Manuscript Click here to view linked References

Neural crest induction requires SALL4-mediated BAF recruitment to lineage specific enhancers

Martina Demurtas[1], Samantha M. Barnada[2], Emma van Domselaar[1], Zoe H. Mitchell[1], Laura Deelen[1], and Marco Trizzino[1,*]

[1]Department of Life Sciences, Imperial College London, London, UK

[2] Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA, USA

*Lead contact: m.trizzino@imperial.ac.uk

SUMMARY

Neural crest induction begins early during neural plate formation, requiring precise transcriptional control to activate lineage-specific enhancers. Here, we demonstrate that SALL4, a transcription factor strongly expressed in cranial neural crest cells (CNCCs) and associated with syndromes featuring craniofacial anomalies, plays a critical role in this early regulatory process. Using a *SALL4*-haploinsufficient human iPSC model that recapitulates clinical haploinsufficiency, we show that the SALL4A isoform directly interacts with the BAF complex subunit DPF2 through its zinc-finger-3 domain. This interaction enables SALL4-mediated recruitment of BAF to CNCC-specific enhancers during early neural plate stages. Without functional SALL4, BAF fails to load at chromatin, leaving CNCC enhancers inaccessible despite normal neuroectodermal progression. Consequently, cells cannot undergo proper CNCC induction and specification due to persistent enhancer repression. Our findings reveal SALL4 as an essential regulator of BAF-dependent enhancer activation during early neural crest development, providing molecular insights into SALL4-associated craniofacial anomalies.

KEYWORDS

SALL4, DPF2, SALL4A, BAF, neural crest, enhancers, chromatin accessibility

INTRODUCTION

Embryonic development is a tightly regulated process that depends on the ability of pluripotent embryonic stem cells (ESCs) to progressively acquire specialized fates and functions[1]. This enables their differentiation into a diverse range of somatic lineages, ultimately giving rise to a complex, fully integrated organism from a single cell[2]. These diverse cell identities are established through epigenetic mechanisms that regulate the spatiotemporal expression of lineage-specific gene regulatory networks, primarily by modulating chromatin accessibility. Chromatin states range from open to closed, shaped by a combination of epigenetic modifications, including histone methylation and acetylation, incorporation of histone variants, DNA methylation, and the activity of chromatin remodeling complexes. These large, multi-subunit complexes utilize ATP hydrolysis to reposition nucleosomes, dynamically regulating access to genes and their regulatory elements[3–6].

Transcription factors also play a crucial role in cell fate determination, acting as molecular switches that trigger mechanisms that ultimately activate or repress genes as a result to their binding to specific regulatory DNA sequences within *cis*-regulatory elements. They influence gene expression by recruiting or inhibiting the transcriptional machinery and regulating 3D chromatin organization[3,4,6–9]. Among these, the embryonic C2H2 zinc-finger transcription factor SALL4 is involved in a wide range of developmental processes[10]. SALL4 is highly expressed in pluripotent cells and throughout embryonic development, gradually decreasing after birth to become undetectable in adult tissues[11].

In humans, the SALL4 gene produces two main isoforms through alternative splicing: SALL4A (1053 aa) and SALL4B (617 aa), which have been proposed to play distinct roles in ESCs. Namely, Rao et al.[12] suggested that SALL4B has function limited to pluripotency regulation in ESCs, while SALL4A is the isoform primarily active during development. Although SALL4 is expressed as early as the two-cell stage, its role in maintaining pluripotency remains debated. Some studies suggest that SALL4 cooperates with core pluripotency transcription factors, such as OCT4 and NANOG, to sustain stem cell identity[13,14]. In contrast, other studies propose that SALL4 plays a more marginal role in directly regulating the pluripotency gene regulatory network[15–17].

During early mouse development, *Sall4* is broadly expressed throughout the embryo, however, as gastrulation progresses, its expression becomes more posteriorly restricted, localizing primarily to the developing tail bud and limb buds[18,19]. Here, it regulates the cell fate commitment of neuromesodermal progenitors by suppressing the neural gene regulatory network in mesodermal cells while simultaneously enhancing mesodermal gene expression[20]. Additionally,

Sall4 remains expressed in mouse craniofacial regions until the E9.5 stage, including the developing frontonasal process as well as the mandibular and maxillary arches[18]. These craniofacial structures originate from cranial neural crest cells (CNCCs), a subpopulation of neural crest cells that migrate dorsolaterally to give rise to the cartilage and bones of the face and anterior skull, as well as cranial neurons and glia. Consistent with this, individuals with *SALL4* haploinsufficiency exhibit a range of phenotypes, including distinctive facial features such as epicanthal folds, widely spaced eyes, and a depressed nasal bridge, along with cranial nerve-related conditions. These include Duane anomaly, a congenital strabismus resulting from abnormal oculomotor nerve innervation[21], and sensorineural hearing loss[22].

Together, these findings suggest that SALL4 may play a critical role in CNCC induction, specification and differentiation. However, its involvement in this developmental process has not yet been investigated. Previous studies have also suggested that SALL4 interacts directly with chromatin remodeling complexes. For example, several studies in mice have identified Sall4 as an interactor of the NuRD chromatin remodeling complex, in both embryonic cells and various developmental processes[15,23–26]. Additionally, recent mass-spectrometry experiments conducted by our group in early stages of human iPSC-to-CNCC differentiation, and specifically at the neuroectodermal stage, revealed that SALL4 interacts with the BAF chromatin remodeling complex during this developmental process[27].

Since this was the first report of a potential direct SALL4-BAF interaction, here we set out to investigate the scope of this interaction, and to determine the role of SALL4 during cranial neural crest cell specification. Given that human SALL4-associated developmental syndromes result from *SALL4* haploinsufficiency, we generated a *SALL4* heterozygous human iPSC line using CRISPR-Cas9-mediated deletion. With this approach, we demonstrate that, at early stages of CNCC specification, SALL4 is required for the recruitment of the BAF complex to open the chromatin to enhancers essential for CNCC fate. In the absence of SALL4, cells progress normally through the neuroectodermal stage but fail to specify into CNCCs, as the lineage-specific enhancers remain repressed. Finally, using *in silico* approaches, we predict that SALL4A isoform directly interacts with the BAF subunit DPF2. Specifically, this interaction occurs between the zinc-finger-cluster-3 (ZFC3) domain of SALL4A and the plant homeodomains (PHDs) of DPF2.

RESULTS

SALL4 interacts with the BAF complex

Cranial neural crest cells can be generate *in vitro* starting from human pluripotent stem cells (hiPSCs). This protocol is approximately two weeks long, and it first goes through a

101 neuroectoderm intermediate [27–31]. In our previous work[27], mass spectrometry (IP-MS) conducted
102 at early stages of the iPSC-to-CNCC differentiation protocol, and specifically Day-5, which
103 corresponds to the neuroectodermal stage, revealed that SALL4 interacts with the BAF chromatin
104 remodeling complex in this developmental cell population. Since this was the first report of a
105 SALL4-BAF interaction, we performed a co-immunoprecipitation (co-IP) at Day-5 to validate it.
106 Specifically, we performed pull-down of ARID1B, which is the only ARID1 subunit active and
107 incorporated in BAF at this stage [27], and blotted for SALL4. This co-IP confirmed that ARID1B
108 interacts with both SALL4 isoforms at this time-point (Fig. 1A).
109 The BAF chromatin remodeling complex is a multi-subunit assembly composed of a dozen
110 components, which are differentially incorporated in a tissue-specific manner[32]. Based on which
111 subunit is present in each complex, in mammals these can be classified into canonical BAF,
112 polybromo-associated BAF, and non-canonical BAF[33]. While our mass-spec and co-IP
113 experiments support the interaction between SALL4 and the canonical BAF (ARID1B is not
114 incorporated in the other BAF configurations), neither of these two approaches has the resolution
115 to identify the exact BAF subunit that mediates the interaction with SALL4. Therefore, we
116 performed an *in silico* analysis using AlphaFold3[34]. We tested various BAF subunit combinations
117 along with four histones and an AT-rich DNA sequence which has been empirically shown to be
118 recognized by SALL4 (5'-GGAAATATTACC-3', PBD: 8A4I). The output was statistically
119 significant, with a 0.45 predicted template modeling (pTM) score and 0.47 interface predicted
120 template modeling (ipTM) score. Through this approach, AlphaFold3 predicted a direct interaction
121 between DPF2, a broadly expressed BAF subunit, and the zinc-fingers 5 and 6 (ZFC3) and 7
122 (ZFC4) of SALL4 (Figs. 1B, 1C, 1D; Supplementary Table 1). *DPF2* is a member of the *BAF45*
123 gene family, which also includes *PHF10*, *DPF1*, and *DPF3*. We were able to rule out its
124 paralogues based on their distinct expression patterns, incorporation into specific BAF
125 complexes, and previous reports that BAF45 subunits are not functionally redundant[35]. In
126 particular, PHF10 is exclusively associated with the pBAF complex, which does not incorporate
127 ARID1B[36], while DPF1 and DPF3 are specifically incorporated into the neuron-specific nBAF
128 complex[37] and are expressed at very low levels in both our iPSCs and neuroectoderm-like cells
129 at Day-5 (Fig. 1E).
130 This finding was particularly interesting given that DPF2 deletion has been linked to early neural
131 ectoderm differentiation in embryonic stem cells (ESCs)[35]. DPF2 exists in two isoforms, resulting
132 from alternative splicing of exon 7 and are spliced in a tissue-specific and developmental-stage-
133 specific manner[38]. DPF2-L is not alternatively spliced and is expressed in ESCs, while DPF2-S is
134 expressed in neurons. Interestingly, both isoforms are expressed in neural stem cells. Notably, of

135 the two main SALL4 isoforms (SALL4A and SALL4B), SALL4B lacks the portion of exon 2 that
136 contains ZFC3. This prompted us to explore whether the SALL4-DPF2 interaction could be
137 isoform-specific. To address this, we reanalysed the interactions in AlphaFold3, incorporating
138 SALL4B to the same BAF subunits as before. To ensure biological relevance, we first excluded
139 two of the five models that did not exhibit interactions with the AT-rich DNA sequence, as SALL4
140 is expected to bind DNA for its function. Among the remaining three models, we observed that
141 SALL4B interacted with DPF2, but unlike the SALL4A isoform, its interaction was not domain
142 specific. Instead, the predicted interactions for SALL4B were characterized by annotated
143 disordered regions and appeared to be largely driven by a non-specific, zinc-finger-to-zinc-finger
144 association (Fig. 1C; Supplementary Table 2), with no particular domain-specific interactions
145 identified. This pattern may be the result of the tendency of AlphaFold3 learning algorithm to
146 associate similar domains, such as zinc fingers, based on their predicted functional specificity
147 rather than spatial conformations. These findings suggest that, during neural ectoderm
148 specification, SALL4 may associate with BAF in an isoform-dependent manner.
149 To accurately interpret our AlphaFold3 predictions, we examined two key confidence metrics: the
150 Predicted Aligned Error (PAE), which assesses the reliability of the relative positioning of residues
151 within the structure, and the Predicted Local Distance Difference Test (pLDDT), which provides a
152 per-residue confidence score for the model. Our SALL4-DPF2 contact analysis revealed 394
153 atomic interactions (center-to-center distance ≤ 8Å) involving 139 residues (Fig. 1B, right;
154 Supplementary Table 1). Encouragingly, the pLDDT values were highest in known functional
155 domains involved in the interaction, particularly ZF5, ZF6, and ZF7 in SALL4A, as well as the
156 PHD1 and PHD2 domains in DPF2, reinforcing the reliability of these predicted structures. To
157 refine our analysis, we focused on residues where at least half of their atomic interactions had a
158 pLDDT > 65 (70 ± 10%) and that were annotated as part of a functional domain, ensuring that
159 only higher-confidence and potentially functional interactions were considered. As a result, ZF7
160 was excluded from further analysis (Supplementary Table 1). In summary, our experiments and
161 analyses confirm that SALL4 interacts with BAF during early stages of human CNCC
162 specification, and predict that this interaction is mediated by the BAF subunit DPF2.
163

164 **Generation of a *SALL4*-haploinsufficient human iPSC line**

165 Our experiments so far have revealed that SALL4 interacts with the BAF complex at early stages
166 of CNCC specification, and specifically at the neuroectodermal stage (Day-5 of our iPSC-to-
167 CNCC differentiation protocol). To investigate the scope of the SALL4-BAF interaction during this
168 developmental process, we generated a *SALL4* haploinsufficient iPSC line (hereafter *SALL4*-het-

169 KO; Fig. 2A). We opted for a heterozygous KO of *SALL4* as this recapitulates the human *SALL4*
170 syndromes, which are all caused by *SALL4* haploinsufficient genetic variants[22,39,40]. The CRISPR-
171 KO was generated by introducing a 11 bp deletion within the 3'-terminus of exon 1 of *SALL4*. This
172 is a conserved region across both *SALL4* isoforms that lacks functionally significant domains.
173 SALL4 immunoblot confirmed that SALL4 protein expression was significantly reduced in the
174 *SALL4*-het-KO iPSCs compared to the unedited wild-type line (*SALL4*-WT; Fig. 2B). We noted
175 that SALL4 expression in the *SALL4*-het-KO line was significantly less than the expected 50% of
176 the wild-type, suggesting potential negative feedback mechanisms triggered by the reduced
177 SALL4 levels, possibly due to the preferentially homodimeric nature of this protein.
178 Next, we investigated whether other SALL paralogs (*SALL1*, *SALL2*, *SALL3*) could compensate
179 for the loss of SALL4. Thus, we investigated their expression at gene level (Fig. 2C) and found
180 that, while *SALL2* and *SALL3* were not differentially expressed at the RNA level, *SALL1* was
181 instead significantly upregulated in the *SALL4*-het-KO cell line, at both RNA and protein level
182 (Figs. 2C and 2D), indicating a potential compensation in pluripotent cells.
183 Importantly, *SALL4*-het-KO iPSCs remain viable and exhibit hallmark iPSC morphology, including
184 a high nucleus-to-cytoplasm ratio and growth in compact colonies with well-defined borders (Fig.
185 2E). Furthermore, these cells show no significant difference in growth rate compared to their
186 isogenic controls (Fig. 2F).
187
188 **Decreased SALL4 expression does not affect the stem cell identity of the iPSCs**
189 Given the debated role of SALL4 in maintaining pluripotency, we investigated whether *SALL4*-
190 het-KO impacted iPSC stem cell identity. Thus, we examined the protein levels of the canonical
191 pluripotency transcription factors NANOG and OCT4 and found that they remained unaltered in
192 the *SALL4*-het-KO cells relative to the *SALL4*-WT control (Fig. 3A). Similarly, the levels of the
193 pluripotency surface markers TRA-1-60-R and SSEA4 were also not significantly affected by
194 decreased SALL4 expression (Fig. 3B). Moreover, we examined the transcriptome of the iPSCs
195 using bulk RNA-seq and identified only 114 differentially expressed genes in *SALL4*-het-KO cells
196 relative to the *SALL4*-WT control (FDR < 0.05; Fig. 3C). Gene Ontology analysis revealed that
197 these genes were not enriched for pluripotency (Fig. 3D and 3E). Finally, by performing tri-lineage
198 differentiation we confirmed that decreased SALL4 expression levels did not impact the ability of
199 the *SALL4*-het-KO cells to successfully differentiate into all three of the germ layers
200 (Supplementary Fig. 1A).
201 Together, these findings suggest that *SALL4* haploinsufficiency does not affect the stem cell
202 identity and self-renewal of human iPSCs, possibly as a result of SALL1-mediated compensation.

203

204 **SALL4 relocates to CNCC developmental enhancers upon CNCC induction cues**

205 After establishing that *SALL4*-het-KO did not affect iPSC stem cell identity, we differentiated the
206 iPSCs into CNCCs using the same aforementioned protocol[27–31] (Fig. 4A). Since the SALL4-BAF
207 interaction was detected at Day-5 of this protocol[27] (neuroectodermal stage), we profiled SALL4
208 binding in *SALL4*-WT cells by performing CUT&RUN at the iPSC stage and at Day-5 of
209 differentiation. We performed two SALL4 CUT&RUN replicates per timepoint, and only retained
210 replicated peaks in each timepoint (FDR < 0.05). With this approach, we identified comparable
211 amounts of SALL4-bound genomic sites at Day-0 and Day-5 (29,088 and 32,707, replicated
212 peaks, respectively, FDR < 0.05). Further analysis of these sites indicated that approximately
213 20% of the Day-0 SALL4 peaks were unique to the iPSC stage and not detected at Day-5
214 (hereafter SALL4 Day-0 specific peaks; n=6,114; Fig. 4B, left). On the other hand, ~50% of the
215 peaks detected at Day-5 were exclusive of this developmental stage and were not detected at
216 Day-0 (hereafter SALL4 Day-5-specific peaks, n= 17,137; Fig. 4B, right). Finally, a subset of
217 ~15,000 sites remained bound by SALL4 throughout exit from pluripotency to neural ectoderm
218 cell fate commitment (i.e. peaks conserved and replicated between Day-0 and Day-5). Together,
219 this data indicate that, upon differentiation cues, there is a major shift in SALL4 binding, resulting
220 in the relocation of this transcription factor at ~17,000 previously unbound regions.

221 To understand the scope of SALL4's relocation, we analysed the positional distribution of the
222 SALL4 peaks by associating each peak with the nearest transcription start site (TSS). Consistent
223 with SALL4's role as an enhancer-binding transcription factor, 86% of SALL4-bound regions at
224 Day-0 were located >1Kb from the nearest TSS (n= 3,729), indicating they could be putative
225 enhancers. This proportion increased even further to 97% (n= 6,860) at Day-5, further
226 emphasizing the preferential binding of SALL4 at enhancer regions (Fig. 4C). To investigate the
227 scope of the newly acquired SALL4 Day-5 peaks, we first looked at the nearest genes. GO terms
228 enriched in the "Day-5 specific" genes included pathways associated with neural crest
229 development, such as neural tube development, face development, cardiac chamber
230 development, and cell-cell adhesion mediated by cadherin (Fig. 4D; FDR < 0.05, clustered using
231 affinity propagation redundancy removal parameter).

232 To further characterise the SALL4 bound enhancers and promoters at Day-0 and Day-5, we
233 performed sequence-based computational motif analysis. At Day-0, motifs associated with core
234 pluripotency transcription factors such as OCT4, NANOG, and SOX2 did not rank among the top
235 20 enriched sequences, further supporting the notion that SALL4 may not be essential for
236 maintaining pluripotency in human cells (Supplementary Figure 2A). Moreover, Day-0 specific

237 SALL4 peaks were enriched for the enhancer box (E-box) motif (Supplementary Figure 2A). The
238 E-box is usually bound by homo- or heterodimers of transcription factors containing a basic Helix-
239 Loop-Helix (bHLH) or a Zn-finger domain to regulate the expression of genes associated to cell
240 lineage commitment, especially in the context of skeletal muscle, cardiovascular and neuronal
241 tissue[41]. On the other hand, in addition to general enhancer activator factors (AP-1, TEAD), the
242 Day-5 specific SALL4 peaks were enriched for many CNCC signature factors, including SOX9,
243 TFAP2A, TWIST1/2, SNAI1, NR2F1/2, TCF4/12 and several others (Fig. 4E). Moreover,
244 neuroectodermal transcription factors, such as ZIC2/3 were also enriched. In summary, this
245 analysis suggests that, as early as the neuroectodermal stage, SALL4 relocates to enhancers of
246 genes associated with craniofacial development.
247
248 **Chromatin accessibility at the CNCC enhancers is SALL4 dependent**
249 Given that we demonstrated that SALL4 interacts with BAF at early stages of CNCC specification,
250 we reasoned that SALL4 binding at the CNCC developmental enhancers could be necessary for
251 their chromatin activation. Thus, we profiled chromatin accessibility by performing ATAC-seq at
252 Day-0 and Day-5 in both *SALL4*-WT and *SALL4*-het-KO conditions (two replicates per timepoint
253 per condition, only peaks present in both replicates were considered; FDR < 0.05). The number
254 of ATAC-seq replicated peaks identified in *SALL4*-WT condition was comparable between Day-0
255 (61,660 peaks; FDR < 0.05) and Day-5 (58,375 peaks; FDR < 0.05). Next, we specifically
256 investigated the chromatin state of the 17,137 genomic sites that gain SALL4 binding at Day-5
257 and found that SALL4 binding at these sites correlates with a marked increase in chromatin
258 accessibility compared to Day-0 (Fig. 4F). Strikingly, the accessibility at these regions was
259 significantly attenuated in *SALL4*-het-KO relative to *SALL4*-WT (Fig. 4G). Overall, these findings
260 suggest that, in response to differentiation signals, SALL4 relocates to CNCC developmental
261 enhancers, which subsequently become activated through chromatin opening, which is SALL4
262 dependent.
263
264 **SALL4 recruits the BAF complex at CNCC developmental enhancers to increase chromatin**
265 **accessibility**
266 Since SALL4 binding at the CNCC enhancers is necessary for chromatin opening, we sought to
267 determine whether the SALL4-bound developmental enhancers were also occupied by BAF, and
268 if BAF binding at these sites could be SALL4 dependent. We performed ChIP-seq for BRG1
269 (SMARCA4), the only catalytic subunit of BAF active at this time point, in *SALL4*-WT cells at Day-
270 5 of the iPSC-to-CNCC differentiation protocol. This experiment revealed that BRG1 consistently

271 binds the same genomic sites as SALL4, supporting the idea that these factors interact and co-
272 occupy regulatory elements during early differentiation (Fig. 5A).
273 To assess whether BRG1 binding depends on SALL4, we performed ChIP-seq for BRG1 in both
274 *SALL4*-WT and *SALL4*-het-KO cells at Day-5. In *SALL4*-WT cells, we detected 32,833 replicated
275 BRG1 peaks (FDR < 0.05), as opposed to only 14,116 peaks in *SALL4*-het-KO cells. This
276 suggested that SALL4 loss leads to a widespread loss of BAF from chromatin. We thus focused
277 on the BRG1 peaks that were lost in *SALL4*-het-KO cells and found that these sites were largely
278 not bound by SALL4 at Day-0, but they gained SALL4 binding at Day-5 (Figs. 5B and 5C). As
279 expected, loss of SALL4 and BRG1 from these sites led to marked decrease of chromatin
280 accessibility (Fig. 5D). Overall, these experiments support that SALL4 is required to recruit the
281 BAF complex at CNCC enhancers. Consistent with this, genes near BRG1 peaks lost in the
282 *SALL4*-het-KO cells are involved in embryonic organ development, epithelial differentiation, and
283 cell migration processes (Fig. 5E).
284 While a large portion of BRG1 peaks was lost in the absence of SALL4, we also identified a subset
285 of novel BRG1 peaks that were absent in *SALL4*-WT cells but gained in the *SALL4*-het-KO cells.
286 These BRG1 peaks gained in the *SALL4*-het-KO cells were associated with genes linked to
287 neurogenesis and nervous system development (Fig. 5F). These findings indicate that, in the
288 absence of SALL4, BRG1 is relocated from CNCC enhancers to neurodevelopmental enhancers,
289 shifting chromatin remodeling activity away from CNCC specification.
290 Together, these results suggest that SALL4 is a key regulator of BAF complex recruitment to
291 CNCC-specific enhancers, ensuring their accessibility during differentiation. In the absence of
292 SALL4, BRG1 fails to localize to these sites and is instead partially redirected to
293 neurodevelopmental enhancers, leading to a misallocation of chromatin remodeling activity,
294 potentially leading to a shift towards neural differentiation at the expense of CNCC induction and
295 specification.
296
297 ***SALL4*-het-KO cells are not able to exit the neural ectoderm stage and to induce the cranial**
298 **neural crest fate**
299 Our data thus far indicate that SALL4 relocates to CNCC enhancers as early as the
300 neuroectodermal stage, recruiting BAF to these regulatory elements for chromatin activation.
301 Next, we investigated the functional consequences of SALL4 loss by generating CNCCs using
302 *SALL4*-WT and *SALL4*-het-KO cells. As expected, the *SALL4*-WT CNCCs expressed the
303 signature CNCC markers at both the RNA (Fig. 6A) and protein (Fig. 6B) level, while *SALL4*-het-
304 KO showed a significant downregulation of these genes (Fig. 6A). This was further supported by

305 RNA-seq conducted at Day-14, which identified 1,106 differentially expressed genes in *SALL4*-
306 het-KO CNCCs relative to *SALL4*-WT (FDR < 0.05; Fig. 6C). This is in sharp contrast with the
307 small number of genes differentially expressed at the iPSC stage (n= 114; Fig. 3C) and at Day-5
308 (n= 170; Supplementary Fig. 2B).
309 Of the 1,106 genes differentially expressed at Day-14, ~60% (n= 659) were downregulated while
310 the remaining were upregulated (Fig. 6C). Remarkably, canonical CNCC markers such as
311 TFAP2A, NR2F1, and SOX9 were significantly downregulated in *SALL4*-het-KO cells.
312 Additionally, genes associated with neural plate border, early neural crest induction and epithelial-
313 to-mesenchymal transition (EMT) were also downregulated (*MSX1*, *GBX2*, *SNAI2*, *SOX5*,
314 *ZEB1/2*; Fig. 6C). In line with these findings, gene ontology analysis revealed that pathways
315 associated with face development, autonomic nervous system, gastrulation, and regulation of
316 morphogenesis were all downregulated (Fig. 6D). The 447 upregulated genes were instead
317 involved in regulatory pathways, especially vesicle-mediated transport of inorganic substances.
318 As aforementioned, genes associated with EMT were also dysregulated. The process of EMT
319 requires a switch in cadherin expression where E-cadherin (*CDH1*) must be downregulated, while
320 N-cadherin (*CDH2*) must be upregulated. Remarkably, our RNA-seq data revealed that this
321 cadherin switch did not occur in the differentiating *SALL4*-het-KO iPSCs, suggesting that the cells
322 remained in an epithelial state. This was further corroborated by the upregulation of several
323 epithelial genes in the differentiating *SALL4*-het-KO cells, including *EpCAM*, *OTX2*, *CDH1*, and
324 *OVOL2* (Fig. 6E). The latter is involved in maintaining epithelial lineages by repressing the
325 expression of the EMT gene *ZEB1*, which was downregulated in the *SALL4*-het-KO condition[42,43].
326 In summary, our data indicate that CNCC specification failed in *SALL4*-het-KO conditions, as
327 reflected by the large number of differentially expressed genes. On the other hand, the
328 transcriptome of the iPSCs and of the neuroectodermal cells (i.e. Day-5) were only modestly
329 affected by SALL4 loss, suggesting that SALL4 may not be essential for neuroectoderm
330 formation, while it is essential for CNCC induction and specification.
331 To investigate this further, we systematically examined the expression of a panel of stage-specific
332 marker genes at Day-0, Day-5 and Day-14 (Fig. 6F) including core pluripotency genes (*POU5F1*
333 and *NANOG*), neural plate border specifiers (*ZIC1, GBX2, MSX1/2, PAX3/7, DLX5/6, GATA2/3*),
334 CNCC specification markers (*SNAI1/2, SOX9, TFAP2A*), and post-migratory CNCC markers
335 (*SOX5, SOX6, SOX10*). In *SALL4*-WT cells, pluripotency markers exhibited a progressive and
336 expected downregulation as differentiation advanced from Day-0 to Day-14. In *SALL4*-het-KO
337 cells, a similar decline in pluripotency markers was observed between Day-0 and Day-5, but
338 instead of continuing to decrease, these markers remained similarly expressed through Day-14

339 (Fig. 6F). Despite this, the two conditions showed similar trends, confirming that decreased levels
340 of SALL4 dId not significantly impact pluripotency markers expression during exit from
341 pluripotency. Similarly, no significant difference was observed in the expression of neural plate
342 border specifiers (Fig. 6G), suggesting that SALL4 is not required for neuroectoderm specification.
343 Conversely, CNCC specification markers and post-migratory CNCC markers all increased in
344 expression after Day-5 in *SALL4*-WT cells, while the SALL4-depleted cells failed to maintain a
345 comparable increase in expression (Fig. 6H and 6I).
346 In summary, *SALL4*-het-KO cells are competent to generate neuroectoderm-like cells (Day-5) but
347 fail to induce the neural crest lineage (Day-14).
348
349 **DISCUSSION**
350 During embryonic development, SALL4 plays a critical role in forming major anatomical
351 structures, including the heart, limbs, cerebral midline, and pituitary gland[44]. Correspondingly,
352 mutations in the haploinsufficient gene *SALL4* are associated with congenital syndromes
353 characterized by malformations in these structures. These syndromes also frequently include
354 distinct craniofacial anomalies, such as epicanthal folds, hypertelorism, a depressed nasal bridge,
355 and hemifacial microsomia[45–47], all of which are hallmark features of disrupted cranial neural crest
356 cell development.
357 Notably, at the murine embryonic stage E9.5, Sall4 expression persists within the frontonasal
358 prominence, which is the most anterior embryonic region[18], implying a potential role for SALL4 in
359 craniofacial development. Given the limited existing knowledge about SALL4 function in this
360 context, we utilized a heterozygous *SALL4* knockout human induced pluripotent stem cell (iPSC)
361 line to investigate its role during cranial neural crest cell (CNCC) specification. Although SALL4
362 expression begins as early as the 2-cell embryonic stage, its role in maintaining pluripotency
363 remains debated. Some studies indicate that SALL4 cooperates closely with key pluripotency
364 factors such as OCT4 and NANOG to sustain stem cell identity[13,14]. Others, however, report that
365 SALL4 has only limited involvement within pluripotency regulatory networks[15–17]. Our findings
366 align with the latter hypothesis, demonstrating that decreased expression of SALL4 does not
367 compromise iPSC identity, possibly due to compensatory mechanisms involving its paralogue,
368 SALL1.
369 Using CUT&RUN, we observed that at early stages of CNCC induction, SALL4 relocates to
370 enhancers of genes involved in CNCC formation, and that SALL4 binding at these regulatory
371 elements is required for their activation. In humans, SALL4 exists in two major isoforms generated
372 by alternative splicing: SALL4A (1,053 amino acids), which retains all eight zinc-finger motifs, and

373 SALL4B (617 amino acids), which contains a partially spliced exon 2. Rao et al.[12] proposed distinct
374 functional roles for these isoforms in murine ESCs, demonstrating that SALL4A homodimers
375 localize predominantly to enhancers of developmental genes, whereas pluripotency-associated
376 enhancers are occupied either by heterodimers comprising both isoforms or by SALL4B alone. In
377 mouse ESCs, approximately 10% of the total Sall4 protein, and primarily the Sall4A isoform, is
378 incorporated into the NuRD complex [15,24]. Here, we identify and characterize a novel interaction
379 between SALL4 and the BAF complex, predicted to be mediated specifically through the SALL4A
380 isoform and the BAF subunit DPF2. Furthermore, we show that upon receiving CNCC induction
381 cues, SALL4 and BAF co-localize to enhancers of genes involved in CNCC specification to prime
382 and activate them, with BAF recruitment being SALL4-dependent.
383 Interestingly, DPF2 and SALL4 exhibit notable functional parallels. Similar to SALL4, DPF2
384 occupancy undergoes significant repositioning during the transition from pluripotency toward early
385 neural commitment in mouse embryoid bodies [48]. Although the role of DPF2 in CNCCs has not
386 been directly investigated, its early recruitment to neuronal enhancers suggests a potential
387 involvement in neuroectodermal differentiation and the subsequent lineage commitment of
388 neuroectoderm-derived cells, including CNCCs. Moreover, genomic sequences bound by DPF2
389 in mouse neuroectodermal cells are AT-rich, resembling the preferred binding motif of
390 SALL4[38,49,50].
391 Our study demonstrated that while cranial neural crest cell induction and specification fail in the
392 absence of SALL4, early neuroectodermal commitment remains unaffected (Fig. 7). This finding
393 is further supported by clinical observations. In fact, phenotypes arising from SALL4
394 haploinsufficiency affect many neural crest derivatives, including craniofacial structures and
395 peripheral neurons. In contrast, phenotypes implicating defects in neural ectoderm or
396 neurogenesis are exceptionally rare[51,52] and usually associated with dysregulation of potassium
397 channels[44]. In the rare cases of intellectual disability reported, such phenotypes were attributed
398 to broader deletions on chromosome 20q13.13-q13.2 rather than *SALL4* loss alone[51].
399 SALL4 was recently identified as a direct target of thalidomide[53], a drug widely prescribed to
400 pregnant women in the late 1950s to alleviate morning sickness. The subsequent alarming rise in
401 births of children with congenital anomalies, including limb malformations (phocomelia), facial
402 abnormalities such as inner ear defects, ocular motility issues, and facial nerve palsy, as well as
403 kidney and heart defects, led to the identification of thalidomide as the causative agent. However,
404 the molecular mechanism by which thalidomide induced these developmental defects remained
405 elusive until nearly six decades after the drug's withdrawal from the market. Building on earlier
406 findings that thalidomide and its derivatives bind to cereblon (CRBN), Matyskiela et al.

407 demonstrated that SALL4 is recruited to the CRL4–CRBN E3 ubiquitin ligase complex following
408 thalidomide-induced ubiquitination, resulting in its proteasomal degradation[54]. More recently,
409 research demonstrated that only the SALL4A isoform, but not SALL4B, is susceptible to
410 degradation by immunomodulatory imide drugs such as thalidomide and its derivative
411 pomalidomide[55]. In this context, our findings might contribute to shed light on the birth
412 malformations caused by thalidomide in the 1960s, potentially revealing some of the underlying
413 mechanisms.
414
415 **RESOURCE AVAILABILITY**
416 ***Data availability***
417 RNA-seq, ATAC-seq, CUT&RUN-seq, and ChIP-seq data have been deposited in the Gene
418 Expression Omnibus (GEO) under accession code GEO: GSE293821 and are publicly available
419 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE293821.
420
421 ***Lead contact***
422 Further information and requests for resources and reagents should be directed to and will be
423 fulfilled by the lead contact, Marco Trizzino (m.trizzino@imperial.ac.uk).
424
425 **ACKNOWLEDGMENTS**
426 We thank the Facility for Imaging by Light Microscopy (FILM) at Imperial College London for
427 providing access to instrumentation and technical support, which is part-supported by funding
428 from the Wellcome Trust (grant 104931/Z/14/Z) and BBSRC (grant BB/L015129/1). We also thank
429 the Genomics Facility at The Wistar Institute for the Next Generation Sequencing support. For
430 this work, MT was funded by the Mathers Foundation. We thank Prof. Brian Hendrich (University
431 of Cambridge) and Dr. Andria Koulle (Imperial College London) for insightful discussion of the
432 data and of early versions of the manuscript.
433
434 **AUTHOR CONTRIBUTIONS**
435 MT designed the project. MD performed most of the experiments. SMB, EVD, ZHM, and LD
436 contributed to some of the experiments. MD and MT analyzed the data and wrote the manuscript.
437 All the authors read and approved the manuscript.
438
439 **DECLARATION OF INTERESTS**
440 Authors declare no competing interests.

441

442

443 **SUPPLEMENTAL INFORMATION**

444 **Document S1. Supplemental figures**

445 **Table S1. List of AlphaFold predicted contacts between SALL4A and DPF2.**

446 **Table S2. List of AlphaFold predicted contacts between SALL4B and DPF2.**

447

448

449 **FIGURE TITLES AND LEGENDS**

450 **Figure 1. SALL4A interacts with the BAF complex through DPF2.**

451 (A) Immunoblot of nuclear protein (5% input) and co-immunoprecipitations (co-IP) of ARID1B and
452 IgG (negative control) from *SALL4*-WT cells at Day-5.

453 (B) (Left) AlphaFold structure model showing the predicted interactions between SALL4A (red),
454 DPF2 (dark blue), and other BAF components, including SMARCA4, SMARCB1, and ACTL6A.
455 (Center) Visual representation of the confidence score (predicted Local Distance Difference Test
456 - pLDDT) for each residue of the AlphaFold-predicted complex, with color gradients indicating
457 prediction confidence (blue: very high confidence, yellow: low confidence, orange/red: very low
458 confidence). (Right) Close-up view of the SALL4A-DPF2 interface highlighting key interacting
459 residues in red.

460 (C) SALL4A and SALL4B protein sequences. C2H2 zinc fingers annotations are highlighted in
461 dark purple and numbered. Visual representation of output models of SALL4A (model 0) and
462 SALL4B (models 1, 3, and 4). Predicted interactions between SALL4 isoforms and DPF2 are
463 highlighted in different colors in each model. SALL4A is predicted to interact with DPF2 through
464 the ZFs 4, 5 and 6 (ZFC3) and 7 (ZFC4), light purple. Made with BioRender.com.

465 (D) Zoomed view of SALL4 (pink) and DPF2 (purple) interacting surfaces. Highlighted are the
466 functional domains involved in the predicted interaction (ZF6 and ZF7 of SALL4 and PHD1 and
467 PHD2 of DPF2). In red, contacts predicted by AlphaFold[34].

468 (E) Violin plots displaying expression (log2[TPM]) of the four members of the DPF protein family
469 (*PHF10*, *DPF1*, and *DPF3*) in iPSCs (left), neuroectodermal-like cells (Day-5 – center), and in
470 CNCCs (Day-14 – right). Differences between conditions were assessed using unpaired student's
471 t-test.

472

473 **Figure 2. Generation of a *SALL4*-haploinsufficient human iPSC line**

474 (A) Visual representation of the CRISPR/Cas9 approach used on the *SALL4* gene model. The
475 star represents the exact location of the deletion generated by Cas9. Sanger sequencing
476 confirming the 19 bp deletion in one allele of *SALL4* is displayed. Made with BioRender.com.

477 (B) Immunoblot for SALL4 in *SALL4*-WT and *SALL4*-het-KO iPSC lines. GAPDH was used as a
478 loading control.

479 (C) Violin plots displaying expression (log2[TPM]) of the *SALL4* paralogues (*SALL1*, *SALL2*,
480 *SALL3*) in *SALL4*-WT and *SALL4*-het-KO iPSCs. Statistical significance was assessed using
481 unpaired student's t-test.

482 (D) Immunoblot for SALL1 in *SALL4*-WT and *SALL4*-het-KO iPSC lines. GAPDH was used as a
483 loading control.
484 (E) Growth curve comparing *SALL4*-WT with *SALL4*-het-KO iPSC lines. N = 2 biologically
485 independent experiments. Error bars display standard error of the mean.
486 (F) Morphology of *SALL4*-WT and *SALL4*-het-KO iPSC colonies at 4x (top) and 10x (bottom).
487
488 **Figure 3. Stem cell identity is unaffected by SALL4 loss**
489 (A) (Left) Immunofluorescence labeling quantifying the expression of the canonical pluripotency
490 markers OCT4 and NANOG in the two iPSC lines. Scale bar: 50 μm. (Right) Corrected Total Cell
491 Fluorescence (CTCF) values were used to quantify and compare markers expression.
492 (B) Flow cytometry quantifying the expression of surface markers for pluripotency (TRA-60-1-R
493 and SSEA-4).
494 (C) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in
495 iPSCs (Day-0). Blue dots (n= 50) represent downregulated genes with p-adj < 0.05 and
496 log2FoldChange < -1.5. Purple dots (n= 63) represent upregulated genes with p-adj < 0.05 and
497 log2FoldChange > 1.5.
498 (D) Top 10 GO term enrichment for downregulated genes in *SALL4*-het-KO iPSCs determined
499 using WebGestalt[56] over-representation pathway analysis, using affinity propagation as
500 parameter for redundancy removal. In bold terms with FDR < 0.05).
501 (E) Top 10 GO term enrichment for upregulated genes in *SALL4*-het-KO iPSCs determined using
502 WebGestalt[56] over-representation pathway analysis, using affinity propagation as parameter for
503 redundancy removal. All terms are non-significant (FDR > 0.05).
504
505 **Figure 4. SALL4 relocates to CNCC enhancers to induce chromatin accessibility**
506 (A) Schematic representation of the iPSC-to-CNCC differentiation protocol. The experimental
507 timeline shows in vitro differentiation of iPSCs into cranial neural crest cells (CNCCs) via a
508 neuroectodermal-like stage. Cells were maintained in pluripotency self-renewing conditions until
509 Day-0, transitioned into differentiation medium starting at Day-0 and further to maintenance
510 medium at ~ Day-7/8. Key analyses were performed at Day-0, Day-5, and Day-14.
511 (B) Heatmaps showing the distribution of SALL4 binding at Day-0 and Day-5 across Day-0
512 specific (left) and Day-5 specific (right) binding regions.
513 (C) Genomic distribution of SALL4 binding sites relative to the nearest transcription start site
514 (TSS). Peaks are categorized by distance into six bins: <1 kb, 1–3 kb, 3–5 kb, 5–10 kb, 10–100

515 kb, and >100 kb. Peaks located at >1kb from the nearest TSS were classified as putative
516 enhancers.
517 (D) Gene ontology analysis of Day-5 specific SALL4-bound regions as determined using
518 WebGestalt[56] over-representation pathway analysis, using affinity propagation as parameter for
519 redundancy removal (FDR<0.05).
520 (E) Table of motifs enriched in Day-5 specific SALL4-bound regions as determined using
521 HOMER[57].
522 (F) Heatmaps showing ATAC-seq signals at Day-0 (left) and Day-5 (right), at Day-5 specific
523 SALL4 CUT&RUN peaks. Each row represents an individual peak. Signal intensity is color-coded,
524 with higher intensity indicating higher chromatin accessibility.
525 (G) Average profiles showing ATAC-seq signal at Day 5-specific SALL4 CUT&RUN peaks in
526 *SALL4*-WT and *SALL4*-het-KO cells in two distinct biological replicates.
527
528 **Figure 5. SALL4 recruits BAF to CNCC enhancers.**
529 (A) Heatmap and average profiles show BRG1 ChIP-seq signal at the SALL4 Day-5 specific
530 CUT&RUN peaks (n = 17,137).
531 (B) Heatmap and average profiles show SALL4 CUT&RUN signal at the genomic sites that lose
532 BRG1 in *SALL4*-het-KO cells at Day-0 and Day-5 (n = 13,505).
533 (C) Heatmap and average profiles show BRG1 ChIP-seq signal at the genomic sites that lose
534 BRG1 in *SALL4*-het-KO cells at day 5 (n = 13,505).
535 (D) ATAC-seq signal at the genomic sites that lose BRG1 in *SALL4*-het-KO cells (n = 13,505).
536 (E) and (F) GO term enrichment for genes near BRG1 ChIP-seq peaks lost (E) and gained (F) in
537 *SALL4*-het-KO cells. GO terms were determined using WebGestalt[56] over-representation
538 pathway analysis, using affinity propagation as parameter for redundancy removal (FDR<0.05).
539
540 **Figure 6. CNCC induction and specification are impaired in *SALL4*-het-KO.**
541 (A) Box plot displaying relative mRNA expression of canonical CNCC markers in *SALL4*-WT and
542 *SALL4*-het-KO cells at Day-14. Statistical significance was assessed using unpaired student's t-
543 test. 18S rRNA was used to normalize samples.
544 (B) Flow cytometry for the CNCC surface marker CD99 in *SALL4*-WT and *SALL4*-het-KO CNCCs
545 at Day-14.
546 (C) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in
547 CNCCs (Day-14). Blue dots (n= 659) represent downregulated genes with p-adj < 0.05 and

548 log2FoldChange < -1.5. Purple dots (n= 447) represent upregulated genes with p-adj < 0.05 and
549 log2FoldChange > 1.5.
550 (D) GO term enrichment for downregulated genes in *SALL4*-het-KO cells determined using
551 WebGestalt[56] over-representation pathway analysis, using affinity propagation as parameter for
552 redundancy removal (FDR < 0.05).
553 (E) Heatmap displaying the expression of canonical epithelial-to-mesenchymal transition (EMT)
554 markers in *SALL4*-WT and *SALL4*-het-KO.
555 (F), (G), (H), and (I). Trajectory of stage-specific markers at Day-0, Day-5, and Day-14. Displayed
556 are the medians of normalized expression (log2[TPMs]) for (F) pluripotency (*POU5F1* and
557 *NANOG*), (G) neural plate border (NPB; *ZIC1, GBX2, MSX1, MSX2, PAX3, PAX7, DLX5, DLX6,*
558 *GATA2, GATA3*), (H) CNCC specification (*SNAI1, SNAI2, SOX9, TFAP2A*), and (I) post-migratory
559 CNCC (*SOX5, SOX6, SOX10*) markers at each time-point. Dotted lines represent simple linear
560 regression.
561
562 **Figure 7. Working model.**
563 In wild-type conditions, SALL4 recruits BAF at neural crest induction enhancers. This activates
564 lineage specific gene expression, allowing for NCC induction and specification. In the absence of
565 SALL4, the NCC induction enhancers remain repressed because BAF is not loaded at chromatin,
566 and the NCC gene program is not launched, and NCC induction is aborted.
567
568

REFERENCES

1. Jia, Y. kang, Yu, Y., and Guan, L. (2024). Advances in understanding the regulation of pluripotency fate transition in embryonic stem cells. Preprint at Frontiers Media SA, https://doi.org/10.3389/fcell.2024.1494398 https://doi.org/10.3389/fcell.2024.1494398.

2. Fagan, M.B. (2017). Stem Cell Lineages: Between Cell and Organism. Philosophy, Theory, and Practice in Biology *9*. https://doi.org/10.3998/ptb.6959004.0009.006.

3. Zheng, Z.H., Sam, T.W., Zeng, Y., Chu, J.J.H., and Loh, Y.H. (2021). Chromatin Regulation in Development: Current Understanding and Approaches. Preprint at Hindawi Limited, https://doi.org/10.1155/2021/8817581 https://doi.org/10.1155/2021/8817581.

4. Gopinathan, G., and Diekwisch, T.G.H. (2022). Epigenetics and Early Development. Preprint at MDPI, https://doi.org/10.3390/jdb10020026 https://doi.org/10.3390/jdb10020026.

5. Moore, L.D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. Preprint, https://doi.org/10.1038/npp.2012.112 https://doi.org/10.1038/npp.2012.112.

6. Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Preprint, https://doi.org/10.1038/nrg887 https://doi.org/10.1038/nrg887.

7. Islam, Z., Ali, A.M., Naik, A., Eldaw, M., Decock, J., and Kolatkar, P.R. (2021). Transcription Factors: The Fulcrum Between Cell Development and Carcinogenesis. Preprint at Frontiers Media S.A., https://doi.org/10.3389/fonc.2021.681377 https://doi.org/10.3389/fonc.2021.681377.

8. Kim, S., and Shendure, J. (2019). Mechanisms of Interplay between Transcription Factors and the 3D Genome. Preprint at Cell Press, https://doi.org/10.1016/j.molcel.2019.08.010 https://doi.org/10.1016/j.molcel.2019.08.010.

9. Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Preprint, https://doi.org/10.1038/nrg3623 https://doi.org/10.1038/nrg3623.

10. Tatetsu, H., Kong, N.R., Chong, G., Amabile, G., Tenen, D.G., and Chai, L. (2016). SALL4, the missing link between stem cells, development and cancer. Preprint at Elsevier B.V., https://doi.org/10.1016/j.gene.2016.02.019 https://doi.org/10.1016/j.gene.2016.02.019.

11. Xiong, J., Todorova, D., Su, N.Y., Kim, J., Lee, P.J., Shen, Z., Briggs, S.P., and Xu, Y. (2015). Stemness factor Sall4 is required for DNA damage response in embryonic stem cells. Journal of Cell Biology *208*, 513–520. https://doi.org/10.1083/jcb.201408106.

12. Rao, S., Zhen, S., Roumiantsev, S., McDonald, L.T., Yuan, G.-C., and Orkin, S.H. (2010). Differential Roles of Sall4 Isoforms in Embryonic Stem Cell Pluripotency. Mol Cell Biol *30*, 5364–5380. https://doi.org/10.1128/mcb.00419-10.

13. Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L., et al. (2006). Sall4 modulates embryonic stem cell pluripotency and early

606 embryonic development by the transcriptional regulation of Pou5f1. Nat Cell Biol *8*, 1114–
607 1123. https://doi.org/10.1038/ncb1481.

608 14. Yang, J. (2018). SALL4 as a transcriptional and epigenetic regulator in normal and
609 leukemic hematopoiesis. Preprint at BioMed Central Ltd., https://doi.org/10.1186/s40364-
610 017-0115-6 https://doi.org/10.1186/s40364-017-0115-6.

611 15. Miller, A., Ralser, M., Kloet, S.L., Loos, R., Nishinakamura, R., Bertone, P., Vermeulen, M.,
612 and Hendrich, B. (2016). Sall4 controls differentiation of pluripotent cells independently of
613 the nucleosome remodelling and deacetylation (NuRD) complex. Development
614 (Cambridge) *143*, 3074–3084. https://doi.org/10.1242/dev.139113.

615 16. Wang, F., Zhao, W., Kong, N., Cui, W., and Chai, L. (2014). The next new target in
616 leukemia: The embryonic stem cell gene SALL4. Mol Cell Oncol *1*.
617 https://doi.org/10.4161/23723548.2014.969169.

618 17. Yuri, S., Fujimura, S., Nimura, K., Takeda, N., Toyooka, Y., Fujimura, Y.I., Aburatani, H.,
619 Ura, K., Koseki, H., Niwa, H., et al. (2009). Sall4 is essential for stabilization, but not for
620 pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene
621 expression. Stem Cells *27*, 796–805. https://doi.org/10.1002/stem.14.

622 18. Tahara, N., Kawakami, H., Chen, K.Q., Anderson, A., Peterson, M.Y., Gong, W., Shah, P.,
623 Hayashi, S., Nishinakamura, R., Nakagawa, Y., et al. (2019). Sall4 regulates
624 neuromesodermal progenitors and their descendants during body elongation in mouse
625 embryos. Development (Cambridge) *146*. https://doi.org/10.1242/DEV.177659.

626 19. Chen, K.Q., Anderson, A., Kawakami, H., Kim, J., Barrett, J., and Kawakami, Y. (2022).
627 Normal embryonic development and neonatal digit regeneration in mice overexpressing a
628 stem cell factor, Sall4. PLoS One *17*. https://doi.org/10.1371/journal.pone.0267273.

629 20. Pappas, M.P., Kawakami, H., Corcoran, D., Chen, K.Q., Scott, E.P., Wong, J., Gearhart,
630 M.D., Nishinakamura, R., Nakagawa, Y., and Kawakami, Y. (2024). Sall4 regulates
631 posterior trunk mesoderm development by promoting mesodermal gene expression and
632 repressing neural genes in the mesoderm. Development (Cambridge) *151*.
633 https://doi.org/10.1242/dev.202649.

634 21. Kekunnaya, R., and Negalur, M. (2017). Duane retraction syndrome: Causes, effects and
635 management strategies. Preprint at Dove Medical Press Ltd,
636 https://doi.org/10.2147/OPTH.S127481 https://doi.org/10.2147/OPTH.S127481.

637 22. Terhal, P., Rösler, B., and Kohlhase, J. (2006). A family with features overlapping Okihiro
638 syndrome, hemifacial microsomia and isolated Duane anomaly caused by a novel SALL4
639 mutation. Am J Med Genet *140 A*, 222–226. https://doi.org/10.1002/ajmg.a.31060.

640 23. Lu, J., Jeong, H., Kong, N., Yang, Y., Carroll, J., Luo, H.R., Silberstein, L.E., Ma, Y., and
641 Chai, L. (2009). Stem cell factor SALL4 represses the transcriptions of PTEN and SALL1
642 through an epigenetic repressor complex. PLoS One *4*.
643 https://doi.org/10.1371/journal.pone.0005577.

24. Bode, D., Yu, L., Tate, P., Pardo, M., and Choudhary, J. (2016). Characterization of two distinct nucleosome remodeling and deacetylase (NuRD) complex assemblies in embryonic stem cells. Molecular and Cellular Proteomics *15*, 878–891. https://doi.org/10.1074/mcp.M115.053207.

25. Liu, B.H., Jobichen, C., Chia, C.S.B., Chan, T.H.M., Tang, J.P., Chung, T.X.Y., Li, J., Poulsen, A., Hung, A.W., Koh-Stenta, X., et al. (2018). Targeting cancer addiction for SALL4 by shifting its transcriptome with a pharmacologic peptide. Proc Natl Acad Sci U S A *115*, E7119–E7128. https://doi.org/10.1073/pnas.1801253115.

26. Wang, B., Li, C., Ming, J., Wu, L., Fang, S., Huang, Y., Lin, L., Liu, H., Kuang, J., Zhao, C., et al. (2023). The NuRD complex cooperates with SALL4 to orchestrate reprogramming. Nat Commun *14*. https://doi.org/10.1038/s41467-023-38543-0.

27. Pagliaroli, L., Porazzi, P., Curtis, A.T., Scopa, C., Mikkers, H.M.M., Freund, C., Daxinger, L., Deliard, S., Welsh, S.A., Offley, S., et al. (2021). Inability to switch from ARID1A-BAF to ARID1B-BAF impairs exit from pluripotency and commitment towards neural crest formation in ARID1B-related neurodevelopmental disorders. Nat Commun *12*. https://doi.org/10.1038/s41467-021-26810-x.

28. Bajpai, R., Chen, D.A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.P., Zhao, Y., Swigut, T., and Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. Nature *463*, 958–962. https://doi.org/10.1038/nature08733.

29. Barnada, S.M., Giner de Gracia, A., Morenilla-Palao, C., López-Cascales, M.T., Scopa, C., Waltrich, F.J., Mikkers, H.M.M., Cicardi, M.E., Karlin, J., Trotti, D., et al. (2024). ARID1A-BAF coordinates ZIC2 genomic occupancy for epithelial-to-mesenchymal transition in cranial neural crest specification. The American Journal of Human Genetics. https://doi.org/10.1016/j.ajhg.2024.07.022.

30. Mitchell, Z.H., den Hoed, J., Claassen, W., Demurtas, M., Deelen, L., Campeau, P.M., Liu, K., Fisher, S.E., and Trizzino, M. (2025). CHD3 regulates BMP signalling response during cranial neural crest cell specification. Preprint, https://doi.org/10.1101/2025.02.14.638260 https://doi.org/10.1101/2025.02.14.638260.

31. Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. Cell *163*, 68–83. https://doi.org/10.1016/j.cell.2015.08.036.

32. Hodges, C., Kirkland, J.G., and Crabtree, G.R. (2016). The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer. Cold Spring Harb Perspect Med *6*. https://doi.org/10.1101/cshperspect.a026930.

33. Zhang, H., Wang, X., Li, J., Shi, R., and Ye, Y. (2021). BAF Complex in Embryonic Stem Cells and Early Embryonic Development. Preprint at Hindawi Limited, https://doi.org/10.1155/2021/6668866 https://doi.org/10.1155/2021/6668866.

683 34. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
684 Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate
685 protein structure prediction with AlphaFold. Nature *596*, 583–589.
686 https://doi.org/10.1038/s41586-021-03819-2.

687 35. Zhang, W., Chronis, C., Chen, X., Zhang, H., Spalinskas, R., Pardo, M., Chen, L., Wu, G.,
688 Zhu, Z., Yu, Y., et al. (2019). The BAF and PRC2 Complex Subunits Dpf2 and Eed
689 Antagonistically Converge on Tbx3 to Control ESC Differentiation. Cell Stem Cell *24*, 138-
690 152.e8. https://doi.org/10.1016/j.stem.2018.12.001.

691 36. Krasteva, V., Crabtree, G.R., and Lessard, J.A. (2017). The BAF45a/PHF10 subunit of
692 SWI/SNF-like chromatin remodeling complexes is essential for hematopoietic stem cell
693 maintenance. Exp Hematol *48*, 58-71.e15. https://doi.org/10.1016/j.exphem.2016.11.008.

694 37. Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H.,
695 Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An Essential Switch in Subunit
696 Composition of a Chromatin Remodeling Complex during Neural Development. Neuron *55*,
697 201–215. https://doi.org/10.1016/j.neuron.2007.06.019.

698 38. Nazim, M., Lin, C.H., Feng, A.C., Xiao, W., Yeom, K.H., Li, M., Daly, A.E., Tan, X., Vu, H.,
699 Ernst, J., et al. (2024). Alternative splicing of a chromatin modifier alters the transcriptional
700 regulatory programs of stem cell maintenance and neuronal differentiation. Cell Stem Cell
701 *31*, 754-771.e6. https://doi.org/10.1016/j.stem.2024.04.001.

702 39. Rgen Kohlhase, J., Heinrich, M., Schubert, L., Liebers, M., Kispert, A., Laccone, F.,
703 Turnpenny, P., Winter, R.M., and Reardon, W. (2002). Okihiro syndrome is caused by
704 SALL4 mutations. Oxford University Press *11*, 2979–2987.
705 https://doi.org/https://doi.org/10.1093/hmg/11.23.2979.

706 40. Al-Baradie, R., Yamada, K., St Hilaire, C., Chan, W.-M., Andrews, C., McIntosh, N.,
707 Nakano, M., Jean Martonyi, E., Raymond, W.R., Okumura, S., et al. (2002). Duane Radial
708 Ray Syndrome (Okihiro Syndrome) Maps to 20q13 and Results from Mutations in SALL4,
709 a New Member of the SAL Family.

710 41. Pan, Y., van der Watt, P.J., and Kay, S.A. (2023). E-box binding transcription factors in
711 cancer. Preprint at Frontiers Media SA, https://doi.org/10.3389/fonc.2023.1223208
712 https://doi.org/10.3389/fonc.2023.1223208.

713 42. Liu, J., Wu, Q., Wang, Y., Wei, Y., Wu, H., Duan, L., Zhang, Q., and Wu, Y. (2018). Ovol2
714 induces mesenchymal–epithelial transition via targeting ZEB1 in osteosarcoma. Onco
715 Targets Ther *11*, 2963–2973. https://doi.org/10.2147/OTT.S157119.

716 43. Haensel, D., Sun, P., MacLean, A.L., Ma, X., Zhou, Y., Stemmler, M.P., Brabletz, S., Berx,
717 G., Plikus, M. V, Nie, Q., et al. (2019). An Ovol2-Zeb1 transcriptional circuit regulates
718 epithelial directional migration and proliferation. EMBO Rep *20*.
719 https://doi.org/10.15252/embr.201846273.

720 44. Kodytková, A., Dušátková, P., Anne Amaratunga, S., Plachý, L., Pruhová, Š., and Lebl, J.
721 (2023). Integrative role of the sall4 gene: From thalidomide embryopathy to genetic defects

722 of the upper limb, internal organs, cerebral midline, and pituitary. Preprint at S. Karger AG,
723 https://doi.org/10.1159/000531452 https://doi.org/10.1159/000531452.

724 45. Miller, M.T. (1991). Thalidomide embryopathy: a model for the study of congenital
725 incomitant horizontal strabismus. Transaction of the American Ophtalmological Society,
726 623–674.

727 46. Kohlhase, J., Chitayat, D., Kotzot, D., Ceylaner, S., Froster, U.G., Fuchs, S., Montgomery,
728 T., and Rösler, B. (2005). SALL4 mutations in Okihiro syndrome (Duane-radial ray
729 syndrome), acro-renal-ocular syndrome, and related disorders. Preprint,
730 https://doi.org/10.1002/humu.20215 https://doi.org/10.1002/humu.20215.

731 47. Kohlhase, J., Chitayat, D., Kotzot, D., Ceylaner, S., Froster, U.G., Fuchs, S., Montgomery,
732 T., and Rösler, B. (2005). SALL4 mutations in Okihiro syndrome (Duane-radial ray
733 syndrome), acro-renal-ocular syndrome, and related disorders. Preprint,
734 https://doi.org/10.1002/humu.20215 https://doi.org/10.1002/humu.20215.

735 48. Zhang, Z., Zhan, X., Kim, B., and Wu, J. (2019). A proteomic approach identifies SAFB-
736 like transcription modulator (SLTM) as a bidirectional regulator of GLI family zinc finger
737 transcription factors. Journal of Biological Chemistry *294*, 5549–5561.
738 https://doi.org/10.1074/jbc.RA118.007018.

739 49. Kong, N.R., Bassal, M.A., Tan, H.K., Kurland, J. V., Yong, K.J., Young, J.J., Yang, Y., Li,
740 F., Lee, J.D., Liu, Y., et al. (2021). Zinc Finger Protein SALL4 Functions through an AT-
741 Rich Motif to Regulate Gene Expression. Cell Rep *34*.
742 https://doi.org/10.1016/j.celrep.2020.108574.

743 50. Pantier, R., Chhatbar, K., Quante, T., Skourti-Stathaki, K., Cholewa-Waclaw, J., Alston, G.,
744 Alexander-Howden, B., Lee, H.Y., Cook, A.G., Spruijt, C.G., et al. (2021). SALL4 controls
745 cell fate in response to DNA base composition. Mol Cell *81*, 845-858.e8.
746 https://doi.org/10.1016/j.molcel.2020.11.046.

747 51. Borozdin, W., Boehm, D., Leipoldt, M., Wilhelm, C., Reardon, W., Clayton-Smith, J.,
748 Becker, K., Mühlendyck, H., Winter, R., Giray, O., et al. (2004). SALL4 deletions are a
749 common cause of Okihiro and acro-renal-ocular syndromes and confirm haploinsufficiency
750 as the pathogenic mechanism. J Med Genet *41*. https://doi.org/10.1136/jmg.2004.019901.

751 52. Ma, X., Huang, R., Li, G., Zhang, T., and Ma, J. (2022). A de novo mutation of SALL4 in a
752 Chinese family with Okihiro syndrome. Mol Med Rep *25*.
753 https://doi.org/10.3892/mmr.2022.12647.

754 53. Donovan KA, A.J.N.R.Y.J.F.E.B.B.E.B.F.E. (2018). Thalidomide promotes degradation of
755 SALL4, a transcription factor implicated in Duane Radial Ray syndrome. Preprint,
756 https://doi.org/10.7554/eLife.38430.001 https://doi.org/10.7554/eLife.38430.001.

757 54. Matyskiela, M.E., Couto, S., Zheng, X., Lu, G., Hui, J., Stamp, K., Drew, C., Ren, Y., Wang,
758 M., Carpenter, A., et al. (2018). SALL4 mediates teratogenicity as a thalidomide-dependent
759 cereblon substrate. Nat Chem Biol *14*, 981–987. https://doi.org/10.1038/s41589-018-0129-
760 x.

761 55. Anh Vu, K.L., Kumari, K., Hui Liu, B., Gao, C., Li, F., Ping Tang, J., Maddalo, D., Auld,
762 D.S., Casalena, D.E., Tian, X., et al. (2023). SALL4B, not targeted by IMiD, is important for
763 SALL4-mediated tumorigenesis. https://doi.org/10.1101/2023.07.07.548071.

764 56. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene
765 set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res *47*, W199–W205.
766 https://doi.org/10.1093/nar/gkz401.

767 57. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C.,
768 Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining
769 Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell
770 Identities. Mol Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

771 58. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T.,
772 Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: An open-source
773 platform for biological-image analysis. Preprint, https://doi.org/10.1038/nmeth.2019
774 https://doi.org/10.1038/nmeth.2019.

775 59. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
776 dispersion for RNA-seq data with DESeq2. Genome Biol *15*.
777 https://doi.org/10.1186/s13059-014-0550-8.

778 60. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing
779 genomic features. Bioinformatics *26*, 841–842.
780 https://doi.org/10.1093/bioinformatics/btq033.

781 61. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S.,
782 Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-
783 sequencing data analysis. Nucleic Acids Res *44*, W160–W165.
784 https://doi.org/10.1093/NAR/GKW257.

785 62. Herbreteau, G., Airaud, F., Pierre-Noël, E., Vallée, A., Bézieau, S., Théoleyre, S., Blons,
786 H., Garinet, S., and Denis, M.G. (2021). MEM: An algorithm for the reliable detection of
787 microsatellite instability (MSI) on a small NGS panel in colorectal cancer. Cancers (Basel)
788 *13*. https://doi.org/10.3390/cancers13164203.

789 63. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum,
790 C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq
791 (MACS). Genome Biol *9*. https://doi.org/10.1186/gb-2008-9-9-r137.

792 64. Polit, L., Kerdivel, G., Gregoricchio, S., Esposito, M., Guillouf, C., and Boeva, V. (2021).
793 CHIPIN: ChIP-seq inter-sample normalization based on signal invariance across
794 transcriptionally constant genes. BMC Bioinformatics *22*. https://doi.org/10.1186/s12859-
795 021-04320-3.

796
797

798 **METHODS**

799 **KEY RESOURCES TABLE**

800

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Brg1 (D1Q7F)	Cell Signaling Technology	AB_2728743
CD99 Antibody, anti-human, Vio® Bright B515, REAfinity™	Miltenyi Biotec	AB_2784331
GAPDH (D16H11) XP Rabbit mAb	Cell Signaling Technology	AB_10622025
Human Nanog Antibody	R&D Systems	AB_355097
Human/Mouse Oct-3/4 Antibody	R&D Systems	AB_354975
SALL1 Polyclonal Antibody	Thermo Fisher Scientific	AB_2900498
Sall4 (EE-30)	Santa Cruz Biotechnology	AB_1129262
Sall4 (D16H12) Rabbit mAb	Cell Signaling Technology	AB_10949321
APC anti-human SSEA-4	Biolegend	AB_2616818
PE anti-human TRA-1-60-R	Biolegend	AB_1279447
Rabbit IgG, monoclonal [EPR25A] Antibody	Abcam	AB_2687931
AP2 alpha Monoclonal Antibody (3B5)	Thermo Fisher Scientific	AB_2199412
CD326 (EpCAM) Monoclonal Antibody (1B7), eBioscience	Thermo Fisher Scientific	AB_795876
Anti-SOX9 antibody	Abcam	AB_2715497
Vimentin Polyclonal Antibody	Thermo Fisher Scientific	AB_2216267
ARID1B antibody	Abcam	AB_2243092
IRDye 800CW Goat anti-Rabbit IgG	LI-COR Biosciences	AB_621843
Goat polyclonal Secondary Antibody to Mouse (IRDye® 800CW)	LI-COR Biosciences	AB_10859231

Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 488	Thermo Fisher Scientific	AB_2762833
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 647	Thermo Fisher Scientific	AB_2762830
Goat anti-Chicken IgY (H+L) Secondary Antibody, Alexa Fluor™ 488	Thermo Fisher Scientific	AB_2534096
Critical commercial assays		
ATAC-Seq Kit	Active Motif	53150
Human Pluripotent Stem Cell Functional Identification Kit	R&D Systems	SC027B
Monarch® Total RNA Miniprep Kit	New England Biolabs	T2010S
Maxima First Strand cDNA Synthesis Kit for RT-qPCR	Thermo Fisher	K1641
NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina®	New England Biolabs	E7760L
CUT&RUN Assay Kit	Cell Signaling Technology	86652
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®	Cell Signaling Technology	E7645L
Experimental models: Cell lines		
PENN123i-SV20	iPSC Core of the University of Pennsylvania	CVCL_EL23
PENN123i-SV20 *SALL4*-het-KO iPSC	Synthego	
Oligonucleotides		
18S	5'-ATACATGCCGACGGGCGCTG-3'	5'-AGGGCTGACCGGGTTGGTT-3'
TFAP2A	5'-GCCTCTCGCTCCTCAGCTCC-3'	5'-CGTTGGCAGCTTTACGTCTCCC-3'
NR2F1	5'-ATCGTGCTGTTCACGTCAGA-3'	5'-GCTCCTCACGTACTCCTCCA-3'

SOX9	5'-GTACCCGCACTTGCACAAC-3'	5'-TCTCFCTCTCGTTCAGAAGTC-3'
SALL1	5'-CCCAGTGTGTGTCCTGTAGT-3'	5'-CCACCTTCACCAACCCTTTG-3'
SALL2	5'-TAGGTGTTGTGTAGGCTGGG-3'	5'-CTCACTTCCATTCCCACCCT-3'
SALL3	5'-GCGCTGCGTAAAGATGAGAA-3'	5'-GGCTACCCTGTCTGTTGAGT-3'
SALL4	5'-GGCATCCAGAGACAGACCTT-3'	5'-GGTCCTCGAGCAGATCTTGT-3'
Software and algorithms		
AlphaFold3	https://alphafoldserver.com/	SCR_025454
TrimGalore!	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	SCR_011847
FlowJo Software version 10.9		SCR_008520
BEDTools		SCR_006646
Deeptools		SCR_016366
R Studio version 4.4.3		
GraphPad Prism 10.2.3	http://www.graphpad.com/faq/viewfaq.cfm?faq=1362	SCR_002798
DESeq2		SCR_015687

METHOD DETAILS

AlphaFold3 structure prediction

FASTA files for each BAF subunits and SALL4 isoforms were collected from the "Sequence & Isoforms" section of Uniprot (www.uniprot.org). The AT-rich DNA sequences was collected from an empirically reported crystal structure of SALL4 zinc finger cluster 4 bound to DNA on RCSB PDB (www.rcsb.org, 8A4I – DOI: https://doi.org/10.2210/pdb8A4I/pdb). FASTA sequences are listed in Supplementary Table 2. AlphaFold outputs were downloaded and .cif files were analysed using UCSF ChimeraX-1.9 (https://www.rbvi.ucsf.edu/chimerax). Contacts were found using default parameters and center-center distance ≤ 8 Å.

Human iPSC culture

Wild-type iPSC lines were obtained from the iPSC Core of the University of Pennsylvania (Control line-1: SV20 line, male, age 43). iPSC lines were cultured in feeder-free, serum-free mTeSR Plus medium (STEMCELL Technologies) with 1% Penicillin-Streptomycin (5,000 U/mL; Gibco

817 #15140122) on tissue culture plates coated overnight with Geltrex™ LDEV-Free hESC-qualified
818 Reduced Growth Factor Basement Membrane Matrix (Fisher-Scientific). Cells were passaged
819 ~1:8 at 60-80% confluency using 0.5mM EDTA to create 50-200 cell clusters.
820
821 **CRISPR/Cas9-mediated SALL4 knock-out**
822 The heterozygous *SALL4* knock-out in the human SV20 iPSC line was generated via
823 CRISPR/Cas9 gene-editing by the biotech company Synthego. Specifically, Cas9 targeted the
824 first exon of the *SALL4* gene in the established SV20 iPSC line producing an 11-base deletion
825 (c.230delGGGCCGCATCT) which generated a premature stop codon.
826
827 **Trilineage differentiation**
828 iPSCs were differentiated into the three germ layers (endoderm, ectoderm and mesoderm) using
829 the Human Pluripotent Stem Cell Functional Identification Kit (R&D Systems, SC027B). The
830 expression of relevant markers was then assessed using immunofluorescence with 10 ug/ml of
831 antibodies provided in the kit (Goat anti-human SOX17, Goat anti-human OTX2 and, Goat anti-
832 human Brachyury).
833
834 **iPSC-to-CNCC differentiation**
835 All iPSC lines were differentiated into CNCCs as previously described (Prescott et al., 2015).
836 Briefly, iPSCs at 60% confluence were treated with CNCC differentiation media (1:1 Neurobasal
837 medium/D-MEM F-12 medium (Invitrogen), 0.5× B-27 supplement with Vitamin A (50× stock,
838 Gibco, #17504044), 0.5× N-2 supplement (100× stock, Gibco; #17502001), 20 ng/ml human FGF
839 (Gibco), 20 ng/ml recombinant EGF (Sigma-Aldrich), 5 µg/ml bovine insulin (Sigma-Aldrich) and
840 1× Glutamax-I supplement (100× stock, Invitrogen). Medium (2 ml) was changed every other day.
841 Around day 7-9 of differentiations, cells were transitioned to CNCC early maintenance media (1:1
842 Neurobasal medium/D-MEM F-12 medium (Invitrogen), 0.5× B-27 supplement with Vitamin A
843 (50× stock, Invitrogen), 0.5× N-2 supplement (100× stock, Invitrogen), 20 ng/ml bFGF
844 (Biolegend), 20 ng/ml EGF (Sigma-Aldrich), 1 mg/ml bovine serum albumin, serum replacement
845 grade (Gemini Bio-Products # 700-104 P) and 1× Glutamax-I supplement (100× stock,
846 Invitrogen). Differentiating cells were passaged ~1:6 with 0.5mM EDTA or Accutase Cell
847 Detachment Solution (Stemcell Technologies; #07920), plated in a fresh, Geltrex-coated 6-well
848 plates. After passaging, cells were treated with 0.05% ROCK-inhibitor (Y27632; Stemcell
849 Technologies) for 24 hours.
850

851 **Real-time quantitative polymerase chain reaction (RT-qPCR)**

852 Cells were harvested as single cells, and the total RNA was extracted using Monarch® Total RNA
853 Miniprep Kit (New England Biolabs) following manufacturer's instructions. 600ng of total RNA was
854 used to synthesize cDNA using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (K1641,
855 Thermo Fisher). For each RT-qPCR reaction, 15 ng of cDNA was used with 0.1 μM of each primer,
856 10 μL of PowerUp™ SYBR™ Green Master Mix (Applied Biosystems) in a final volume of 20 μl.
857 Samples were analyzed with CFX Connect Real-Time PCR Detection System (Bio-Rad) using
858 the following thermal cycling parameters: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C,
859 20 s at 63 °C followed by 30 s at 72 °C. Each sample was run in triplicate. 18 S rRNA was used
860 as normalizer.

861

862 **Nuclear Fractionation and Co-immunoprecipitation**

863 Cells were washed twice with ice-cold PBS, scraped, and collected by centrifugation at 10,000 ×
864 g for 1 minute. Pellets were resuspended in 500 μL of Buffer A (10mM HEPES, 1.5mM
865 Mg2Cl,10mM KCl, 1mM EDTA, 0.5mM DTT) and cytoplasmic fractions were obtained by
866 incubating the cells on ice for 15 minutes, adding of 0.05% NP-40, followed by centrifugation at
867 17,000 × g for 10 minutes at 4 °C. The resulting pellet was gently washed with Buffer A and
868 resuspended in Buffer C (10mM HEPES, 1.5mM Mg2Cl, 300mM NaCl, 0.2mM EDTA, 0.5mM
869 DTT, 26% glycerol). Samples were incubated at 4 °C with shaking at 3,400 rpm for 1 hour, with
870 vortexing every 10 minutes. Nuclear fractions were collected by centrifugation at 17,000 × g for
871 10 minutes at 4 °C.

872 For co-immunoprecipitation (co-IP), 25 μL of Dynabeads™ Protein A (ThermoFisher, Cat#
873 10002D) were washed twice with PBS containing 0.1% Tween-20 (PBS-T), resuspended in 1 mL
874 of PBS-T, and incubated with 2 μg of the appropriate antibody for 1 hour at 4 °C with rotation.
875 Bead-antibody complexes were washed twice with PBS-T, resuspended in 125 μL PBS-T, and
876 incubated with 100 μg of nuclear extract for 3 hours at 4 °C with rotation. Beads were then washed
877 five times with Buffer C and eluted in 25 μL of Buffer C supplemented with 4.1 μL of 6× reducing
878 Laemmli SDS sample buffer (ThermoFisher Scientific). Eluates and 5% input controls (nuclear
879 extract diluted 1:10 in PBS and mixed with 1:6 volume of 6× Laemmli buffer) were boiled for 5
880 minutes prior to SDS-PAGE. For Western blotting, 20 μL of each sample were loaded as
881 described below.

882

883 **Western Blot**

884 For each condition, cells were harvested and lysed with cold radioimmunoprecipitation assay
885 buffer (Pierce™ RIPA buffer, Thermo Fisher) with Pierce™ Protease Inhibitor Mini Tablets, EDTA-
886 free (Thermo Fisher). After centrifugation at 4°C at 16,000x g for 10 minutes, the total extracted
887 protein was quantified using Pierce™ BCA Protein Assay Kit, following manufacturer's
888 indications. Thirty micrograms of total protein were loaded in Novex WedgeWell 4–20% Tris-
889 Glycine Gel (Invitrogen) and separated through gel electrophoresis (SDS–PAGE) Tris-Glycine-
890 SDS buffer (Invitrogen). The proteins were then transferred to nitrocellulose membrane (Thermo
891 Fisher Scientific) and incubated for one hour in Intercept® (PBS) Blocking Buffer (Licor). After
892 blocking, membranes were incubated overnight at 4°C (while rocking) in suitable antibody diluted
893 in blocking buffer and 0.2% Tween-20, at recommended concentrations. Membranes were
894 washed in 0.1% PBS-T before and after being incubated for one hour in the appropriate secondary
895 antibody (1:15,000 dilution). The antibodies were then visualized using Odyssey XF Imaging
896 System (Licor) or ChemiDocTM MP Imaging System and imaged with Image Studio Software
897 (Licor) or Fiji[58].
898
899 **Flow Cytometry analysis of surface markers**
900 To obtain a single-cell suspension for flow cytometry analysis, cells were treated with Accutase
901 for 5 min at 37°C. Cells were then washed with cold phosphate-buffered saline (PBS)-2% fetal
902 bovine serum (FBS) and live cells were counted. 1 × 106 cells/condition were resuspended in
903 100 µL PBS-2% FBS and stained. For pluripotency evaluation, 4 µl of the respective antibodies
904 were used. For analysis of differentiation, 2 µl of the respective antibodies were used. Cells were
905 incubated for 15 min on ice and protected from light, before transferring them into FACS tubes
906 containing an additional 300 µL PBS-2% FBS. Flow cytometry data were acquired using
907 NovoCyte Penteon Flow Cytometer Systems 5 Lasers and analyzed with FlowJo Software version
908 10.9.
909
910 **Immunofluorescence**
911 Cells were passaged onto Geltrex-coated coverslips and fixed at required confluence or time-
912 point. In detail, cells were incubated in 4% PFA at 37°C for 15 minutes, followed by three 2-minute
913 washes in 1× PBS. For permeabilization, cells were treated with 0.1% (v/v) Triton X-100 (T8787,
914 Sigma-Aldrich) at room temperature for 10 minutes before three 2-minute washes in 1× PBS +
915 0.1% (v/v) Tween 20 and a 1-hour blocking upon 10% Donkey serum (D9663, Sigma-Aldrich) at
916 room temperature. Cells were incubated with corresponding PBS-diluted primary antibodies
917 protected from direct light at 4°C overnight. After three 5-minute washes with 1× PBS + 0.1%

918 Tween 20, cells were incubated for 1 hour in PBS-diluted secondary antibodies and 15 minutes
919 with DAPI (1:5,000; 422801, BioLegend). Cells were then washed three times with 1× PBS +
920 0.1% Tween for 5 minutes and once with 1× PBS. Coverslips were then fetched and mounted
921 onto slides using Fluorescence Mounting Medium (S3023, Agilent Technologies). Slides were
922 imaged using confocal microscope (LSM780, Zeiss).
923
924 **RNA-seq**
925 Total RNA was extracted using Monarch® Total RNA Miniprep Kit (New England Biolabs,
926 T2010S) according to the manufacturer's instructions. RNA was quantified using NanoDrop 2000
927 UV Visible Spectrophotometer, while the RNA integrity was checked on Tapestation 2200
928 (Agilent). Only samples with RIN value above 8.0 were used for transcriptome analysis. RNA
929 libraries were prepared using 1 µg of total RNA input using NEBNext® Poly(A) mRNA Magnetic
930 Isolation Module, NEBNext® UltraTM II Directional RNA Library Prep Kit for Illumina® according
931 to the manufacturer's instructions (New England Biolabs).
932
933 **RNA-sequencing analysis**
934 After removing the adapters with TrimGalore!
935 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) , Kallisto was used to quantify
936 reads mapping to each gene. Differential gene expression levels were determined using
937 DESeq2[59]. All statistical analyses were performed using the latest versions of BEDTools[60],
938 deepTools[61], R Studio (version 4.4.3; Trophy Case), and GraphPad Prism software version 10.2.3
939 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San Diego, CA).
940
941 **CUT&RUN-seq**
942 For each condition, 1x 10^5 cells were processed using CUT&RUN Assay Kit #86652 (Cell
943 Signaling Technology). Briefly, each sample was resuspended in Concanavalin A Beads and
944 treated with an Antibody Binding Buffer and the suitable antibody. After the antibody incubation,
945 each reaction underwent a step of pAG-MNase Enzyme binding, and subsequent DNA Digestion
946 and diffusion. The diffused DNA was then purified using Phenol/Chloroform Extraction and
947 Ethanol Precipitation. The purified DNA was quantified using QuantiFluor dsDNA System, using
948 the ONE dsDNA protocol (Promega). Libraries for sequencing were prepared using NEBNext
949 Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) and the libraries size distribution
950 was assessed through automated electrophoresis using the Agilent 2200 Tapestation System
951 (Agilent Technology).

952

953 **ChIP-seq**

954 All samples from different conditions were processed together to prevent batch effects.
955 Approximately 15 million cells were double cross-linked with 2mM disuccinimidyl glutarate (DSG)
956 for 20 minutes followed by 1% formaldehyde for 10 minutes at room temperature, quenched with
957 125 mM glycine, harvested, and washed twice with 1X PBS. The fixed cell pellet was resuspended
958 in ChIP lysis buffer (150 mM NaCl, 1% Triton X-100, 0.7% SDS, 500 µM DTT, 10 mM Tris-HCl, 5
959 mM EDTA) and chromatin was sheared to an average length of 200–900 base-pairs, using a
960 Covaris S220 Ultrasonicator at 5% duty factor for 7 minutes for 3 times. The chromatin lysate was
961 diluted with SDS-free ChIP lysis buffer. 10µg of antibody was used for BRG1. The antibody was
962 added to at least 5µg of sonicated chromatin along with Dynabeads Protein G magnetic beads
963 (Invitrogen) and incubated with rotation at 4°C overnight. The beads were washed twice with each
964 of the following buffers: Mixed Micelle Buffer (150 mM NaCl, 1% Triton X- 100, 0.2% SDS, 20 mM
965 Tris-HCl, 5 mM EDTA, 65% sucrose), Buffer 200 (200 mM NaCl, 1% Triton X-100, 0.1% sodium
966 deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), LiCl detergent wash (250 mM LiCl,
967 0.5% sodium deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA) and a final wash was
968 performed with cold 0.1X TE. Finally, beads were resuspended in 1X TE containing 1% SDS and
969 incubated at 65°C for 10 min to elute immunocomplexes. The elution was repeated twice, and the
970 samples were incubated overnight at 65°C to reverse cross-linking, along with the input (5% of
971 the starting material). The DNA was digested with 0.5 mg/ml Proteinase K for 3 hours at 55°C
972 and then purified using the ChIP DNA Clean & Concentrator kit (Zymo) and quantified with QUBIT.
973 Barcoded libraries were made with NEBNext Ultra II DNA Library Prep Kit for Illumina) using
974 NEBNext Multiplex Oligos Dual Index Primers for Illumina (New England BioLabs) and sequenced
975 on NovaSeq X Plus instrument (Illumina).

976

977 **ATAC-seq**

978 One hundred thousand cells/condition were harvested by enzymatic dissociation using
979 StemPro™ Accutase™ (Thermo Fisher Scientific). Each cell pellet was processed using ATAC-
980 Seq Kit (Active Motif), following the manufacturer's directions. Briefly, the fresh cell pellets were
981 washed in 1X PBS and lysate in ATAC Lysis Buffer. Each lysate underwent a process of
982 tagmentation and purification, before being amplified through PCR, using the following thermal
983 cycler parameters: 72°C for 5 min, 98°C for 30 sec, followed by 10 cycles of 98°C for 10 sec,
984 63°C for 30 sec, 72°C for 1 min. Each PCR reaction was cleaned up using SPRI beads (1.2X of

985 the sample volume). The Agilent 2200 Tapestation System (Agilent Technologies) was used as
986 quality control tool before performing next generation sequencing (NGS).
987
988 **Bioinformatics analysis of CHIP-seq, ATAC-seq and CUT&RUN data**
989 Adapters were removed with TrimGalore!
990 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), the sequences were aligned
991 to the reference hg19, using Burrows-Wheeler Alignment tool, with the MEM algorithm[62]. Uniquely
992 mapping aligned reads were filtered based on mapping quality (MAPQ > 10) to restrict our
993 analysis to higher quality and uniquely mapped reads, and PCR duplicates were removed.
994 MACS2[63] was used to call peaks using the default parameters at 5% FDR.
995
996 **Motif analysis**
997 BED files for the regions of interest were produced using BEDTools[60]. Motif analysis was
998 performed using HOMER[57]. Shuffled input sequences were used as background. E-values <
999 0.001 were used as a threshold for significance.
1000
1001 **QUANTIFICATION AND STATISTICAL ANALYSIS**
1002 RT-qPCR and image quantification data were analyzed using Prism software version 10.2.3
1003 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San Diego, CA). Data is
1004 represented as mean ± standard error of mean (SEM). GO enrichment analysis was performed
1005 using Webgestalt 2019[56]. All statistical analyses were performed using the latest versions of
1006 BEDTools[60], Deeptools[61], R version 4.4.3 (Throphy Case; https://www.R-project.org/), and Prism
1007 software version 10.2.3 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San
1008 Diego, CA). In particular, Bigwigs were generated with Deeptools and normalized by sequencing
1009 depth using the same software, and upper-quantile normalized using CHIPIN[64].
1010
1011

FIG 1

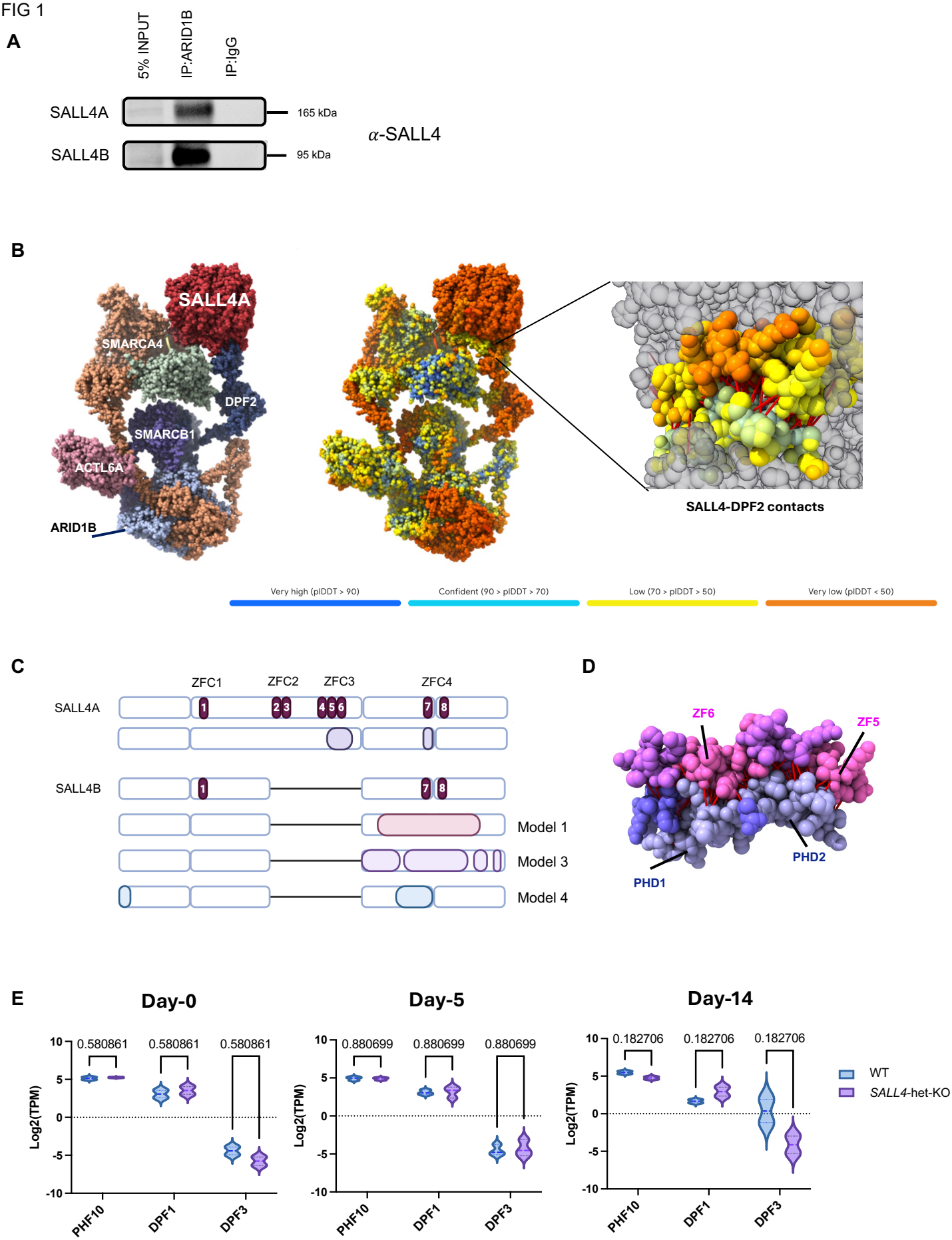
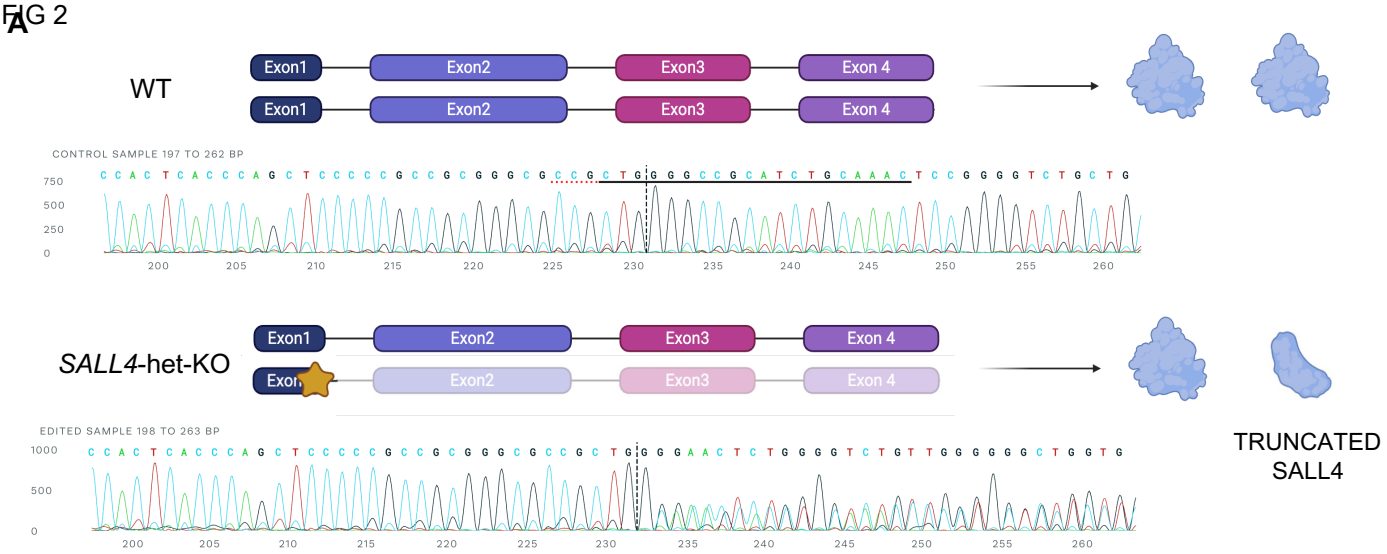
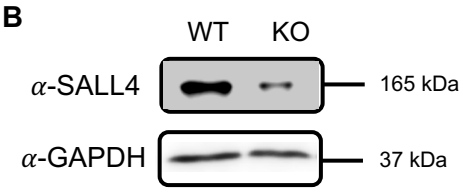
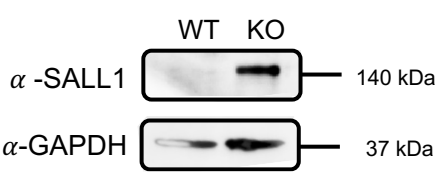
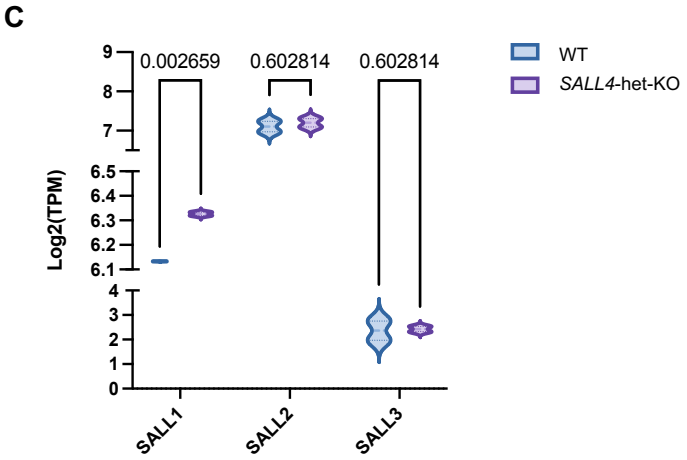

A
5% INPUT
IP:ARID1B
IP:IgG
SALL4A
SALL4B
165 kDa
95 kDa
α-SALL4
B
SALL4A
SMARCA4
DPF2
SMARCB1
ACTL6A
ARID1B
SALL4-DPF2 contacts
Very high (pIDDT > 90)
Confident (90 > pIDDT > 70)
Low (70 > pIDDT > 50)
Very low (pIDDT < 50)
C
ZFC1
ZFC2
ZFC3
ZFC4
SALL4A
SALL4B
Model 1
Model 3
Model 4
D
ZF6
ZF5
PHD2
PHD1
E
Day-0
Day-5
Day-14
Log2(TPM)
0.580861
0.880699
0.182706
PHF10
DPF1
DPF3
WT
SALL4-het-KO

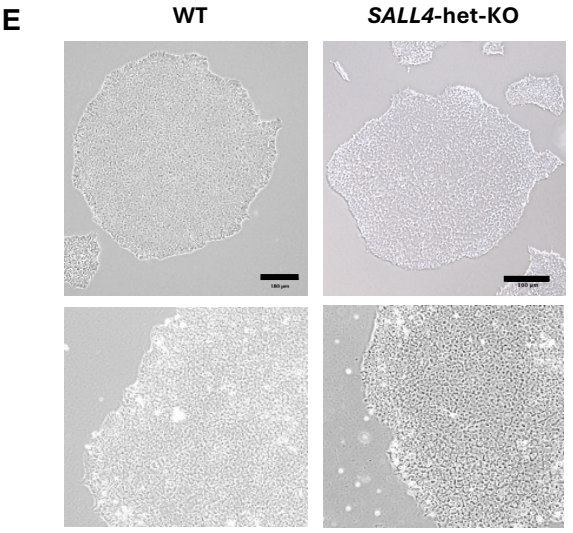
FIG 2


A
WT
Exon1
Exon2
Exon3
Exon 4
CONTROL SAMPLE 197 TO 262 BP
SALL4-het-KO
EDITED SAMPLE 198 TO 263 BP
TRUNCATED SALL4


B
WT
KO
α-SALL4
165 kDa
α-GAPDH
37 kDa


D
WT
KO
α -SALL1
140 kDa
α-GAPDH
37 kDa


C
0.002659
0.602814
0.602814
Log2(TPM)
SALL1
SALL2
SALL3
WT
SALL4-het-KO


E
WT
SALL4-het-KO

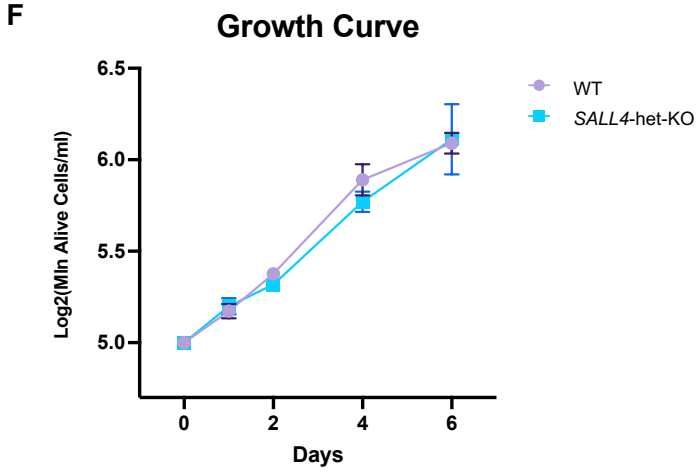
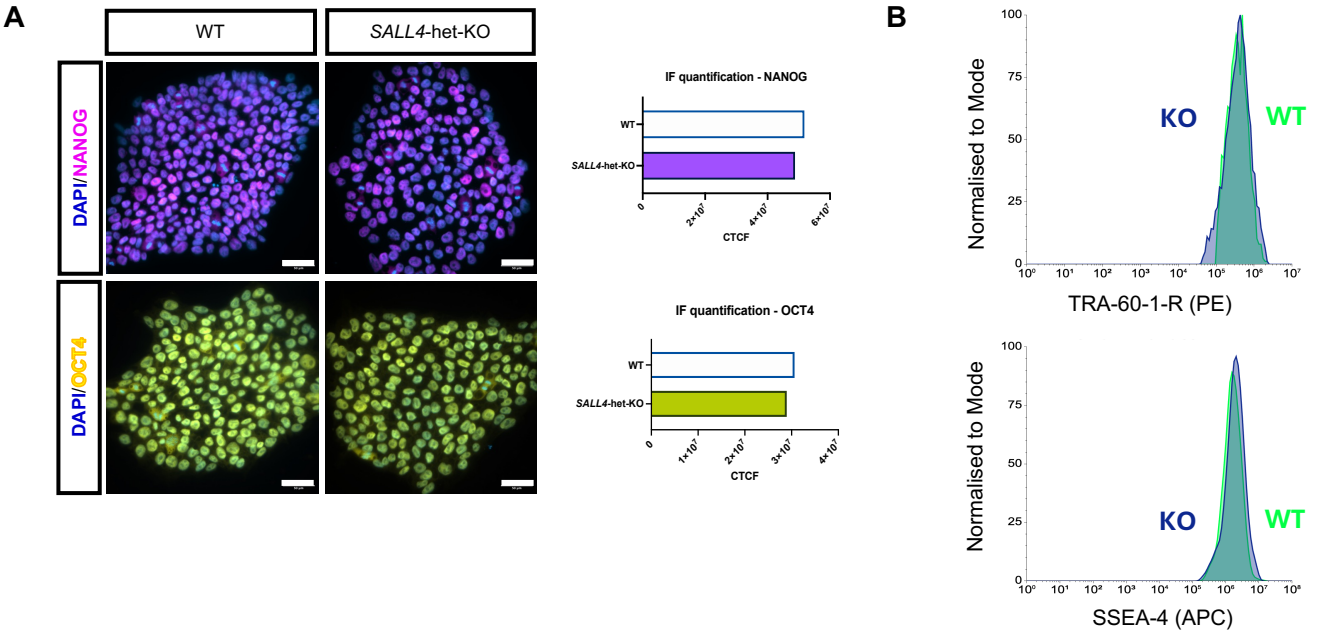
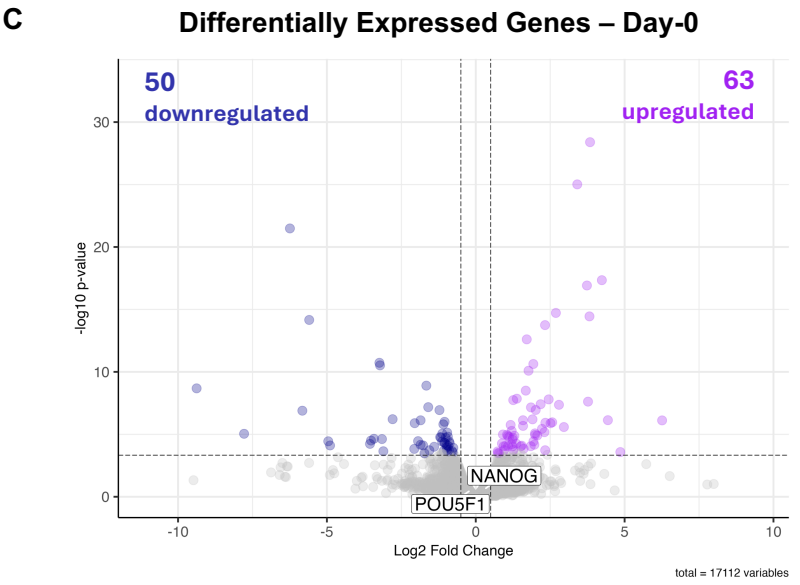

F
Growth Curve
Log2(MIn Alive Cells/ml)
Days
WT
SALL4-het-KO

FIG 3





Top 10 Gene Ontology Terms – Day-0

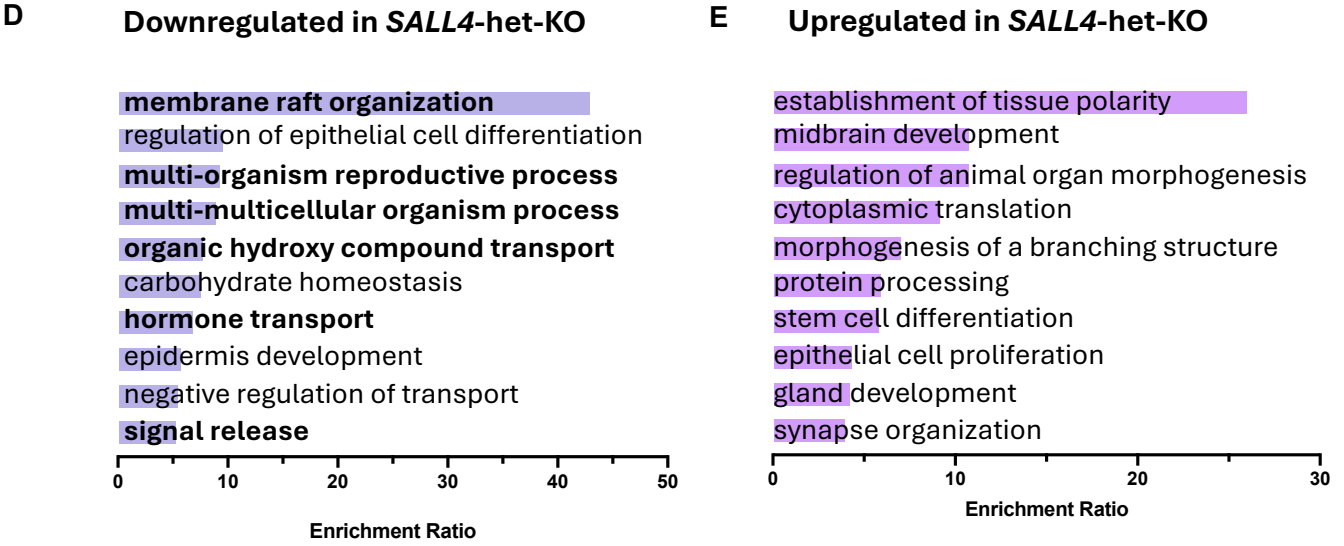
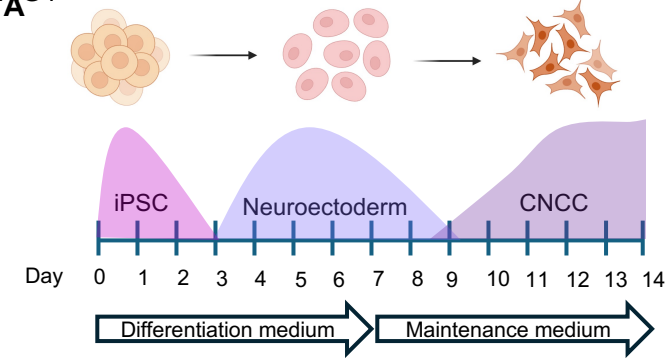
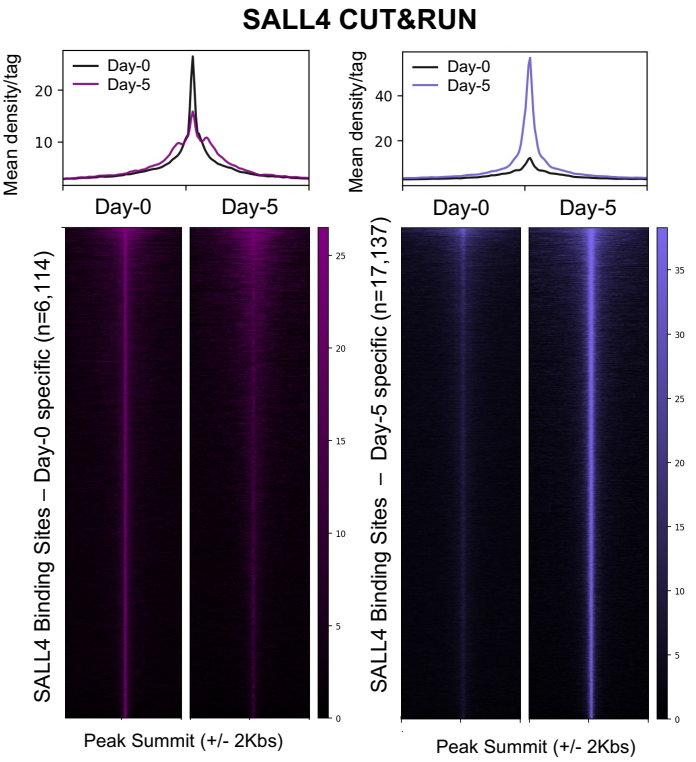


FIG4

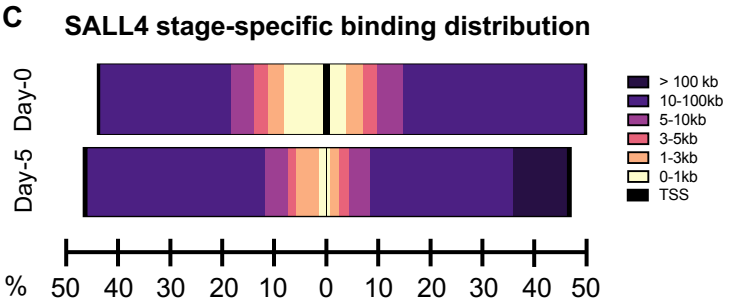
A



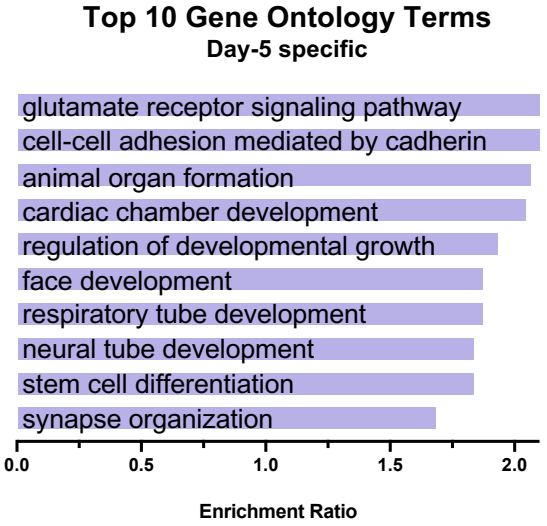
B



C



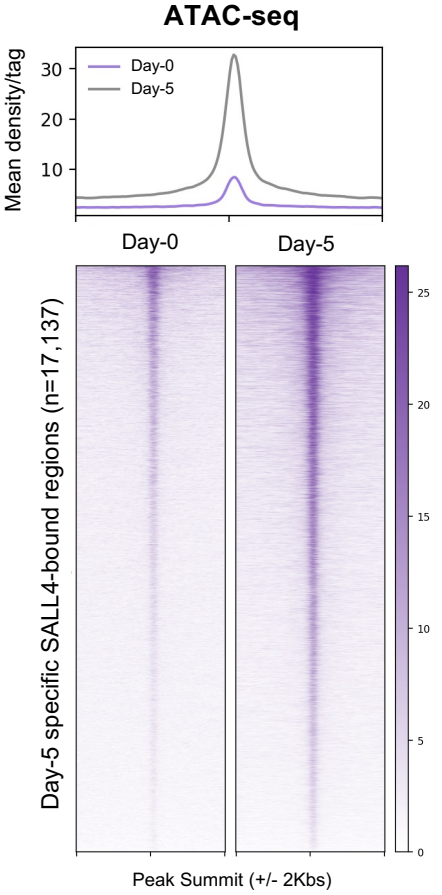
D



E

SALL4 CUT&RUN Day-5 specific motifs		
TF	Motif Consensus	p-value
TEAD		$1e^{-1356}$
JUN/FOS		$1e^{-938}$
SOX9		$1e^{-817}$
ZIC		$1e^{-498}$
TCF/TWIST		$1e^{-132}$
TFAP2A		$1e^{-90}$

F



G

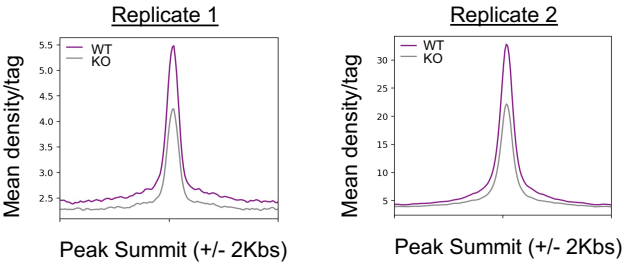
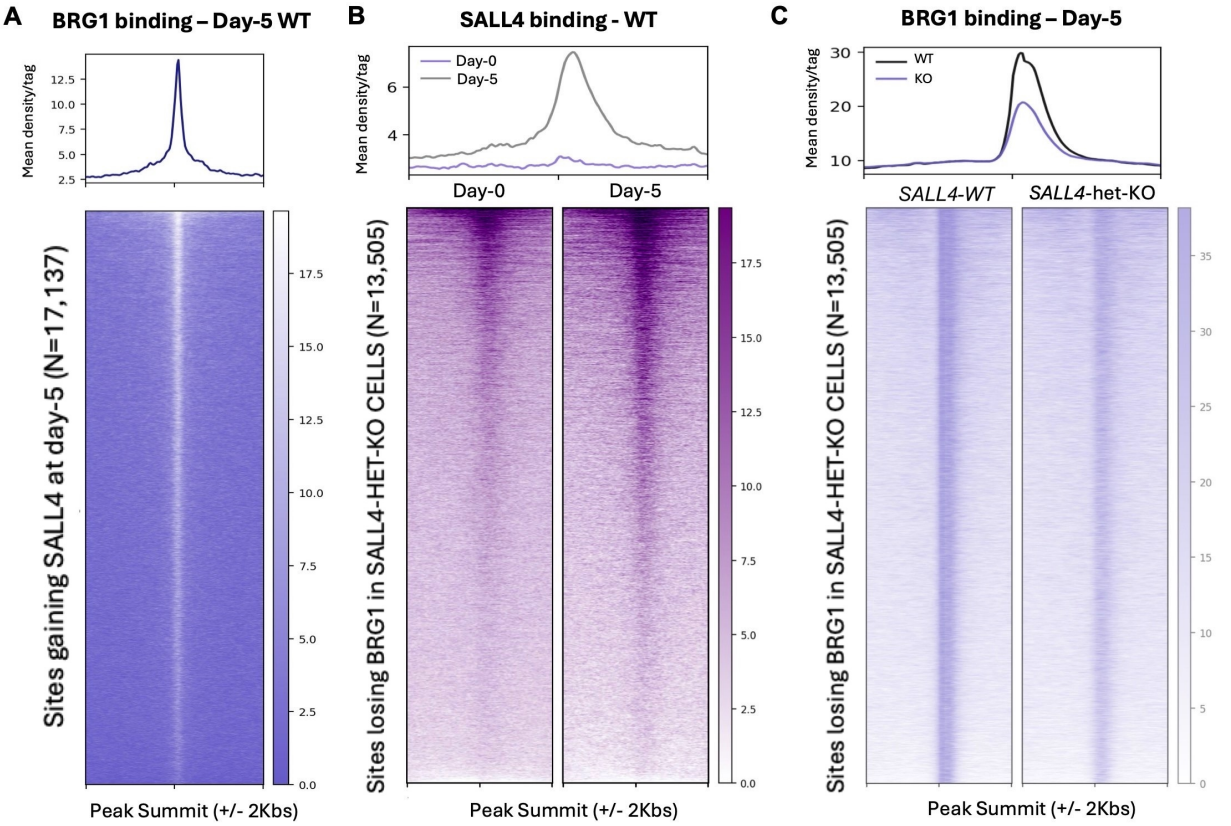
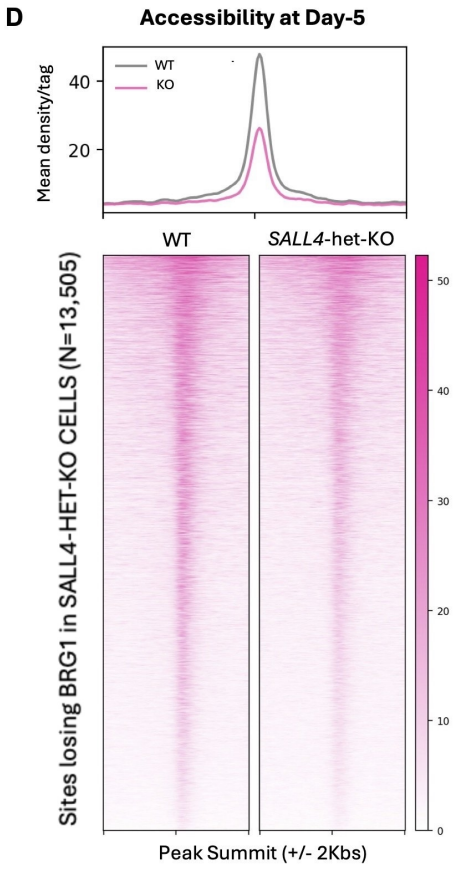


FIG. 5


A
BRG1 binding – Day-5 WT
Mean density/tag
Sites gaining SALL4 at day-5 (N=17,137)
Peak Summit (+/- 2Kbs)
B
SALL4 binding - WT
Day-0
Day-5
Sites losing BRG1 in SALL4-HET-KO CELLS (N=13,505)
Peak Summit (+/- 2Kbs)
C
BRG1 binding – Day-5
WT
KO
SALL4-WT
SALL4-het-KO
Sites losing BRG1 in SALL4-HET-KO CELLS (N=13,505)
Peak Summit (+/- 2Kbs)


D
Accessibility at Day-5
WT
KO
Mean density/tag
WT
SALL4-het-KO
Sites losing BRG1 in SALL4-HET-KO CELLS (N=13,505)
Peak Summit (+/- 2Kbs)

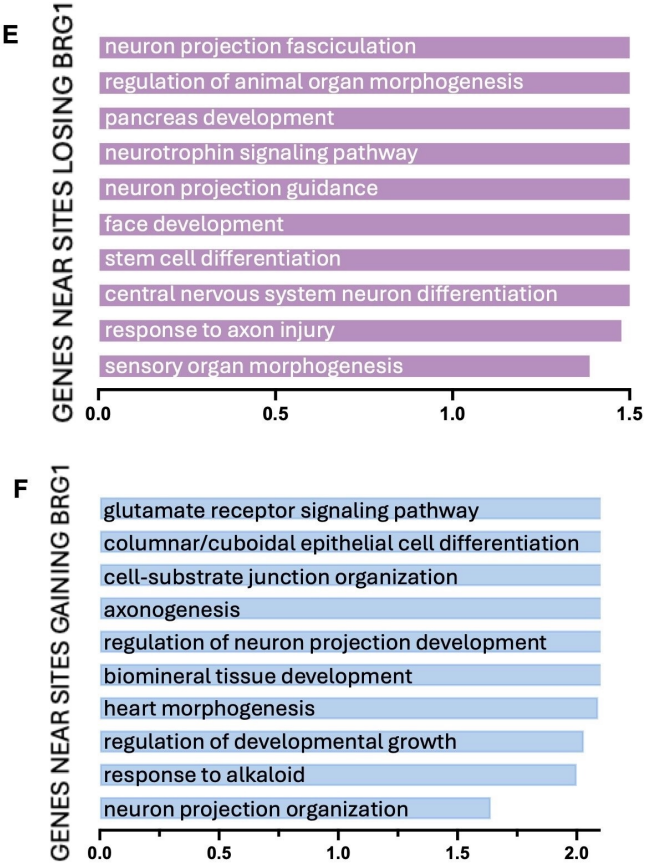

E
GENES NEAR SITES LOSING BRG1
neuron projection fasciculation
regulation of animal organ morphogenesis
pancreas development
neurotrophin signaling pathway
neuron projection guidance
face development
stem cell differentiation
central nervous system neuron differentiation
response to axon injury
sensory organ morphogenesis
0.0
0.5
1.0
1.5
F
GENES NEAR SITES GAINING BRG1
glutamate receptor signaling pathway
columnar/cuboidal epithelial cell differentiation
cell-substrate junction organization
axonogenesis
regulation of neuron projection development
biomineral tissue development
heart morphogenesis
regulation of developmental growth
response to alkaloid
neuron projection organization
0.0
0.5
1.0
1.5
2.0

FIG 6

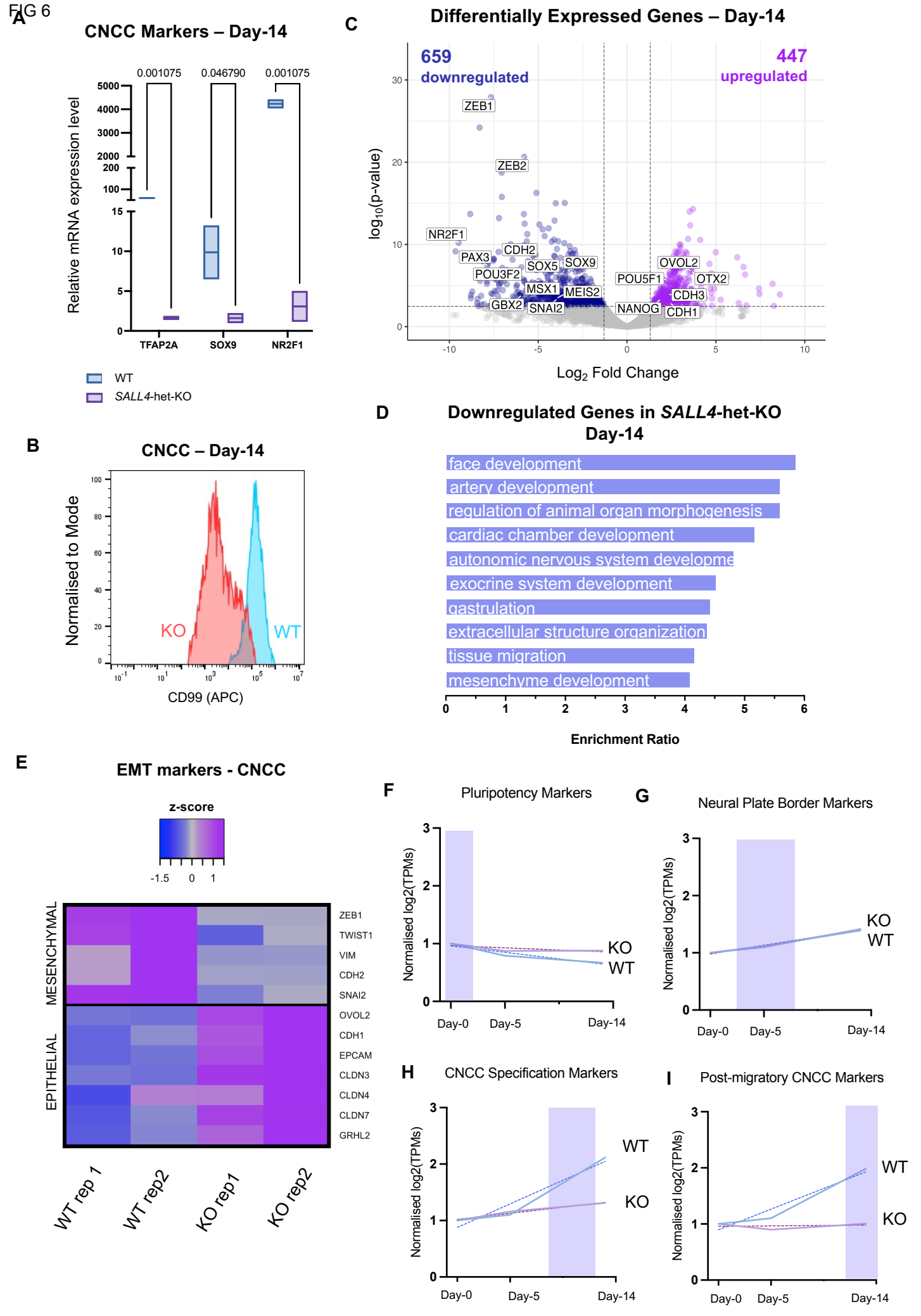
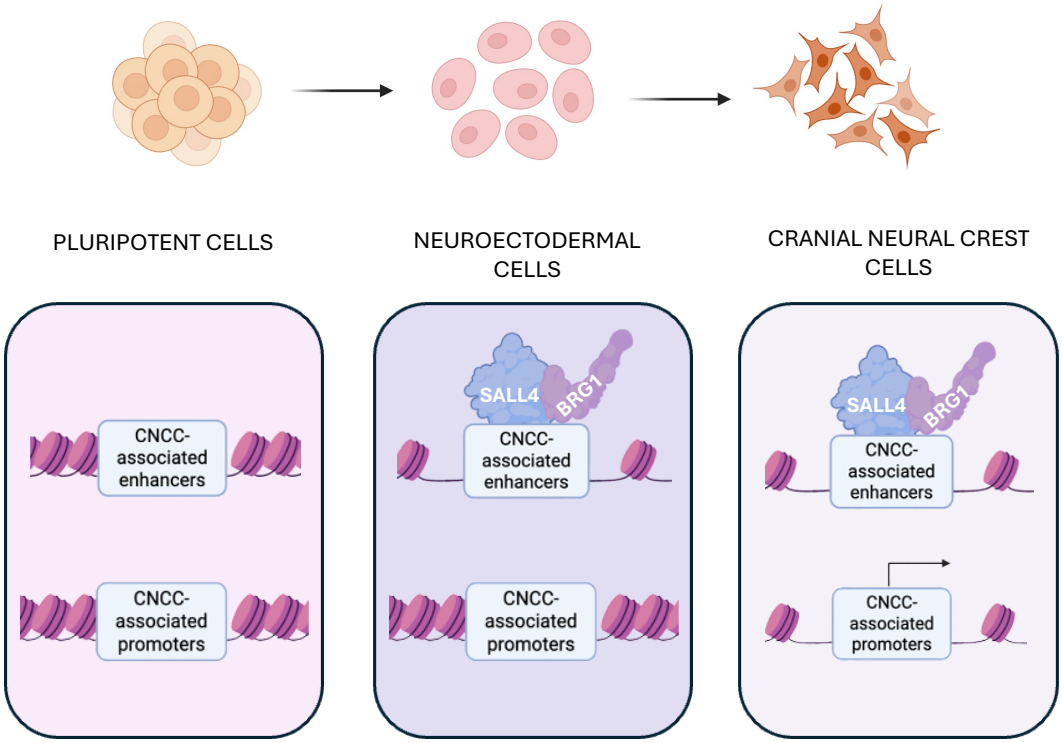
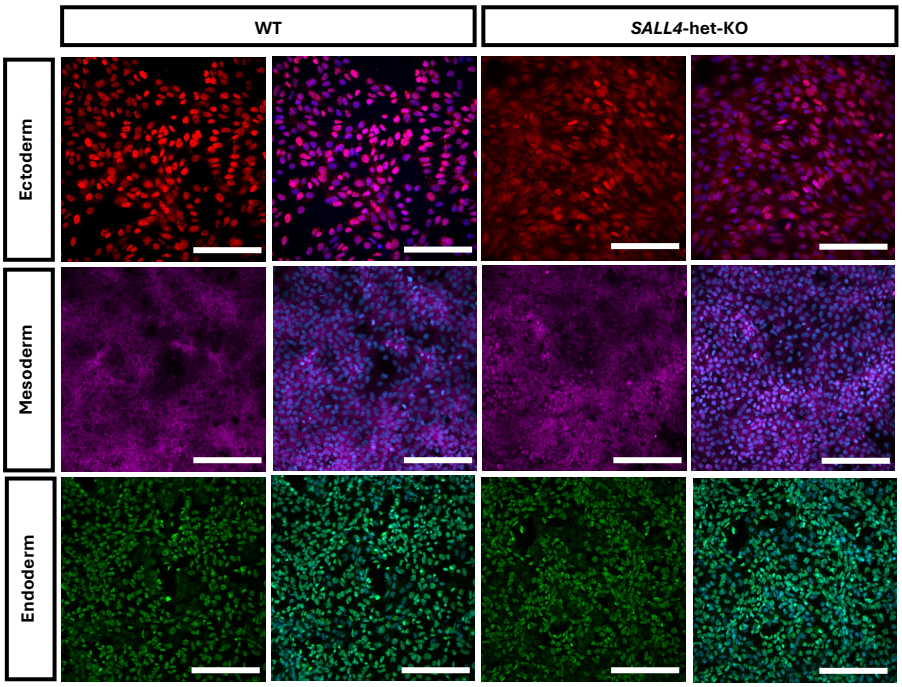

A
CNCC Markers – Day-14
0.001075
0.046790
0.001075
Relative mRNA expression level
TFAP2A
SOX9
NR2F1
WT
SALL4-het-KO
B
CNCC – Day-14
Normalised to Mode
KO
WT
CD99 (APC)
C
Differentially Expressed Genes – Day-14
659
downregulated
447
upregulated
log10(p-value)
Log2 Fold Change
ZEB1
ZEB2
NR2F1
PAX3
CDH2
SOX5
SOX9
POU3F2
MSX1
MEIS2
GBX2
SNAI2
POU5F1
OVOL2
OTX2
CDH3
NANOG
CDH1
D
Downregulated Genes in SALL4-het-KO Day-14
face development
artery development
regulation of animal organ morphogenesis
cardiac chamber development
autonomic nervous system developme
exocrine system development
gastrulation
extracellular structure organization
tissue migration
mesenchyme development
Enrichment Ratio
E
EMT markers - CNCC
z-score
-1.5
0
1
MESENCHYMAL
EPITHELIAL
ZEB1
TWIST1
VIM
CDH2
SNAI2
OVOL2
CDH1
EPCAM
CLDN3
CLDN4
CLDN7
GRHL2
WT rep 1
WT rep2
KO rep1
KO rep2
F
Pluripotency Markers
Normalised log2(TPMs)
KO
WT
Day-0
Day-5
Day-14
G
Neural Plate Border Markers
Normalised log2(TPMs)
KO
WT
Day-0
Day-5
Day-14
H
CNCC Specification Markers
Normalised log2(TPMs)
WT
KO
Day-0
Day-5
Day-14
I
Post-migratory CNCC Markers
Normalised log2(TPMs)
WT
KO
Day-0
Day-5
Day-14

FIG 7


PLURIPOTENT CELLS
NEUROECTODERMAL CELLS
CRANIAL NEURAL CREST CELLS
CNCC-associated enhancers
CNCC-associated promoters
SALL4
BRG1
CNCC-associated enhancers
CNCC-associated promoters
SALL4
BRG1
CNCC-associated enhancers
CNCC-associated promoters

Supplementary Figures

A



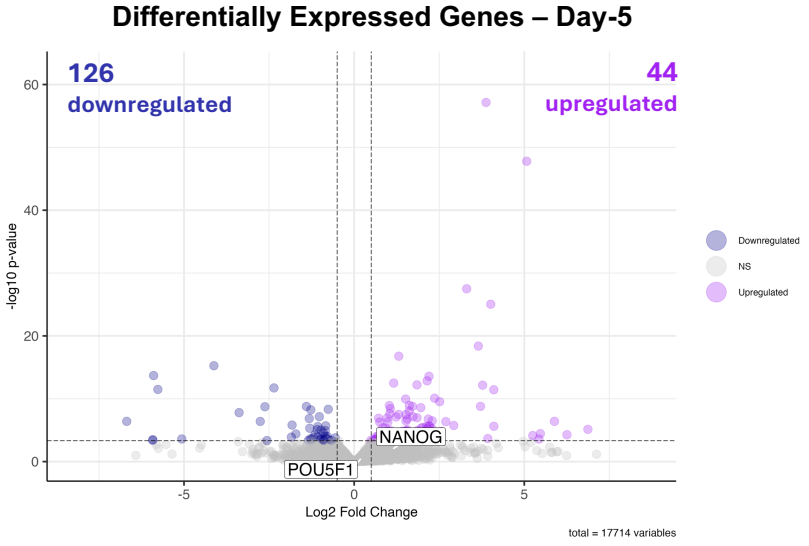
Supplementary Figure 1.

A. Immunofluorescence for ectodermal (OTX2), mesodermal (Brachyury), and endodermal (SOX17) markers in *SALL4*-WT and *SALL4*-het-KO after trilineage differentiation. Scale bar: 100 μm.

A

SALL4 CUT&RUN Day-0 specific motifs		
Transcription Factor	**Motif Consensus**	**p-value**
CTCF/BORIS		$1e^{-461}$
bHLH		$1e^{-106}$
OCT:OCT (MORE)		$1e^{-103}$
ASCL1/2		$1e^{-67}$
OCT4-NANOG-SOX2		$1e^{-36}$

B



Supplementary Figure 2.
(A) Table of motifs enriched in Day-0 specific SALL4-bound regions as determined using HOMER[63].
(B) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in neuroectodermal-like cells (Day-5). Blue dots (n= 126) represent downregulated genes with p-adj < 0.05 and log2FoldChange < -1.5. Purple dots (n= 44) represent upregulated genes with p-adj < 0.05 and log2FoldChange > 1.5.