (A)** Region heatmap displays CDX2 ChIP-seq binding at
1659 open chromatin sites recovered from the self-organising map (pMN; Mazzoni et al.,
1660 2013). **(B-E)** Gene ontology enrichment analysis for CDX2-bound regions shown in
1661 (A). In hindbrain accessible regions (C), CDX2 binding is associated with neural
1662 genes in contrast to either the NMP and spinal cord (NMP/SC) shared or SC-specific
1663 sites (E), which target genes involved in anterior-posterior patterning.

1664

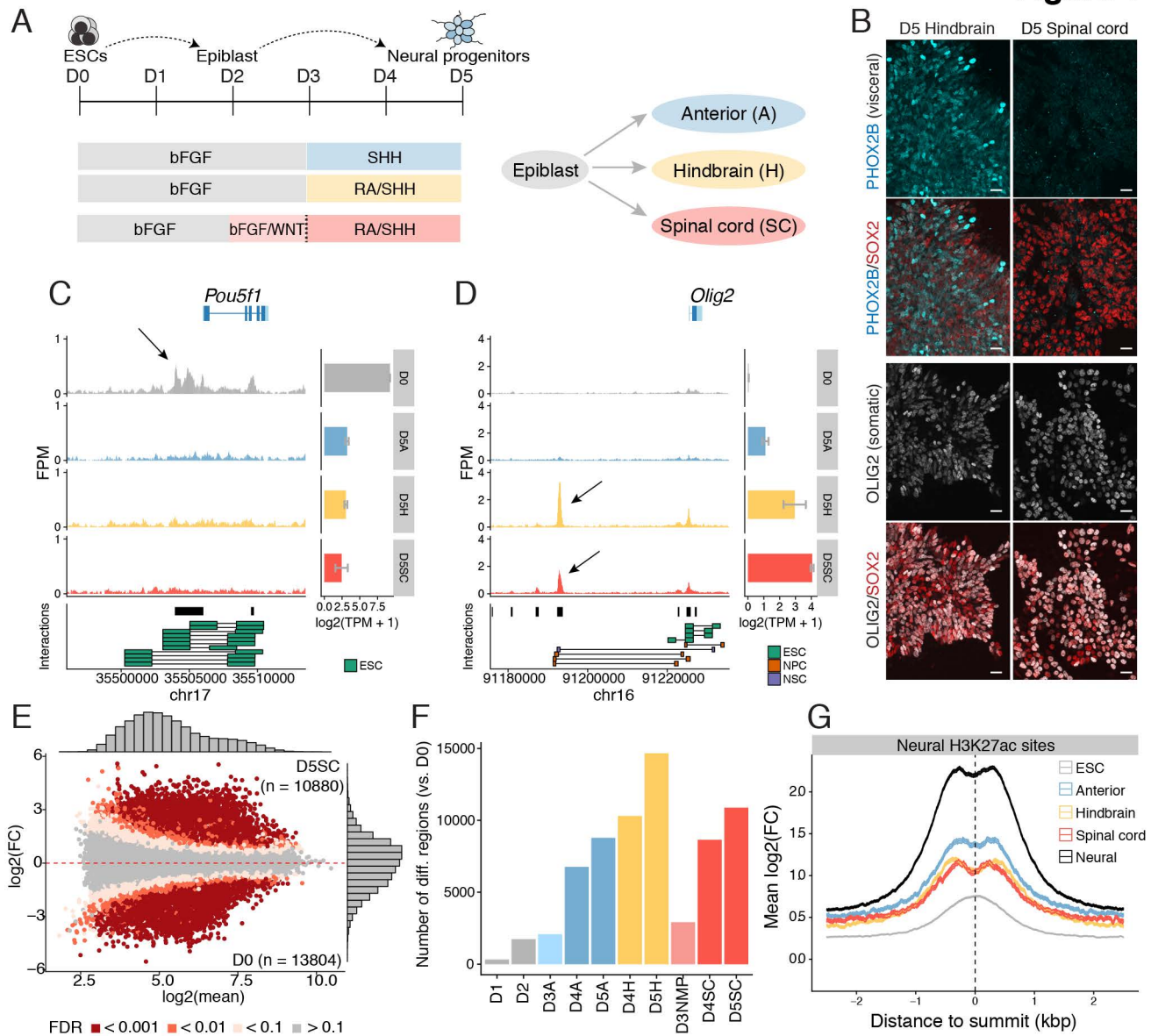
1665 **Supplementary Tables**

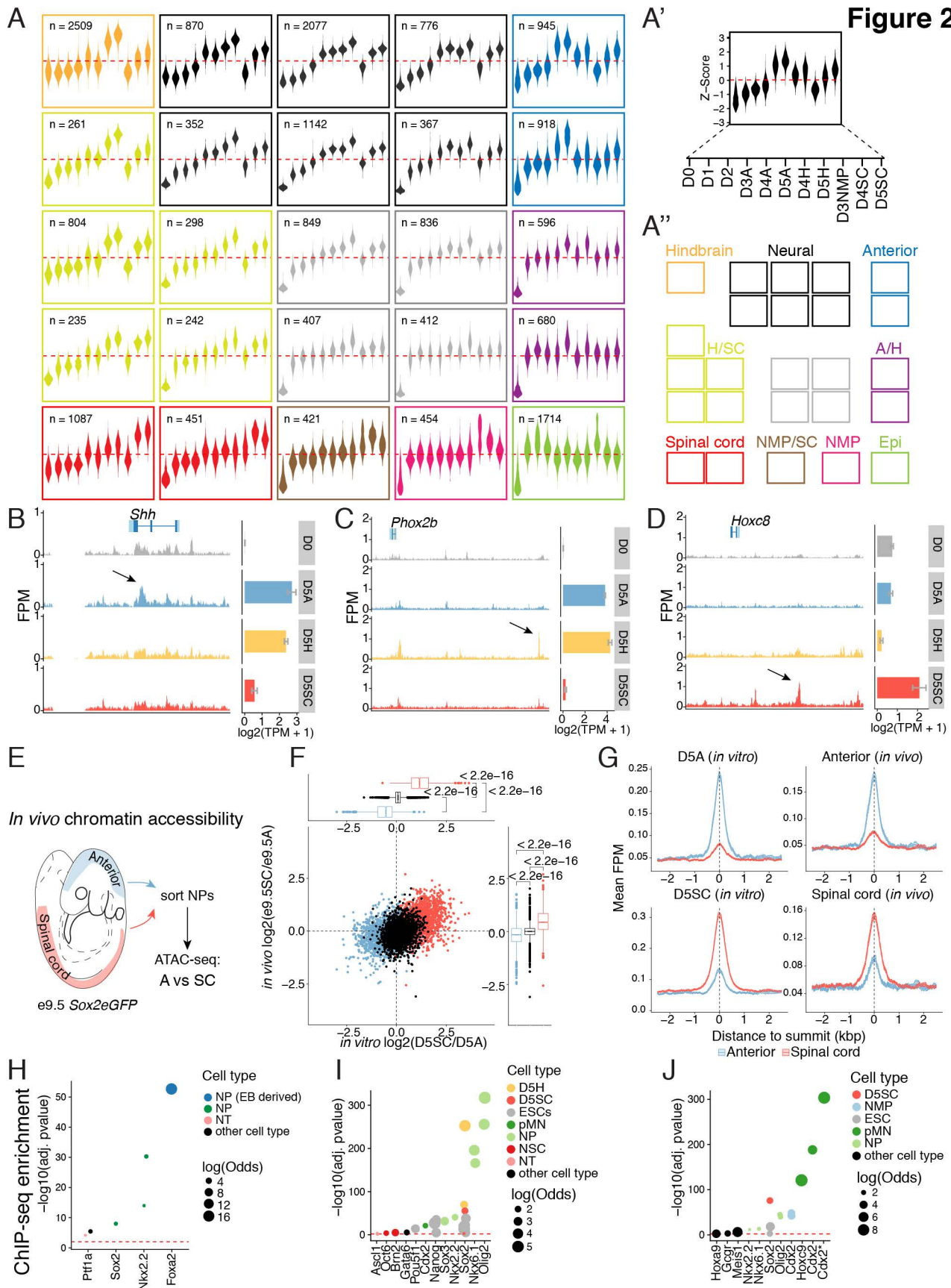
1666 **Table S1.** Peak to gene annotation of region-specific sites identified in this study.

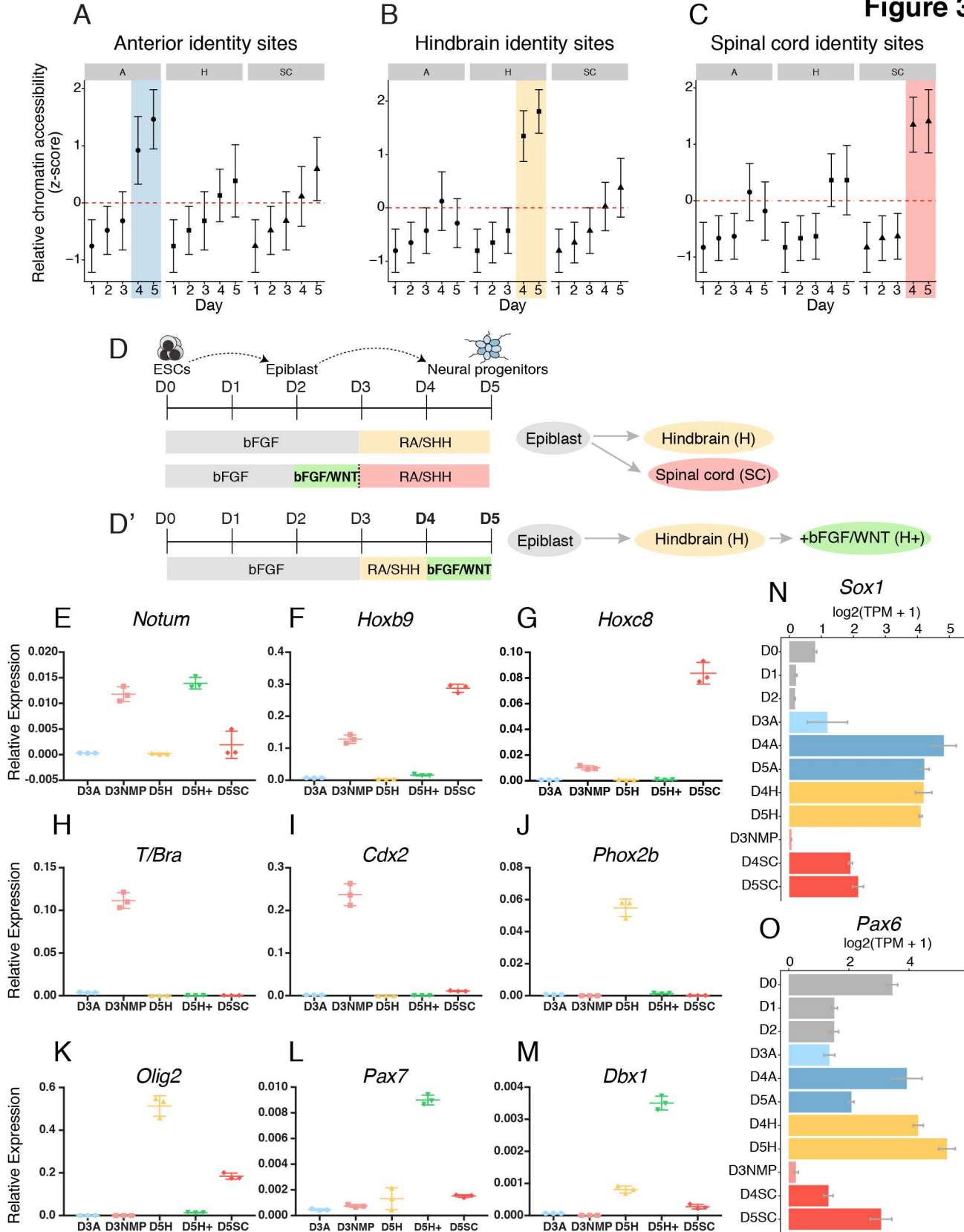
1667 **Table S2.** List of all datasets used for ChIP-seq, ATAC-seq, and mRNA-seq analysis.

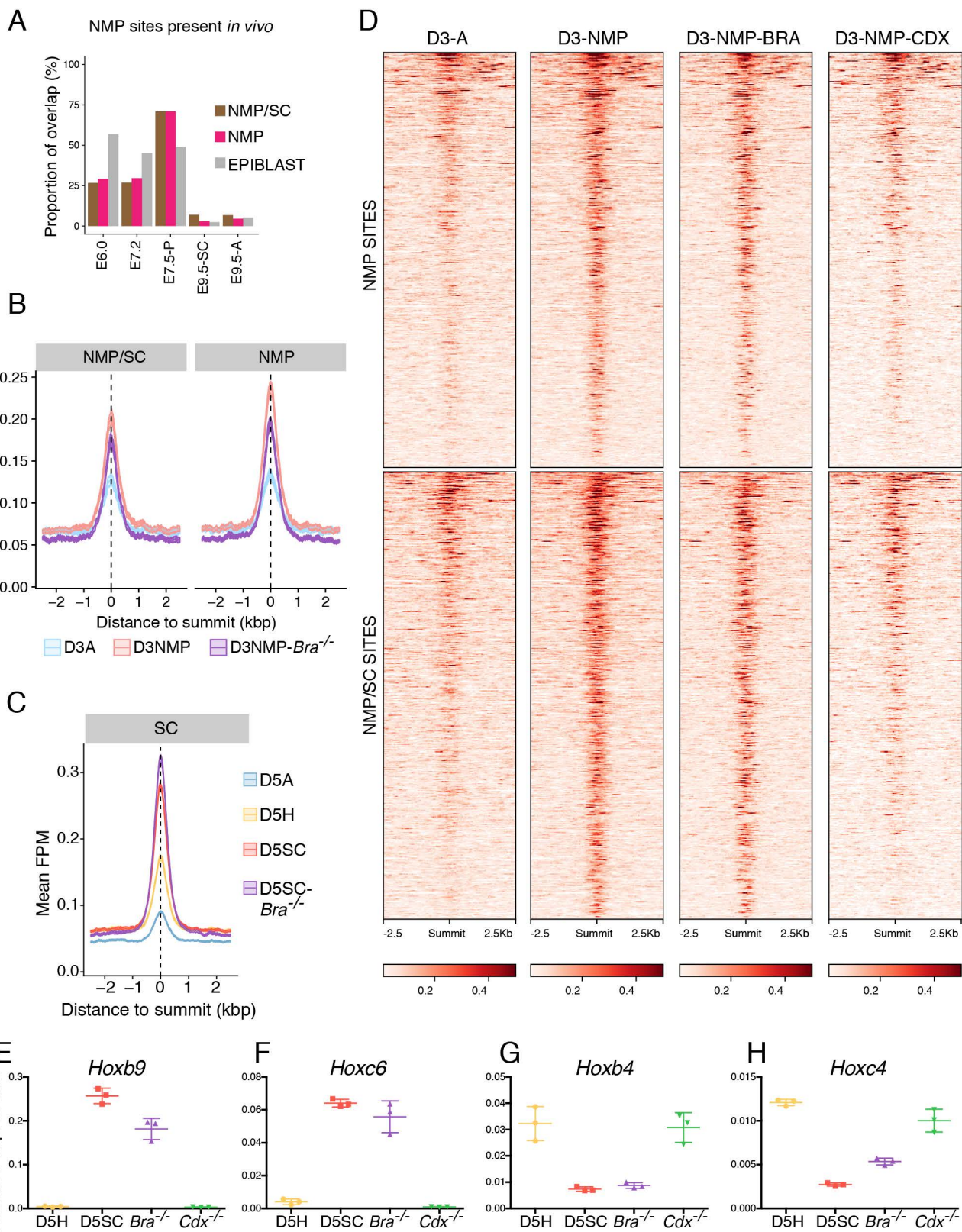
1668 **Table S3.** List of NMP sites, NMP/SC shared and SC sites.

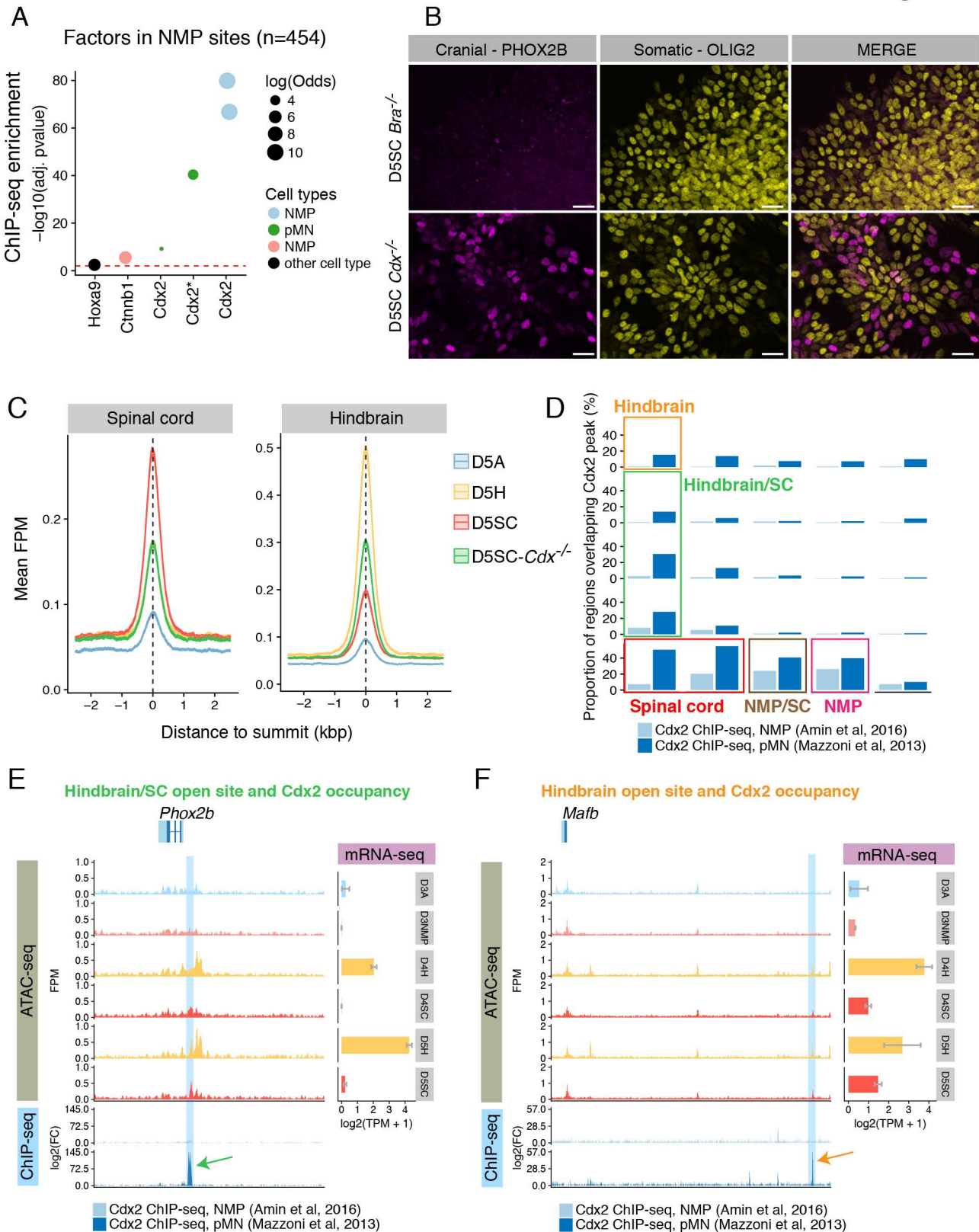
1669 **Table S4.** Primers used for qPCR.



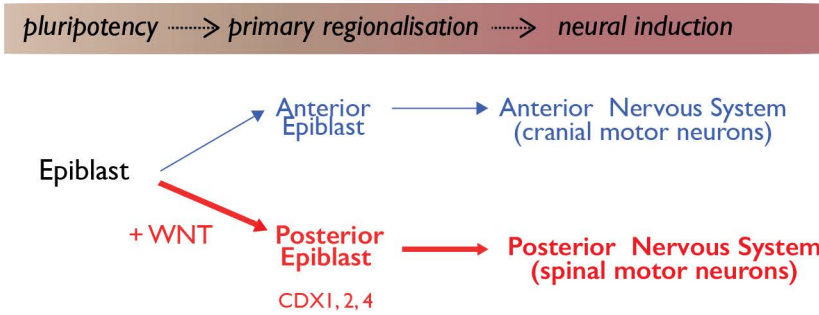




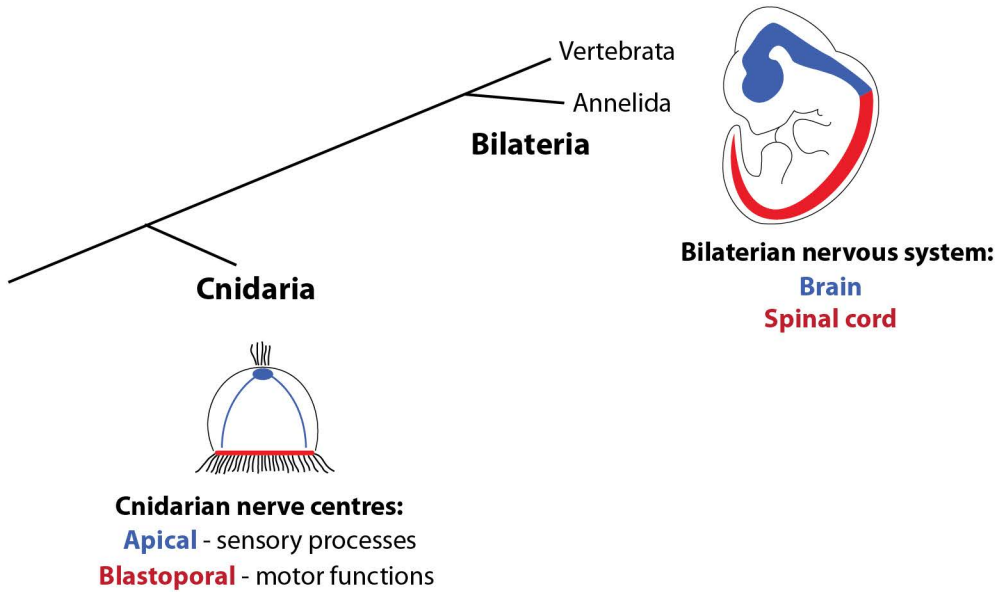


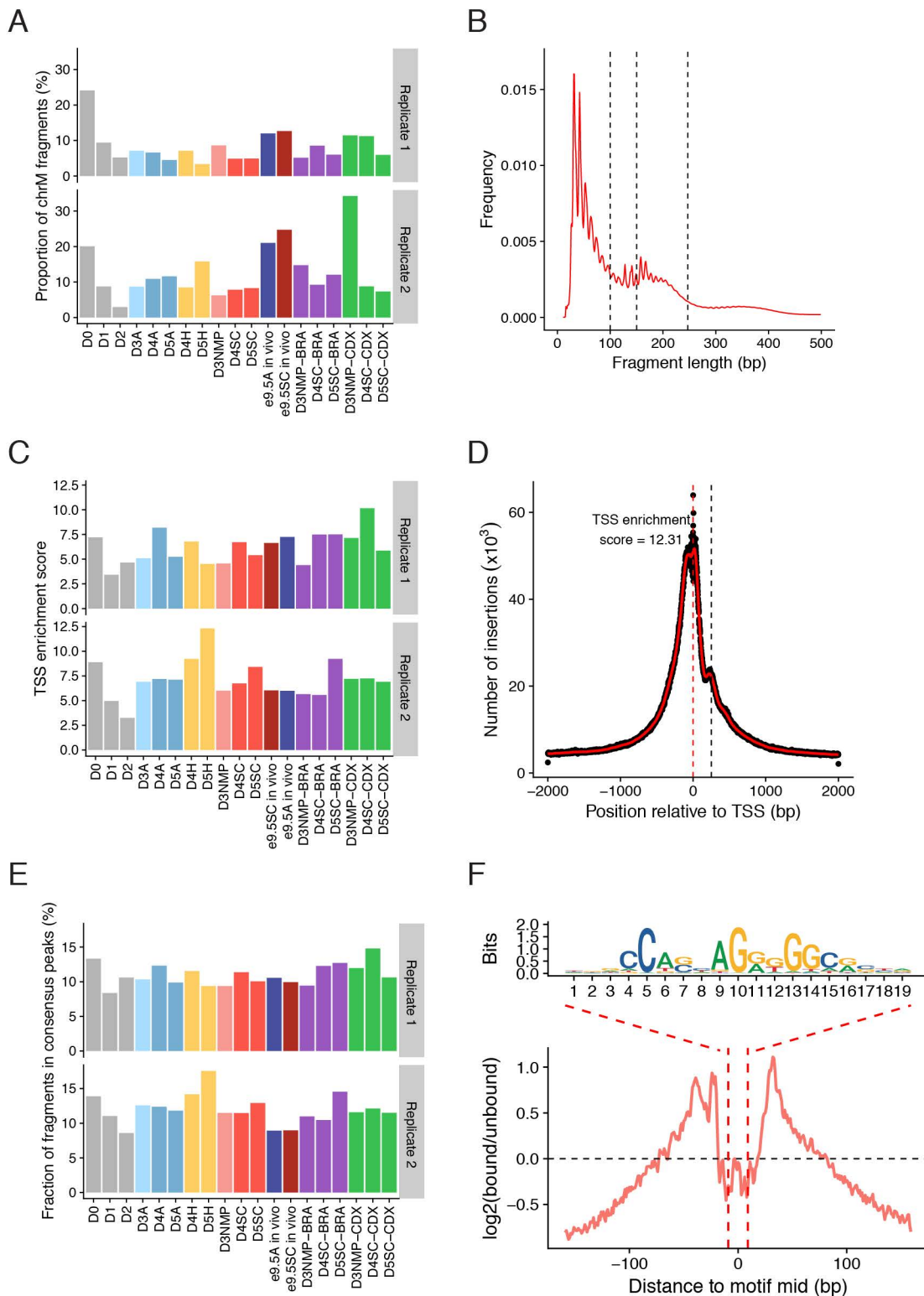


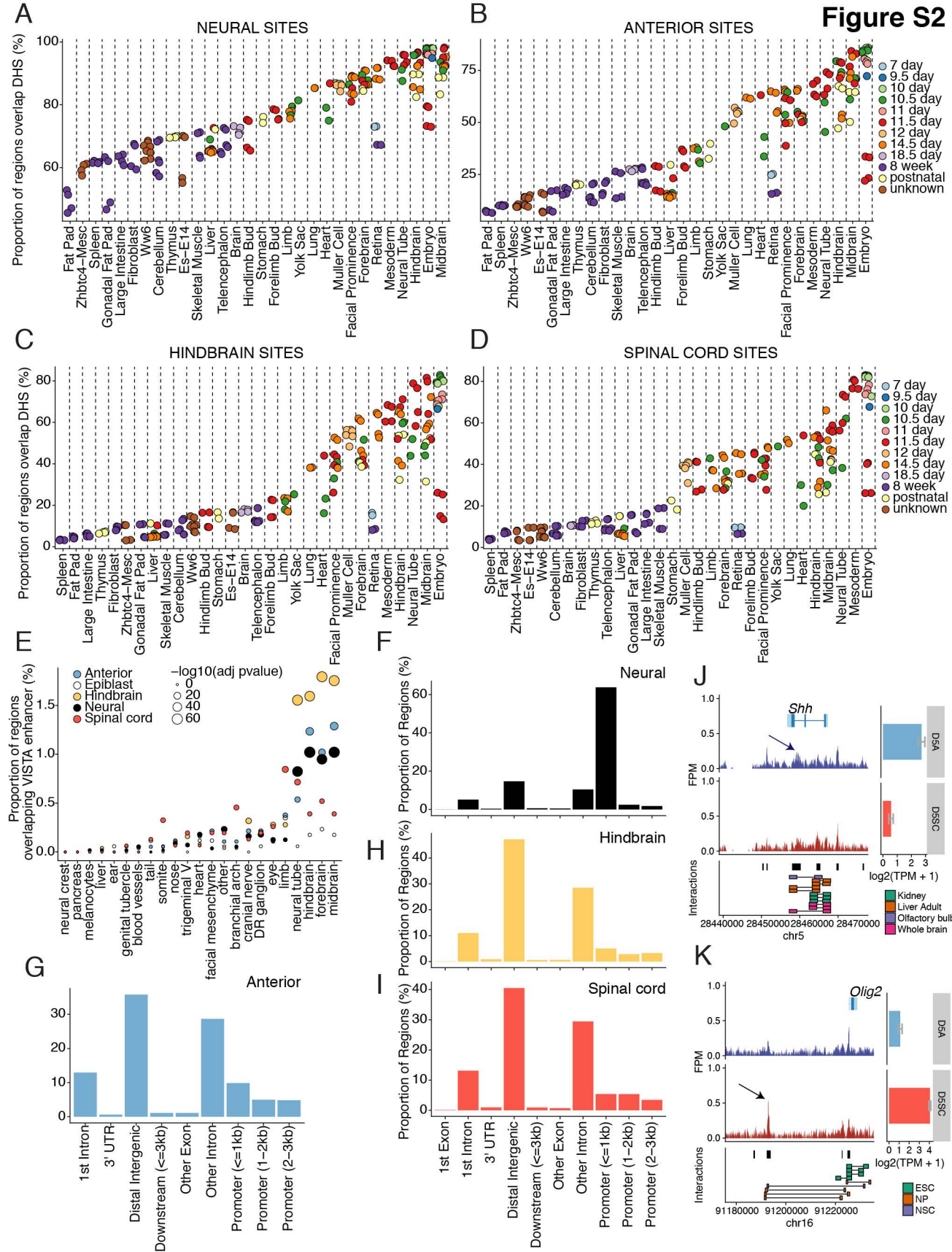
A



B

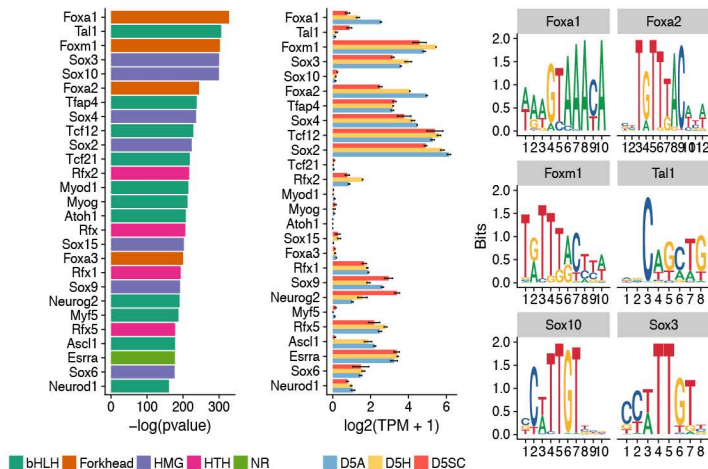






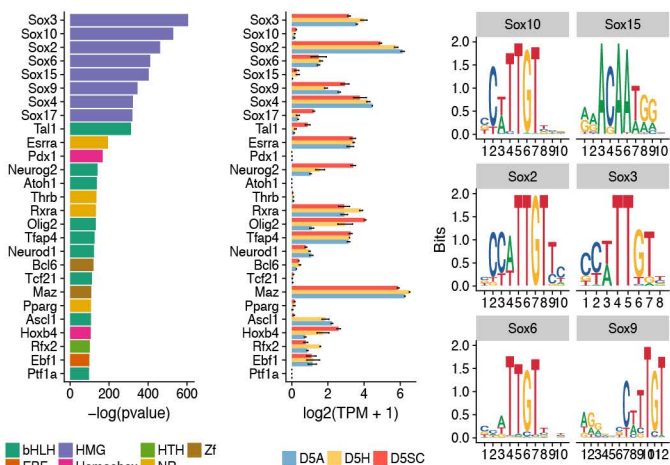
A

Anterior genomic regions



B

Hindbrain genomic regions



C

Spinal cord genomic regions

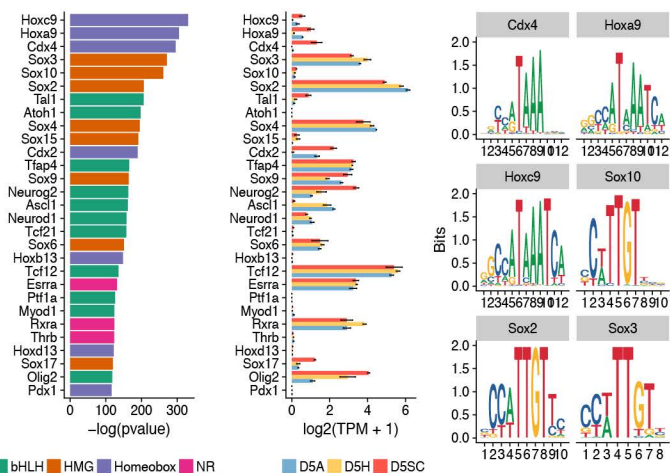
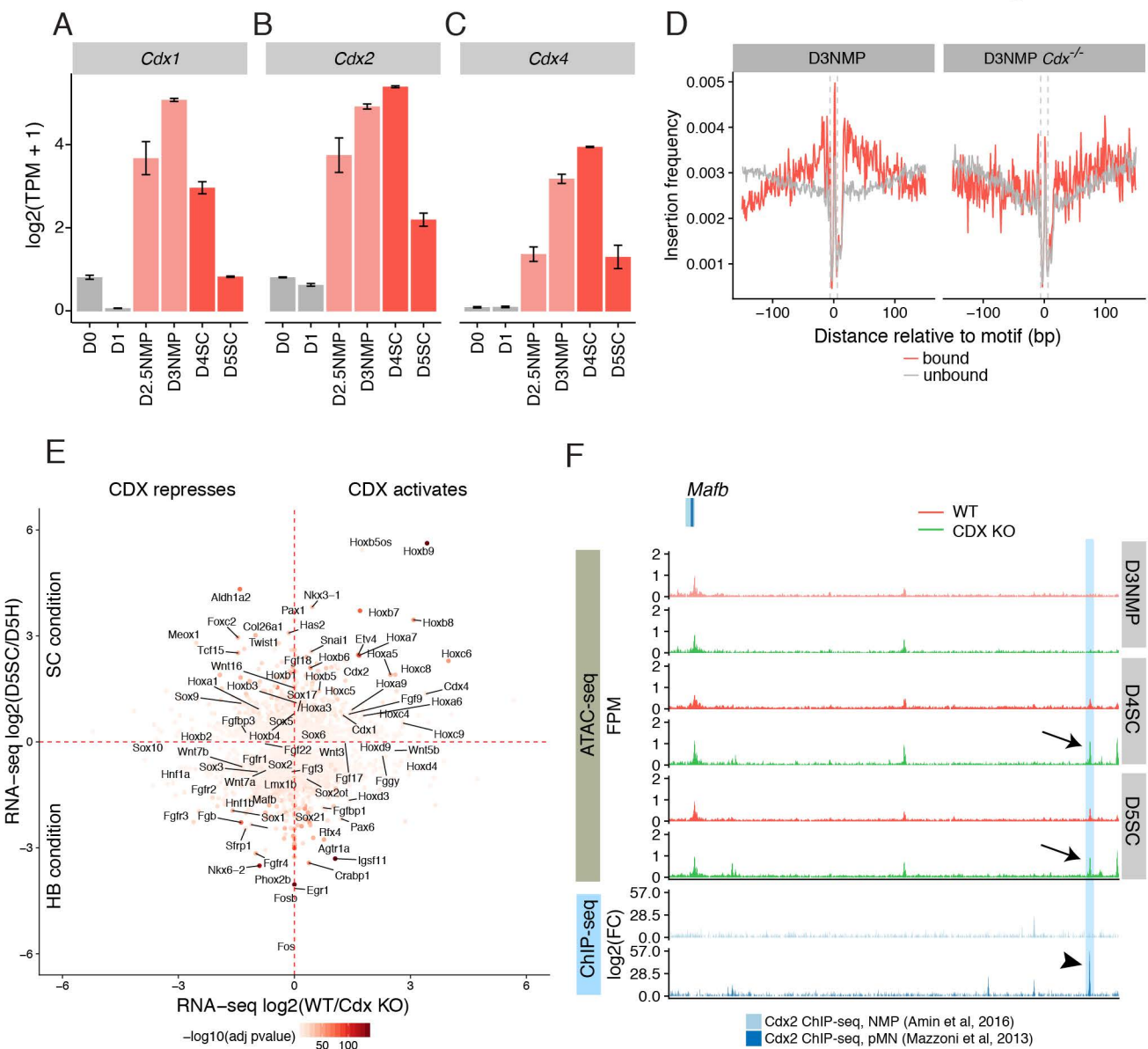
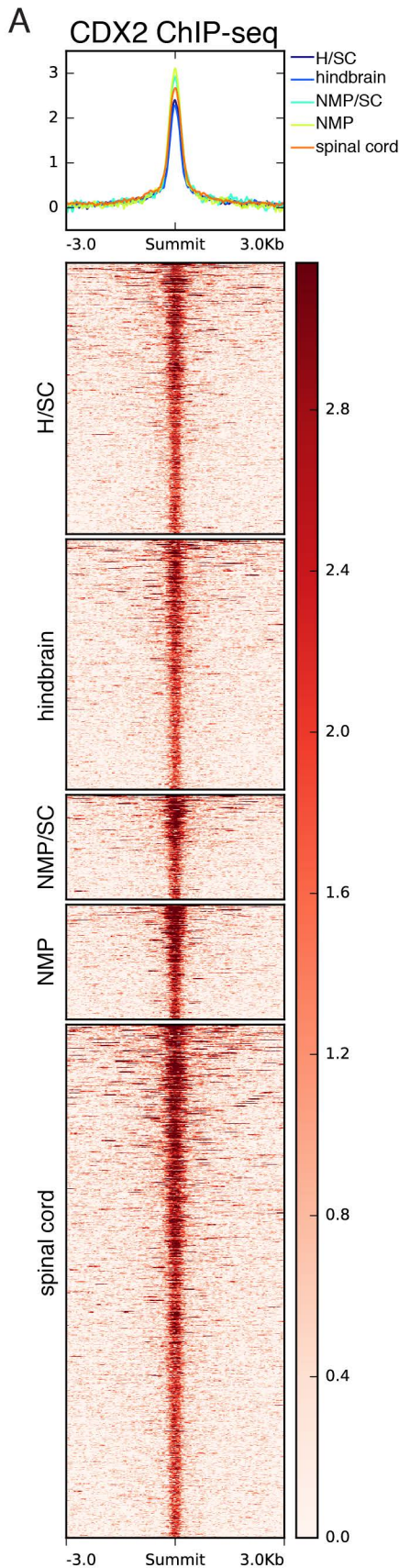
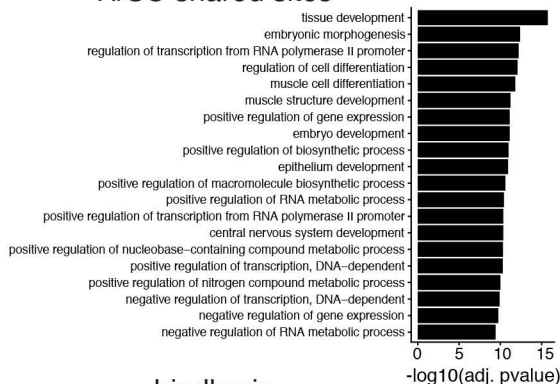


Figure S4

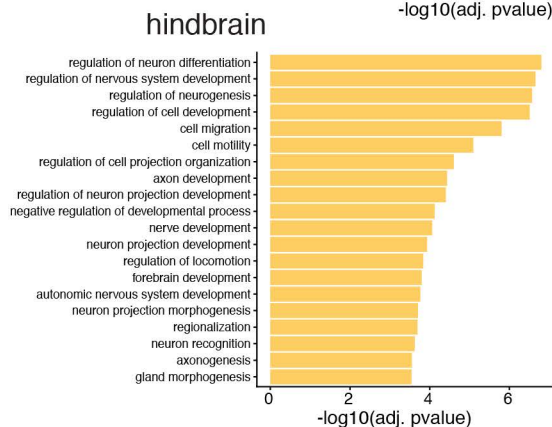




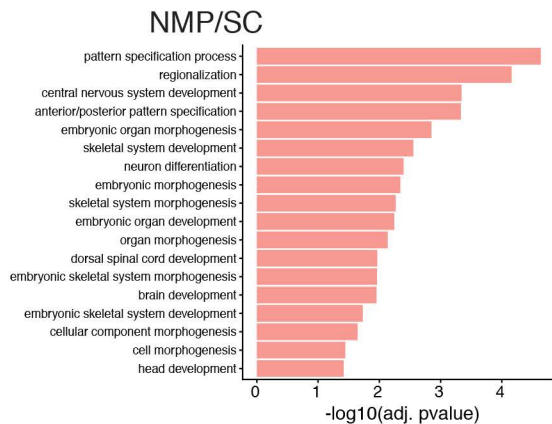
B GO Biological processes
H/SC shared sites



C



D



E

