1 TITLE

2 PRE-BORDER GENE FOXB1 REGULATES THE DIFFERENTIATION TIMING

3 AND AUTONOMIC NEURONAL POTENTIAL OF HUMAN NEURAL CREST

4 CELLS

5

6 AUTHORS

7 Alan W. Leung[1]*[#], Francesc López-Giráldez[2], Cayla Broton[1], Kaixuan Lin[1], Maneeshi S.

8 Prasad[3], Jacqueline C. Hernández[3], Andrew Z. Xiao[1]*, Martín I. Garcia-Castro[3]*

9

10 AFFILIATIONS

11 [1]Department of Genetics and Yale Stem Cell Center, Yale School of Medicine, New

12 Haven, CT 06520, USA

13 [2]Yale Center for Genome Analysis, Yale University School of Medicine, New Haven, CT

14 06520, USA

15 [3]School of Medicine Division of Biomedical Sciences, University of California,

16 Riverside, CA 92521, USA

17

18 *Correspondence

19 A.W.L. - leungwailun@gmail.com

20 A.Z.X. - andrew.xiao@yale.edu

21 M.I.G.C. - martin.garcia-castro@ucr.edu

22

23 [#]Lead contact

24 A.W.L.

25   SUMMARY

26   What are the factors that are induced during the transitory phases from pluripotent stem

27   cells to lineage specified cells, how are they regulated, and what are their functional

28   contributions are fundamental questions for basic developmental biology and clinical

29   research. Here, we uncover a set of pre-border (pB) gene candidates, including forkhead

30   box B1 (FOXB1), induced during human neural crest (NC) cell development. We

31   characterize their associated enhancers that are bound by pluripotency factors and rapidly

32   activated by β-catenin-mediated signaling during differentiation. Surprisingly, the

33   endogenous transient expression of FOXB1 directly regulates multiple early NC and

34   neural progenitor loci including *PAX7*, *MSX2*, *SOX1*, and *ASCL1*, controls the timing of

35   NC fate acquisition, and differentially activates autonomic neurogenic versus

36   mesenchymal fates in mature NC cells. Our findings provide further insight into the

37   concept of the less characterized pB state and clearly establishes FOXB1 as a key

38   regulator in early cell fate decisions during human pluripotent stem cell differentiation.

39   KEYWORDS

40   Differentiation; Pluripotent Stem Cells; Human Development; Forkhead box B1;

41   Transcription Factors; Neural border; Neural Crest; Autonomic Neurons

42   INTRODUCTION

43   A fundamental question in developmental biology is how pluripotent cells transit

44   from pluripotency towards lineage specification (Kalkan and Smith, 2014). Recent advent

45   of time-series single cell RNA-sequencing has revealed dynamic transcriptional features

46   and cell states during pluripotent cell differentiation *in vivo* and *in vitro* (Pijuan-Sala et

47   al., 2018). Gain of function by overexpression, and loss or reduction of function

48   (knockdown and knockout) approaches, such as CRISPR/CRISPR

49   interference/siRNA/miRNA screening and reporter assays, have also been used to

50   identify genes that are important for initial cell fate decision from pluripotent cells

51   (Arduini and Brivanlou, 2012; Betschinger et al., 2013; Genga et al., 2019; Hackett et al.,

52   2018; Ma et al., 2015; Wang et al., 2012). To functionally characterize novel cell states,

53   there is an urgency to develop experimental pipelines that will integrate the various

54   molecular inputs, effectors, and feedback regulations to infer an increasingly

55   sophisticated road map of cellular differentiation.

56   Neural crest (NC) is an embryonic cell population that emerges from the dorsal

57   neural tube, migrates extensively, and generates peripheral nervous system neurons and

58   glia, and ectomesenchyme-derived craniofacial bone and cartilage, amongst many other

59   derivatives (Prasad et al., 2019; Stuhlmiller and Garcia-Castro, 2012). This wide range of

60   differentiation potential defies canons of sequential segregation of potential because NC

61   has been considered to emerge from the ectoderm, and bone and cartilage derivatives are

62   normally only associated with mesoderm, not with ectoderm. This issue remains

63   unresolved, with a general perception that the ectodermally-restricted NC lineage

64   acquires ectomesenchymal capacity. Hall suggested that NC constituted a fourth germ

65   layer, but provided little mechanistic evidence (Hall, 2000). Work in chick and rabbit

66   embryos has challenged the classic perception of NC ontogeny, postulating that the early,

67   and thus anterior NC, are specified during gastrulation, independently of mesoderm and

68   or neural interactions and thus postulating a pre-germ layer origin (Basch et al., 2006;

69   Betters et al., 2018). Our unpublished work has further identified in chick blastula

70   embryos with matching experiments in human embryonic stem (ES) cells a NC

71   specification status in pre-streak epiblast within chick embryos and within hours after

72   initiation of NC induction from human ES cells that is distinct molecularly and

4

73   functionally from pluripotent stem cells (Prasad et al 2019, submitted).

74   Previously we established a robust model in human pluripotent stem cells, which

75   produce NC cells with anterior character, and allowed us to interrogate the ontogeny of

76   human anterior NC (Leung et al., 2016). In this process of NC induction, we uncovered a

77   class of genes, which we dubbed as pre-border (pB) genes (including *GBX2*, *SP5*, *ZIC3*

78   and *ZEB2*) that are induced prior to, and with distinct signaling requirements than, classic

79   neural plate border genes (Leung et al., 2016). The biological significance of these pB

80   genes is currently not clear. Evidence from vertebrate embryos and cell reprogramming

81   supports the existence of a 'pB-like' cell state prior to the emergence of neural border or

82   definitive neurectoderm (Basch et al., 2006; Betters et al., 2018; Thier et al., 2019;

83   Trevers et al., 2018). The concept of pB cell state however is poorly characterized. In our

84   previous study, we proposed the idea that NC specification takes place prior and

85   independently of neurectoderm commitment (Leung et al., 2016). Here we aim to

86   scrutinize the transcriptome and dissect the molecular functions of pB candidate genes

87   and aim to provide further insights into the earliest lineage specification events during the

88   exit of pluripotency towards NC lineage.

89   Forkhead box (FOX) family proteins are important factors in disease and

90   development (Golson and Kaestner, 2016). They contain a DNA binding domain known

91   as a forkhead or wing helix domain and these proteins can act either as pioneer factors to

92   open local chromatin structures, as classic transcription factors, or both (Iwafuchi-Doi

93   and Zaret, 2016; Lalmansingh et al., 2012). They play essential roles in the regulation of

94   mammalian pluripotency (Krishnakumar et al., 2016; Respuela et al., 2016) and in the

95   specification and further differentiation of neural crest (Kos et al., 2001; Lukoseviciute et

96   al., 2018; Sasai et al., 2001; Seo et al., 2017; Seo and Kume, 2006; Teng et al., 2008; Xu

97   et al., 2018).

98   Foxb1 is strongly expressed in epiblast, neural plate, neuroepithelium, and

99   midbrain neural folds including NC progenitors (Labosky et al., 1997). Lineage tracing in

100   mice suggests contributions to multiple lineages, including NC derivatives (Zhao et al.,

101   2007). *Foxb1*$^{-/-}$ mice displayed developmental delay, posterior truncation, and open

102   neural tube phenotype (Labosky et al., 1997). These mice also lacked a subgroup of

103   medial mammillary body neurons implicated in spatial memory and behaviors (Wehr et

104   al., 1997). Foxb1 could act as a survival factor for these mammillary neurons during the

105   formation of mammillo-thalamic axonal tract (Alvarez-Bolado et al., 2000). In frog

106   embryos, forced expression of Foxb1 led to increased neural induction mediated via

107   cooperation with a Pou transcription factor (Takebayashi-Suzuki et al., 2011) It is of note

108   that a human patient with intellectual disability and distinctive facial features was found

109   with large 15q22.2 deletion spanning the *FOXB1* locus (Yamamoto et al., 2014). The

110   function of FOXB1 during human ES differentiation is currently unknown.

111   In this work, by studying the transcriptomic changes, enhancer utility, molecular

112   functions, and genomic binding patterns of pB candidates, we aim to reveal the

113   importance of pB intermediate during the exit of pluripotency and the specification of NC

114   cells. Our data uncovers 68 candidates for pB intermediate and reveals that these gene

115   candidates are regulated by elements pre-occupied by pluripotency factors and β-catenin

116   co-factors. Our data also reveals major functions of FOXB1, a key pB candidate, in

117   controlling the timing of NC differentiation, and regulating the differentiation potential of

118   NC precursors towards autonomic neuronal fates. Finally, we provide genomic-binding

119   data of FOXB1 showing that these functions are carried out by direct transcriptional

120   controls on key neurogenic and NC gene loci. Altogether, our data provides important

121   new insights into the poorly characterized pB cellular state and reveals FOXB1 as an

122   essential regulator of human ES cell differentiation. Our data further highlights the

123   critical function of FOXB1 in conferring autonomic neurogenic potential to pre-NC cells.

124

125   RESULTS

126   ES cell transcriptional response to canonical WNT signaling during NC induction

127   Human NC cell induction using our established protocol takes 5 days (Leung et

128   al., 2016) and progresses via a pB and a neural border stage (Fig. 1A) to generate

129   SOX10$^+$ NC cells that do not express the definitive neurectoderm marker PAX6 along

130   their course of induction (Fig. 1B) (Leung et al., 2016). To trace the transcriptomic

131   trajectory of early human NC progenitors, we performed time-course mRNA sequencing

132   (RNA-seq) of human ES cells, cells derived from differentiation day 3 (pre-NC,

133   representing a mixture of pB and neural border progenitors) and differentiation day 5

134   (representing NC progenitors). RNA-seq samples were classified by unsupervised

135   hierarchical clustering according to their differentiation status (Fig. 1C). Differential

136   expression (p<0.05) was respectively detected at 3338 (ES cells versus day 3), 4424 (ES

137   cells versus day 5) and 2009 (day 3 versus day 5) loci. The RPKM values and fold

138   changes comparing different time points for all mapped transcripts are provided in table

139   S1. Factors known to regulate pluripotency and self-renewal including *POU5F1* and

140   *NANOG*, except *SOX2,* and components for FGF (*FGF2*, *DUSP6*), and NODAL

141   (*NODAL*, *GDF3*, *LEFTY1*, *LEFTY2*, *TDGF1*) signaling were downregulated during

142   differentiation (Fig. 1D). Previously identified pB transcripts *GBX2*, *ZIC3*, *ZEB2*, and

143   *SP5*, as well as *MEIS2* (a known NC factor), were dramatically upregulated at day 3 (Fig.

144   1E). This was followed by a gradual induction of neural border and NC transcripts

145   including *PAX3*, *PAX7*, *TFAP2A*, *NR2F1*, *MSX1*, *MSX2*, *SNAI2, NR2F1*, *CDH6*, *SOX9*

146   and *SOX10*, while *FOXD3* and *ETS1*, displayed a biphasic expression pattern (Fig. 1D,

147   E). As expected for their essential roles in human NC cell induction (Leung et al., 2016),

148   ligands for WNT and BMP pathways, as well as their receptors (*FZDs*) and extracellular

149   regulators (*CHRD*, *NOG*, *BMPER*), were upregulated synchronously during

150   differentiation (Fig. S1A). Signaling components for Hedgehog (*HHIP, GLI1, GLI3,*

151   *SMO, SUFU*), NOTCH (*JAG1*, *HES3/4*), and insulin growth factor (IGF)(*IGF1*, *IGF1R,*

152   *IGFBP2/5* and *PAPPA*) pathways were either induced or repressed with their general

153   trends indicative of either a blockade (Hedgehog) or activation (Notch and IGF) of the

154   pathways (Fig. S1B). Notably, increased transcription of imprinting loci *IGF2* and *H19*

155   was observed at both day 3 and day 5 (Fig. S1C). Few selected *HOX* genes, except

7

156   *HOXA1* and *HOXB1/2* (Fig. S1C), were activated, suggestive of cranial character of our

157   NC progenitors.

158   To systematically identify a complete pB gene set, we subjected day 3

159   transcriptome data to further analyses based on p values (<0.05), RPKM values (>8.2,

160   expressing at a higher absolute transcript level than *GBX2*, a known pB gene), and fold

161   changes (>10 folds over ES cells) of individual transcripts (Fig. 1F). Based on these

162   criteria, 68 transcripts were identified (Fig. 1F'). These transcripts were expressed and

163   induced at much higher level than neural border and NC transcripts. To examine the cell

164   type specificity of these pB candidates, we performed RT-qPCR on 4 cell types: ES cells,

165   pre-NC, prospective anterior neuroectoderm (NE) and non-neural ectoderm (NNE) (Fig.

166   S1D). We found that 45.6% (n=31/68) and 19.1% (n=13/68) of pB candidates were "NC-

167   specific" (Fig S1E) and "NC-enriched" (Fig S1F) respectively and were expressed at

168   significantly higher levels in pre-NC cultures than those under ES cells, NE or NNE

169   condition. For "NC-specific" pB candidates, no significant transcriptional induction was

170   detected under NE and NNE conditions compared to ES (Fig S1E). These two categories

171   together constitute 64.7% of all pB candidates. The rest of the candidates were either

172   "NC-biased" (n=7/68, Fig S1G, induced under NC condition but no consistent increase in

173   expression was observed compared NC versus NNE and NE), assuming a "broad

174   ectoderm expression" pattern (n=7/68, Fig S1H, significant induction detected in all 3

175   ectoderm lineages), or "highly variable in expression or not specific to NC" (n=10/68,

176   Fig S1I, no significant induction detected under NC condition). In conclusion,

177   transcription of a majority of pB candidates was highly specific and enriched in early NC

178   cultures.

179

180   <u>β-catenin dependent activation of early NC response genes</u>

181   pB genes including *GBX2*, *ZEB2*, *ZIC3*, and *SP5* were known to be regulated by

182   β-catenin (CTNNB1) signaling (Leung et al., 2016). As expected, β-catenin was the top

183   predicted upstream regulator for differentially expressed genes found in our

184   transcriptome (p=4.6e-27 and 1.36e-27 for day 3 and day 5 transcriptomes respectively,

185   Fig. S2A). To test if pB candidates were β-catenin transcriptional targets, we examined

186   the expression of 12 selected pB candidates, identified above as showing "NC-specific"

8

187   induction (Fig S1E), in ES and differentiating cells that were transduced with either a

188   *luciferase* shRNA or a *β-catenin* shRNA (Fig 2A). First we found that *β-catenin* shRNA

189   downregulated *β-catenin* (*CTNNB1*) RNA expression at ES and day 1, but did not affect

190   expression of pluripotency factor *POU5F1* (Fig. 2B). At day 1, a majority of pB

191   candidates (7 out of 12) however were repressed by *β-catenin* shRNA (Fig. S2B).

192   Notably, no change in pB candidate expression was detected in *β-catenin* shRNA cell

193   line. By day 3, all 12 pB candidates were downregulated in *β-catenin* shRNA-transduced

194   cultures (Fig. S2C). To confirm the expression dynamics of the rest of the pB candidates,

195   we performed RNA-seq on cultures transduced with control *luciferase* or *β-catenin*

196   shRNAs derived from ES cells, at day 1 and at day 3 (Fig. 2C, table S2). In line with our

197   earlier time-course transcriptome analysis, 95.6% pB candidates (n=65/68) were induced

198   in control *luciferase* shRNA cultures at day 3. Out of these candidates, 73.8% (n=48/65)

199   displayed a 2-fold or more downregulation in *β-Catenin* shRNA cultures. A lower, yet

200   significant, proportion of pB candidates (44.1%, n=30/68) were already induced in

201   control day 1 cultures. Almost all of them (96.7%, n=29/30) were downregulated in *β-*

202   *catenin* shRNA-transduced cultures.

203   FGF and BMP signaling are known to promote NC induction. To test if these

204   signaling pathways contributed to the induction of pB candidates, we activated these

205   pathways under NC condition using FGF2 (Fig. 2D) or BMP4 (Fig. 2E) ligand. Out of

206   the 12 "NC-specific" pB candidates we tested, a majority of them were repressed to basal

207   transcriptional level by either of these ligands suggesting that high FGF or BMP

208   signaling did not favor pB candidate induction from ES cell cultures.

209   Altogether, these data establishes canonical WNT signaling is the main

210   contributor among major signaling pathways for pB candidate induction (Fig. 2F) and β-

211   catenin is a positive regulator for pB candidate induction during NC induction.

212

213   Early NC response genes are Poised for Transcriptional Activation in Pluripotent Stem

214   Cells

215   Expression of many pB candidates were robustly induced within 24 hours from

216   the start of differentiation (Fig. S1C, E). It is possible that pluripotency factors expressed

217   in ES cells might prime developmental genes such as pB candidates for which a rapid β-

218   catenin activation was found.

219   By visual inspection of published ChIP-seq datasets for key pluripotency factors

220   (NANOG, POU5F1, SOX2, BCL11A)(Gertz et al., 2013; Tsankov et al., 2015), β-

221   catenin cofactors (NIPBL, LEF1, TCF4)(Estaras et al., 2015; Tsankov et al., 2015), and

222   chromatin signature marks (Consortium, 2012; Rada-Iglesias et al., 2011), we identified

223   putative regulatory elements close to pB candidate genes *GBX2* (Fig. S3A) and *FOXB1*

224   (Fig. S3B). These 2 elements displayed strong binding of pluripotency factors and β-

225   catenin co-factors, open chromatin configuration (ATAC-seq), presence of enhancer

226   marks (EP300, H3K4me1), and β-catenin (CTNNB1) binding upon CHIR 99021(Funa et

227   al., 2015) or WNT ligand (Estaras et al., 2015)-mediated ES cell differentiation. To

228   identify other putative pB candidate regulatory elements, we tested the co-occupancy of

229   pluripotency factors and β-catenin cofactors by performing pairwise comparisons of the

230   ChIP-seq peak sets for pluripotency factors and β-catenin cofactors (Fig. S3C). We found

231   significant overlaps for all pairwise comparisons ($p<10^{-4}$) (Fig. S3C, Table S3). Among

232   these comparisons, NANOG and NIPBL had the largest overlapping binding regions

233   (n=37,481, Fig. S3E). Notably, these NANOG-NIPBL co-bound elements with

234   characteristics of poised or inactive enhancers (Fig. S3F) were significantly associated

235   with developmental processes for NC cells (NC differentiation/development, PNS

236   development, abnormal craniofacial morphology) and gene expression in ectoderm

237   (TS11/12/14 embryonic ectoderm) (Fig. S3G). These poised and inactive enhancers were

238   also represented in high percentage (>50%) within the 199 NANOG-NIPBL co-bound

239   elements associated with the 68 pB candidate loci (Fig. S3H). Among these 199 pB

240   associated NANOG-NIPBL co-bound elements, a high percentage was also bound by β-

241   catenin upon canonical WNT-mediated differentiation. Such elements were found at

242   genomic location upstream (Fig 3A), downstream (Fig 3B), or within intronic sequences

243   (Fig 3C) of pB candidate loci. To conclude, these *in silico* analysis supports the model

244   that β-catenin signaling during NC induction unlocks poised or inactive enhancers for pB

245   genes thereby leading to their increased transcription.

246   To confirm the above *in silico* analysis that NANOG pre-occupied pB candidate

247   regulatory elements and to visualize NANOG binding dynamics during NC induction, we

248   performed NANOG CHIP-qPCR on ES and differentiation day 3 cells. We confirmed

249   that 76.9% of all elements tested (n=10/13) have enriched NANOG binding at human ES

250   cell stage. NANOG is known to prime developmental gene loci. During ES cell

251   differentiation, NANOG is also known to repress NC gene transcription (Wang et al.,

252   2012). While a smaller subset (n=6/13) still had significant NANOG binding at day 3,

253   NANOG binding was much attenuated at day 3 as we detected an overall decreasing

254   trend of NANOG binding in 53.8% of tested elements (n=7/13) at day 3 as compared to

255   day 0 (Fig. 3D).

256   To demonstrate these enhancers were activated upon NC induction, ChIP-qPCR

257   analysis the enhancer mark H3K27ac was carried on pB candidate regulatory elements in

258   human ES cells (day 0) and differentiation day 3 cells. The data revealed that a majority

259   of (76.9%, n=10/13) of the tested elements displayed increased accumulation of

260   H3K27ac mark (Fig. 3C), demonstrating that these elements became more active during

261   differentiation.

262   Lastly, to functionally test these pB enhancer elements, we used CRISPR

263   approach to knock out one of them, the FOXB1-E1 (asterisk, Fig. 3B and 3E), a

264   NANOG-NIPBL co-bound and β-catenin bound element that resides immediately

265   upstream of the *FOXB1* gene. Deletion of this 495-bp E1 element (top panel, Fig. 3F) led

266   to reduction of *FOXB1* transcription in differentiating cells in a day wise time course

267   experiment (Fig. 3G). Further differentiation assay with additional knockout cell lines

268   (FOXB1-E1-KO1, -KO2, -KO3) confirmed the time course experimental result in that

269   removal of FOXB1-E1 element led to significantly reduced transcription of FOXB1 gene

270   at differentiation day 1 (47%±30% of control level, $p<0.01$), day 3 (30%±14% of control

271   level, $p<0.001$), and day 5 (21%±11% of control level, $p<0.01$) (Fig. 3H). The

272   downregulation of *FOXB1* in differentiating cells strongly indicated that *E1* element

273   acted as a functional enhancer for FOXB1 induction and its continued expression. These

274   data together suggest that distinct regulatory elements primed by NANOG and NIPBL,

275   and bound by β-catenin during early differentiation of human ES cells positively

276   regulated pB candidate transcription.

277

278   <u>Pre-Border Candidates are expressed in chick epiblast prior to overt neural plate border</u>

279   <u>formation</u>

280   As a first approach to decipher the function of pB genes, we focus on FOXB1

281 which has a known role in lineage differentiation. Our RNA-seq data suggests that it is

282 expressed at high level at differentiation day 3 during NC induction but not at ES cells or

283 day 5 cultures (Fig 4A, S4A). This result is consistent with lineage tracing of *Foxb1* in

284 murine embryos revealing its expression in NC precursors but not in mature NC cells or

285 their derivatives (Zhao et al., 2007). By RT-PCR we confirmed that during human NC

286 induction *FOXB1* was transiently induced and peaked at day 2 or day 3 similar to other

287 documented pB genes (Leung et al., 2016), and its expression complemented two highly

288 expressing FOX factors, *FOXD3* and *FOXH1,* which either rapidly downregulated upon

289 differentiation (*FOXH1*) or assumed a bi-phasic expression pattern (*FOXD3*) (Fig. 3G,

290 4B). Whereas most other FOX factors *FOXA3*, *FOXI3*, *FOXN3*, and *FOXO4* showed

291 decreased expression in day 3 compared to ESC, the only two other FOX factors that

292 showed increased transcription at day 3 were *FOXOB3* and *FOXP4* but their fold

293 increase in RPKM was limited to around 2 fold compared to more than 10 fold for

294 *FOXB1*. In agreement to transcript analysis, FOXB1 protein was detected at day 3, but

295 not in human ES cells (Fig. S4B). We reasoned that FOXB1 carried a unique function

296 during the early differentiation window, when other FOX factors were not as highly

297 expressed.

298   Expression of *Foxb1* is first detected in gastrulating mouse embryos in the

299 primitive streak and the surrounding epiblast (Labosky et al., 1997). In chick

300 embryos, *Foxb1* was expressed at the same location at Hamburger-Hamilton (HH) stage

301 4 (Fig. 4C), posterior to the caudal boundary of the anterior neural marker *Ganf*

302 (Fernandez-Garre et al., 2002), and just before *Pax7* induction at the prospective neural

303 border at HH4+ (Basch et al., 2006). *In situ* hybridization for other pB candidates with

304 chicken homologs including *Cntn2*, *Prmt8*, *Nrcam*, *Plagl1*, *Stmn2*, *Enpp2*, and *Enc1*

305 showed similar expression patterns as *Foxb1* at HH stage 4 (n=7, Fig. S4C). Notably,

306 *Foxb1* robust expression is restricted to the ectoderm/epiblast layer (arrows, Fig. 4C').

307 Upon development, *Foxb1* expression progressively restricted medially as neural plate

308 developed (Fig. 4D), and was visibly extinct from the neural border at HH8

309 (asterisks, Fig 4D'). The broad expression domain of *Foxb1* in early epiblast cells could

310   represent a competent zone for NC development, and hinted at a function for Foxb1 in

311   prospective NC populations prior to the formation of the neural border.

312

313   <u>Loss of FOXB1 De-Repressed Neural Crest Development</u>

314   To interrogate the function of FOXB1 in human ES cell differentiation, we

315   attempted to reduce expression level of FOXB1 induced by administration of a *FOXB1-*

316   siRNA. Surprisingly, FOXB1-siRNA resulted in upregulation of NC marker expression

317   at day 5 as indicated by increased transcription of NC markers including *SOX10* (Fig 4E)

318   and proportion of SOX10$^+$ cells (Fig. 4F). We then took advantage of CRISPR/CAS9

319   technologies. To this end, we removed the sequence encoding the Forkhead DNA-

320   binding domain of FOXB1 using a pair of guide RNAs (dgRNA) (Fig. 4G). PCR

321   genotyping and Sanger sequencing revealed specific deletion of the 278-bp genomic

322   region containing the sequence encoding the FOXB1 DNA-binding domain (Fig. 4H and

323   data not shown). Analysis with Western blot confirmed the loss of FOXB1 protein in

324   differentiating *FOXB1-KO* cell lines (Fig. S4B). We then assessed possible effects during

325   NC formation in KO lines and isogenic controls. Cultures from the parent human ES cell

326   line, 3 isogenic and 3 KO cell lines were processed for RT-qPCR analysis at ES cell

327   stage, day 3, and day 5. Confirming siRNA data, expression of neural border and NC

328   genes *PAX3*, *TFAP2A*, and *CHD6* were significantly upregulated at day 5 in KO over

329   control lines (Fig. 4I). Increases in expression for these genes however were already

330   observed at day 3, suggesting that loss of FOXB1 affected the early phase of NC

331   induction. Consistent with qPCR data, immunostaining of dgRNA-KO cells at day 5 also

332   revealed increased SOX10+ cells (data not shown). To confirm the dgRNA experiment,

333   we designed another experiment that targeted the translational start site of FOXB1 using

334   a single guide RNA (sgRNA) (Fig. S4D). We established a knockout (sgRNA-KO) and a

335   heterozygote (sgRNA-Het) cell line for analysis. We characterized the mutations in these

336   2 cell lines by TA-cloning (Fig. S4E) and verified the absence of FOXB1 protein in

337   sgRNA-KO line by western blotting (Fig. S4F). To provide a broader perspective of the

338   influence of FOXB1 in NC development we performed RNA-seq on these samples.

339   Principal component analysis of sgRNA-KO and sgRNA-Het transcriptome data of

340   differentiating cells collected at day 3 and day 5 revealed clear segregation of the

13

341   isogenic control, sgRNA-Het and sgRNA-KO cell lines according to their differentiation

342   time points by PC1, as well as tight correlations according to their genotypes at PC2,

343   revealing a potential dosage effect of FOXB1 function (Fig. S4G). Similar to dgRNA-KO

344   cell lines, we found that neural border and NC transcripts were upregulated in sgRNA-

345   Het and sgRNA-KO at both day 3 and day 5 (asterisks, Fig. S4H, table S4). A few pB

346   candidates were also upregulated (*DMBX1*, *OLFM3*, *PAX2, PKNOX2*) or downregulated

347   (*NRTK2*, *SOX21*, *VGLL3*) indicating that FOXB1 might feedback control selected pB

348   genes (double asterisks, Fig. S4H). Intriguingly, we also observed activation of

349   imprinting loci (*H19*, *IGF2*, *XIST*)(Fig. S4I), PRC2-regulated loci (*HOXA* and *HOXB*

350   clusters, as well as the neighboring homeodomain gene *EVX1*) (Fig. S4J and data not

351   shown), and loci related to increased cell proliferation (*EGFR, MYC*, and *S100A11*) (Fig.

352   S4K).

353   To investigate more precisely when does FOXB1 affects NC induction, we

354   performed time-series analyses (Fig 4J). We found that dgRNA-KO cells underwent

355   accelerated NC induction, as revealed by a 24-hour earlier induction of neural border

356   genes *PAX3*, *PAX7*, and *TFAP2A* (Fig. 4K). Instead of an accelerated loss of pluripotency

357   marker expression as would be expected, expression dynamics for pluripotency gene

358   *POU5F1* appear similar in isogenic control and dgRNA-KO cells (Fig. 4L). Similarly,

359   expression of another pB gene *GBX2*, seem normal from day 1 to 3 in control and

360   dgRNA-KO cells (Fig. 4M). These results suggest that FOXB1 may regulate the temporal

361   acquisition of border specifiers and later NC markers, independently of POU5F1, and

362   GBX2. Then we interrogated if FOXB1 depletion modulates other fates. To this end we

363   deployed a BMP and Activin inhibitor cocktail on ES cell cultures to promote

364   neurectoderm fate (Fig. 4N). Control and dgRNA-KO FOXB1 cells appear to attain

365   neuroectoderm precursor status and displayed no significant expression differences in

366   neurectoderm markers *PAX6* and *HES5* (Fig. 4O), even though *HES5* dgRNA-KO

367   displayed a reduced transcription trend.

368   As indicated by comparable expression, loss of *FOXB1* also did not alter the

369   ability of dgRNA-KO cells line or sgRNA-KO cells to form mesoderm (T, Fig. 4P) or

370   definitive endoderm (SOX17, Fig. 4Q)(ES and ES-de, *EOMES*, *CER1*, and *SOX17*, Fig.

371   S4L, M) from ES cell stage, suggesting that FOXB1 effect on ES cell differentiation was

372   specific to NC lineage. These data altogether demonstrated that loss of FOXB1 promoted

373   the acquisition of NC.

374

375   <u>FOXB1 Directly Regulated the Transcription of Key Neurogenic and Neural Crest Loci</u>

376   To interrogate the molecular mechanism by which FOXB1 acted on the NC

377   differentiation program, we examined the genome-binding pattern of FOXB1. We

378   performed ChIP-seq of day 3 cultures when FOXB1 protein could be detected using a

379   specific anti-FOXB1 antibody (Fig. S4B, S4F). Peak calling against input sample

380   uncovered 3348 and 4388 high-confidence peaks for 2 wildtype ES cell samples

381   (FDR<0.1%, Day3WT1 and Day3WT2, Fig. 5A) that had significant overlaps with each

382   other (enrichment fold $\log_2$=9.3, p<10^{-5}, Fig 5A') , but only ~10 peaks each for 2 dgRNA-

383   KO samples (FDR<0.1%, Day3KO1 and Day3KO2, Fig. 5A). The 2158 day-3 common

384   peaks had enrichment of GC-rich repeats and forkhead transcription factor-binding motif

385   (data not shown and Fig. 5A"), and were specifically bound by FOXB1 as shown by heat

386   maps of aggregating ChIP-seq signals for wild type, dgRNA-KO, and input samples (Fig.

387   5B). Specificity of FOXB1 binding was additionally verified by ChIP-qPCR assay

388   against IgG and dgRNA-KO controls at selected FOXB1 bound regions (n=13, Fig.

389   S5A).

390   Volcano plot analysis of direct FOXB1 targets combining our above FOXB1

391   sgRNA-KO RNA-seq data and the FOXB1 CHIP-seq data set revealed 3 classes of gene

392   targets (Fig. 5C, Table S5). The first class included neural progenitor genes that were

393   downregulated (>2-fold in RPKM) in the FOXB1 sgRNA-KO day 3 transcriptome such

394   as *ASCL1*, *CRB2*, *RGMA* and *SOX1* (bolded, Fig. 5C). Of interest, *CRB2* and *SOX1* were

395   also identified as pB candidates (Fig. 1F'). The second class included FOXB1 direct

396   targets whose transcription were not affected in FOXB1 sgRNA-KO cells at day 3.

397   Among them, there were genes whose expression was normally increased in our

398   differentiating cultures such as *IGF1R* and *JAG1* (Fig. S1B), and several pB candidates

399   *BOC*, *CNTFR*, *GBX2*, *GREB1L,* *MAP2*, *MEIS2*, *OTX1* and *ZNF503* (Fig. 1F'). Neural

400   progenitor and differentiation genes (*FEZF2*, *ISL1*, *MSI1*, *POU6F1*), placodal ectoderm

401   regulator (*SIX1*), and general DNA methylation machinery components (*DNMT1* and

402   *DNMT3A*) were among FOXB1 targets whose expression was not affected at day 3

15

403   (Table S4). Lastly, the third class included upregulated genes (>2-fold in RPKM) such as

404   *BDNF*, *CCL2, COL1A2, DMBX1, HHIP, LINC01405, MSX2, PAPPA, PAX7, FAT1,*

405   *SALL3, TFAP2B,* and *TLE4* (bolded, Fig. 5C). Among the upregulated gene loci, some of

406   them were known NC genes such as *EDN1*, *MSX2*, *PAX7*, and *TFAP2B*, whereas others

407   were pB candidates (*DMBX1*, Fig 1F') or normally upregulated during NC induction

408   including *CCL2*, *COL1A2*, *FAT1*, *HHIP*, *PAPPA*, and *TLE4* (Fig. S1B and table S1).

409   By visualizing FOXB1-binding sites within target gene loci, we found that for

410   neural progenitor gene loci, FOXB1-binding sites could be found upstream of *ASCL1*,

411   *SOX1*, and *MSI1* genes (asterisks, Fig. 6A-C). Among the FOXB1-bound elements in the

412   upregulated NC gene loci, some were intronic or located within gene bodies (*PAPPA,*

413   *PAX7*, *SALL3*, *TLE4*, *FAT1*, *HHIP*, and *TFAP2B* loci)(asterisks, Fig. 6D, E, S6A); others

414   were located either upstream (*BDNF*, *CCL2*, *COL1A2*, and *EDN1* loci) (asterisks, Fig.

415   6F, S6B) or downstream (asterisks, *MSX2* locus) (Fig. 6G).

416   Genome annotation revealed a significant proportion of intronic FOXB1-bound

417   elements and an enrichment of FOXB1-bound elements locating within promoters (9.6%)

418   and upstream regions (14.8%) (Fig. 6H). We found that many FOXB1-bound elements

419   within 5' UTR exons and promoters were either close to (Fig. 6A-B) or contained (Fig.

420   6C) CpG islands. We performed gene ontogeny analysis on a shortlisted FOXB1-bound

421   element set containing only 5' UTR exons, promoters and upstream elements (n=545).

422   We found significant enrichment of FOXB1 direct targets with gene sets correlated to

423   mammalian-specific phenotype, neuronal differentiation, and developmental processes

424   (Theiler stages 13 to 21) at peripheral (NC-derived) and central (neuroectoderm-derived)

425   nervous system structures (Fig. 6J) consistent with the model that FOXB1 directly

426   regulate human NC cell differentiation. In conclusion, our CHIP-seq data confirmed that

427   FOXB1 directly regulate NC gene induction by directly binding to NC and related gene

428   loci.

429

430   Loss of FOXB1 Function Leads to Long-term Suppression of Autonomic Neuronal Gene

431   Expression

432   Transient expression of a transcription factor can confer developmental

433   competence to differentiating cells to generate subsequent lineages as has been

16

434   demonstrated for other transcription factors such as Neurod1 (Pataskar et al., 2016) and

435   Foxa1/2 (Wang et al., 2015). The binding of FOXB1 in day 3 pre-NC cells to neural

436   progenitor gene loci could therefore carry such developmental function. One of these

437   neural progenitor loci, ASCL1, its gene product is a master regulator for neurogenic

438   potential as well as autonomic neuron differentiation (Lo et al., 1998; Oh et al., 2016).

439   NC contributes to mesenchymal progenitors and peripheral neurons including sensory

440   and autonomic neurons. We therefore speculated that removal of FOXB1, which directly

441   regulated *ASCL1*, would have a far-reaching effect on the differentiation program of

442   human NC cells towards autonomic lineages. While we also hypothesize that effect on

443   *ASCL1* expression or other direct targets may affect the balance of mesenchymal and

444   other neurogenic potentials in FOXB1 KO cell lines.

445   To test these hypotheses, we performed differentiation assay of day 5 isogenic

446   wildtype and dgRNA-KO cells (Fig. 7A) using a tri-inhibitor cocktail (CHIR 99021,

447   SU5402 and DAPT) to promote peripheral neuron (PN) formation (Chambers et al.,

448   2012; Leung et al., 2016), or FGF2 ligand to promote mesenchymal progenitor (MP)

449   formation (Fig. S7A-B). We observed that autonomic neuronal markers *ASCL1* (Fig. 7B),

450   *PHOX2A* (Fig. 7C), and *TH* (Fig. 7D), sensory neuron marker *POU4F1* (Fig. 7E) and

451   general neuronal marker *TUBB3* (Fig. 7F) were induced in day 11 isogenic control cells

452   under PN, but not under MP, condition, confirming that PN condition promotes both

453   sensory and autonomic neuron generation. Notably, *ASCL1* (Fig. 7B) and its downstream

454   targets in autonomic neuron development, the homeodomain transcription factor

455   (*PHOX2A*)(Fig. 7C), and the enzyme synthesizing the neurotransmitter noradrenaline

456   (*TH*) (Fig. 7D), were either not induced to the same extent as isogenic control cells or not

457   induced at all in dgRNA-KO under PN differentiation condition. Expression of general or

458   sensory peripheral neuron markers, such as *POU4F1* (Fig. 7E) and *TUBB3* (Fig. 7F),

459   however, were not changed in dgRNA-KO cells. Intriguingly, we observed that

460   mesenchymal markers *TWIST1* (Fig.7G) and *NT5E* (Fig.7H) were upregulated in day 5

461   and/or day 11-MP dgRNA-KO cells upon MP induction.

462   These data suggested that loss of FOXB1 function promoted expression of

463   mesenchymal genes but had a long-lasting effect on *ASCL1* gene transcription and might

17

464   therefore affect the ability of mature NC to generate autonomic neuronal progenitors as

465   indicated by inability of FOXB1 KO to upregulate other autonomic markers.

466   DISCUSSION

467   Identification and characterization of novel cell intermediates, especially those

468   that exist transiently in early embryos, is a challenging process. Here, we present a

469   pipeline for such research, which utilizes a range of techniques including time-course

470   transcriptomic analysis to identify key transcripts, signaling pathway perturbations to

471   characterize signal dependency, knockout of enhancers to demonstrate regulatory

472   functions, gene targeting of candidate factors and differentiation assays for phenotypic

473   characterization, and genome-wide chromatin binding data to pinpoint genomic

474   occupancy and to infer direct regulations and genetic networks. This pipeline could be

475   applied to studies of other cell types in stem cell and regenerative biology to identify

476   novel cellular states and intermediates. Our data here supports the existence of pB state,

477   however more definitive answers awaits further evidence, perhaps from the use of

478   approaches such as single cell RNA-sequencing and lineage tracing experiments would

479   provide more definitive proof. In this study, we also provide valuable resources and

480   genomic data sets for FOXB1 binding, β-catenin immediate targets during ES cell

481   differentiation, putative novel enhancer location for pB candidates/β-catenin and

482   pluripotent factor regulated loci, as well as transcriptome of NC induction for data mining

483   and for generation of further testable hypotheses.

484   Our pioneer discovery of pB as a cellular intermediate during NC induction

485   (Leung et al., 2016) is gaining momentum and increasing attention in recent years in

486   embryo and reprogramming studies. For instance, a recent work describes a novel

487   population of 'neural border stem cells' reprogrammed from blood cells that bear

488   potential to generate central nervous system and NC derivatives (Thier et al., 2019). In

489   another study, Trevers et al., identifies embryonic 'pB' cells from pre-gastrula avian

490   embryos that bear resembling molecular signatures and differentiation potentials to our *in*

491   *vitro*-derived pB cells (Trevers et al., 2018). Our current manuscript serves as a follow-up

492   study for our previous work and others, and provides the molecular framework to derive

493   pB-like cells from embryos and other differentiated cell types for developmental, stem

494   cell, and therapeutic studies.

495   While inhibition of bone morphogenetic protein (BMP) signaling abolished

496   human NC induction and neural border gene expression, expression of pB genes was

497   preserved (Leung et al., 2016). In our expanded list of 68 pB candidates, we also found

498   that in almost all candidates tested including FOXB1, induction of these genes were

499   immune to inhibition of BMP signaling (data not shown). Such difference in responses to

500   BMP signaling hints at a potential strategic function of pB candidates in NC

501   development.

502   Our data identifies FOXB1 as a key factor limiting NC induction. Here, we show

503   that FOXB1 is expressed transiently in human pre-NC cells before the specification of

504   NC. We additionally show that *Foxb1* expresses in chick embryos prior to the appearance

505   of the neural plate border. Lineage tracing in mouse embryos supports the notion that

506   Foxb1 is expressed in pre-NC cells and there is no evidence to suggest that it is present in

507   mature NC cells or their derivatives.

508   Our findings from FOXB1 ChIP-seq and the phenotypic analyses of FOXB1 KO

509   cells suggest a broader function of FOXB1 in directly regulating NC development.

510   FOXB1 function however is distinct from FOXD3, a pluripotent stem cell and NC factor,

511   whose expression complements FOXB1 during NC induction, and acts as an activator for

512   NC development. The fact that the neural border gene *PAX7* is directly regulated by

513   FOXB1 strongly indicates that FOXB1 acts upstream and represses the development of

514   the neural border stage. Sequence conservation analysis by MultiZ (Emera et al., 2016)

515   reveals that our FOXB1-bound elements are less evolutionary-conserved. More than 5%

516   of these elements are originated from the human clade (compared to <1% in control

517   elements) and up to 25% are primate or younger sequences (data not shown). We

518   postulate that evolution of these young elements could regulate NC developmental

519   processes, thereby shaping species-specific features during human and primate evolution.

520   Indeed, we found strong primate-specific FOXB1 bound elements within key NC loci

521   such as *MSX2* (+41K), *EDN1* (-131K), and *NR2F1* (-91K), which could potentially

522   contribute to the discrepancy between our data and murine embryo data that shows no

523   alteration of border/NC markers such as *Pax3* and *Msx1* in *FOXB1* mutant mice

524   (Labosky et al., 1997).

525   It is assumed that NC cells retain neurogenic potential from neurectoderm

526   progenitors, which NC was believed to derive from. But, here and in our previous work

527   (Leung et al., 2016), we have shown that NC and neural lineages segregated from each

20

528   other before the appearance of definitive neuroectoderm. NC cells therefore have to

529   establish its neurogenic potential in a mechanism either shared with or independent of

530   pre-neuroectoderm progenitors. We found that FOXB1 KO cells acquired a heightened

531   differentiation potential to generate mesenchymal progenitors. This was accompanied by

532   a deficiency in their ability to specify autonomic neuronal fate. Such defect in

533   differentiation potential could be due to a change of axial identity in FOXB1 KO cells

534   (which show an upregulation of anterior *HOX* genes), as migrating NC populations from

535   different axial identity of the rhombomeres are known to differentially contribute to

536   autonomic neuronal lineages (Lumb et al., 2014). Our CHIP-seq data however shows that

537   FOXB1 is directly activating and maintaining the competency of NC cells to express

538   *ASCL1*, an autonomic neuron master regulator (Lo et al., 1998). FOXB1 may act via

539   maintaining the open chromatin structures at *ASCL1* proximal promoter and/or other

540   unknown epigenetic mechanisms to maintain NC cell competency to induce *ASCL1*

541   expression. On the other hand, it could act as a classic transcription factor via recruitment

542   of co-activators or co-repressors to activate or repress target gene expression similar to

543   other FOX family proteins (Lalmansingh et al., 2012), or specific domains within the

544   FOX factor could specify these distinct transcriptional activities as have previously

545   shown for other FOX factors (Neilson et al., 2012).

546   Interestingly, we also find that a significant portion of our FOXB1 peaks (>30%)

547   are occupied by the neuron subtype-specific transcription factor and an autism

548   susceptible gene FOXP1 in human neural progenitor cell lines (Araujo et al., 2015)(data

549   not shown) suggesting that FOXB1 may carry unknown functions in regulating latter

550   neural development.

551   The early requirement of WNT/β-catenin signaling for human NC induction

552   (Leung et al., 2016; Menendez et al., 2011; Mica et al., 2013) has been reported in

553   vertebrate model organisms such as birds, amphibians and fish (Basch et al., 2006; Chang

554   and Hemmati-Brivanlou, 1998; Garcia-Castro et al., 2002; Patthey et al., 2009; Saint-

555   Jeannet et al., 1997)(Prasad et al., submitted). Our data supports a model in which

556   WNT/β-catenin signaling directly activates a panel of pB genes including FOXB1 and

557   other factors such as GBX2 and SP5 that in turn regulate the specification, lineage

558   competency, and differentiation timing of differentiating ES cells to acquire NC fate.

21

559   Previous works have demonstrated pre-broder genes GBX2 and SP5 acting upstream of

560   neural border genes to induce the latter expression (Li et al., 2009; Park et al., 2013). We

561   propose that pro-neural crest factors such as GBX2 and SP5 could act parallel to FOXB1

562   to influence NC fate choice.

563   This work further suggest that the proposed transient pB cell stage has unique

564   transcriptional, epigenetic and functional features, different from the pluripotent cell state

565   from which it emerges, and from later stages NC progenitors found at the neural plate

566   border and neural folds. We have confirmed early expression for several pB gene

567   candidates prior to expression of neural plate border genes in chick embryos, and

568   demonstrated in our human system their early expression to be canonical WNT

569   dependent and insensitive to BMP inhibition. Focusing on FOXB1 as a pB gene example,

570   we showed that it directly regulates NC loci, temporally controls progression of

571   differentiating cells towards more advanced NC state, and carries a remarkable impact in

572   terminal differentiation driven by its early transient expression. These findings lend

573   strong support to our proposed pB state and its significance to NC ontogeny.

574

585   AUTHOR CONTRIBUTIONS

586   Conceptualization: A.W.L., A.Z.X., M.I.G.C.; Methodology: A.W.L., C.B.; Software:

587   F.L-G; Writing - Original Draft: A.W.L.; Writing – Review & Editing: A.W.L., A.Z.X.,

588   M.I.G.C; Formal analysis: A.W.L., F.L-G; Investigation: A.W.L., B.C., M.S.P., J.C.;

589   Visualization: A.W.L., F.L-G, K.L.; Funding acquisition: A.Z.X., M.I.G.C.

590

591   DECLARATION OF INTERESTS

592   The authors declare no conflict of interests.

593   REFERENCE

594   Alvarez-Bolado, G., Zhou, X., Voss, A.K., Thomas, T., and Gruss, P. (2000). Winged
595   helix transcription factor Foxb1 is essential for access of mammillothalamic axons to
596   the thalamus. Development (Cambridge, England) *127*, 1029-1038.
597   Araujo, D.J., Anderson, A.G., Berto, S., Runnels, W., Harper, M., Ammanuel, S., Rieger,
598   M.A., Huang, H.C., Rajkovich, K., Loerwald, K.W.*, et al.* (2015). FoxP1 orchestration of
599   ASD-relevant signaling pathways in the striatum. Genes & development *29*, 2081-
600   2096.
601   Arduini, B.L., and Brivanlou, A.H. (2012). Modulation of FOXD3 activity in human
602   embryonic stem cells directs pluripotency and paraxial mesoderm fates. Stem cells
603   *30*, 2188-2198.
604   Basch, M.L., Bronner-Fraser, M., and Garcia-Castro, M.I. (2006). Specification of the
605   neural crest occurs during gastrulation and requires Pax7. Nature *441*, 218-222.
606   Betschinger, J., Nichols, J., Dietmann, S., Corrin, P.D., Paddison, P.J., and Smith, A.
607   (2013). Exit from pluripotency is gated by intracellular redistribution of the bHLH
608   transcription factor Tfe3. Cell *153*, 335-347.
609   Betters, E., Charney, R.M., and Garcia-Castro, M.I. (2018). Early specification and
610   development of rabbit neural crest cells. Dev Biol.
611   Chambers, S.M., Qi, Y., Mica, Y., Lee, G., Zhang, X.J., Niu, L., Bilsland, J., Cao, L., Stevens,
612   E., Whiting, P.*, et al.* (2012). Combined small-molecule inhibition accelerates
613   developmental timing and converts human pluripotent stem cells into nociceptors.
614   Nature biotechnology *30*, 715-720.
615   Chang, C., and Hemmati-Brivanlou, A. (1998). Neural crest induction by Xwnt7B in
616   Xenopus. Dev Biol *194*, 129-134.
617   Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human
618   genome. Nature *489*, 57-74.
619   Emera, D., Yin, J., Reilly, S.K., Gockley, J., and Noonan, J.P. (2016). Origin and
620   evolution of developmental enhancers in the mammalian neocortex. Proceedings of
621   the National Academy of Sciences of the United States of America *113*, E2617-2626.
622   Estaras, C., Benner, C., and Jones, K.A. (2015). SMADs and YAP compete to control
623   elongation of beta-catenin:LEF-1-recruited RNAPII during hESC differentiation.
624   Molecular cell *58*, 780-793.
625   Fernandez-Garre, P., Rodriguez-Gallardo, L., Gallego-Diaz, V., Alvarez, I.S., and
626   Puelles, L. (2002). Fate map of the chicken neural plate at stage 4. Development
627   (Cambridge, England) *129*, 2807-2822.
628   Funa, N.S., Schachter, K.A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honore, C.,
629   Hansen, K., and Semb, H. (2015). beta-Catenin Regulates Primitive Streak Induction
630   through Collaborative Interactions with SMAD2/SMAD3 and OCT4. Cell Stem Cell
631   *16*, 639-652.
632   Garcia-Castro, M.I., Marcelle, C., and Bronner-Fraser, M. (2002). Ectodermal Wnt
633   function as a neural crest inducer. Science *297*, 848-851.
634   Genga, R.M.J., Kernfeld, E.M., Parsi, K.M., Parsons, T.J., Ziller, M.J., and Maehr, R.
635   (2019). Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular
636   Drivers of Early Human Endoderm Development. Cell Rep *27*, 708-718 e710.

637 Gertz, J., Savic, D., Varley, K.E., Partridge, E.C., Safi, A., Jain, P., Cooper, G.M., Reddy,
638 T.E., Crawford, G.E., and Myers, R.M. (2013). Distinct properties of cell-type-specific
639 and shared transcription factor binding sites. Molecular cell *52*, 25-36.
640 Golson, M.L., and Kaestner, K.H. (2016). Fox transcription factors: from development
641 to disease. Development (Cambridge, England) *143*, 4558-4570.
642 Hackett, J.A., Huang, Y., Gunesdogan, U., Gretarsson, K.A., Kobayashi, T., and Surani,
643 M.A. (2018). Tracing the transitions from pluripotency to germ cell fate with CRISPR
644 screening. Nature communications *9*, 4292.
645 Hall, B.K. (2000). The neural crest as a fourth germ layer and vertebrates as
646 quadroblastic not triploblastic. Evolution & development *2*, 3-5.
647 Iwafuchi-Doi, M., and Zaret, K.S. (2016). Cell fate control by pioneer transcription
648 factors. Development (Cambridge, England) *143*, 1833-1837.
649 Kalkan, T., and Smith, A. (2014). Mapping the route from naive pluripotency to
650 lineage specification. Philosophical transactions of the Royal Society of London
651 Series B, Biological sciences *369*.
652 Kos, R., Reedy, M.V., Johnson, R.L., and Erickson, C.A. (2001). The winged-helix
653 transcription factor FoxD3 is important for establishing the neural crest lineage and
654 repressing melanogenesis in avian embryos. Development (Cambridge, England)
655 *128*, 1467-1479.
656 Krishnakumar, R., Chen, A.F., Pantovich, M.G., Danial, M., Parchem, R.J., Labosky, P.A.,
657 and Blelloch, R. (2016). FOXD3 Regulates Pluripotent Stem Cell Potential by
658 Simultaneously Initiating and Repressing Enhancer Activity. Cell stem cell *18*, 104-
659 117.
660 Labosky, P.A., Winnier, G.E., Jetton, T.L., Hargett, L., Ryan, A.K., Rosenfeld, M.G.,
661 Parlow, A.F., and Hogan, B.L. (1997). The winged helix gene, Mf3, is required for
662 normal development of the diencephalon and midbrain, postnatal growth and the
663 milk-ejection reflex. Development (Cambridge, England) *124*, 1263-1274.
664 Lalmansingh, A.S., Karmakar, S., Jin, Y., and Nagaich, A.K. (2012). Multiple modes of
665 chromatin remodeling by Forkhead box proteins. Biochim Biophys Acta *1819*, 707-
666 715.
667 Leung, A.W., Murdoch, B., Salem, A.F., Prasad, M.S., Gomez, G.A., and Garcia-Castro,
668 M.I. (2016). WNT/beta-catenin signaling mediates human neural crest induction via
669 a pre-neural border intermediate. Development *143*, 398-410.
670 Li, B., Kuriyama, S., Moreno, M., and Mayor, R. (2009). The posteriorizing gene Gbx2
671 is a direct target of Wnt signalling and the earliest factor in neural crest induction.
672 Development (Cambridge, England) *136*, 3267-3278.
673 Lo, L., Tiveron, M.C., and Anderson, D.J. (1998). MASH1 activates expression of the
674 paired homeodomain transcription factor Phox2a, and couples pan-neuronal and
675 subtype-specific components of autonomic neuronal identity. Development
676 (Cambridge, England) *125*, 609-620.
677 Lukoseviciute, M., Gavriouchkina, D., Williams, R.M., Hochgreb-Hagele, T.,
678 Senanayake, U., Chong-Morrison, V., Thongjuea, S., Repapi, E., Mead, A., and Sauka-
679 Spengler, T. (2018). From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically
680 Remodels Neural Crest Regulatory Landscape In Vivo. Dev Cell *47*, 608-628 e606.

681 Lumb, R., Wiszniak, S., Kabbara, S., Scherer, M., Harvey, N., and Schwarz, Q. (2014).
682 Neuropilins define distinct populations of neural crest cells. Neural development *9*,
683 24.
684 Ma, Y., Yao, N., Liu, G., Dong, L., Liu, Y., Zhang, M., Wang, F., Wang, B., Wei, X., Dong, H.,
685 *et al.* (2015). Functional screen reveals essential roles of miR-27a/24 in
686 differentiation of embryonic stem cells. The EMBO journal *34*, 361-378.
687 Menendez, L., Yatskievych, T.A., Antin, P.B., and Dalton, S. (2011). Wnt signaling and
688 a Smad pathway blockade direct the differentiation of human pluripotent stem cells
689 to multipotent neural crest cells. Proc Natl Acad Sci U S A *108*, 19240-19245.
690 Mica, Y., Lee, G., Chambers, S.M., Tomishima, M.J., and Studer, L. (2013). Modeling
691 neural crest induction, melanocyte specification, and disease-related pigmentation
692 defects in hESCs and patient-specific iPSCs. Cell Rep *3*, 1140-1152.
693 Neilson, K.M., Klein, S.L., Mhaske, P., Mood, K., Daar, I.O., and Moody, S.A. (2012).
694 Specific domains of FoxD4/5 activate and repress neural transcription factor genes
695 to control the progression of immature neural ectoderm to differentiating neural
696 plate. Dev Biol *365*, 363-375.
697 Oh, Y., Cho, G.S., Li, Z., Hong, I., Zhu, R., Kim, M.J., Kim, Y.J., Tampakakis, E., Tung, L.,
698 Huganir, R., *et al.* (2016). Functional Coupling with Cardiac Muscle Promotes
699 Maturation of hPSC-Derived Sympathetic Neurons. Cell stem cell *19*, 95-106.
700 Park, D.S., Seo, J.H., Hong, M., Bang, W., Han, J.K., and Choi, S.C. (2013). Role of Sp5 as
701 an essential early regulator of neural crest specification in xenopus. Developmental
702 dynamics : an official publication of the American Association of Anatomists *242*,
703 1382-1394.
704 Pataskar, A., Jung, J., Smialowski, P., Noack, F., Calegari, F., Straub, T., and Tiwari, V.K.
705 (2016). NeuroD1 reprograms chromatin and transcription factor landscapes to
706 induce the neuronal program. The EMBO journal *35*, 24-45.
707 Patthey, C., Edlund, T., and Gunhaga, L. (2009). Wnt-regulated temporal control of
708 BMP exposure directs the choice between neural plate border and epidermal fate.
709 Development (Cambridge, England) *136*, 73-83.
710 Pijuan-Sala, B., Guibentif, C., and Gottgens, B. (2018). Single-cell transcriptional
711 profiling: a window into embryonic cell-type specification. Nat Rev Mol Cell Biol *19*,
712 399-412.
713 Prasad, M.S., Charney, R.M., and Garcia-Castro, M.I. (2019). Specification and
714 formation of the neural crest: Perspectives on lineage segregation. Genesis *57*,
715 e23276.
716 Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J.
717 (2011). A unique chromatin signature uncovers early developmental enhancers in
718 humans. Nature *470*, 279-283.
719 Respuela, P., Nikolic, M., Tan, M., Frommolt, P., Zhao, Y., Wysocka, J., and Rada-
720 Iglesias, A. (2016). Foxd3 Promotes Exit from Naive Pluripotency through Enhancer
721 Decommissioning and Inhibits Germline Specification. Cell stem cell *18*, 118-133.
722 Saint-Jeannet, J.P., He, X., Varmus, H.E., and Dawid, I.B. (1997). Regulation of dorsal
723 fate in the neuraxis by Wnt-1 and Wnt-3a. Proceedings of the National Academy of
724 Sciences of the United States of America *94*, 13713-13718.

725   Sasai, N., Mizuseki, K., and Sasai, Y. (2001). Requirement of FoxD3-class signaling for
726   neural crest determination in Xenopus. Development (Cambridge, England) *128*,
727   2525-2536.
728   Seo, S., Chen, L., Liu, W., Zhao, D., Schultz, K.M., Sasman, A., Liu, T., Zhang, H.F., Gage,
729   P.J., and Kume, T. (2017). Foxc1 and Foxc2 in the Neural Crest Are Required for
730   Ocular Anterior Segment Development. Investigative ophthalmology & visual
731   science *58*, 1368-1377.
732   Seo, S., and Kume, T. (2006). Forkhead transcription factors, Foxc1 and Foxc2, are
733   required for the morphogenesis of the cardiac outflow tract. Dev Biol *296*, 421-436.
734   Stuhlmiller, T.J., and Garcia-Castro, M.I. (2012). Current perspectives of the signaling
735   pathways directing neural crest induction. Cell Mol Life Sci *69*, 3715-3737.
736   Takebayashi-Suzuki, K., Kitayama, A., Terasaka-Iioka, C., Ueno, N., and Suzuki, A.
737   (2011). The forkhead transcription factor FoxB1 regulates the dorsal-ventral and
738   anterior-posterior patterning of the ectoderm during early Xenopus embryogenesis.
739   Dev Biol *360*, 11-29.
740   Teng, L., Mundell, N.A., Frist, A.Y., Wang, Q., and Labosky, P.A. (2008). Requirement
741   for Foxd3 in the maintenance of neural crest progenitors. Development (Cambridge,
742   England) *135*, 1615-1624.
743   Teo, A.K., Arnold, S.J., Trotter, M.W., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J.,
744   Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm
745   specification through eomesodermin. Genes & development *25*, 238-250.
746   Thier, M.C., Hommerding, O., Panten, J., Pinna, R., Garcia-Gonzalez, D., Berger, T.,
747   Worsdorfer, P., Assenov, Y., Scognamiglio, R., Przybylla, A., *et al.* (2019).
748   Identification of Embryonic Neural Plate Border Stem Cells and Their Generation by
749   Direct Reprogramming from Adult Human Blood Cells. Cell stem cell *24*, 166-182
750   e113.
751   Trevers, K.E., Prajapati, R.S., Hintze, M., Stower, M.J., Strobl, A.C., Tambalo, M.,
752   Ranganathan, R., Moncaut, N., Khan, M.A.F., Stern, C.D., *et al.* (2018). Neural
753   induction by the node and placode induction by head mesoderm share an initial
754   state resembling neural plate border and ES cells. Proceedings of the National
755   Academy of Sciences of the United States of America *115*, 355-360.
756   Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., and
757   Meissner, A. (2015). Transcription factor binding dynamics during human ES cell
758   differentiation. Nature *518*, 344-349.
759   Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y.,
760   Wang, J., *et al.* (2015). Epigenetic priming of enhancers predicts developmental
761   competence of hESC-derived endodermal lineage intermediates. Cell stem cell *16*,
762   386-399.
763   Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage
764   specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell
765   stem cell *10*, 440-454.
766   Wehr, R., Mansouri, A., de Maeyer, T., and Gruss, P. (1997). Fkh5-deficient mice show
767   dysgenesis in the caudal midbrain and hypothalamic mammillary body.
768   Development (Cambridge, England) *124*, 4447-4456.

769   Xu, P., Balczerski, B., Ciozda, A., Louie, K., Oralova, V., Huysseune, A., and Crump, J.G.
770   (2018). Fox proteins are modular competency factors for facial cartilage and tooth
771   specification. Development (Cambridge, England) *145*.
772   Yamamoto, T., Mencarelli, M.A., Di Marco, C., Mucciolo, M., Vascotto, M., Balestri, P.,
773   Gerard, M., Mathieu-Dramard, M., Andrieux, J., Breuning, M., *et al.* (2014).
774   Overlapping microdeletions involving 15q22.2 narrow the critical region for
775   intellectual disability to NARG2 and RORA. Eur J Med Genet *57*, 163-168.
776   Zhao, T., Zhou, X., Szabo, N., Leitges, M., and Alvarez-Bolado, G. (2007). Foxb1-driven
777   Cre expression in somites and the neuroepithelium of diencephalon, brainstem, and
778   spinal cord. Genesis *45*, 781-787.
779

780   FIGURE LEGENDS

781   **Figure 1.** Identification of pre-border candidates by transcriptomic analysis. (A)

782   Schematic diagram depicting the differentiation program for human NC induction.

783   Asterisks indicate time points when samples were harvested for RNA-seq. (B) SOX10

784   and PAX6 immunostaining of day 5 NC cells. (C) Hierarchical clustering of all expressed

785   genes from ESC, day 3 and day 5 transcriptomes. (D, E) K-means clustering of the

786   RPKM values of transcripts mapped in the ESC, day 3 and day 5 transcriptomes that are

787   functionally associated with pluripotency and self renewal of human ES cells (D), or

788   neural plate border, pre-migratory neural crest and differentiating neural crest cells (E).

789   Relative intensities were displayed with blue being lowest in expression level to red

790   being highest. (F) Plot of all transcripts identified from RNA-seq with day 3 RPKM

791   values plotted against day 3 fold change over day 0 (ES cells) (F') Pink boxed-region is

792   magnified on the right to display the 68 pB candidates transcripts. The 4 pre-border

793   candidates identified in (Leung et al., 2016) are displayed in bold.

794

795   **Figure 2.** β-catenin positively regulates pB candidate gene induction. (A) Schematic

796   diagram showing lentiviral constructs containing luciferase and β-catenin shRNAs and

797   the procedures for generating stably infected shRNA expressing human ES cell lines. (B)

798   qPCR data of *CTNNB1* and *POU5F1* comparing control (luciferase shRNA) against

799   knockdown (*β-catenin* shRNA) in ES and day 1 cultures. (C) Heat map of the expression

800   dynamics of the 68 pB candidates in control and knockdown cell lines derived from

801   RNA-seq data of ES cell, day 1 and day 3 cells generated using Morpheus. (D)

802   Differentiation day 3 cultures treated with or without FGF2 ligand at the indicated

803   concentration. Dunnett's multiple comparison tests were carried to calculate statistics

804   between control group with ligand-treated groups. (E) Differentiation day 3 cultures

805   treated with or without BMP4 ligand. Fisher's LSD test was used. (F) A step-wise

806   induction model for human NC and a gene regulatory network of human NC with

807   signaling inputs from canonical WNT, FGF and BMP signaling.

808

809   **Figure 3.** Functional characterization of pB gene regulatory elements. (A-C) Loci for pB

810   candidate embedded with published ChIP-seq and ATAC-seq data sets in bigwig formats

31

811   were displayed on UCSC browsers. 'ESC stage' indicates the sequencing data were

812   derived from human ES cells. 'WNT-Dif' indicates the data were from human ES cells

813   treated with CHIR 99021 or WNT3A ligand. qPCR measurements on ChIP DNA pulled

814   down using anti-NANOG (D) and anti-H3K27ac (E) antibodies. Isotype-matched IgGs

815   were used as controls. Asterisks directly above standard deviation bars represent

816   comparisons between IgG and H3K27ac groups at their corresponding time points. One-

817   way ANOVA tests were carried out with Bonferroni's corrections for multiple testing.

818   (F) CRISPR targeting strategy for FOXB1-E1 element. (G) PCR genotyping of isogenic

819   and FOXB1-E1 targeted clones. Expected sizes of wildtype and knockout PCR products

820   are shown on the right. (H) Day wise time course quantification of *FOXB1* expression

821   during NC induction in isogenic-WT1 and FOXB1-E1-KO1 cell lines. Three culture

822   samples were collected for each cell lines at each differentiation day. Unpaired t-tests

823   with Holm-Sidak corrections for multiple testing were carried out on the indicated time

824   points. (I) Normalized expression values and statistics at each time points were displayed.

825

826   **Figure 4.** Phenotypic characterization of a human *FOXB1* loss-of-function cell model.

827   (A) RNA-seq reads on human *FOXB1* locus converted to bigwig format and displayed on

828   igv software. Sample labeling refers to Fig 1C. (B) Time course-qPCR analysis of

829   selected FOX transcription factor transcripts during ES cell differentiation. (C-D) *Foxb1*

830   *in situ* hybridization in HH4 and HH8 stage chick embryos. Transverse sections are

831   displayed to the right (C' and D'). Arrows are pointing to the ectoderm cell layer.

832   Asterisks indicate the position of the neural plate border. (E) qPCR data of NC factors on

833   day 5 cultures treated with control or FOXB1 siRNA. A representative of 3 biological

834   replicates is displayed. (F) DAPI staining and SOX10 immunostaining of day 5 cultures

835   treated with control or FOXB1 siRNA. Magnified images for SOX10+ cells are displayed

836   in the insets. (G) CRISPR targeting strategy for CRISPR-mediated mutations at the DNA

837   binding domain sequence of *FOXB1* gene using two gRNAs. (H) PCR genotyping of

838   isogenic and dgRNA targeted cell lines. (I) Time-course qPCR measurements of NC

839   factors on three isogenic and dgRNA-KO cell lines. Two-way ANOVA was performed

840   with Dunnett correction for multiple testing. (J) Differentiation and sample collection

841   schematics for NC induction. (K) Time-course qPCR data of NC genes for isogenic

842   wildtype and dgRNA-KO cell lines during NC induction. Multiple t-tests with Holm-

843   Sidak corrections were performed. (L, M). *POU5F1* and *GBX2* time-course qPCR

844   comparing isogenic control and *FOXB1* dgRNA-KO cell lines. Multiple t-tests with

845   Holm-Sidak corrections were performed. (N) Differentiation and sample collection

846   schematics for neurectoderm competency test. (O) Time-course qPCR data of

847   neurectoderm genes during differentiation using a schematic displayed in panel N.

848   Multiple t-tests with Holm-Sidak corrections were performed. Isogenic wildtype and

849   sgRNA-KO ES cells were treated with FGF2 or Activin to promote mesoderm (P) and

850   definitive endoderm (Q) induction as indicated by T and SOX17 immunostaining

851   respectively.

852

853   **Figure 5.** Global genomic binding analysis of FOXB1 protein during human neural crest

854   induction. (A) Peak numbers called from FOXB1 ChIP-seq on 2 wildtype and 2 dgRNA-

855   KO samples. (A') Venn diagram showing overlapping elements for the 2 wildtype ChIP-

856   seq samples. (A") Motif analysis of 2158 FOXB1-bound elements. (B) Heat map of

857   signals centered on the 2158 FOXB1-bound elements for 2 wildtype, 2 knockout and an

858   input samples. (C) Volcano plot of all FOXB1 targets within a 50kb window from TSS.

859   Purple dots indicate genes that are differentially expressed in our earlier ES, day 3 and

860   day 5 transcriptome (Fig 1). Bolded genes are key candidates for neural and NC
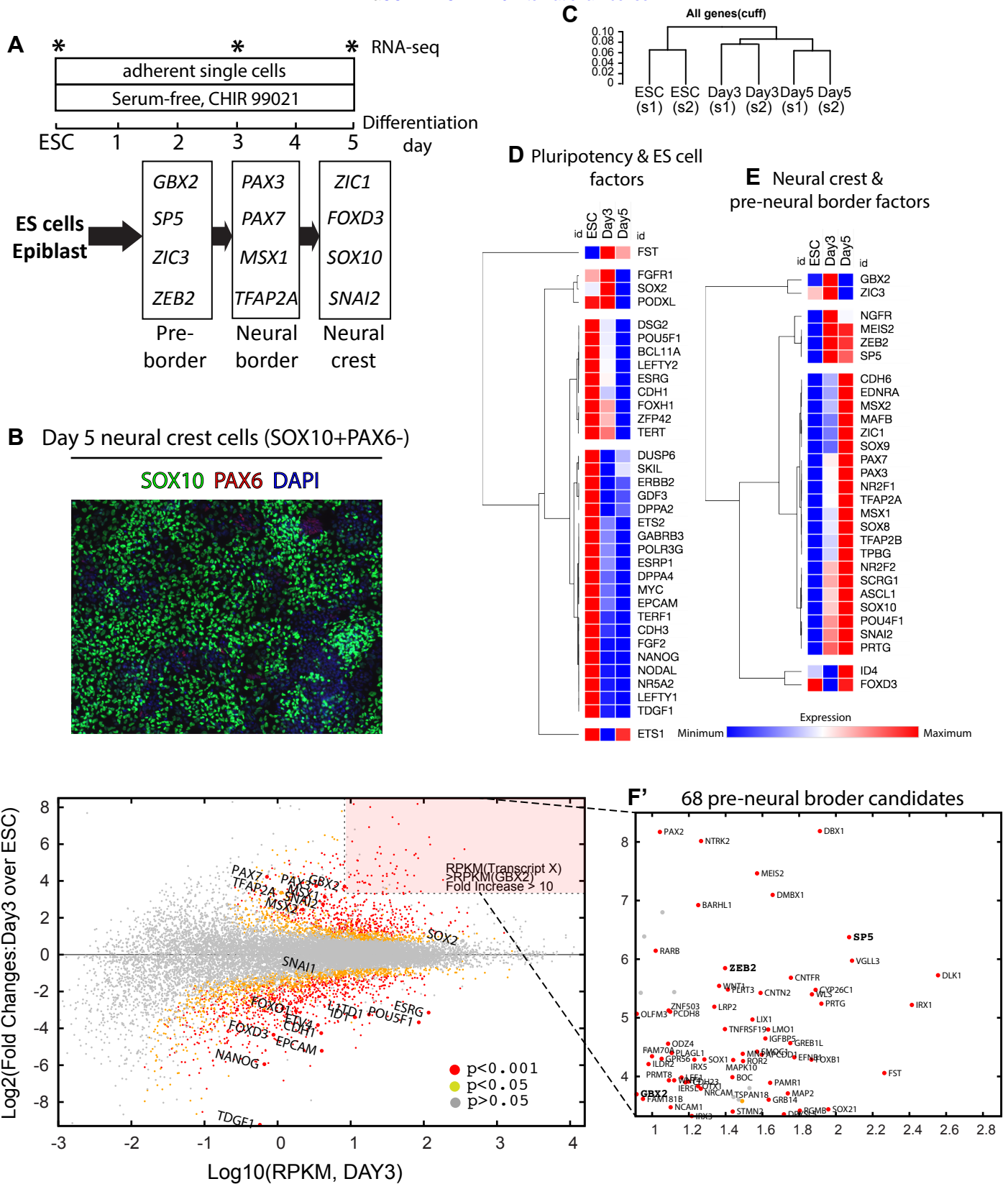
861   development.

862

863   **Figure 6.** FOXB1 directly regulates target neural and neural crest genes. IGV tracks

864   displaying ChIP-seq signals for the 2 wildtype, 2 knockout samples, and an optional CpG

865   annotation at neuronal progenitor (A-C) and neural crest (D-G) gene loci. Asterisks

866   indicate location of called peaks. (H) Genome annotation of the 2158 common FOXB1

867   peak regions using NCBI Ref. Seq. build hg19. Priority was set as follows: Promoters,

868   5'UTR exons, upstream sequences, 3'UTR exons, coding exons, introns, downstream

869   sequences and intergenic regions. Actual peak numbers are shown in brackets. (I) The

870   545 selected FOXB1-bound elements, contained within upstream elements, promoters,

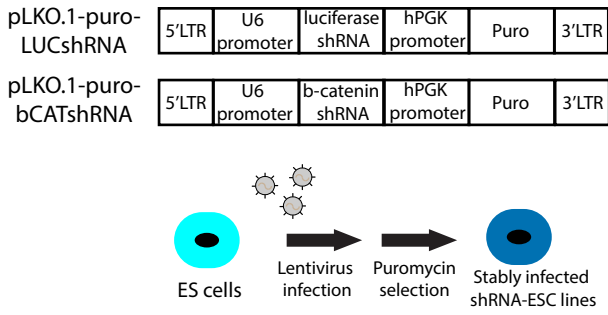871   and 5' UTR exons, were further classified according to the presence or absence of CpG

33

872   islands. (J) GREAT gene ontogeny analysis of the 545 selected FOXB1 peak set using a

873   20 kb genomic window.

874

875   **Figure 7.** *FOXB1* regulates autonomic neuron differentiation potential in NC cells. (A)

876   Terminal differentiation schematic for NC cells. qPCR analysis of autonomic neuronal

877   (B-D), sensory and general neuronal (E-F) and mesenchymal progenitor (G-H) markers

878   on ES, day 5 and differentiated day 11 isogenic control and dgRNA-KO cells.
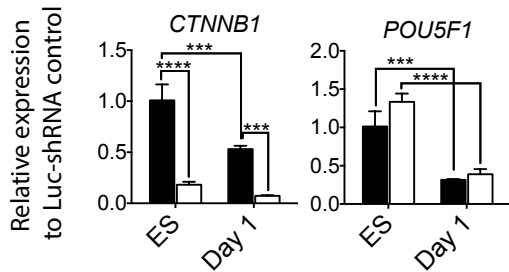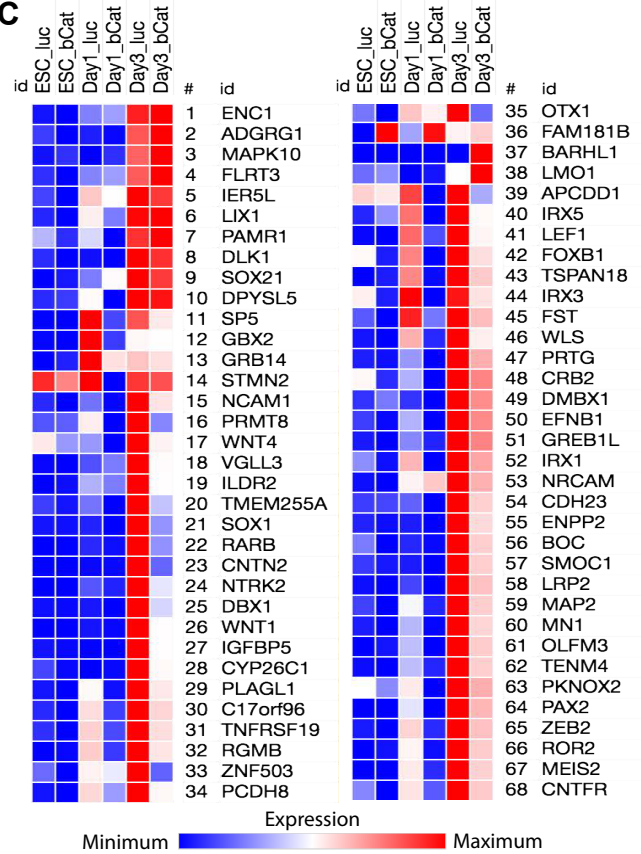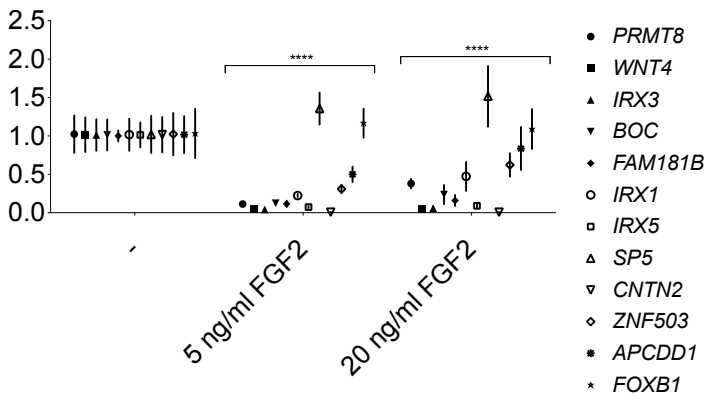
A — RNA-seq timeline; adherent single cells, Serum-free, CHIR 99021; ESC to Differentiation day 5. ES cells/Epiblast → Pre-border (GBX2, SP5, ZIC3, ZEB2) → Neural border (PAX3, PAX7, MSX1, TFAP2A) → Neural crest (ZIC1, FOXD3, SOX10, SNAI2).

B — Day 5 neural crest cells (SOX10+PAX6-); SOX10 PAX6 DAPI

C — All genes(cuff): ESC (s1), ESC (s2), Day3 (s1), Day3 (s2), Day5 (s1), Day5 (s2)

D — Pluripotency & ES cell factors

E — Neural crest & pre-neural border factors

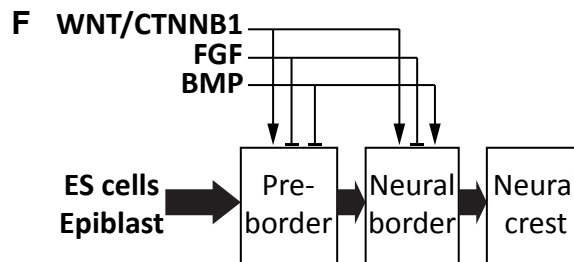**F** — Log2(Fold Changes:Day3 over ESC) vs Log10(RPKM, DAY3); p<0.001 (red), p<0.05 (yellow), p>0.05 (grey)

F' — 68 pre-neural broder candidates
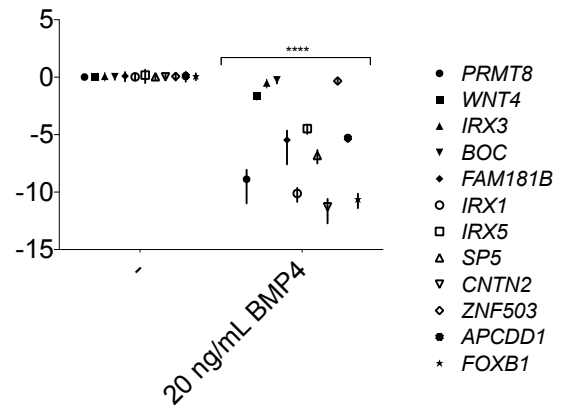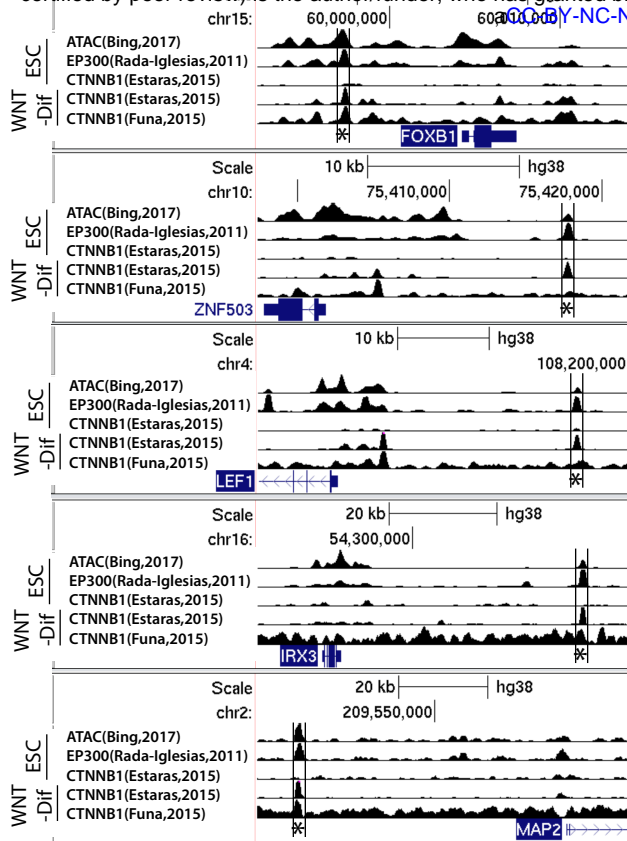
A Upstream pB enhancers

C Intronic pB enhancers

A Number of called FOXB1-CHIP-seq peaks

A' Day3WT1 Day3WT2 — 1190 **2158** 2230

A" Forkhead motif p=6.2e-11

B Common FOXB1 peaks (n=2158)

C Day 3, Direct Targets (n=1145) of FOXB1 bound regions (n=2158)

Downregulated >2-fold in KO — Upregulated >2-fold in KO

Differential genes identified in human NC induction time course transcriptome

Not differentially expressed during human NC induction transcriptome

Neural progenitor loci

Neural crest loci

A — Day3WT1, Day3WT2, Day3KO1, Day3KO2, CpG Islands — 6,811 bp — *ASCL1*

B — Day3WT1, Day3WT2, Day3KO1, Day3KO2, CpG Islands — 11 kb — *LINC00403* *SOX1*

C — Day3WT1, Day3WT2, Day3KO1, Day3KO2, CpG Islands — 6,175 bp — *MSI1*

D — Day3WT1, Day3WT2, Day3KO1, Day3KO2 — 23 kb — *PAX7*

E — Day3WT1, Day3WT2, Day3KO1, Day3KO2 — 15 kb — *TFAP2B*

F — Day3WT1, Day3WT2, Day3KO1, Day3KO2 — 11 kb — *EDN1*

G — Day3WT1, Day3WT2, Day3KO1, Day3KO2 — 50 kb — *MSX2* *MIR4634*

H — Percentage of FOXB1 common Peaks (abs number): Promoters, 3kb (206) 9.6%; 5'UTR Exons (20) 0.9%; Upstream Elements, 20kb (319) 14.8%; 3'UTR Exons (24) 1.1%; Coding Exons (52) 2.4%; Introns (818) 37.9%; Downstream Elements, 20kb (132) 6.1%; Intergenic Regions (570) 26.4%

I — CpG / Non-CpG; Promoters, 3kb: 94 / 112; 5'UTR Exons: 9 / 11; Upstream Elements, 20kb: 26 / 293; Percentage of Elements

J — FOXB1 promoter/upstream bound regions (n=545); Binom. FDR Q-Val