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

ABSTRACT

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 associated with syndromes featuring craniofacial anomalies, plays a critical role in early cranial neural crest (CNCC) specification. Using *SALL4*-het-KO human iPSCs to model clinical haploinsufficiency, we show that SALL4 directly recruits BAF to CNCC-lineage specific enhancers at the neuroectodermal stage, specifically when neural crest gene expression is induced at the neural plate border. Without functional SALL4, BAF is not loaded at chromatin, leaving CNCC enhancers inaccessible. Consequently, the cells cannot undergo proper CNCC induction and specification due to persistent enhancer repression, despite normal neuroectodermal and neural plate progression. Moreover, by performing SALL4 isoform-specific depletion, we demonstrate that the SALL4A is the isoform essential for CNCC induction and specification, and that SALL4B cannot compensate for SALL4A loss in this developmental process.

In summary, our findings reveal SALL4 as essential regulator of BAF-dependent enhancer activation during early stages of neural crest development, providing molecular insights into SALL4-associated craniofacial anomalies.

KEYWORDS

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

CONFIDENTIAL

INTRODUCTION

Transcription factors and chromatin remodelers play crucial roles in cell fate determination (Gopinathan and Diekwisch, 2022; Islam et al., 2021; Kim and Shendure, 2019; Li, 2002; Voss and Hager, 2014; Zheng et al., 2021). Among these, the C2H2 zinc-finger SALL4 is involved in a wide range of developmental processes (Tatetsu et al., 2016). SALL4 is highly expressed in pluripotent cells and throughout embryonic development, gradually decreasing after birth to become undetectable in adult tissues (Xiong et al., 2015).

In humans, *SALL4* produces two main isoforms: SALL4A (1053 aa) and SALL4B (617 aa), which have been proposed to play distinct roles in ESCs. Namely, Rao et al. (2010) suggested that SALL4B has function limited to pluripotency regulation in ESCs, while SALL4A is 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 factors to sustain stem cell identity (Yang, 2018; Zhang et al., 2006). In contrast, others propose that SALL4 plays a more marginal role in pluripotency regulation (Miller et al., 2016; Wang et al., 2014; Yuri et al., 2009).

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 (Chen et al., 2022; Tahara et al., 2019). Here, it regulates cell fate commitment of neuromesodermal progenitors by suppressing neural genes in mesodermal cells while simultaneously enhancing mesodermal gene expression (Pappas et al., 2024). Additionally, *Sall4* remains expressed in mouse craniofacial regions until E9.5, including the developing frontonasal process and the mandibular and maxillary arches (Tahara et al., 2019). These craniofacial structures originate from cranial neural crest cells (CNCCs), a subpopulation of the neural crest cells that gives rise to the cartilage and bones of the face and anterior skull, as well as cranial neurons and glia (Dash and Trainor, 2020; Milet and Monsoro-Burq, 2012; Pla and Monsoro-Burq, 2018; Roth et al., 2021; Szabó et al., 2018). 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 and flattened nasal bridge, along with cranial nerve-related conditions and underdevelopment of one side of the face (Kekunnaya and Negalur, 2017). Many of these distinctive features are classic indicators of impaired CNCC development. *SALL4*-associated phenotypes include Duane anomaly, a congenital strabismus resulting from abnormal oculomotor nerve innervation (Kekunnaya and Negalur, 2017), and sensorineural hearing loss (Terhal et al., 2006), but also upper limb anomalies (Duane-radial ray syndrome) and, in some cases, renal anomalies.

CONFIDENTIAL

Together, these findings suggest that SALL4 may play a critical role in CNCC development, which has not yet been investigated.

Previous studies have also suggested that SALL4 interacts directly with chromatin remodeling complexes. For example, Sall4 is an interactor of the NuRD complex, in mESCs and during development (Bode et al., 2016; Liu et al., 2018b; Lu et al., 2009; Miller et al., 2016; Wang et al., 2023). 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 complex at this developmental stage (Pagliaroli et al., 2021). 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 determine the role of SALL4 during CNCC development.

RESULTS

SALL4 interacts with the BAF complex

CNCCs can be generated *in vitro* using human induced pluripotent stem cells (hiPSCs). This protocol is two weeks long, and it first goes through a neuroectoderm intermediate (Bajpai et al., 2010; Barnada et al., 2024; Mitchell et al., 2025; Pagliaroli et al., 2021; Prescott et al., 2015). In our previous work (Pagliaroli et al., 2021), mass spectrometry (IP-MS) conducted at early stages of iPSC-to-CNCC differentiation (Day-5, corresponding to neuroectoderm), revealed that SALL4 interacts with BAF at this developmental stage. Since this was the first report of a SALL4-BAF interaction, we performed a co-immunoprecipitation (co-IP) at Day-5 to validate it, pulling-down ARID1B, which is the only ARID1 subunit active and incorporated in BAF at this stage (Pagliaroli et al., 2021), and blotting for SALL4. This co-IP confirmed that ARID1B interacts with both SALL4 isoforms at this time-point (Fig. 1A).

BAF is a multi-subunit assembly composed of a dozen subunits, which are differentially incorporated in a tissue-specific manner (Hodges et al., 2016). Based on which subunit is present in each complex, in mammals these can be classified into canonical BAF, polybromo-associated BAF, and non-canonical BAF (Zhang et al., 2021). While our mass-spec and co-IP experiments supported the interaction between SALL4 and canonical BAF (ARID1B is not incorporated in other BAF configurations), neither of these two approaches has the resolution to identify the exact BAF subunit that mediates the interaction with SALL4. Therefore, we performed an *in silico* analysis using AlphaFold3 (Jumper et al., 2021). We tested various BAF subunit combinations along with four histones and an AT-rich DNA sequence which has been empirically shown to be recognized by SALL4 (5'-GGAAATATTACC-3'; PDB: 8A4I) and SALL4A. The output was statistically

CONFIDENTIAL

101 significant, with 0.45 predicted template modeling (pTM) score, and 0.47 interface predicted
102 template modeling (ipTM) score (Figs. S1A, B). Through this approach, AlphaFold3 predicted a
103 direct interaction between DPF2, a broadly expressed BAF subunit, and the zinc-fingers 5 and 6
104 (ZFC3) and 7 (ZFC4) of SALL4A (Figs. 1B, C, D; Figs. S1A, B; Table 1). *DPF2* is a member of
105 the *BAF45* gene family, which also includes *PHF10*, *DPF1*, and *DPF3*. We were able to rule out
106 its paralogues based on their distinct expression patterns, incorporation into specific BAF
107 complexes, and previous reports that BAF45 subunits are not functionally redundant (Zhang et
108 al., 2019a). In particular, PHF10 is exclusively associated with the pBAF, which does not
109 incorporate ARID1B (Krasteva et al., 2017), while DPF1 and DPF3 are incorporated in BAF only
110 in neurons (Lessard et al., 2007). Moreover, we compared the gene expression of each *DPF*
111 paralogue at both Day-0 and Day-5 and observed no significant difference in expression between
112 *SALL4*-WT and *SALL4*-het-KO for any of them (Figs. 1E, F).
113 This prediction was particularly interesting given that DPF2 deletion has been linked to early
114 neural ectoderm differentiation in embryonic stem cells (ESCs) (Zhang et al., 2019a). Notably, of
115 the two main SALL4 isoforms (SALL4A and SALL4B), SALL4B lacks the portion of exon 2 that
116 contains ZFC3. This prompted us to explore whether the SALL4-DPF2 interaction could be
117 isoform-specific. To address this, we reanalysed the interactions in AlphaFold3, incorporating
118 SALL4B to the same BAF subunits as before. To ensure biological relevance, we first excluded
119 two of the five models that did not exhibit interactions with the AT-rich DNA sequence, as SALL4
120 is expected to bind DNA for its function. Among the remaining three models, we observed that
121 SALL4B interacted with DPF2, but unlike the SALL4A isoform, its interaction was not domain
122 specific. Instead, the predicted interactions for SALL4B were characterized by annotated
123 disordered regions and appeared to be largely driven by a non-specific, zinc-finger-to-zinc-finger
124 association (Figs. S1C, D; Table S2), with no particular domain-specific interactions identified.
125 To accurately interpret our AlphaFold3 predictions, we examined two key confidence metrics: the
126 Predicted Aligned Error (PAE), which assesses the reliability of the relative positioning of residues
127 within the structure, and the Predicted Local Distance Difference Test (pLDDT), which provides a
128 per-residue confidence score for the model (Table S1). Our SALL4-DPF2 contact analysis
129 revealed 394 atomic interactions (center-to-center distance ≤ 8Å) involving 139 residues (Fig. 1D
130 and Table S1). Encouragingly, the pLDDT values were highest in known functional domains
131 involved in the interaction, particularly ZF5, ZF6, and ZF7 in SALL4A, as well as the PHD1 and
132 PHD2 domains in DPF2, reinforcing the reliability of these predicted structures. Consistently, our
133 model also aligned closely with experimentally resolved structures at 3 Å resolution available for
134 the BAF base module (Fig. 1E). To refine our analysis, we focused on residues where at least

CONFIDENTIAL

135 half of their atomic interactions had a pLDDT > 65 (70 ± 10%) and that were annotated as part of
136 a functional domain, ensuring that only higher-confidence and potentially functional interactions
137 were considered. As a result, ZF7 was excluded from further analysis (Table S1).
138 In summary, our experiments and analyses confirmed that SALL4 interacts with BAF during early
139 stages of human CNCC specification, and predicted that this interaction might be mediated by
140 the BAF subunit DPF2 through ZFC3.
141
142 **Generation of a *SALL4*-haploinsufficient human iPSC lines**
143 Our experiments so far revealed that SALL4 interacts with BAF at early stages of CNCC
144 specification, and specifically at the neuroectodermal stage. To investigate the scope of this
145 interaction, we generated *SALL4* haploinsufficient iPSC lines (hereafter *SALL4*-het-KO; Fig. 2A).
146 We opted for heterozygous KO of *SALL4* as this recapitulates the human *SALL4* syndromes,
147 which are all caused by *SALL4* haploinsufficient variants (Al-Baradie et al., 2002; Rgen Kohlhase
148 et al., 2002; Terhal et al., 2006). Two different CRISPR/Cas9 clones were generated and used
149 as biological replicates. Specifically, the CRISPR-KO was generated by introducing a 11 bp (clone
150 C4) deletion and a 19 bp (clone F1) deletion, respectively (Fig. S2A), both within the 3'-terminus
151 of exon 1 of *SALL4*. This is a conserved region across both *SALL4* isoforms that lacks functionally
152 significant domains. Therefore, both *SALL4* isoforms were equally impacted by the CRISPR
153 deletions.
154 SALL4 immunoblot and immunofluorescence assays confirmed that SALL4 protein expression
155 was significantly reduced in the *SALL4*-het-KO iPSCs compared to controls (*SALL4*-WT; Figs.
156 2B, C; Fig. S2B, C). We noted that SALL4A protein expression in the *SALL4*-het-KO line was
157 significantly less than the expected 50% of the wild-type, suggesting potential negative feedback
158 mechanisms triggered by the reduced SALL4 levels, possibly due to the preferentially
159 homodimeric nature of this protein. On the other hand, SALL4B protein was barely detectable in
160 both *SALL4*-WT and *SALL4*-het-KO conditions (Fig. 2B).
161 In contrast, at the gene expression level we did not observed significant differences in *SALL4*
162 expression between *SALL4*-WT and *SALL4*-het-KO, for any isoform (Fig. 2D). This may reflect
163 autoregulatory mechanisms, where reduced SALL4 protein levels in the knockout trigger
164 compensatory transcriptional upregulation of the gene which however does not result in increased
165 protein level because of nonsense-mediated decay or other post-transcriptional regulatory
166 mechanisms.
167 Next, we investigated whether other *SALL* paralogs (*SALL1*, *SALL2*, *SALL3*) could compensate
168 for SALL4 loss. First, we investigated their mRNA expression (Fig. 2E) and found *SALL1* as the

CONFIDENTIAL

169 only paralog significantly upregulated in *SALL4*-het-KO (Fig. 2E). SALL1 upregulation was ever
170 greater at protein level (Fig. 2F), indicating a potential compensation in pluripotent cells by post-
171 transcriptional regulation.
172 Importantly, *SALL4*-het-KO iPSCs remained viable and exhibited hallmarks iPSC morphology
173 (Fig. S2D). Furthermore, these cells showed no significant difference in growth rate compared to
174 controls (Fig. 2G).
175
176 **Decreased SALL4 expression does not affect the stem cell identity of the iPSCs**
177 Given the debated role of SALL4 in maintaining pluripotency, we investigated whether *SALL4*-
178 het-KO impacted iPSC stem cell identity. Protein levels of the pluripotency transcription factors
179 NANOG and OCT4 remained unaltered in the *SALL4*-het-KO cells relative to controls (Fig. 3A).
180 Similarly, the levels of the pluripotency surface markers TRA-1-60-R and SSEA4 were also not
181 significantly affected by decreased SALL4 (Fig. 3B; Fig. S4A). Moreover, we examined the
182 transcriptome of the iPSCs and identified only 114 differentially expressed genes in *SALL4*-het-
183 KO cells relative to controls (FDR < 0.05; Fig. 3C and Table 3). Notably, canonical pluripotency
184 markers, such as *POU5F1, NANOG, SOX2, KLF4,* and *MYC*, were not significantly differentially
185 expressed (Fig. 3C and Table S3). Consistent with this, gene ontology analysis confirmed that
186 neither up- (Fig. S3A) or downregulated (Fig. S3B) genes were enriched for pluripotency-
187 associated terms. Finally, by performing tri-lineage differentiation we confirmed that SALL4
188 depletion did not impact the ability of the *SALL4*-het-KO cells to successfully differentiate into all
189 three of the germ layers (Fig. S3C).
190 Together, these findings suggest that *SALL4* haploinsufficiency did not affect IPSC stemness
191 and self-renewal, possibly as a result of SALL1-mediated compensation.
192
193 **SALL4 relocates to CNCC developmental enhancers upon CNCC induction cues**
194 After establishing that *SALL4*-het-KO did not affect iPSC stemness, we differentiated the iPSCs
195 into CNCCs using the same aforementioned protocol (Bajpai et al., 2010; Barnada et al., 2024;
196 Mitchell et al., 2025; Pagliaroli et al., 2021; Prescott et al., 2015) (Fig. 4A). Since the SALL4-BAF
197 interaction was detected at Day-5 (Pagliaroli et al., 2021), corresponding to neuroectodermal
198 spheres, we performed CUT&RUN to profile SALL4 binding in *SALL4*-WT cells in iPSCs and at
199 Day-5 of differentiation. We performed two SALL4 CUT&RUN replicates per timepoint, and only
200 retained replicated peaks in each timepoint (FDR < 0.05). With this approach, we identified
201 comparable amounts of SALL4-bound genomic sites at Day-0 and Day-5 (29,088 and 32,707,
202 replicated peaks, respectively, FDR < 0.05; Fig. 4B). Further analysis of these sites indicated that

CONFIDENTIAL

203 approximately 20% of the Day-0 SALL4 peaks were unique to iPSCs and not detected at Day-5
204 (hereafter SALL4 Day-0 specific peaks; n = 6,114; Fig. 4B, 4C, left). On the other hand, ~50% of
205 the replicated peaks detected at Day-5 were exclusive of this developmental stage and were not
206 detected in iPSCs (hereafter SALL4 Day-5-specific peaks, n = 17,137; Fig. 4B and 4C). Finally, a
207 subset of ~15,000 sites remained bound by SALL4 at both time points.
208 These data indicate that, upon neural crest differentiation cues, there is a major shift in SALL4
209 binding, resulting in the relocation of this transcription factor at ~17,000 previously unbound
210 regions.
211 To characterize the scope of SALL4 relocation, we analyzed the positional distribution of SALL4
212 peaks by associating each peak with its nearest transcription start site (TSS), identifying 4,362
213 genes at Day-0 and 6,861 genes at Day-5 (Table S4). Consistent with SALL4's established role
214 as an enhancer-binding transcription factor, 87% of SALL4-bound regions at Day-0 were
215 located >1 kb from the nearest TSS, suggesting these represent putative enhancers. This
216 proportion increased substantially to 97% at Day-5 (Fig. 4D).
217 GO terms enriched in the “Day-5 specific” genes included pathways associated with neural crest
218 development, such as neural tube development, face development, and cardiac chamber
219 development (Fig. 4E; FDR < 0.05).
220 To further characterise the SALL4 bound elements at Day-0 and Day-5, we performed sequence-
221 based computational motif analysis. At Day-0, motifs associated with core pluripotency
222 transcription factors such as OCT4, NANOG, and SOX2 did not rank among the top 20 enriched
223 sequences, further supporting that SALL4 may not be essential for maintaining pluripotency in
224 human cells (Fig. S5A).
225 On the other hand, in addition to general enhancer activator factors (AP-1, TEAD), the Day-5
226 specific SALL4 peaks were enriched for many CNCC signature factors, including SOX9, TFAP2A,
227 TWIST1/2, SNAI1, NR2F1/2, TCF4/12 and several others (Fig. 4F). Moreover, neuroectodermal
228 transcription factors, such as ZIC2/3 were also enriched. Additionally, *de novo* motif analysis in
229 iPSCs revealed enrichment for the CTCF-like motifs (Fig. S5B), while at Day-5 there was
230 enrichment for enhancer activators TEADS/ETS and AP-1 (Fig. S5C). Finally, AT-rich motifs
231 resembling the SALL4 annotated motif were found only in a small minority of peaks in iPSCs (Fig.
232 S5B). This might suggest that SALL4 binding at the CNCC enhancers is not dictated by
233 recognition of its cognate motif, and possibly mediated by other lineage-specific transcription
234 factors.
235 In summary, our data so far indicate that SALL4 relocates to CNCC developmental enhancers
236 upon differentiation cues.

CONFIDENTIAL

237

238 **Chromatin accessibility at the CNCC enhancers is SALL4 dependent**

239 Given that we demonstrated that SALL4 interacts with BAF at early stages of CNCC specification,
240 we reasoned that SALL4 binding at the CNCC developmental enhancers could be necessary for
241 their chromatin activation. We profiled chromatin accessibility by ATAC-seq at Day-0 and Day-5
242 in both *SALL4*-WT and *SALL4*-het-KO conditions (two biological replicates per timepoint per
243 condition, only peaks present in both replicates were considered; FDR < 0.05). The number of
244 ATAC-seq replicated peaks identified in *SALL4*-WT condition was comparable between Day-0
245 (61,660 peaks; FDR < 0.05) and Day-5 (58,375 peaks; FDR < 0.05; Fig. 5A). Next, we specifically
246 investigated the chromatin state of the 17,137 genomic sites that gained SALL4 binding at Day-
247 5, and found that SALL4 binding at these sites correlates with a marked increase in chromatin
248 accessibility compared to Day-0 (Fig. 5B). Strikingly, the accessibility at these regions was
249 significantly attenuated in *SALL4*-het-KO relative to *SALL4*-WT (Fig. 5C). This suggests that, in
250 response to differentiation signals, SALL4 relocates to CNCC developmental enhancers, which
251 subsequently become activated through chromatin opening, which is SALL4 dependent.

252 Motif analysis of regions that lost accessibility in *SALL4*-het-KO cells (n = 1,649; Fig. 5A) showed
253 again enrichment for CNCC factors such as ZIC2, ZEB1/2, and SNAIL, along with BMP signalling
254 factors such as SMADs, and enhancer activators such as AP-1 (Fig. 5D). On the other hand,
255 regions aberrantly accessible in the *SALL4*-het-KO cells (n = 3,460; Fig. 5A) showed enrichment
256 for pluripotency factors, CTCF, and neural plate markers (Fig. 5E).

257

258 **SALL4 recruits the BAF complex at CNCC developmental enhancers to increase chromatin**
259 **accessibility**

260 Since SALL4 binding at the CNCC enhancers is necessary for chromatin opening, we investigated
261 whether the SALL4-bound developmental enhancers were also occupied by BAF, and if BAF
262 binding at these sites could be SALL4 dependent. We performed ChIP-seq for BRG1
263 (SMARCA4), which is the only catalytic BAF subunit active at this sage, in *SALL4*-WT cells at
264 Day-5. We found that BRG1 consistently binds the same genomic sites as SALL4, supporting the
265 idea that these factors interact and co-occupy regulatory elements during early differentiation (Fig.
266 6A).

267 To assess whether BRG1 binding depends on SALL4, we performed ChIP-seq for BRG1 in both
268 *SALL4*-WT and *SALL4*-het-KO cells at Day-5. In *SALL4*-WT, we detected 32,833 replicated
269 BRG1 peaks (FDR < 0.05), as opposed to only 14,116 peaks in *SALL4*-het-KO. This suggests
270 that SALL4 loss triggered widespread loss of BAF from chromatin. We focused on BRG1 peaks

CONFIDENTIAL

271 lost in *SALL4*-het-KO cells and found that these sites were largely not bound by SALL4 at Day-0,
272 but they gained SALL4 binding at Day-5 (Figs. 6B–D). As expected, loss of SALL4 and BRG1
273 from these sites led to marked decrease of chromatin accessibility (Fig. 6E). Moreover, we
274 explored the identity of some of these putative enhacers by confirming that these overlapped with
275 some of those reported on databases such as VISTA Enhancer Browser (Kosicki et al., 2025) and
276 GeneHancer (Fishilevich et al., 2017; Fig. 6F). Overall, these experiments support that SALL4 is
277 required to recruit BAF at CNCC enhancers.
278 To further confirm that these regions are enhancers, we leveraged publicly available H3K4me1
279 and H3K4me3 ChIP-seq datasets generated using a comparable protocol at our Day-5 equivalent
280 (Prescott et al., 2015; Rada-Iglesias et al., 2011). Moreover, we performed ChIP-seq for H3K27ac
281 at Day-5, Day-11 and Day-14, roughly corresponding to CNCC-induction, early CNCC, late CNCC
282 stages. This analysis revealed that nearly all the sites bound by SALL4 and BRG1 are marked by
283 the enhancer mark H3K4me1, and only a small subset of them has the promoter mark H3K4me3
284 (Fig. S6A). Moreover, nearly all these sites have the active enhancer mark H3K27ac at Day-5,
285 and the acetylation levels progressively increase from CNCC induction to CNCC stage,
286 supporting that these are *bona fide* CNCC enhancers that increase in activation after SALL4
287 binding (Figs. S6B, C)
288 While a large portion of BRG1 peaks was lost in the absence of SALL4, we also identified a subset
289 of novel BRG1 peaks that were absent in *SALL4*-WT cells but gained in the *SALL4*-het-KO cells.
290 These BRG1 peaks gained in the *SALL4*-het-KO cells were associated with genes linked to
291 neurogenesis and nervous system development (Fig. S7A and Table S4). In contrast, the genes
292 located near BRG1 peaks lost in the *SALL4*-het-KO cells were involved in embryonic organ
293 development, epithelial differentiation, and cell migration processes (Fig. S7B and Table S4).
294 Together, these results suggest that SALL4 is a regulator of BAF recruitment to CNCC enhancers,
295 ensuring their accessibility during differentiation. In the absence of SALL4, BRG1 fails to localize
296 to these sites and is partially redirected to neurodevelopmental enhancers, leading to a
297 misallocation of chromatin remodeling, and to a shift towards default neural differentiation at the
298 expense of CNCC induction and specification.
299

300 ***SALL4*-het-KO cells are not able to exit the neural ectoderm stage and to induce the cranial**
301 **neural crest fate**

302 Our data thus far indicate that SALL4 relocates to CNCC enhancers as early as the
303 neuroectodermal stage, recruiting BAF to these regulatory elements for chromatin activation.
304 Next, we investigated the functional consequences of SALL4 loss by generating CNCCs using

CONFIDENTIAL

305 *SALL4*-WT and *SALL4*-het-KO cells. As expected, *SALL4*-WT CNCCs expressed the signature
306 CNCC markers at both the mRNA (Fig. 7A) and protein (Figs. 7B and S4B) level, while *SALL4*-
307 het-KO failed to induce these genes (Fig. 7A). This was further supported by RNA-seq conducted
308 at Day-14, which identified 1,106 genes differentially expressed in *SALL4*-het-KO CNCCs relative
309 to *SALL4*-WT (FDR < 0.05; Fig. 7C and Table S3). This was in sharp contrast with the small
310 number of genes differentially expressed at the iPSC stage (n = 114; Fig. 3C and Table S3) and
311 at Day-5 (n = 170; Fig. S7C; Table S3).
312 Of the 1,106 genes differentially expressed at Day-14, ~60% (n = 659) were downregulated while
313 the remaining were upregulated (Fig. 7C). Canonical CNCC markers such as *TFAP2A*, *NR2F1*,
314 and *SOX9* were significantly downregulated in *SALL4*-het-KO cells. Additionally, genes
315 associated with neural plate border, early neural crest induction and epithelial-to-mesenchymal
316 transition (EMT) were also downregulated (Fig. 7D, F). In line with this, gene ontology analysis
317 revealed that pathways associated with face development, autonomic nervous system,
318 gastrulation, and regulation of morphogenesis were all downregulated (Fig. 7D). The 447
319 upregulated genes were instead involved in cell regulatory pathways (Fig. 7E).
320 As aforementioned, genes associated with EMT were dysregulated. The process of EMT requires
321 a switch in cadherin expression where class-1 cadherins must be downregulated, while class-2
322 cadherins must be upregulated. Related to this, our RNA-seq data revealed that this cadherin
323 switch did not occur in the differentiating *SALL4*-het-KO iPSCs, suggesting that the cells remained
324 in an epithelial state. This was further corroborated by the upregulation of several epithelial genes
325 in the differentiating *SALL4*-het-KO cells, including *EpCAM*, *CDH1*, and *OVOL2* (Fig. 7F). The
326 latter is involved in maintaining epithelial lineages by repressing the expression of the EMT gene
327 *ZEB1*, which was downregulated in the *SALL4*-het-KO condition (Haensel et al., 2019; Liu et al.,
328 2018a).
329 In summary, our data indicate that CNCC specification failed in *SALL4*-het-KO conditions, as
330 reflected by the large number of differentially expressed genes. On the other hand, the
331 transcriptomes of iPSCs and neuroectodermal cells (i.e. Day-5) were only modestly affected by
332 SALL4 loss, suggesting that SALL4 may not be essential for neuroectoderm formation, while it is
333 essential for CNCC induction and specification.
334 To investigate this further, we systematically examined the expression of a panel of stage-specific
335 marker genes at Day-0, Day-5 and Day-14 (Fig. S8A–D) including core pluripotency genes
336 (*POU5F1* and *NANOG*), neural plate border specifiers (*ZIC1, GBX2, MSX1/2, PAX3/7, GATA2/3*),
337 CNCC specification markers (*SNAI1/2, SOX9, TFAP2A*), and post-migratory CNCC markers
338 (*SOX5, SOX6, SOX10*). In *SALL4*-WT cells, pluripotency markers exhibited a progressive and

CONFIDENTIAL

339 expected downregulation as differentiation advanced from Day-0 to Day-14. In *SALL4*-het-KO
340 cells, a similar decline in pluripotency markers was observed between Day-0 and Day-5, but
341 instead of continuing to decrease, these markers remained similarly expressed through Day-14
342 (Fig. S8A). Despite this, the two conditions showed similar trends, confirming that decreased
343 levels of SALL4 did not significantly impact pluripotency markers expression during exit from
344 pluripotency. Similarly, no significant difference was observed in the expression of neural plate
345 border specifiers (Fig. S8B), suggesting that SALL4 is not required for human neuroectoderm
346 specification. Conversely, CNCC specification markers and post-migratory CNCC markers all
347 increased in expression after Day-5 in *SALL4*-WT cells, while the SALL4-depleted cells failed to
348 maintain a comparable increase in expression (Fig. S8C, D).
349 In summary, *SALL4*-het-KO cells are competent to generate neuroectoderm-like cells but fail to
350 induce the neural crest lineage.
351
352 **SALL4A but not SALL4B is necessary for CNCC specification**
353 Given previous reports suggesting distinct roles for the two SALL4 isoforms in ESCs (Rao et al.,
354 2010), and our prediction that the interaction between SALL4 and DPF2 may be isoform-specific,
355 we investigated whether CNCC differentiation exhibits preferential dependence on a particular
356 SALL4 isoform.
357 Time-course mRNA expression in *SALL4*-WT cells revealed that *SALL4B* is consistently lowly
358 expressed throughout differentiation, whereas *SALL4A* expression increases upon differentiation
359 cues, peaking around Day-5 and then declining over time (Figs. S9A, B). On the other hand, the
360 paralog SALL1 is expressed at low levels in iPSCs and barely detectable throughout the entire
361 differentiation protocol (Fig. S9C).
362 To investigate potential isoform-specific requirements for CNCC development, we selectively
363 degraded SALL4A while preserving SALL4B expression. To achieve this, we treated cells with
364 pomalidomide, a thalidomide derivative that selectively degrades SALL4A but not SALL4B (Anh
365 Vu et al., 2023; Liu et al., 2024; Fig. 8A).
366 To test the effects of SALL4A degradation, we differentiated *SALL4*-WT iPSCs into CNCCs. Cells
367 were treated with either DMSO (control) or 5 μM pomalidomide (5 μM POM) every 24 hours (Fig.
368 8B). To confirm consistent degradation of SALL4A throughout the differentiation process, we
369 performed immunoblotting at key time points (Day-5 and Day-14). As expected, SALL4A was
370 successfully degraded, while SALL4B expression remained intact (Fig. 8C). Importantly,
371 pomalidomide degrades several proteins beyond SALL4, and its protein degradation activity is
372 typically cell-type specific (Sievers et al., 2018). Two of the known pomalidomide targets have

CONFIDENTIAL

373 suggested functions in neural crest development. These are ARID2 (Bajpai et al., 2010;
374 Yamamoto et al., 2020) and ZMYM2 (Jourdeuil et al., 2025; Renneville et al., 2021). Crucially,
375 immunoblot analysis for these two proteins performed at Day 5 following treatment with 5 μM
376 pomalidomide revealed that the levels of both ARID2 and ZMYM2 remained unchanged
377 compared to the DMSO control (Figs. S9D, E), further supporting that pomalidomide affinity for
378 its targets is cell-type specific, as previously suggested (Bajpai et al., 2010; Yamamoto et al.,
379 2020). This control experiments allowed to rule out potential off-target effects caused by unwanted
380 degradation of one of these two proteins. The other known pomalidomide targets, including
381 IKAROS, AIOLOS, PLZF and ZNFX1 have no established contribution to neural crest
382 specification and in fact they are not highly expressed in our system.
383 Next, we assessed the downstream effects of SALL4A degradation by performing RT-qPCR and
384 immunofluorescence for key neural crest markers, such as *TFAP2A*, *NR2F1*, and *SOX9* (Figs.
385 8D, E), and EMT markers VIM and EpCAM (Fig. 8F). Degradation of SALL4A recapitulated what
386 we previously observed in *SALL4*-het-KO cells. Specifically, the cells treated with pomalidomide
387 failed to differentiate successfully into CNCCs despite the persistent expression of SALL4B.
388 Taken together, our results highlight a specific and essential role for SALL4A in orchestrating
389 chromatin dynamics during CNCC specification and suggest that SALL4B might function in a
390 different regulatory capacity rather than acting as a compensatory factor for SALL4A.

391

392

393 **DISCUSSION**

394 Mutations in *SALL4* are associated with congenital syndromes which frequently include distinct
395 craniofacial anomalies(Kohlhase et al., 2005a; Kohlhase et al., 2005b; Miller, 1991), suggesting
396 disrupted CNCC development.
397 Notably, at the murine embryonic stage E9.5, Sall4 expression persists within the frontonasal
398 prominence (Tahara et al., 2019), implying a potential role for SALL4 in craniofacial development.
399 Given the limited existing knowledge about SALL4 function in this context, we utilized a
400 heterozygous *SALL4* knockout hiPSC line to investigate its role CNCC specification.
401 At early stages of CNCC induction, SALL4 relocates to enhancers of genes involved in CNCC
402 formation, and SALL4 binding at these regulatory elements is required for their activation. In
403 mESCs, approximately 10% of the total Sall4 protein, and primarily the Sall4A isoform, is
404 incorporated into the NuRD complex (Bode et al., 2016; Miller et al., 2016). Here, we identify and
405 characterize a novel interaction between SALL4 and the BAF complex, predicted to be mediated
406 specifically through the SALL4A isoform and the BAF subunit DPF2. Interestingly, DPF2 and

CONFIDENTIAL

407 SALL4 exhibit notable functional parallels. Similar to SALL4, DPF2 occupancy undergoes
408 significant repositioning during the transition from pluripotency toward early neural commitment
409 in mouse embryoid bodies (Zhang et al., 2019b). Although the role of DPF2 in CNCCs has not
410 been directly investigated, its early recruitment to neuronal enhancers suggests a potential
411 involvement in neuroectodermal differentiation and lineage commitment of neuroectoderm-
412 derived cells, including CNCCs. Moreover, genomic sequences bound by DPF2 in mouse
413 neuroectoderm are AT-rich, resembling the preferred binding motif of SALL4 (Kong et al., 2021;
414 Nazim et al., 2024; Pantier et al., 2021). It is important to note that AlphaFold3 is built on deep
415 learning rather than reinforcement learning. Consequently, the absence of more comprehensive
416 experimental SALL4 data could significantly affect the reliability of the proposed SALL4–DPF2
417 interaction model. Additionally, AlphaFold 5,000-token limit per prediction poses a constraint
418 when modeling large multi-subunit complexes such as BAF. Fortunately, BAF has been
419 extensively characterized, and our predicted model match experimentally resolved structures at
420 3 Å resolution.
421 Our study demonstrated that while CNCC induction fails without SALL4, early neuroectodermal
422 commitment and neural plate border identity remain unaffected. Our findings demonstrate that,
423 during the earliest stages of CNCC specification, SALL4 occupies CNCC-associated enhancers
424 and recruits BAF. Together, these observations support previous studies that proposed that
425 neural crest cell fate commitment is initiated well before overt gene expression changes occur
426 (Basch et al., 2006; Patthey et al., 2008; Stuhlmiller and García-Castro, 2012). This finding is
427 further supported by clinical observations. In fact, while phenotypes arising from *SALL4*
428 haploinsufficiency affect many neural crest derivatives, phenotypes implicating defects in
429 neurodevelopment are exceptionally rare (Borozdin et al., 2004; Ma et al., 2022) and usually
430 associated with dysregulation of potassium channels (Kodytková et al., 2023). In the rare cases
431 of intellectual disability reported, such phenotypes were attributed to broader deletions on
432 chromosome 20q13.13-q13.2 rather than *SALL4* loss alone (Borozdin et al., 2004).
433 In the absence of SALL4, after two weeks of differentiation the cells acquire an aberrant fate,
434 expressing a mix of pluripotency genes (suggesting lack of cell differentiation), and neuronal
435 genes (suggesting derailed cell fate). This indicates that most cells safely reach the
436 neuroectodermal stage but fail to go beyond that, while some of the cells simply do not
437 differentiate or regress back to pluripotency.
438 Interestingly, only a small minority of the CNCC enhancers bound by SALL4 has the AT-rich motif
439 that is thought to be recognized by SALL4 (Kong et al., 2021; Pantier et al., 2021). This suggests
440 that SALL4 might bind these enhancers in cooperation with other lineage-specific factors. SALL1

CONFIDENTIAL

441 also recognizes AT-rich motifs, but similar to SALL4 it also binds unspecifically. Since SALL1 is
442 upregulated in the *SALL4*-het-KO cells, it might possible that some of the sites associated with a
443 gain of BRG1 upon loss of SALL4 could be associated with SALL1 being upregulated and
444 hijacking BAF recruitment to distinct sites.
445
446 *Limitations*
447 SALL4 was recently identified as a direct target of thalidomide (Donovan KA, 2018), a drug widely
448 prescribed to pregnant women in the late 1950s to alleviate morning sickness. The subsequent
449 alarming rise in births of children with congenital anomalies, including limb malformations, facial
450 abnormalities such as inner ear defects, ocular motility issues, and facial nerve palsy, as well as
451 kidney and heart defects, led to the identification of thalidomide as the causative agent. However,
452 the molecular mechanism by which thalidomide induced these developmental defects remained
453 elusive until nearly six decades later. Building on earlier findings that thalidomide and its
454 derivatives bind to cereblon (CRBN), Matyskiela et al. demonstrated that SALL4 is recruited to
455 the CRL4–CRBN E3 ubiquitin ligase complex following thalidomide-induced ubiquitination,
456 resulting in its proteasomal degradation (Matyskiela et al., 2018). More recently, research
457 demonstrated that only the SALL4A isoform, but not SALL4B, is susceptible to degradation by
458 immunomodulatory imide drugs such as pomalidomide (Anh Vu et al., 2023). Our data suggests
459 that SALL4 role in CNCC specification might be isoform-specific and these findings might
460 contribute to shed light on the birth malformations caused by thalidomide. However, while we
461 ensured that known pomalidomide targets important for CNCC developments were not degraded
462 in our system, we cannot rule out degradation of previously uncharacterized and unknown
463 pomalidomide targets with key roles in CNCC specification. While highly unlikely, this is an
464 important limitation of our study, and therefore the findings of our pomalidomide experiments,
465 should be taken with caution.
466
467 **ACKNOWLEDGMENTS**
468 We thank the Facility for Imaging by Light Microscopy (FILM) at Imperial College London for
469 providing access to instrumentation and technical support, which is part-supported by funding
470 from the Wellcome Trust (grant 104931/Z/14/Z) and BBSRC (grant BB/L015129/1). We thank
471 Prof. Brian Hendrich (University of Cambridge) and Dr. Andria Koulle (Imperial College London)
472 for insightful discussion of the data and of early versions of the manuscript.
473

CONFIDENTIAL

FUNDING

For this work, MT was funded by the Mathers Foundation and by Biotechnology and Biological Sciences Research Council (BBSRC, grant BB/Y000854/1).

COMPETING INTERESTS

None declared.

DATA AND RESOURCE AVAILABILITY

RNA-seq, ATAC-seq, CUT&RUN-seq, and ChIP-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession code GEO: GSE293821 and are publicly available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE293821.

AUTHOR CONTRIBUTIONS

MT designed the project. MD performed most of the experiments. SMB, EVD, ZHM, and LD contributed to some of the experiments. MD and MT analyzed the data and wrote the manuscript. All the authors read and approved the manuscript.

FIGURE LEGENDS

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

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

(B) AlphaFold structure model showing the predicted interactions between SALL4A (red), DPF2 (dark blue), and other BAF components, including SMARCA4, SMARCB1, and ACTL6A.

(C) SALL4A protein sequence and visual representation of output model of SALL4A (model 0). C2H2 zinc-fingers annotations are highlighted in dark purple and numbered. Interactions between SALL4A and DPF2 through the ZFs 4, 5 and 6 (ZFC3) and 7 (ZFC4) are highlighted in light purple. Made with BioRender.com.

(D) Zoomed view of SALL4 (pink) and DPF2 (purple) interacting surfaces. Highlighted are the functional domains involved in the predicted interaction. In red, contacts predicted by AlphaFold(Jumper et al., 2021).

(E) Box plot displaying expression (TPMs) of the four members of the DPF protein family in iPSCs (n = 2). Differences between conditions were assessed using unpaired student's t-test.

CONFIDENTIAL

506 (F) Box plot displaying expression (TPMs) of the four members of the DPF protein at Day-5 (n =
507 3). Differences between conditions were assessed using unpaired student's t-test.

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

509 (A) Visual representation of the CRISPR/Cas9 approach used on for *SALL4*-het-KO. The star
510 represents the exact location of the deletion. Sanger sequencing confirming the 11-bp deletion in
511 one allele of *SALL4* is displayed (clone F1). Made with BioRender.com.
512 (B) Immunoblot for SALL4 in *SALL4*-WT and *SALL4*-het-KO (clone F1) iPSC lines.
513 (C) Bar plots showing densitometry analysis of SALL4A and SALL4B expression in WT and
514 *SALL4*-het-KO iPSC lines from immunoblot assay.
515 (D) Bar plots showing RNA expression in TPMs for SALL4A and SALL4B in WT and *SALL4*-het-
516 KO iPSCs. Individual biological replicates (n = 2) are displayed as circles. Statistical significance
517 was assessed using unpaired student's t-test. Error bars display standard error.
518 (E) Bar plots displaying expression TPMs of all *SALL* paralogues in *SALL4*-WT and *SALL4*-het-
519 KO iPSCs. Individual biological replicates (n = 2) are displayed as circles. Statistical significance
520 was assessed using unpaired student's t-test. Error bars display standard error of the mean.
521 (F) Immunoblot for SALL1 in *SALL4*-WT and *SALL4*-het-KO (clone F1) iPSC lines.
522 (G) Growth curve comparing *SALL4*-WT with *SALL4*-het-KO iPSC lines. N = 2 biologically
523 independent experiments. Error bars display standard error.
524

525 **Figure 3. Stem cell identity is unaffected by SALL4 loss**

526 (A) (Left) Immunofluorescence labeling quantifying the expression of the canonical pluripotency
527 markers OCT4 and NANOG in *SALL4*-WT and *SALL4*-het-KO (clone F1) iPSCs. Scale bar: 100
528 μm. (Right) Corrected Total Cell Fluorescence (CTCF) values were used to quantify and compare
529 markers expression (OCT4 n=34; NANOG n=37). Error bars display standard error.
530 (B) Flow cytometry quantifying the expression of surface markers for pluripotency (TRA-60-1-R
531 and SSEA-4).
532 (C) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in
533 iPSCs (Day-0). Blue dots (n = 50) represent downregulated genes with p-adj < 0.05 and
534 log2FoldChange < -1.5. Purple dots (n = 63) represent upregulated genes with p-adj < 0.05 and
535 log2FoldChange > 1.5.
536

537 **Figure 4. SALL4 relocates to CNCC enhancers to induce chromatin accessibility**

CONFIDENTIAL

538 (A) (Bottom) Schematic representation of the iPSC-to-CNCC differentiation protocol. The
539 experimental timeline shows in vitro differentiation of iPSCs into CNCCs via a neuroectodermal-
540 like stage. Immunofluorescence labeling quantifying the expression of the canonical stage-
541 specific markers (iPSCs: OCT4; neuroectoderm: OTX2; CNCCs: SOX9) showing ratio of cells
542 successfully undergoing stage-specific specifications during CNCC generation. Scale bar: 100
543 μm.
544 (B) Diagram showing the number of regions bound by SALL4 uniquely at Day-0 ("Day-0 specific";
545 n = 6,114), at Day-5 ("Day-5 specific"; n = 17,137), and regions that maintain SALL4 binding at
546 both time points ("shared", n = 15,570).
547 (C) Heatmaps showing the distribution of SALL4 binding at Day-0 and Day-5 across Day-0
548 specific (left) and Day-5 specific (right) binding regions. CUT&RUN signal intensity was plotted
549 across ±2 kb regions centered on peak summits. Rows represent individual peaks ranked by
550 signal intensity.
551 (D) Genomic distribution of SALL4 binding sites relative to the nearest TSS. Peaks are
552 categorized by distance into six bins: <1 kb, 1–3 kb, 3–5 kb, 5–10 kb, 10–100 kb, and >100 kb.
553 Peaks located at >1kb from the nearest TSS were classified as putative enhancers. Percentages
554 of peaks for each bin are represented by bars to the left (downstream) or right (upstream) of the
555 TSS.
556 (E) Gene ontology analysis of downregulated nearest genes to Day-5 specific SALL4-bound
557 regions as determined using WebGestalt (Liao et al., 2019) over-representation pathway analysis,
558 using affinity propagation as parameter for redundancy removal (FDR < 0.05).
559 (F) Table of motifs enriched in Day-5 specific SALL4-bound regions as determined using HOMER
560 (Heinz et al., 2010).
561
562 **Figure 5. Chromatin accessibility at the CNCC enhancers is SALL4 dependent**
563 (A) Venn diagram showing chromatin accessibility in *SALL4*-WT (ATAC-seq peaks n = 58,375)
564 and *SALL4*-het-KO (n = 60,186). It also depicts the number of regions uniquely accessible in
565 *SALL4*-WT ("Sites losing accessibility in KO"; n = 1,649), in *SALL4*-het-KO ("Sites gaining
566 accessibility in KO"; n = 3,460), and regions that remain accessible in both conditions ("Sites
567 maintaining accessibility in WT and KO", n = 60,186).
568 (B) Heatmaps showing ATAC-seq signals at Day-0 (left) and Day-5 (right), at Day-5 specific
569 SALL4 CUT&RUN peaks. Signal intensity was plotted across ±2 kb regions centered on peak
570 summits. Rows represent individual peaks ranked by signal intensity.

CONFIDENTIAL

571 (C) Average profiles showing ATAC-seq signal at Day 5-specific SALL4 CUT&RUN peaks (n =
572 17,137) in *SALL4*-WT and *SALL4*-het-KO cells in two distinct biological replicates.
573 (D) Table of motifs enriched in regions gaining or maintaining accessibility at Day-5 in *SALL4*-het-
574 KO regions as determined using HOMER (Heinz et al., 2010).
575 (E) Table of motifs enriched in regions losing accessibility at Day-5 in *SALL4*-het-KO regions as
576 determined using HOMER (Heinz et al., 2010).
577
578 **Figure 6. SALL4 recruits BAF to CNCC enhancers.**
579 (A) Heatmap showing BRG1 ChIP-seq signal at the genomic sites bound by SALL4 in WT cells
580 at Day-5 (n = 17,137). Signal intensity was plotted across ± 2 kb regions centered on peak
581 summits. Rows represent individual peaks ranked by signal intensity.
582 (B) Venn diagram showing peaks bound by SALL4 uniquely at Day-5 and that lose accessibility
583 in *SALL4*-het-KO.
584 (C) Heatmap and average profiles show BRG1 ChIP-seq signal at Day-5 at the BRG1 peaks lost
585 in *SALL4*-het-KO (clone F1; n = 13,505). Signal intensity was plotted across ±2 kb regions
586 centered on peak summits. Rows represent individual peaks ranked by signal intensity.
587 (D) Heatmap and average profiles show SALL4 CUT&RUN signal at the genomic sites that lose
588 BRG1 in *SALL4*-het-KO cells at Day-0 and Day-5 (n = 13,505).
589 (E) Heatmaps and average profile show ATAC-seq signal at the genomic sites that lose BRG1 in
590 *SALL4*-het-KO cells (clone F1; n = 13,505). Signal intensity was plotted across ±2 kb regions
591 centered on peak summits. Rows represent individual peaks ranked by signal intensity.
592 (F) Two representative examples of enhancers bound by SALL4 at Day 5 that lose BRG1 co-
593 occupancy and chromatin accessibility in the *SALL4*-het-KO (clone F1) cells. Both correspond to
594 ChIP-seq peaks annotated as enhancer regions, as reported in GeneHancer (left) and VISTA
595 Enhancer Browser (right).
596
597 **Figure 7. CNCC induction and specification are impaired in *SALL4*-het-KO.**
598 (A) Bar plot displaying relative mRNA expression of canonical CNCC markers in *SALL4*-WT and
599 *SALL4*-het-KO cells at Day-14 (n = 2). Statistical significance was assessed using unpaired
600 student's t-test. 18S rRNA was used to normalize.
601 (B) Flow cytometry for the CNCC surface marker CD99 in *SALL4*-WT and *SALL4*-het-KO (clone
602 F1) CNCCs at Day-14.
603 (C) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in
604 CNCCs (Day-14). Blue dots (n= 659) represent downregulated genes with p-adj < 0.05 and

CONFIDENTIAL

605 log2FoldChange < -1.5. Purple dots (n= 447) represent upregulated genes with p-adj < 0.05 and
606 log2FoldChange > 1.5.
607 (D) and (E) GO term enrichment for upregulated (D) and downregulated (E) genes in *SALL4*-het-
608 KO cells (FDR < 0.05).
609 (F) Heatmap displaying the expression of canonical epithelial-to-mesenchymal transition (EMT)
610 markers in *SALL4*-WT and *SALL4*-het-KO cells (Day-14).
611
612 **Figure 8. SALL4A is the only SALL4 isoform necessary for CNCC specification**

613 (A) Schematic of pomalidomide-mediated ubiquitin ligation. Ub, ubiquitin. POM, pomalidomide.
614 CRBN, DDB1, Cul4 and Rbx1 are subunits of CRL4/CRBN E3 ubiquitin ligase.
615 (B) Schematic representation of pomalidomide treatment during iPSC-to-CNCC differentiation
616 protocol.
617 (C) Time-course immunoblot for SALL4 in *SALL4*-WT at Day-5 and Day-14 after daily treatment
618 with DMSO (control) or 5μM pomalidomide (5μM POM).
619 (D) Box plot displaying the relative mRNA expression of canonical CNCC markers (SOX9,
620 TFAP2A, and NR2F1) in *SALL4*-WT at Day-14 treated with 5μM pomalidomide or DMSO (control)
621 daily for 14 days. Individual biological replicates (n = 2) are displayed as circles. Statistical
622 significance was assessed using unpaired student's t-test. Error bars display standard error.
623 (E) Immunofluorescence labeling quantifying the expression of the canonical CNCC markers
624 SOX9 and TFAP2A in *SALL4*-WT treated with 5μM pomalidomide or DMSO (control) daily for 14
625 days. Scale bar: 100 μm.
626 (F) Immunofluorescence labeling quantifying the expression of the canonical EMT markers
627 EpCAM and VIM in *SALL4*-WT treated with 5uM pomalidomide or DMSO (control) daily for 14
628 days. Scale bar: 100 μm.
629
630 **REFERENCES**

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

634 **Anh Vu, K. L., Kumari, K., Hui Liu, B., Gao, C., Li, F., Ping Tang, J., Maddalo, D., Auld, D. S., Casalena, D.**
635 **E., Tian, X., et al.** (2023). SALL4B, not targeted by IMiD, is important for SALL4-mediated
636 tumorigenesis.

CONFIDENTIAL

637 **Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C. P., Zhao, Y., Swigut, T.**
638 **and Wysocka, J.** (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation.
639 *Nature* **463**, 958–962.

640 **Barnada, S. M., Giner de Gracia, A., Morenilla-Palao, C., López-Cascales, M. T., Scopa, C., Waltrich, F.**
641 **J., Mikkers, H. M. M., Cicardi, M. E., Karlin, J., Trotti, D., et al.** (2024). ARID1A-BAF coordinates
642 ZIC2 genomic occupancy for epithelial-to-mesenchymal transition in cranial neural crest
643 specification. *The American Journal of Human Genetics*.

644 **Basch, M. L., Bronner-Fraser, M. and García-Castro, M. I.** (2006). Specification of the neural crest occurs
645 during gastrulation and requires Pax7. *Nature* **441**, 218–222.

646 **Bode, D., Yu, L., Tate, P., Pardo, M. and Choudhary, J.** (2016). Characterization of two distinct
647 nucleosome remodeling and deacetylase (NuRD) complex assemblies in embryonic stem cells.
648 *Molecular and Cellular Proteomics* **15**, 878–891.

649 **Borozdin, W., Boehm, D., Leipoldt, M., Wilhelm, C., Reardon, W., Clayton-Smith, J., Becker, K.,**
650 **Mühlendyck, H., Winter, R., Giray, O., et al.** (2004). SALL4 deletions are a common cause of
651 Okihiro and acro-renal-ocular syndromes and confirm haploinsufficiency as the pathogenic
652 mechanism. *J Med Genet* **41**,.

653 **Chen, K. Q., Anderson, A., Kawakami, H., Kim, J., Barrett, J. and Kawakami, Y.** (2022). Normal
654 embryonic development and neonatal digit regeneration in mice overexpressing a stem cell factor,
655 Sall4. *PLoS One* **17**,.

656 **Dash, S. and Trainor, P. A.** (2020). The development, patterning and evolution of neural crest cell
657 differentiation into cartilage and bone. *Bone* **137**,.

658 **Donovan KA, A. J. N. R. Y. J. F. E. B. B. E. B. F. E.** (2018). Thalidomide promotes degradation of SALL4, a
659 transcription factor implicated in Duane Radial Ray syndrome. *Elife*.

660 **Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Stein, T. I., Rosen, N., Kohn, A., Twik,**
661 **M., Safran, M., et al.** (2017). GeneHancer: Genome-wide integration of enhancers and target
662 genes in GeneCards. *Database* **2017**,.

663 **Gopinathan, G. and Diekwisch, T. G. H.** (2022). Epigenetics and Early Development. *J Dev Biol* **10**,.

664 **Haensel, D., Sun, P., MacLean, A. L., Ma, X., Zhou, Y., Stemmler, M. P., Brabletz, S., Berx, G., Plikus, M.**
665 **V, Nie, Q., et al.** (2019). An Ovol2-Zeb1 transcriptional circuit regulates epithelial directional
666 migration and proliferation. *EMBO Rep* **20**,.

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

670 **Herbreteau, G., Airaud, F., Pierre-Noël, E., Vallée, A., Bézieau, S., Théoleyre, S., Blons, H., Garinet, S.**
671 **and Denis, M. G.** (2021). MEM: An algorithm for the reliable detection of microsatellite instability
672 (MSI) on a small NGS panel in colorectal cancer. *Cancers (Basel)* **13**,.

CONFIDENTIAL

673 **Hodges, C., Kirkland, J. G. and Crabtree, G. R.** (2016). The many roles of BAF (mSWI/SNF) and PBAF
674 complexes in cancer. *Cold Spring Harb Perspect Med* **6**,.

675 **Islam, Z., Ali, A. M., Naik, A., Eldaw, M., Decock, J. and Kolatkar, P. R.** (2021). Transcription Factors: The
676 Fulcrum Between Cell Development and Carcinogenesis. *Front Oncol* **11**,.

677 **Jourdeuil, K., Neilson, K. M., Tavares, A. L. P. and Moody, S. A.** (2025). Zmym2 Alters Expression of Key
678 Craniofacial Genes. *Genesis (United States)* **63**,.

679 **Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates,**
680 **R., Žídek, A., Potapenko, A., et al.** (2021). Highly accurate protein structure prediction with
681 AlphaFold. *Nature* **596**, 583–589.

682 **Kekunnaya, R. and Negalur, M.** (2017). Duane retraction syndrome: Causes, effects and management
683 strategies. *Clinical Ophthalmology* **11**, 1917–1930.

684 **Kim, S. and Shendure, J.** (2019). Mechanisms of Interplay between Transcription Factors and the 3D
685 Genome. *Mol Cell* **76**, 306–319.

686 **Kodytková, A., Dušátková, P., Anne Amaratunga, S., Plachý, L., Pruhová, Š. and Lebl, J.** (2023).
687 Integrative role of the sall4 gene: From thalidomide embryopathy to genetic defects of the upper
688 limb, internal organs, cerebral midline, and pituitary. *Horm Res Paediatr* **97**, 106–112.

689 **Kohlhase, J., Chitayat, D., Kotzot, D., Ceylaner, S., Froster, U. G., Fuchs, S., Montgomery, T. and Rösler,**
690 **B.** (2005a). SALL4 mutations in Okihiro syndrome (Duane-radial ray syndrome), acro-renal-ocular
691 syndrome, and related disorders. *Hum Mutat* **26**, 176–183.

692 **Kohlhase, J., Chitayat, D., Kotzot, D., Ceylaner, S., Froster, U. G., Fuchs, S., Montgomery, T. and Rösler,**
693 **B.** (2005b). SALL4 mutations in Okihiro syndrome (Duane-radial ray syndrome), acro-renal-ocular
694 syndrome, and related disorders. *Hum Mutat* **26**, 176–183.

695 **Kong, N. R., Bassal, M. A., Tan, H. K., Kurland, J. V., Yong, K. J., Young, J. J., Yang, Y., Li, F., Lee, J. D.,**
696 **Liu, Y., et al.** (2021). Zinc Finger Protein SALL4 Functions through an AT-Rich Motif to Regulate
697 Gene Expression. *Cell Rep* **34**,.

698 **Kosicki, M., Baltoumas, F. A., Kelman, G., Boverhof, J., Ong, Y., Cook, L. E., Dickel, D. E., Pavlopoulos,**
699 **G. A., Pennacchio, L. A. and Visel, A.** (2025). VISTA Enhancer browser: an updated database of
700 tissue-specific developmental enhancers. *Nucleic Acids Res* **53**, D324–D330.

701 **Krasteva, V., Crabtree, G. R. and Lessard, J. A.** (2017). The BAF45a/PHF10 subunit of SWI/SNF-like
702 chromatin remodeling complexes is essential for hematopoietic stem cell maintenance. *Exp*
703 *Hematol* **48**, 58-71.e15.

704 **Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., Wu, H., Aebersold, R., Graef,**
705 **I. A. and Crabtree, G. R.** (2007). An Essential Switch in Subunit Composition of a Chromatin
706 Remodeling Complex during Neural Development. *Neuron* **55**, 201–215.

707 **Li, E.** (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nat*
708 *Rev Genet* **3**, 662–673.

CONFIDENTIAL

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199–W205.

Liu, J., Wu, Q., Wang, Y., Wei, Y., Wu, H., Duan, L., Zhang, Q. and Wu, Y. (2018a). Ovol2 induces mesenchymal–epithelial transition via targeting ZEB1 in osteosarcoma. *Onco Targets Ther* **11**, 2963–2973.

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. (2018b). Targeting cancer addiction for SALL4 by shifting its transcriptome with a pharmacologic peptide. *Proc Natl Acad Sci U S A* **115**, E7119–E7128.

Liu, B. H., Liu, M., Radhakrishnan, S., Jaladanki, C. K., Gao, C., Tang, J. P., Kumari, K., Go, M. L., Vu, K. A. L., Seo, H.-S., et al. (2024). Targeting transcription factors through an IMiD independent zinc finger domain.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**,.

Lu, J., Jeong, H., Kong, N., Yang, Y., Carroll, J., Luo, H. R., Silberstein, L. E., Ma, Y. and Chai, L. (2009). Stem cell factor SALL4 represses the transcriptions of PTEN and SALL1 through an epigenetic repressor complex. *PLoS One* **4**,.

Ma, X., Huang, R., Li, G., Zhang, T. and Ma, J. (2022). A de novo mutation of SALL4 in a Chinese family with Okihiro syndrome. *Mol Med Rep* **25**,.

Matyskiela, M. E., Couto, S., Zheng, X., Lu, G., Hui, J., Stamp, K., Drew, C., Ren, Y., Wang, M., Carpenter, A., et al. (2018). SALL4 mediates teratogenicity as a thalidomide-dependent cereblon substrate. *Nat Chem Biol* **14**, 981–987.

Milet, C. and Monsoro-Burq, A. H. (2012). Neural crest induction at the neural plate border in vertebrates. *Dev Biol* **366**, 22–33.

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

Miller, A., Ralser, M., Kloet, S. L., Loos, R., Nishinakamura, R., Bertone, P., Vermeulen, M. and Hendrich, B. (2016). Sall4 controls differentiation of pluripotent cells independently of the nucleosome remodelling and deacetylation (NuRD) complex. *Development (Cambridge)* **143**, 3074–3084.

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.

Nazim, M., Lin, C. H., Feng, A. C., Xiao, W., Yeom, K. H., Li, M., Daly, A. E., Tan, X., Vu, H., Ernst, J., et al. (2024). Alternative splicing of a chromatin modifier alters the transcriptional regulatory programs of stem cell maintenance and neuronal differentiation. *Cell Stem Cell* **31**, 754-771.e6.

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

CONFIDENTIAL

746 exit from pluripotency and commitment towards neural crest formation in ARID1B-related
747 neurodevelopmental disorders. *Nat Commun* **12**,.

748 **Pantier, R., Chhatbar, K., Quante, T., Skourti-Stathaki, K., Cholewa-Waclaw, J., Alston, G., Alexander-**
749 **Howden, B., Lee, H. Y., Cook, A. G., Spruijt, C. G., et al.** (2021). SALL4 controls cell fate in response
750 to DNA base composition. *Mol Cell* **81**, 845-858.e8.

751 **Pappas, M. P., Kawakami, H., Corcoran, D., Chen, K. Q., Scott, E. P., Wong, J., Gearhart, M. D.,**
752 **Nishinakamura, R., Nakagawa, Y. and Kawakami, Y.** (2024). Sall4 regulates posterior trunk
753 mesoderm development by promoting mesodermal gene expression and repressing neural genes
754 in the mesoderm. *Development (Cambridge)* **151**,.

755 **Patthey, C., Gunhaga, L. and Edlund, T.** (2008). Early development of the central and peripheral nervous
756 systems is coordinated by Wnt and BMP signals. *PLoS One* **3**,.

757 **Pla, P. and Monsoro-Burq, A. H.** (2018). The neural border: Induction, specification and maturation of
758 the territory that generates neural crest cells. *Dev Biol* **444**, S36–S46.

759 **Polit, L., Kerdivel, G., Gregoricchio, S., Esposito, M., Guillouf, C. and Boeva, V.** (2021). CHIPIN: ChIP-seq
760 inter-sample normalization based on signal invariance across transcriptionally constant genes. *BMC*
761 *Bioinformatics* **22**,.

762 **Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F. H., Swigut,**
763 **T. and Wysocka, J.** (2015). Enhancer Divergence and cis-Regulatory Evolution in the Human and
764 Chimp Neural Crest. *Cell* **163**, 68–83.

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

767 **Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. and Wysocka, J.** (2011). A unique
768 chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–285.

769 **Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F. and**
770 **Manke, T.** (2016). deepTools2: a next generation web server for deep-sequencing data analysis.
771 *Nucleic Acids Res* **44**, W160–W165.

772 **Rao, S., Zhen, S., Roumiantsev, S., McDonald, L. T., Yuan, G.-C. and Orkin, S. H.** (2010). Differential
773 Roles of Sall4 Isoforms in Embryonic Stem Cell Pluripotency. *Mol Cell Biol* **30**, 5364–5380.

774 **Renneville, A., Gasser, J. A., Grinshpun, D. E., Jean Beltran, P. M., Udeshi, N. D., Matyskiela, M. E.,**
775 **Clayton, T., McConkey, M., Viswanathan, K., Tepper, A., et al.** (2021). Avadomide Induces
776 Degradation of ZMYM2 Fusion Oncoproteins in Hematologic Malignancies. *Blood Cancer Discov* **2**,
777 251–265.

778 **Rgen Kohlhase, J., Heinrich, M., Schubert, L., Liebers, M., Kispert, A., Laccone, F., Turnpenny, P.,**
779 **Winter, R. M. and Reardon, W.** (2002). Okihiro syndrome is caused by SALL4 mutations. *Oxford*
780 *University Press* **11**, 2979–2987.

781 **Roth, D. M., Bayona, F., Baddam, P. and Graf, D.** (2021). Craniofacial Development: Neural Crest in
782 Molecular Embryology. *Head Neck Pathol* **15**,.

CONFIDENTIAL

783 **Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden,**
784 **C., Saalfeld, S., Schmid, B., et al.** (2012). Fiji: An open-source platform for biological-image analysis.
785 *Nat Methods* **9**, 676–682.

786 **Sievers, Q. L., Petzold, G., Bunker, R. D., Renneville, A., Słabicki, M., Liddicoat, B. J., Abdulrahman, W.,**
787 **Mikkelsen, T., Ebert, B. L. and Thomä, N. H.** (2018). Defining the human C2H2 zinc finger degrome
788 targeted by thalidomide analogs through CRBN. *Science (1979)* **362**,.

789 **Stuhlmiller, T. J. and García-Castro, M. I.** (2012). FGF/MAPK signaling is required in the gastrula epiblast
790 for avian neural crest induction. *Development* **139**, 289–300.

791 **Szabó, A., Szabó, S. and Mayor, R.** (2018). Mechanisms of Neural Crest Migration.

792 **Tahara, N., Kawakami, H., Chen, K. Q., Anderson, A., Peterson, M. Y., Gong, W., Shah, P., Hayashi, S.,**
793 **Nishinakamura, R., Nakagawa, Y., et al.** (2019). Sall4 regulates neuromesodermal progenitors and
794 their descendants during body elongation in mouse embryos. *Development (Cambridge)* **146**,.

795 **Tatetsu, H., Kong, N. R., Chong, G., Amabile, G., Tenen, D. G. and Chai, L.** (2016). SALL4, the missing link
796 between stem cells, development and cancer. *Gene* **584**, 111–119.

797 **Terhal, P., Rösler, B. and Kohlhase, J.** (2006). A family with features overlapping Okihiro syndrome,
798 hemifacial microsomia and isolated Duane anomaly caused by a novel SALL4 mutation. *Am J Med*
799 *Genet* **140 A**, 222–226.

800 **Voss, T. C. and Hager, G. L.** (2014). Dynamic regulation of transcriptional states by chromatin and
801 transcription factors. *Nat Rev Genet* **15**, 69–81.

802 **Wang, F., Zhao, W., Kong, N., Cui, W. and Chai, L.** (2014). The next new target in leukemia: The
803 embryonic stem cell gene SALL4. *Mol Cell Oncol* **1**,.

804 **Wang, B., Li, C., Ming, J., Wu, L., Fang, S., Huang, Y., Lin, L., Liu, H., Kuang, J., Zhao, C., et al.** (2023). The
805 NuRD complex cooperates with SALL4 to orchestrate reprogramming. *Nat Commun* **14**,.

806 **Xiong, J., Todorova, D., Su, N. Y., Kim, J., Lee, P. J., Shen, Z., Briggs, S. P. and Xu, Y.** (2015). Stemness
807 factor Sall4 is required for DNA damage response in embryonic stem cells. *Journal of Cell Biology*
808 **208**, 513–520.

809 **Yamamoto, J., Suwa, T., Murase, Y., Tateno, S., Mizutome, H., Asatsuma-Okumura, T., Shimizu, N.,**
810 **Kishi, T., Momose, S., Kizaki, M., et al.** (2020). ARID2 is a pomalidomide-dependent CRL4CRBN
811 substrate in multiple myeloma cells. *Nat Chem Biol* **16**, 1208–1217.

812 **Yang, J.** (2018). SALL4 as a transcriptional and epigenetic regulator in normal and leukemic
813 hematopoiesis. *Biomark Res* **6**,.

814 **Yuri, S., Fujimura, S., Nimura, K., Takeda, N., Toyooka, Y., Fujimura, Y. I., Aburatani, H., Ura, K., Koseki,**
815 **H., Niwa, H., et al.** (2009). Sall4 is essential for stabilization, but not for pluripotency, of embryonic
816 stem cells by repressing aberrant trophectoderm gene expression. *Stem Cells* **27**, 796–805.

817 **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.**
818 (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the
819 transcriptional regulation of Pou5f1. *Nat Cell Biol* **8**, 1114–1123.

CONFIDENTIAL

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

822 **Zhang, W., Chronis, C., Chen, X., Zhang, H., Spalinskas, R., Pardo, M., Chen, L., Wu, G., Zhu, Z., Yu, Y., et**
823 **al.** (2019a). The BAF and PRC2 Complex Subunits Dpf2 and Eed Antagonistically Converge on Tbx3
824 to Control ESC Differentiation. *Cell Stem Cell* **24**, 138-152.e8.

825 **Zhang, Z., Zhan, X., Kim, B. and Wu, J.** (2019b). A proteomic approach identifies SAFB-like transcription
826 modulator (SLTM) as a bidirectional regulator of GLI family zinc finger transcription factors. *Journal*
827 *of Biological Chemistry* **294**, 5549–5561.

828 **Zhang, H., Wang, X., Li, J., Shi, R. and Ye, Y.** (2021). BAF Complex in Embryonic Stem Cells and Early
829 Embryonic Development. *Stem Cells Int* **2021**,.

830 **Zheng, Z. H., Sam, T. W., Zeng, Y., Chu, J. J. H. and Loh, Y. H.** (2021). Chromatin Regulation in
831 Development: Current Understanding and Approaches. *Stem Cells Int* **2021**,.

832
833

CONFIDENTIAL

834 **METHODS**

835 **REAGENT TABLE**

836

REAGENT	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
PAX6 antibody	Abcam	AB_305110
ARID1B antibody	Abcam	AB_2243092
ARID2 Polyclonal Antibody	Bethyl Laboratories	AB_1731041
ZMYM2 Polyclonal Antibody	Thermo Fisher Scientific	AB_11152210
IRDye 800CW Goat anti-Rabbit IgG	LI-COR Biosciences	AB_621843

CONFIDENTIAL

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		

CONFIDENTIAL

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'-TCTCGCTCTCGTTCAGAAGTC-3'
NANOG	5'-ATGCCTCACACGGAGACTGT-3'	5'-AAGTGGGTTGTTTGCCTTTG-3'
POU5F1	5'-TCGAGAACCGAGTGAGAGG-3'	5'-GAACCACACTCGGACCACA-3'
MSX1	5'-GAAGATGCGCTCGTCAAAGC-3'	5'-CTTACGGTTCGTCTTGTGTTTG-3'
PAX3	5'-ACAACGCCTGACGTGGAG-3'	5'-ATTTACTTCTCAGGATGCGGCT-3'
TWIST1	5'-GCCAGGTACATCGACTTCCTCT-3'	5'-TCCATCCTCCAGACCGAGAAGG-3'
Software and algorithms		
AlphaFold3	https://alphafoldserver.com/	SCR_025454
TrimGalore!	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	SCR_011847

CONFIDENTIAL

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

837

838

839 **METHOD DETAILS**

840 **AlphaFold3 structure prediction**

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

848

849 **Human iPSC culture**

850 Wild-type iPSC lines were obtained from the iPSC Core of the University of Pennsylvania (Control
851 line-1: SV20 line, male, age 43). iPSC lines were cultured in feeder-free, serum-free mTeSR Plus
852 medium (STEMCELL Technologies) with 1% Penicillin-Streptomycin (5,000 U/mL; Gibco
853 #15140122) on tissue culture plates coated overnight with Geltrex™ LDEV-Free hESC-qualified
854 Reduced Growth Factor Basement Membrane Matrix (Fisher-Scientific). Cells were passaged
855 ~1:8 at 60-80% confluency using 0.5mM EDTA to create 50-200 cell clusters.

856

857 **CRISPR/Cas9-mediated SALL4 knock-out**

858 The heterozygous *SALL4* knock-out in the human SV20 iPSC line was generated via
859 CRISPR/Cas9 gene-editing by the biotech company Synthego. Specifically, two clones were
860 used. In both clones Cas9 targeted the first exon of the *SALL4* gene in the established SV20 iPSC
861 line producing an 11-base deletion (c.230delGGGCCGCATCT; clone C4, named KO1 in the
862 manuscript) and a 19-base deletion (c.del226GCTGGGGCCGCATCT; clone F1, named KO2 in

CONFIDENTIAL

863 the manuscript) respectively, which in both cases generated a premature stop codon. The two
864 clones were used as biological replicates in all experiments.
865
866 **Trilineage differentiation**
867 iPSCs were differentiated into the three germ layers (endoderm, ectoderm and mesoderm) using
868 the Human Pluripotent Stem Cell Functional Identification Kit (R&D Systems, SC027B). The
869 expression of relevant markers was then assessed using immunofluorescence with 10 ug/ml of
870 antibodies provided in the kit (Goat anti-human SOX17, Goat anti-human OTX2 and, Goat anti-
871 human TBXT).
872
873 **iPSC-to-CNCC differentiation**
874 All iPSC lines were differentiated into CNCCs as previously described (Prescott et al., 2015).
875 When iPS cells reached ~60% confluence in Geltrex-coated 6-well plates, cells were dissociated
876 with accutase in small suspension clusters and were treated with CNCC differentiation media (1:1
877 Neurobasal medium/D-MEM F-12 medium (Invitrogen), 0.5× B-27 supplement with Vitamin A
878 (50× stock, Gibco, #17504044), 0.5× N-2 supplement (100× stock, Gibco; #17502001), 20 ng/ml
879 human FGF (Gibco), 20 ng/ml recombinant EGF (Sigma-Aldrich), 5 µg/ml bovine insulin (Sigma-
880 Aldrich) and 1× Glutamax-I supplement (100× stock, Invitrogen). Medium (2 ml) was changed
881 every other day. At Day-7 of differentiation, adhering neurospheres and early migrating CNCCs
882 were transitioned to CNCC early maintenance media until Day-14 (1:1 Neurobasal medium/D-
883 MEM F-12 medium (Invitrogen), 0.5× B-27 supplement with Vitamin A (50× stock, Invitrogen),
884 0.5× N-2 supplement (100× stock, Invitrogen), 20 ng/ml bFGF (Biolegend), 20 ng/ml EGF (Sigma-
885 Aldrich), 1 mg/ml bovine serum albumin, serum replacement grade (Gemini Bio-Products # 700-
886 104 P) and 1× Glutamax-I supplement (100× stock, Invitrogen). Cell medium (2 ml) was changed
887 every other day. Differentiating cells were passaged with Accutase Cell Detachment Solution
888 (Stemcell Technologies; #07920) for all the subsequent passages when cells reached 80%
889 confluence, plated in a fresh, Geltrex-coated 6-well plates. After passaging, cells were treated
890 with 0.05% ROCK-inhibitor (Y27632; Stemcell Technologies) for 24 hours.
891
892 **Real-time quantitative polymerase chain reaction (RT-qPCR)**
893 Cells were harvested as single cells, and the total RNA was extracted using Monarch® Total RNA
894 Miniprep Kit (New England Biolabs) following manufacturer's instructions. 600ng of total RNA was
895 used to synthesize cDNA using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (K1641,
896 Thermo Fisher). For each RT-qPCR reaction, 15 ng of cDNA was used with 0.1 µM of each primer,

CONFIDENTIAL

897 10 µL of PowerUp™ SYBR™ Green Master Mix (Applied Biosystems) in a final volume of 20 µl.
898 Samples were analyzed with CFX Connect Real-Time PCR Detection System (Bio-Rad) using
899 the following thermal cycling parameters: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C,
900 20 s at 63 °C followed by 30 s at 72 °C. Each sample was run in triplicate. 18 S rRNA was used
901 as normalizer.
902
903 **Nuclear Fractionation and Co-immunoprecipitation**
904 Cells were washed twice with ice-cold PBS, scraped, and collected by centrifugation at 10,000 ×
905 g for 1 minute. Pellets were resuspended in 500 µL of Buffer A (10mM HEPES pH7.9, 1.5mM
906 $MgCl_2$, 10mM KCl, 1mM PMSF, 0.5mM DTT, protease inhibitor cocktail) and cytoplasmic fractions
907 were obtained by incubating the cells on ice for 15 minutes, adding of 0.05% NP-40, followed by
908 centrifugation at 17,000 × g for 10 minutes at 4 °C. The resulting pellet was gently washed with
909 Buffer A and resuspended in Buffer C (10mM HEPES pH7.9, 12.5% glycerol, 0.75mM $MgCl_2$,
910 10mM KCl, 0.1mM EDTA, 400mM NaCl, 1mM PMSF, 0.5mM DTT, protease inhibitor cocktail).
911 Samples were incubated at 4 °C with shaking at 3,400 rpm for 1 hour, with vortexing every 10
912 minutes. Nuclear fractions were collected by centrifugation at 17,000 × g for 10 minutes at 4 °C.
913 For co-immunoprecipitation (co-IP), 25 µL of Dynabeads™ Protein A (ThermoFisher, Cat#
914 10002D) were washed twice with PBS containing 0.1% Tween-20 (PBS-T), resuspended in 1 mL
915 of PBS-T, and incubated with 2 µg of the appropriate antibody for 1 hour at 4 °C with rotation.
916 Bead-antibody complexes were washed twice with PBS-T, resuspended in 125 µL PBS-T, and
917 incubated with 100 µg of nuclear extract for 3 hours at 4 °C with rotation. Beads were then washed
918 five times with Buffer C and eluted in 25 µL of Buffer C supplemented with 4.1 µL of 6× reducing
919 Laemmli SDS sample buffer (ThermoFisher Scientific). Eluates and 5% input controls (nuclear
920 extract diluted 1:10 in PBS and mixed with 1:6 volume of 6× Laemmli buffer) were boiled for 5
921 minutes prior to SDS-PAGE. For Western blotting, 20 µL of each sample were loaded as
922 described below.
923
924 **Western Blot**
925 For each condition, cells were harvested and lysed with cold radioimmunoprecipitation assay
926 buffer (Pierce™ RIPA buffer, Thermo Fisher) with Pierce™ Protease Inhibitor Mini Tablets, EDTA-
927 free (Thermo Fisher). After centrifugation at 4°C at 16,000x g for 10 minutes, the total extracted
928 protein was quantified using Pierce™ BCA Protein Assay Kit, following manufacturer's
929 indications. Thirty micrograms of total protein were loaded in Novex WedgeWell 4–20% Tris-
930 Glycine Gel (Invitrogen) and separated through gel electrophoresis (SDS–PAGE) Tris-Glycine-

CONFIDENTIAL

931 SDS buffer (Invitrogen). The proteins were then transferred to nitrocellulose membrane (Thermo
932 Fisher Scientific) and incubated for one hour in Intercept® (PBS) Blocking Buffer (Licor). After
933 blocking, membranes were incubated overnight at 4°C (while rocking) in suitable antibody diluted
934 in blocking buffer and 0.2% Tween-20, at recommended concentrations. Membranes were
935 washed in 0.1% PBS-T before and after being incubated for one hour in the appropriate secondary
936 antibody (1:15,000 dilution). The antibodies were then visualized using Odyssey XF Imaging
937 System (Licor) or ChemiDocTM MP Imaging System and imaged with Image Studio Software
938 (Licor) or Fiji(Schindelin et al., 2012).
939
940 **Flow Cytometry analysis of surface markers**
941 To obtain a single-cell suspension for flow cytometry analysis, cells were treated with Accutase
942 for 5 min at 37°C. Cells were then washed with cold phosphate-buffered saline (PBS)-2% fetal
943 bovine serum (FBS) and live cells were counted. 1×10^6 cells/condition were resuspended in
944 100 µL PBS-2% FBS and stained. For pluripotency evaluation, 4 µl of the respective antibodies
945 were used. For analysis of differentiation, 2 µl of the respective antibodies were used. Cells were
946 incubated for 15 min on ice and protected from light, before transferring them into FACS tubes
947 containing an additional 300 µL PBS-2% FBS. Flow cytometry data were acquired using
948 NovoCyte Penteon Flow Cytometer Systems 5 Lasers and analyzed with FlowJo Software version
949 10.9.
950
951 **Immunofluorescence**
952 Cells were passaged onto Geltrex-coated coverslips and fixed at required confluence or time-
953 point. In detail, cells were incubated in 4% PFA at 37°C for 15 minutes, followed by three 2-minute
954 washes in 1× PBS. For permeabilization, cells were treated with 0.1% (v/v) Triton X-100 (T8787,
955 Sigma-Aldrich) at room temperature for 10 minutes before three 2-minute washes in 1× PBS +
956 0.1% (v/v) Tween 20 and a 1-hour blocking upon 10% Donkey serum (D9663, Sigma-Aldrich) at
957 room temperature. Cells were incubated with corresponding PBS-diluted primary antibodies
958 protected from direct light at 4°C overnight. After three 5-minute washes with 1× PBS + 0.1%
959 Tween 20, cells were incubated for 1 hour in PBS-diluted secondary antibodies and 15 minutes
960 with DAPI (1:5,000; 422801, BioLegend). Cells were then washed three times with 1× PBS +
961 0.1% Tween for 5 minutes and once with 1× PBS. Coverslips were then fetched and mounted
962 onto slides using Fluorescence Mounting Medium (S3023, Agilent Technologies). Slides were
963 imaged using confocal microscope (LSM780, Zeiss). The intensity of fluorescent signal was
964 measured using Corrected Total Cell Fluorescence (CTCF) using Image J. Briefly, for each cell,

CONFIDENTIAL

965 a region of interest (ROI) was drawn around the cell boundary, and the Integrated Density (sum
966 of all pixel intensity values within the ROI), area of ROI (total number of pixels contained in the
967 ROI), and the mean background fluorescence (average pixel intensity measured from a
968 background region of the image lacking fluorescent signal) were measured.
969 The corrected fluorescence intensity was then calculated as:

970 $$CTCF = Integrated\ Density - (Area\ of\ ROI\ \times\ Mean\ Background\ Fluorescence)$$

971 For colony-forming cells (e.g., iPSCs), we analyzed segmented nuclei located within a section
972 along the horizontal diameter of each colony, encompassing approximately 30–40 cells. This
973 approach allowed us to capture differences between the colony center and edges while enabling
974 comparisons across colonies of similar size.
975
976 **RNA-seq**
977 Total RNA was extracted using Monarch® Total RNA Miniprep Kit (New England Biolabs,
978 T2010S) according to the manufacturer's instructions. RNA was quantified using NanoDrop 2000
979 UV Visible Spectrophotometer, while the RNA integrity was checked on Tapestation 2200
980 (Agilent). Only samples with RIN value above 8.0 were used for transcriptome analysis. RNA
981 libraries were prepared using 1 µg of total RNA input using NEBNext® Poly(A) mRNA Magnetic
982 Isolation Module, NEBNext® UltraTM II Directional RNA Library Prep Kit for Illumina® according
983 to the manufacturer's instructions (New England Biolabs).
984
985 **RNA-sequencing analysis**
986 After removing the adapters with TrimGalore!
987 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) , Kallisto was used to quantify
988 reads mapping to each gene. Differential gene expression levels were determined using
989 DESeq2(Love et al., 2014). All statistical analyses were performed using the latest versions of
990 BEDTools(Quinlan and Hall, 2010), deepTools(Ramírez et al., 2016), R Studio (version 4.4.3;
991 Trophy Case), and GraphPad Prism software version 10.2.3
992 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San Diego, CA).
993
994 **CUT&RUN**
995 For each condition, 1x 10^5 cells were processed using CUT&RUN Assay Kit #86652 (Cell
996 Signaling Technology). Briefly, each sample was resuspended in Concanavalin A Beads and
997 treated with an Antibody Binding Buffer and the suitable antibody. After the antibody incubation,
998 each reaction underwent a step of pAG-MNase Enzyme binding, and subsequent DNA Digestion

CONFIDENTIAL

999 and diffusion. The diffused DNA was then purified using Phenol/Chloroform Extraction and
1000 Ethanol Precipitation. The purified DNA was quantified using QuantiFluor dsDNA System, using
1001 the ONE dsDNA protocol (Promega). Libraries for sequencing were prepared using NEBNext
1002 Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) and the libraries size distribution
1003 was assessed through automated electrophoresis using the Agilent 2200 Tapestation System
1004 (Agilent Technology).
1005
1006 **ChIP-seq**
1007 All samples from different conditions were processed together to prevent batch effects.
1008 Approximately 15 million cells were double cross-linked with 2mM disuccinimidyl glutarate (DSG)
1009 for 20 minutes followed by 1% formaldehyde for 10 minutes at room temperature, quenched with
1010 125 mM glycine, harvested, and washed twice with 1X PBS. The fixed cell pellet was resuspended
1011 in ChIP lysis buffer (150 mM NaCl, 1% Triton X-100, 0.7% SDS, 500 μM DTT, 10 mM Tris-HCl, 5
1012 mM EDTA) and chromatin was sheared to an average length of 200–900 base-pairs, using a
1013 Covaris S220 Ultrasonicator at 5% duty factor for 7 minutes for 3 times. The chromatin lysate was
1014 diluted with SDS-free ChIP lysis buffer. 10μg of antibody was used for BRG1. The antibody was
1015 added to at least 5μg of sonicated chromatin along with Dynabeads Protein G magnetic beads
1016 (Invitrogen) and incubated with rotation at 4°C overnight. The beads were washed twice with each
1017 of the following buffers: Mixed Micelle Buffer (150 mM NaCl, 1% Triton X- 100, 0.2% SDS, 20 mM
1018 Tris-HCl, 5 mM EDTA, 65% sucrose), Buffer 200 (200 mM NaCl, 1% Triton X-100, 0.1% sodium
1019 deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), LiCl detergent wash (250 mM LiCl,
1020 0.5% sodium deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA) and a final wash was
1021 performed with cold 0.1X TE. Finally, beads were resuspended in 1X TE containing 1% SDS and
1022 incubated at 65°C for 10 min to elute immunocomplexes. The elution was repeated twice, and the
1023 samples were incubated overnight at 65°C to reverse cross-linking, along with the input (5% of
1024 the starting material). The DNA was digested with 0.5 mg/ml Proteinase K for 3 hours at 55°C
1025 and then purified using the ChIP DNA Clean & Concentrator kit (Zymo) and quantified with QUBIT.
1026 Barcoded libraries were made with NEBNext Ultra II DNA Library Prep Kit for Illumina) using
1027 NEBNext Multiplex Oligos Dual Index Primers for Illumina (New England BioLabs) and sequenced
1028 on NovaSeq X Plus instrument (Illumina).
1029
1030 **ATAC-seq**
1031 One hundred thousand cells/condition were harvested by enzymatic dissociation using
1032 StemPro™ Accutase™ (Thermo Fisher Scientific). Each cell pellet was processed using ATAC-

CONFIDENTIAL

1033 Seq Kit (Active Motif), following the manufacturer's directions. Briefly, the fresh cell pellets were
1034 washed in 1X PBS and lysate in ATAC Lysis Buffer. Each lysate underwent a process of
1035 tagmentation and purification, before being amplified through PCR, using the following thermal
1036 cycler parameters: 72°C for 5 min, 98°C for 30 sec, followed by 10 cycles of 98°C for 10 sec,
1037 63°C for 30 sec, 72°C for 1 min. Each PCR reaction was cleaned up using SPRI beads (1.2X of
1038 the sample volume). The Agilent 2200 Tapestation System (Agilent Technologies) was used as
1039 quality control tool before performing next generation sequencing (NGS).
1040

1041 **Bioinformatics analysis of CHIP-seq, ATAC-seq and CUT&RUN data**

1042 Adapters were removed with TrimGalore!
1043 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), the sequences were aligned
1044 to the reference hg19, using Burrows-Wheeler Alignment tool, with the MEM
1045 algorithm(Herbreteau et al., 2021). Uniquely mapping aligned reads were filtered based on
1046 mapping quality (MAPQ > 10) to restrict our analysis to higher quality and uniquely mapped reads,
1047 and PCR duplicates were removed. MACS2(Zhang et al., 2008) was used to call peaks using the
1048 default parameters at 5% FDR.
1049

1050 **Motif analysis**

1051 BED files for the regions of interest were produced using BEDTools(Quinlan and Hall, 2010). Motif
1052 analysis was performed using HOMER(Heinz et al., 2010). Shuffled input sequences were used
1053 as background. E-values < 0.001 were used as a threshold for significance.
1054

1055 **QUANTIFICATION AND STATISTICAL ANALYSIS**

1056 RT-qPCR and image quantification data were analyzed using Prism software version 10.2.3
1057 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San Diego, CA). Data is
1058 represented as mean ± standard error of mean (SEM). GO enrichment analysis was performed
1059 using Webgestalt 2019(Liao et al., 2019). All statistical analyses were performed using the latest
1060 versions of BEDTools(Quinlan and Hall, 2010), Deeptools(Ramírez et al., 2016), R version 4.4.3
1061 (Throphy Case; https://www.R-project.org/), and Prism software version 10.2.3
1062 (http://www.graphpad.com/faq/viewfaq.cfm?faq=1362; GraphPad, San Diego, CA). In particular,
1063 Bigwigs were generated with Deeptools and normalized by sequencing depth using the same
1064 software, and upper-quantile normalized using CHIPIN(Polit et al., 2021).

CONFIDENTIAL

Figure 1

Click here to access/download;Figure;Figure_1_final.jpg

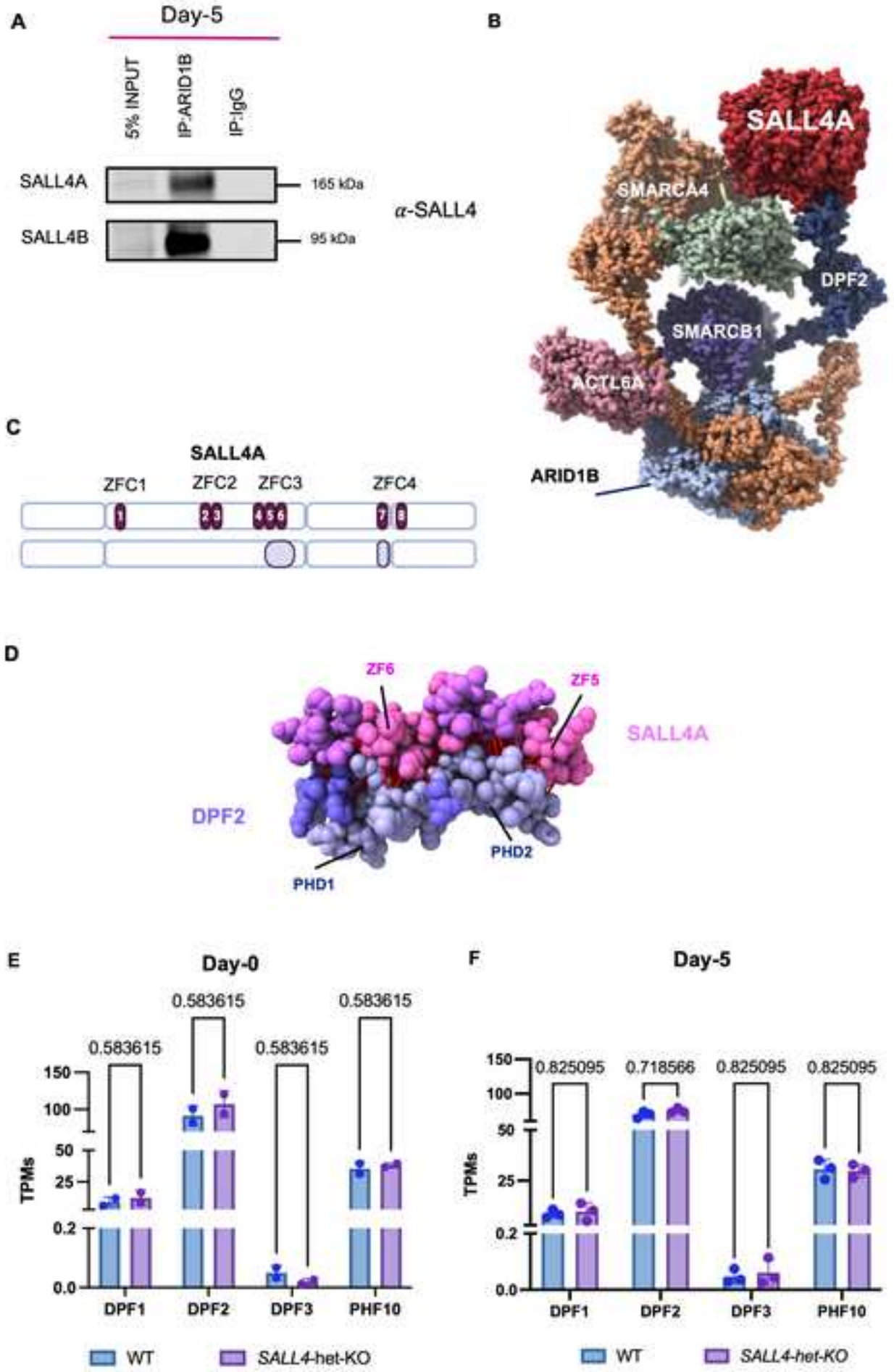


Figure 2

Click here to access/download;Figure;Figure_2_final.jpg

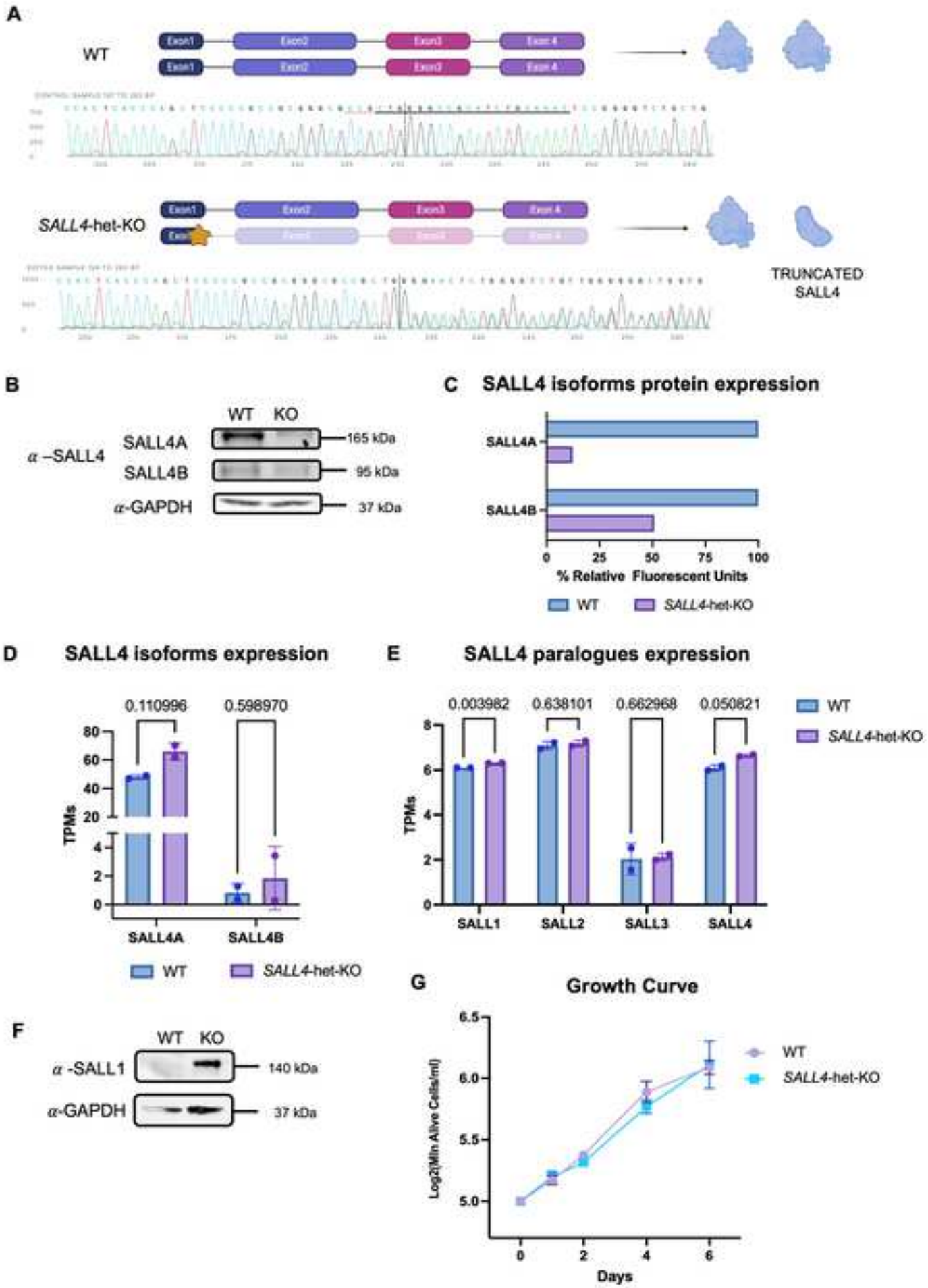


Figure 3

Click here to access/download;Figure;Figure_3_final_NOV25.jpg

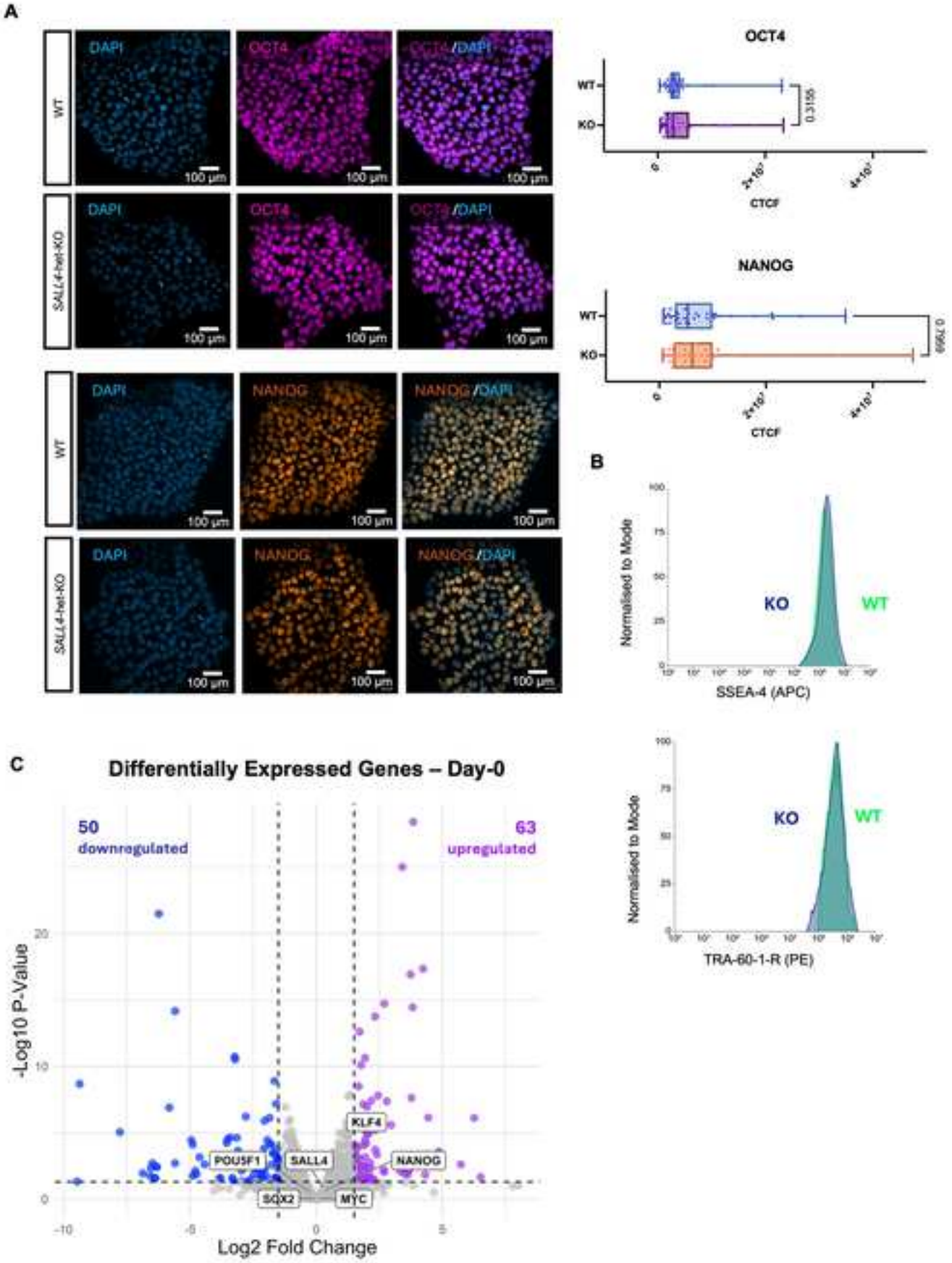


Figure 4

Click here to access/download;Figure;Figure_4_final.jpg

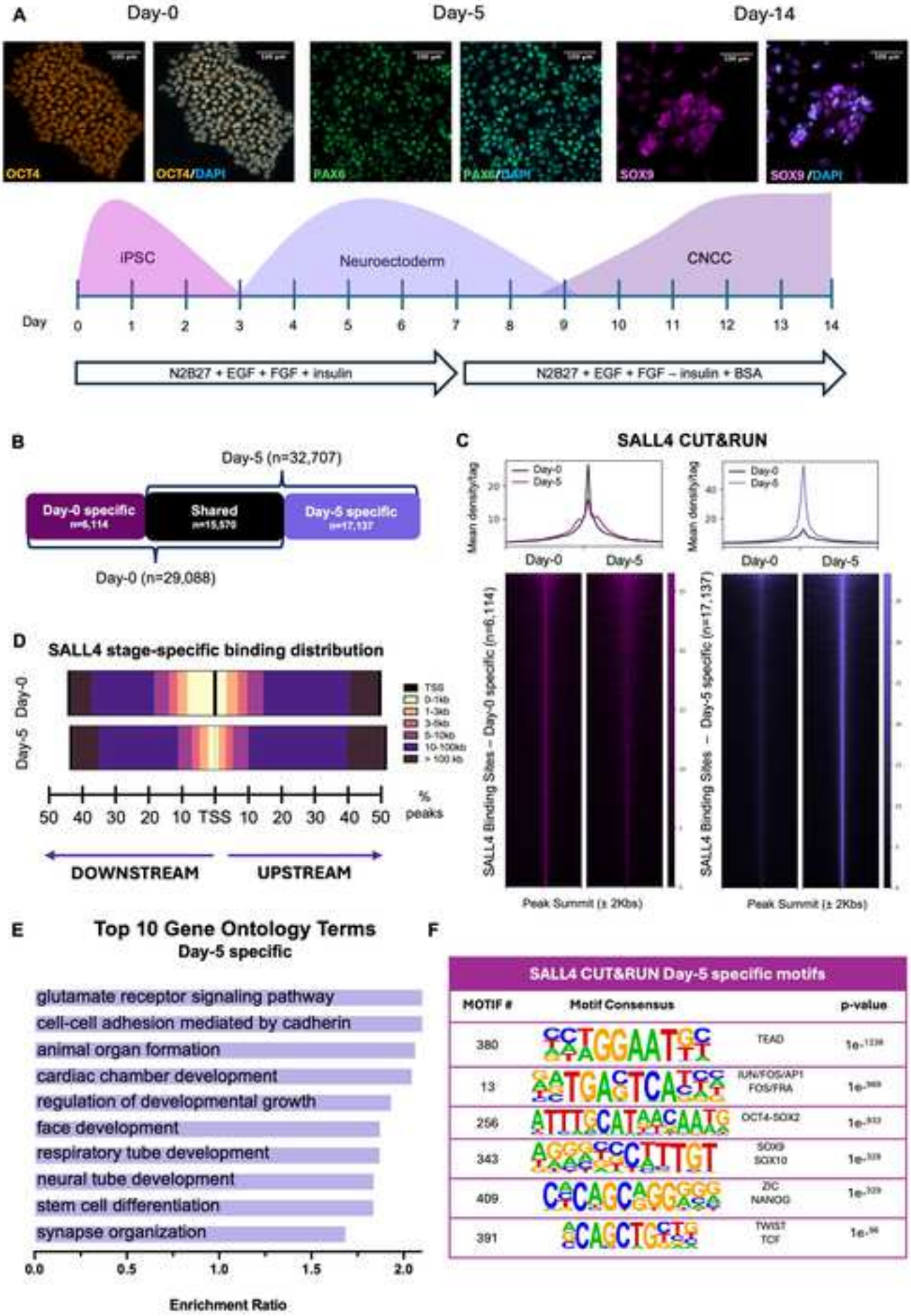
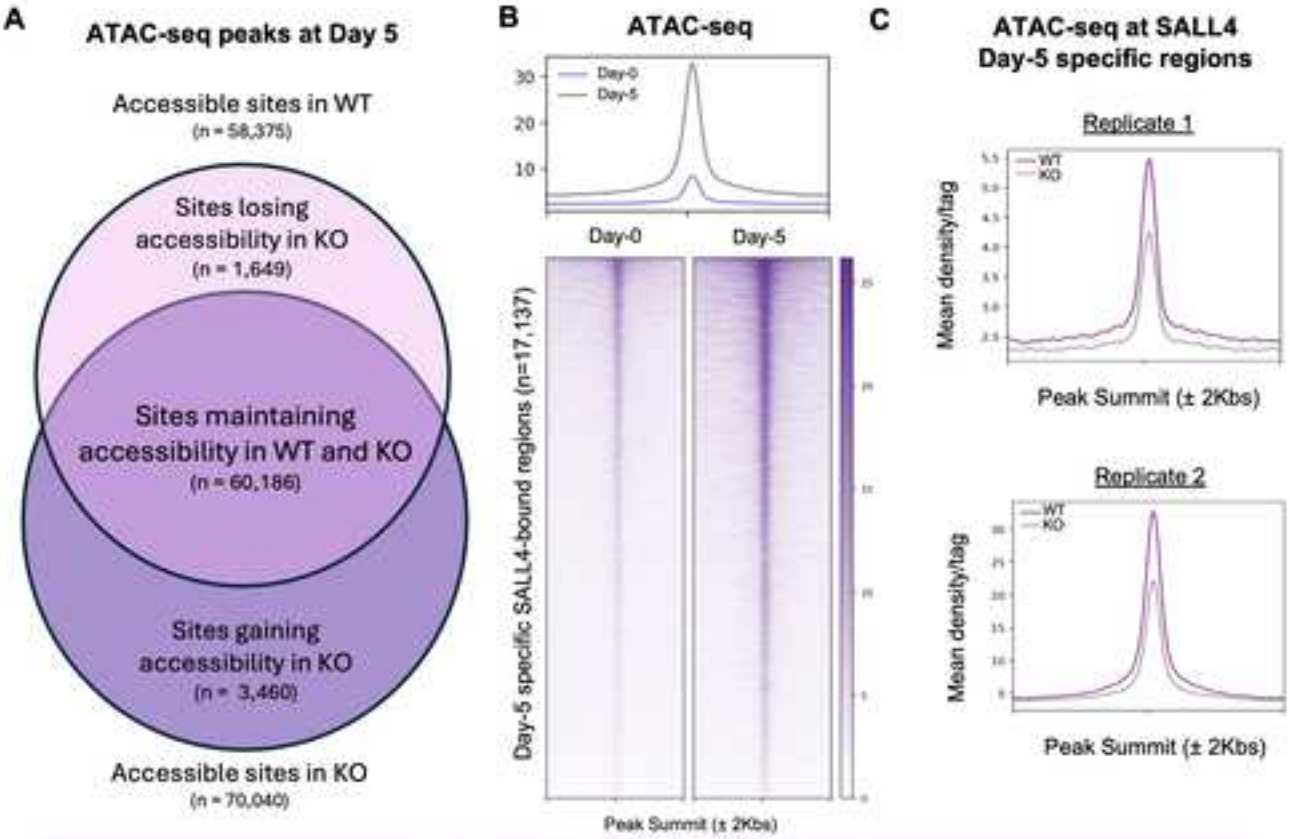


Figure 5

Click here to access/download;Figure;Figure_5_final.jpg



D

Gaining accessibility in KO Day 5 motifs			
MOTIF #	Motif Consensus		p-value
51		CTCF BORIS	$1e^{-270}$
256		POU5F1/SOX2/ NANOG POU2F3 (epidermis) POU3F1 (nervous s.) POU2F2 (metastases)	$1e^{-103}$
410 (8%)		ZIC3 ZIC2 NANOG	$1e^{-84}$

E

Losing accessibility in KO Day 5 motifs			
MOTIF #	Motif Consensus	TF	p-value
410 (11%)		ZIC3 ZIC2 NANOG	$1e^{-59}$
92		ERG ETS	$1e^{-58}$
402		ZEB1 ZEB2 SNAIL1	$1e^{-50}$
407		SMAD4 SMAD2 P53	$1e^{-47}$
107		AP-1 JUN FOS ATF3	$1e^{-37}$

Figure 6

Click here to access/download;Figure;Figure_6_final.jpg

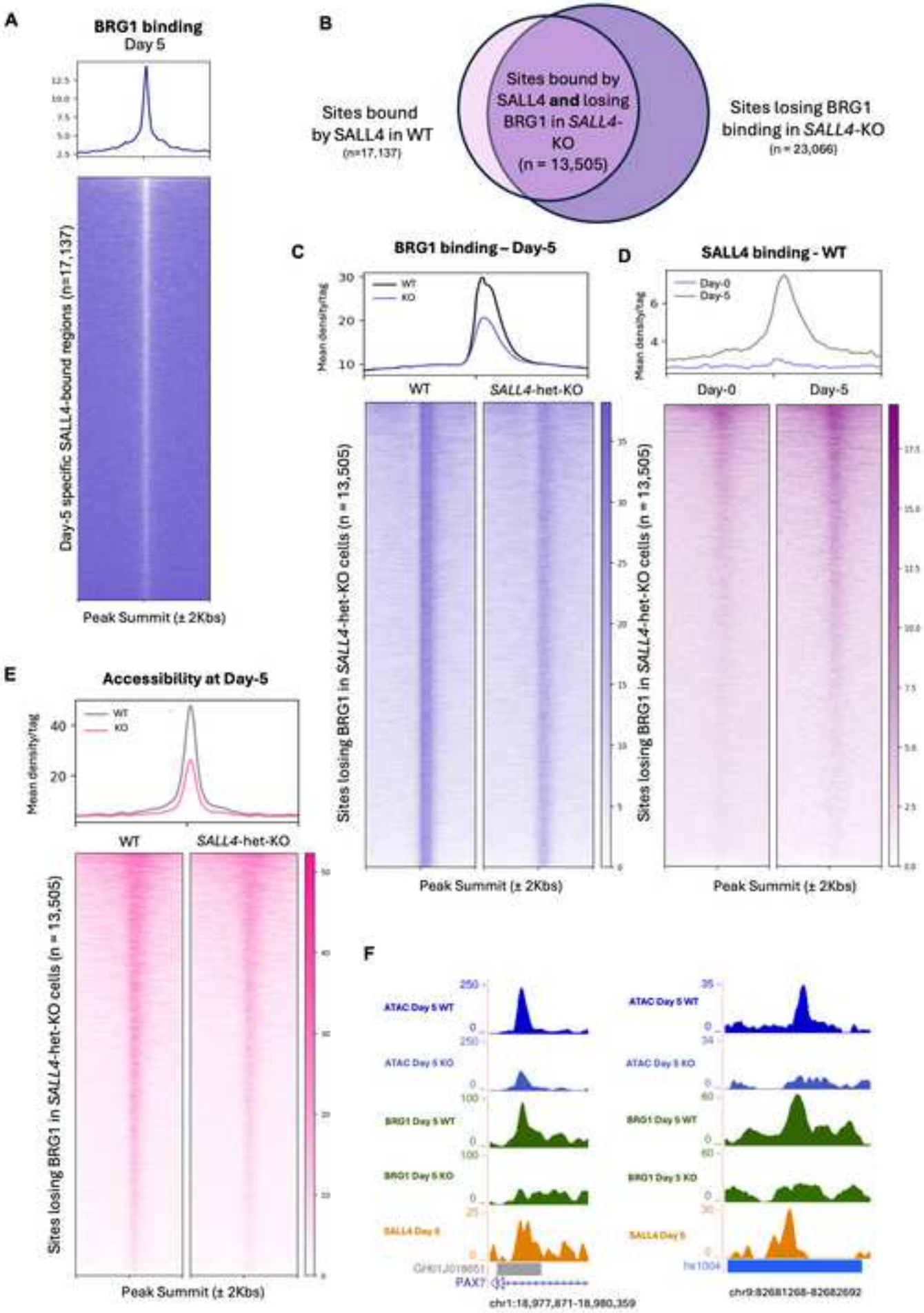


Figure 7

Click here to access/download;Figure;Figure_7_final.jpg

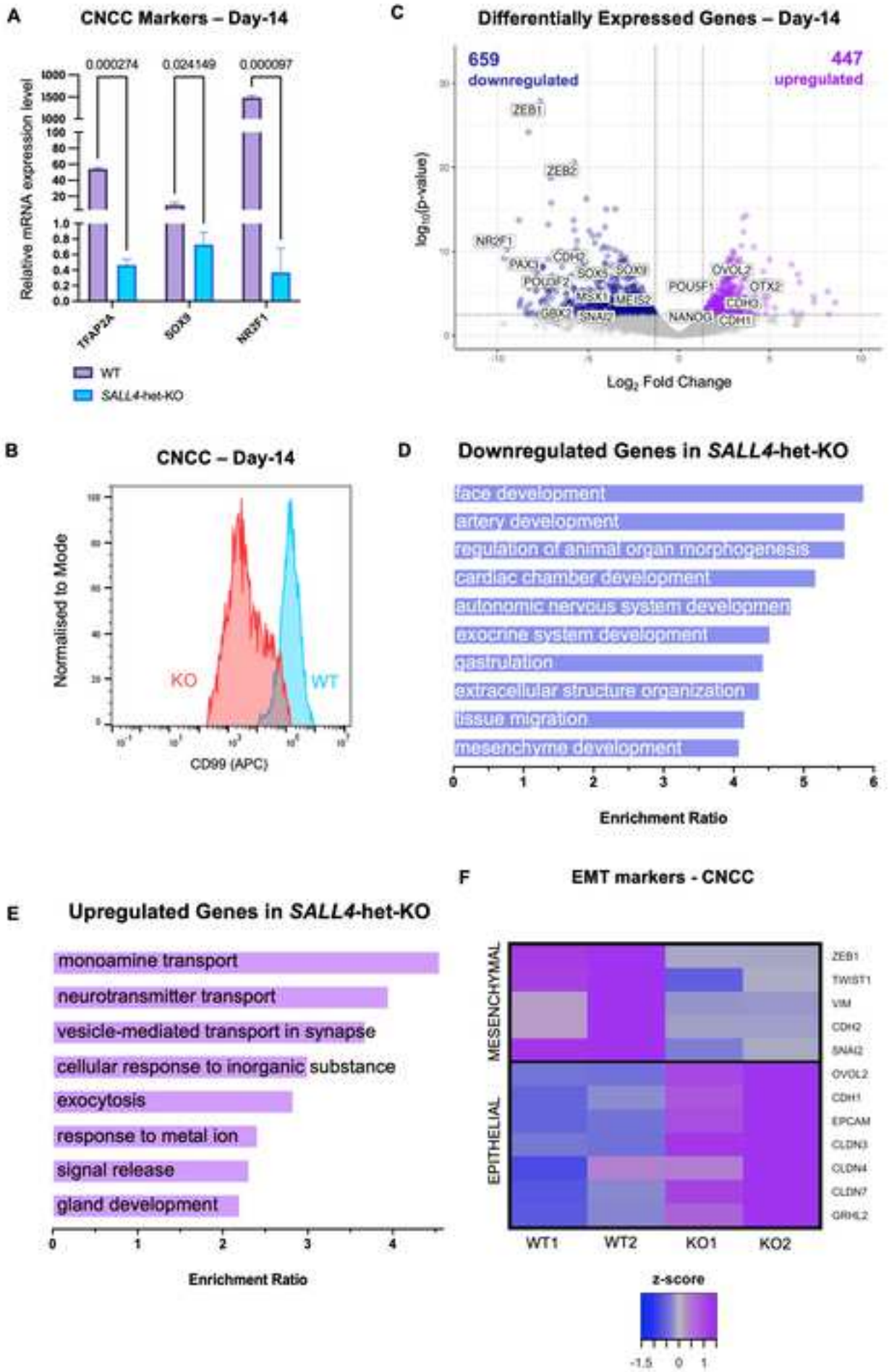
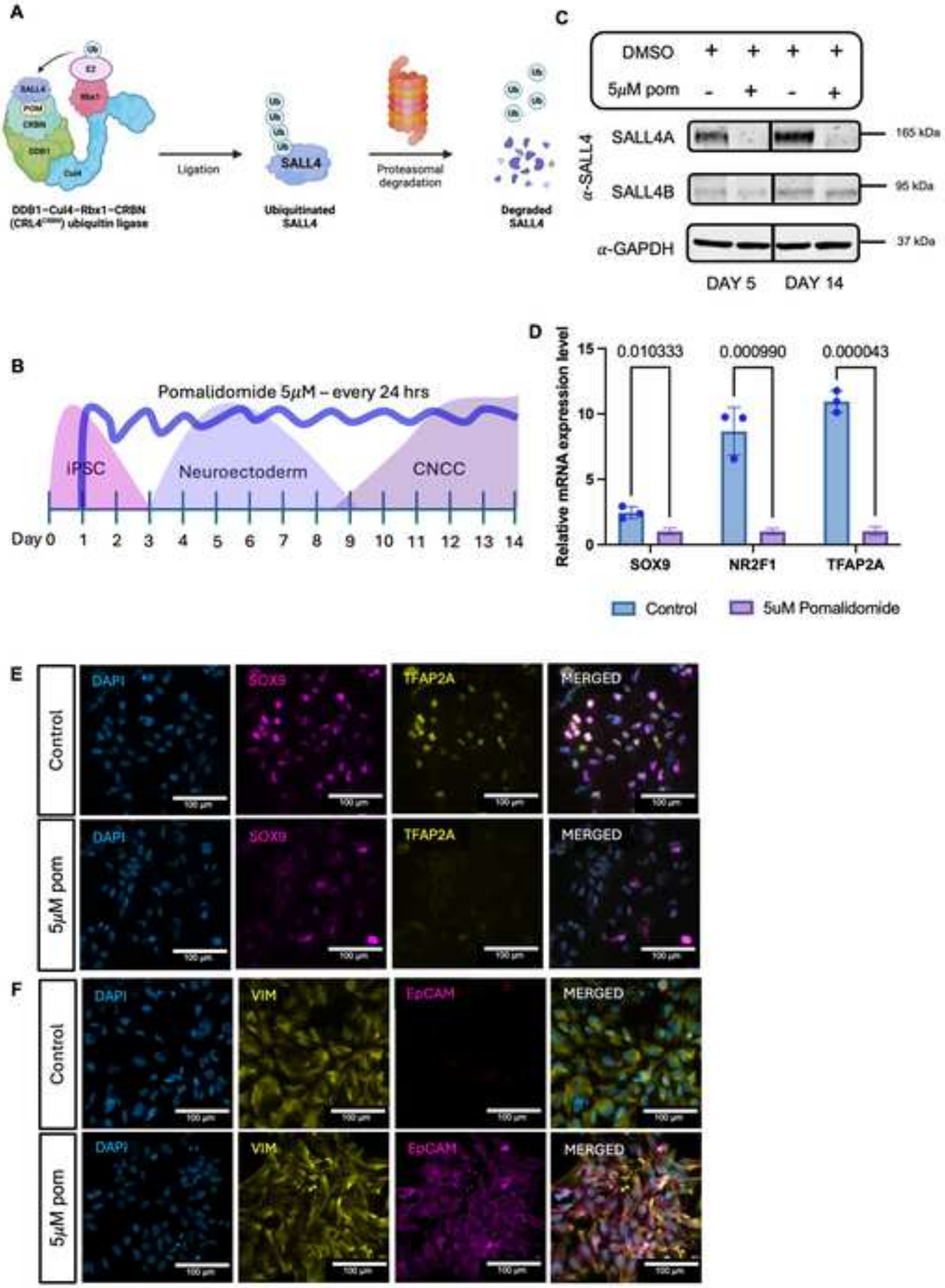
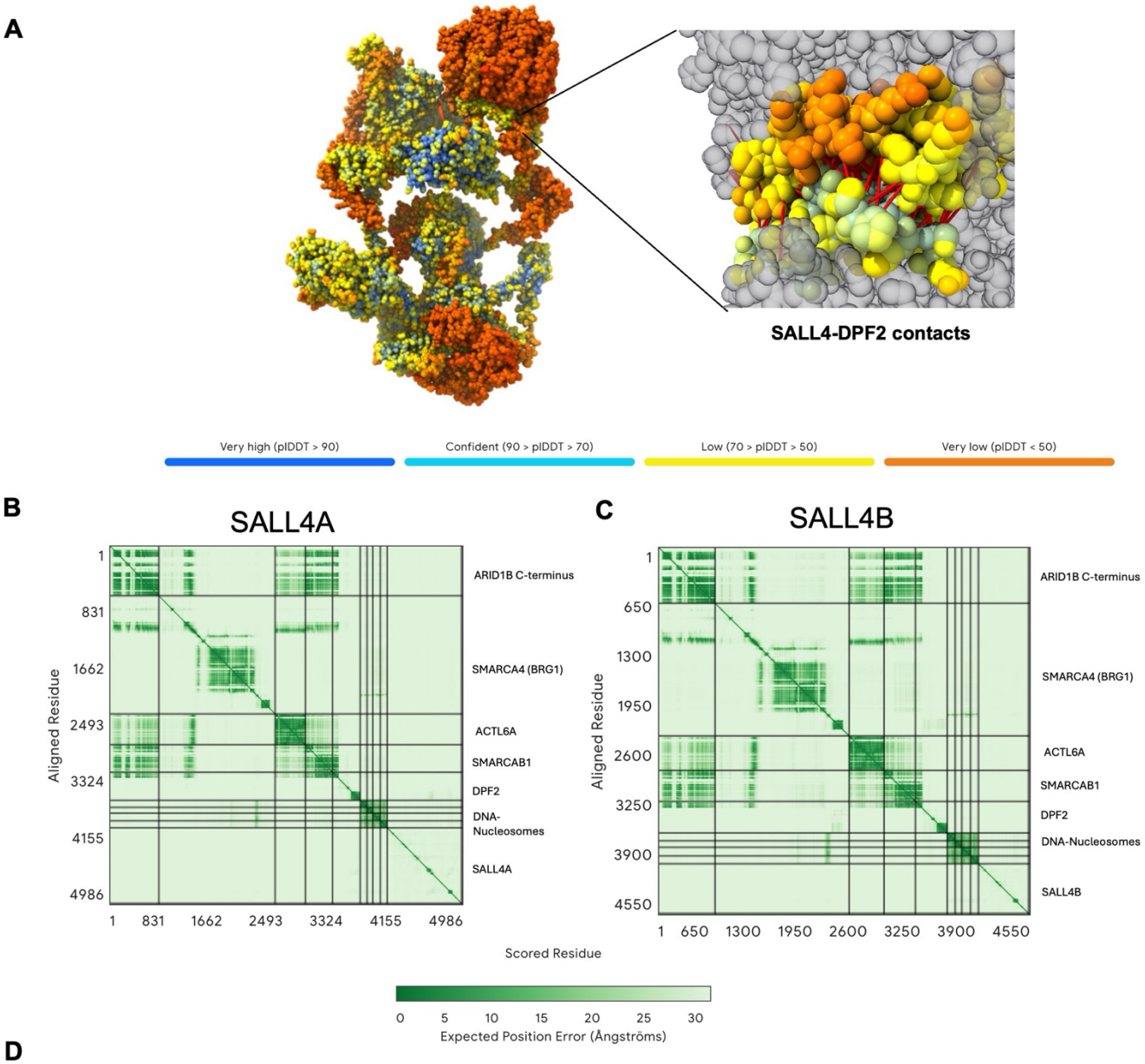


Figure 8

Click here to access/download;Figure;Figure_8_final.jpg




A
SALL4-DPF2 contacts
Very high (pIDDT > 90)
Confident (90 > pIDDT > 70)
Low (70 > pIDDT > 50)
Very low (pIDDT < 50)
B
SALL4A
Aligned Residue
1
831
1662
2493
3324
4155
4986
ARID1B C-terminus
SMARCA4 (BRG1)
ACTL6A
SMARCAB1
DPF2
DNA-Nucleosomes
SALL4A
Scored Residue
C
SALL4B
Aligned Residue
1
650
1300
1950
2600
3250
3900
4550
ARID1B C-terminus
SMARCA4 (BRG1)
ACTL6A
SMARCAB1
DPF2
DNA-Nucleosomes
SALL4B
Scored Residue
0
5
10
15
20
25
30
Expected Position Error (Ångströms)
D
SALL4B
1
7
8
Model 1
Model 3
Model 4
E
PREDICTED
SALL4
DPF2
SMARCA4
SMARCB1
ACTL6
ARID1B
EM - 3Å resolution
SMARCA4
DPF2
SMARCB
ACTL6
ARID1A

Supplementary Figure 1.

(A) (Left) Visual representation of the confidence score pLDDT (predicted Local Distance Difference Test) for each residue of the AlphaFold-predicted complex, with color gradients indicating prediction confidence (blue: very high confidence, yellow: low confidence, orange/red: very low confidence). (Right) Close-up view of the SALL4A-DPF2 interface highlighting key interacting residues in red.

(B) AlphaFold predicted aligned error (PAE) plot for model 0 of SALL4A. The plot shows the expected positional error in Ångströms for each residue pair.

(C) AlphaFold predicted aligned error (PAE) plot for model 0 of SALL4B. The plot shows the expected positional error in Ångströms for each residue pair.

(D) SALL4B protein sequences. C2H2 zinc fingers annotations are highlighted in dark purple and numbered. Visual representation of output models of SALL4B (models 1, 3, and 4). Interactions between SALL4B and DPF2 are highlighted in pink, light purple, and light blue. Made with BioRender.com.

(E) Comparison between the predicted BAF-SALL4A in silico structure and an experimentally identified structure of the human BAF Base module (PDB: 6LTH).

CONFIDENTIAL

A

CLONE F1

Status	Guide Target	PAM Sequence	Indel %	Model Fit (R^2)	Knockout-Score	Nuclease
Succeeded	GTTTGCAGATGCGGCCCCAG	CGG	51	0.97	51	

POWERED BY EditCo ICE

RELATIVE CONTRIBUTION OF EACH SEQUENCE (NORMALIZED)

	INDEL	CONTRIBUTION	SEQUENCE
	-19	51%	AGCTCCCCCGCCGCGGGCGCC----|----------ACTCCGGGGTCTGCTGCTGCGGCTGCTGCTCGCCC
+	0	46%	AGCTCCCCCGCCGCGGGCGCCGCTG|GGGCCGCATCTGCAAACTCCGGGGTCTGCTGCTGCGGCTGCTGCTCGCCC

CLONE C4

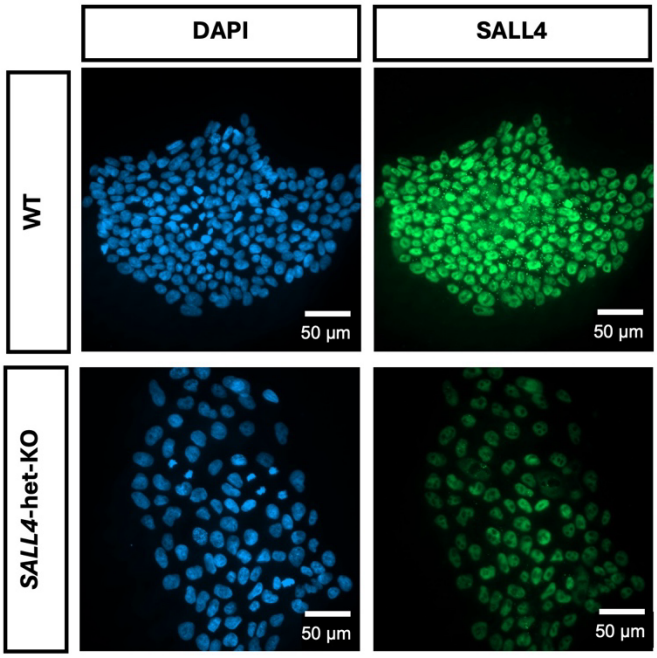
Status	Guide Target	PAM Sequence	Indel %	Model Fit (R^2)	Knockout-Score	Nuclease
Succeeded	GTTTGCAGATGCGGCCCCAG	CGG	52	0.98	52	

POWERED BY EditCo ICE

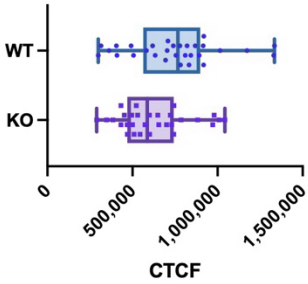
RELATIVE CONTRIBUTION OF EACH SEQUENCE (NORMALIZED)

	INDEL	CONTRIBUTION	SEQUENCE
	-11	52%	AGCTCCCCCGCCGCGGGCGCCGCTG|----------GCAAACTCCGGGGTCTGCTGCTGCGGCTGCTGCTCGCCC
+	0	46%	AGCTCCCCCGCCGCGGGCGCCGCTG|GGGCCGCATCTGCAAACTCCGGGGTCTGCTGCTGCGGCTGCTGCTCGCCC

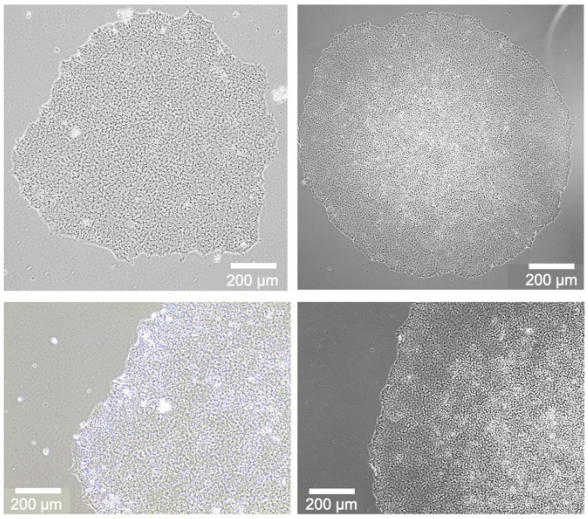
B



C



D



Supplementary Figure 2.

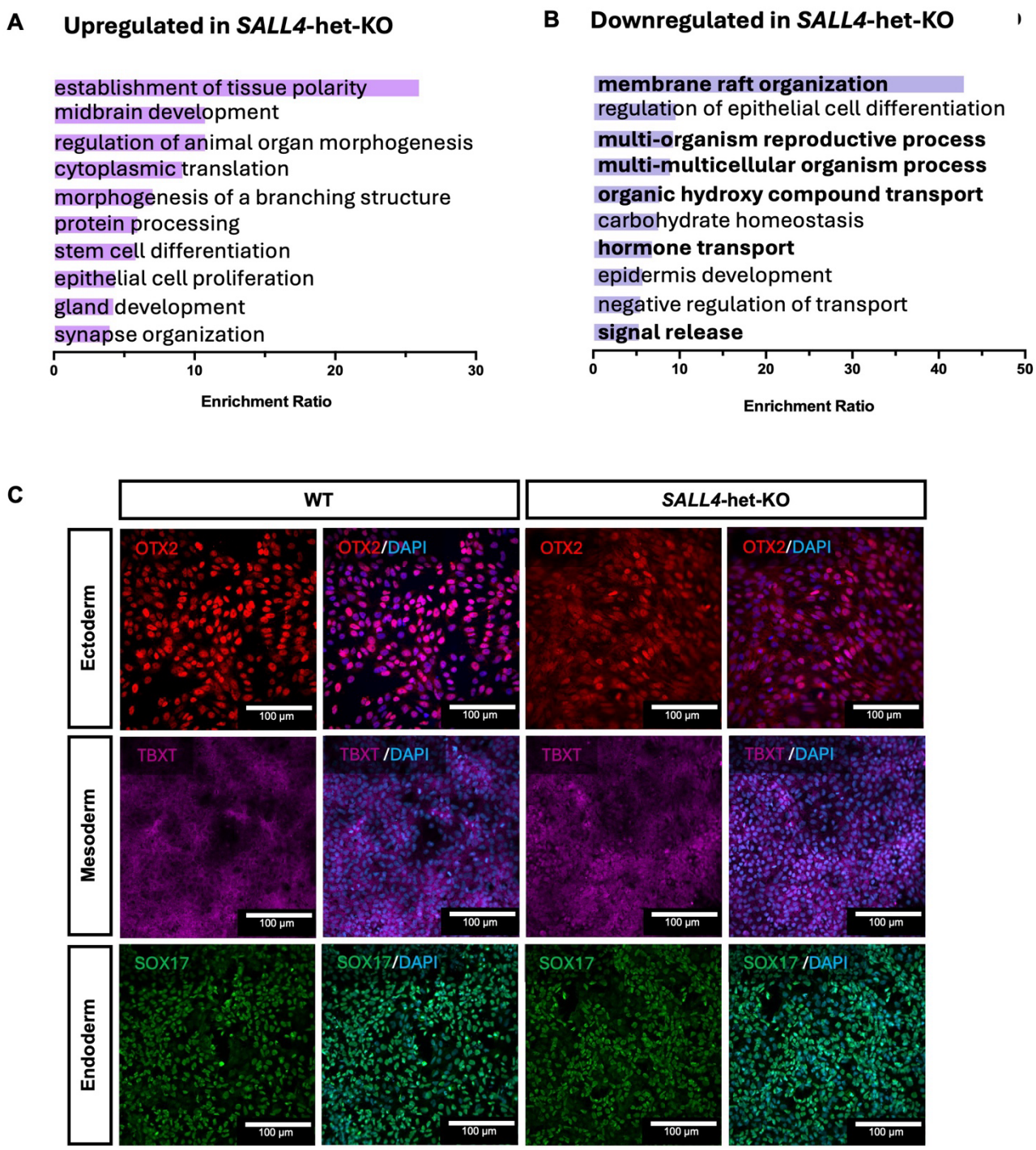
(A) Sanger sequencing reports of the two iPSC *SALL4*-het-KO clones used in this study: clone F1 (19-bp deletion, top) and clone C4 (11-bp deletion, bottom). The data were analyzed using the ICE (Inference of CRISPR Edits) tool to quantify CRISPR-Cas9 editing efficiency. Editing resulted in an overall indel frequency of 51% and 52%, respectively. The sequence traces show the relative contribution of each allele, normalized across the population, with deleted bases indicated by dashed lines. Cut sites are represented by black vertical dotted lines, and the wild-type sequence is marked by a "+" symbol on the far left.

(B) Immunofluorescence labeling quantifying the expression of SALL4 in WT and clone F1 iPSC lines. Scale bar: 50 µm.

(C) Box plot showing Corrected Total Cell Fluorescence (CTCF) values quantifying and comparing SALL4 expression in WT and SALL4-het-KO (clone F1) in immunofluorescence assay (sample n= 30; background n= 5). Error bars display standard error of the mean.

(D) Phase-contrast microscopic images showing the morphology of *SALL4*-WT and *SALL4*-het-KO (clone F1) iPSC colonies at 4x (top) and 10x (bottom).

CONFIDENTIAL



Supplementary Figure 3.

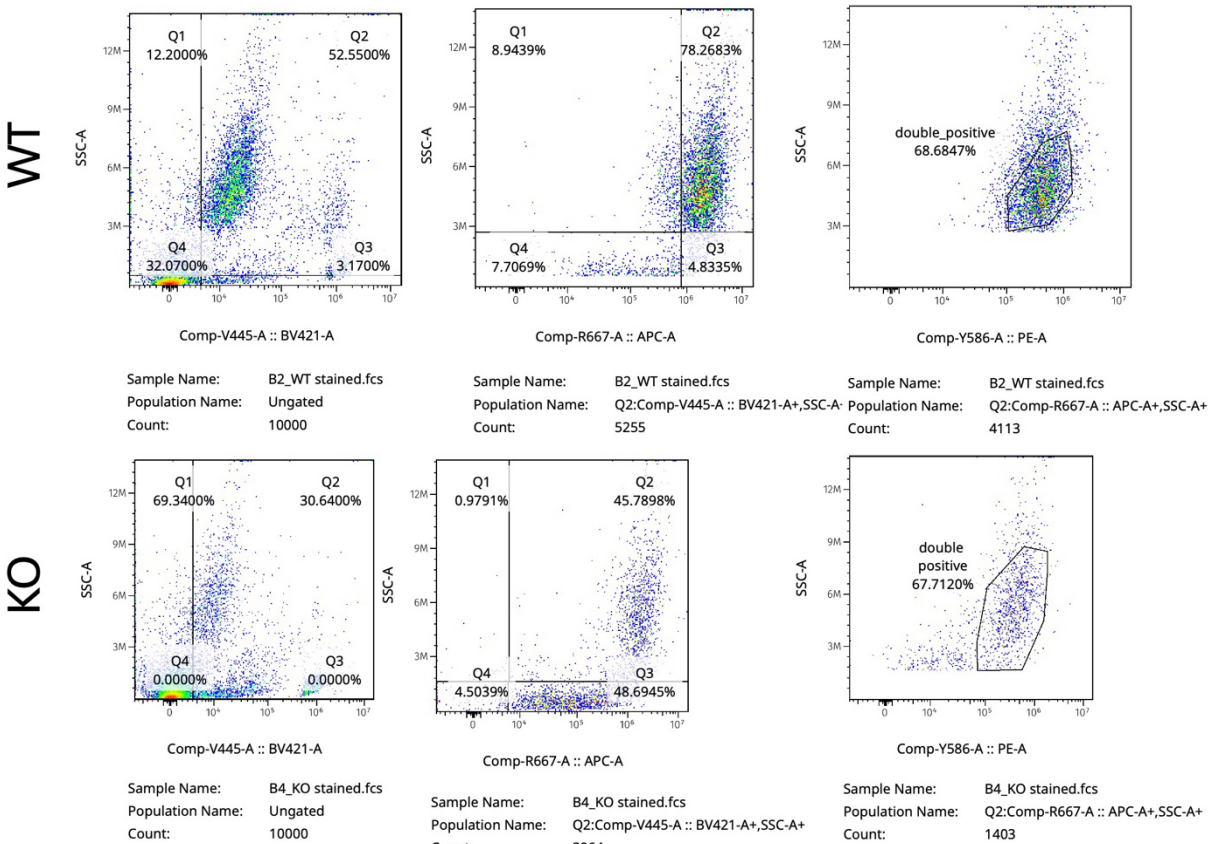
(A) Top 10 GO term enrichment for upregulated genes in *SALL4*-het-KO iPSCs determined using WebGestalt[71] over-representation pathway analysis, using affinity propagation as parameter for redundancy removal. All terms are non-significant (FDR > 0.05).

(B) Top 10 GO term enrichment for downregulated genes in *SALL4*-het-KO iPSCs determined using WebGestalt[71] over-representation pathway analysis, using affinity propagation as parameter for redundancy removal. In bold terms with FDR < 0.05.

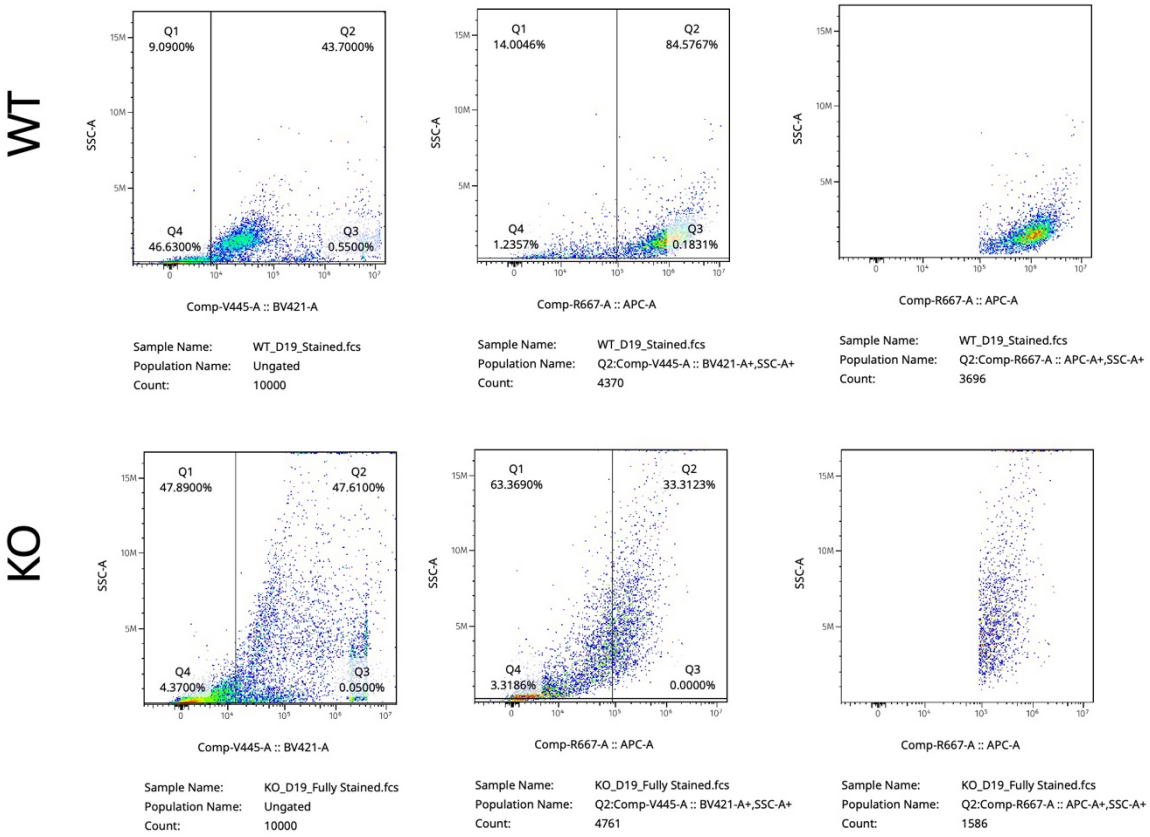
(C) Immunofluorescence for ectodermal (OTX2), mesodermal (TBXT), and endodermal (SOX17) markers in *SALL4*-WT and *SALL4*-het-KO (clone F1) after trilineage differentiation. Scale bar: 100 µm.

CONFIDENTIAL

A. GATING STRATEGY FIG. 3B



B. GATING STRATEGY FIG. 7B



Supplementary Figure 4.

Gating strategy for flow cytometry data, as reported in (A) figure 3B and (B) figure 7B.

A

SALL4 CUT&RUN Day-0 specific motifs			
MOTIF #	**Motif Consensus**		**p-value**
51		CTCF BORIS	$1e^{-270}$
260		OCT:OCT	$1e^{-106}$

B

Day-0 specific SALL4 binding sites – *de novo* motifs				
Motif Consensus	**% of Target**	**% of Background**	**p-value**	**Closest match**
	20.17%	2.23%	$1e^{-486}$	CTCF BORIS
	4.63%	0.40%	$1e^{-118}$	
	2.60%	0.17%	$1e^{-73}$	SALL4 (ZFC4) ARID/HMG

C

Day-5 specific SALL4 binding sites – *de novo* motifs				
Motif Consensus	**% of Target**	**% of Background**	**p-value**	**Closest match**
	51.16%	17.48%	$1e^{-1613}$	TEAD
	34.74%	10.61%	$1e^{-1108}$	AP-1
	26.53%	9.00%	$1e^{-698}$	NANOG

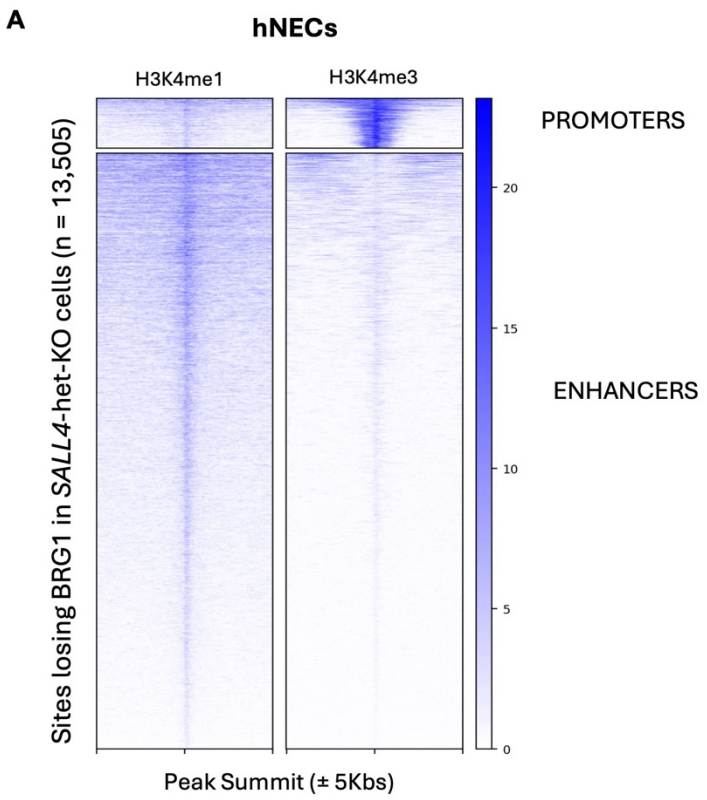
Supplementary Figure 5.

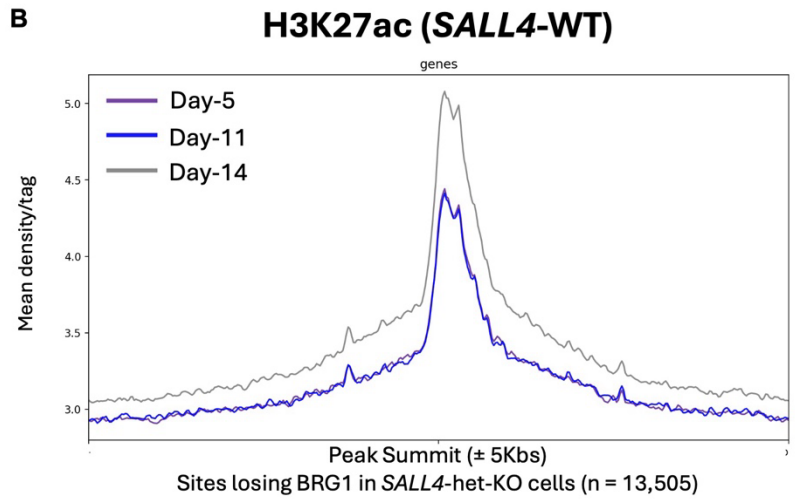
(A) Table of motifs enriched in Day-0 specific SALL4-bound regions as determined using HOMER[68].

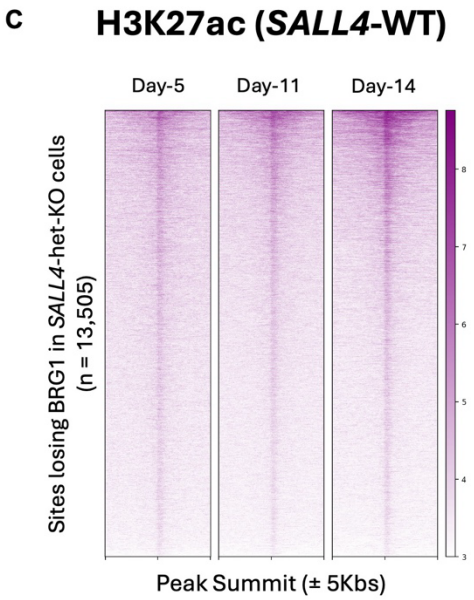
(B) Table of de novo motifs enriched in Day-0 specific SALL4-bound regions as determined using HOMER[68].

(C) Table of de novo motifs enriched in Day-5 specific SALL4-bound regions as determined using HOMER[68].

CONFIDENTIAL







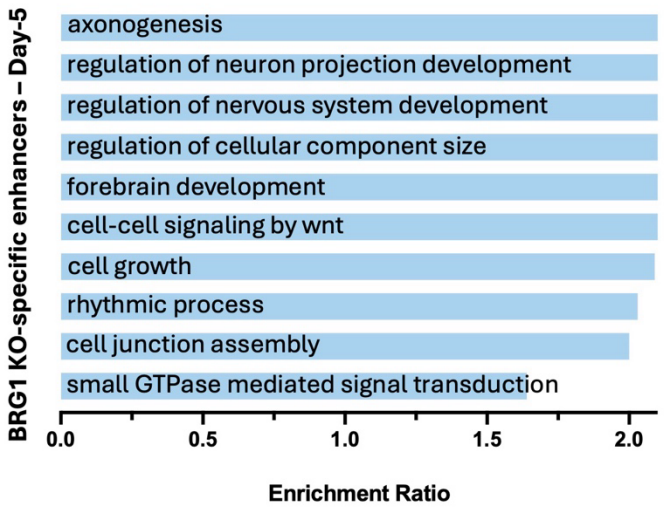
Supplementary Figure 6.

(A) ChIP-seq signal intensity for H3K4me1 and H3K4me3 plotted across ±5 kb of sites losing BRG1 in SALL4-het-KO cells regions (n= 13,505) in human neuroectoderm-like cells (hNECs), centered on peak summits. Rows represent individual peaks ranked by signal intensity.

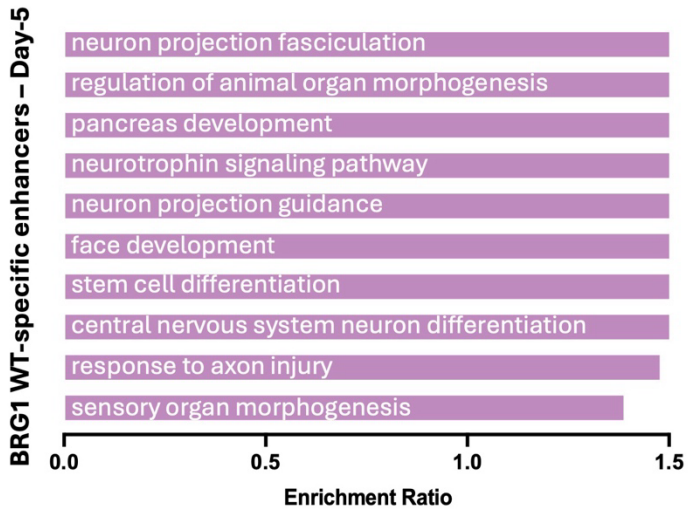
(B) Average profiles and (C) heatmaps showing H3K27ac ChIP-seq signal genomic sites plotted across ±5 kb of sites losing BRG1 in *SALL4*-het-KO cells regions (n= 13,505) at Day-5, Day-11, and Day-14 of the iPSC-to-CNCC specification. Centered on peak summits.

CONFIDENTIAL

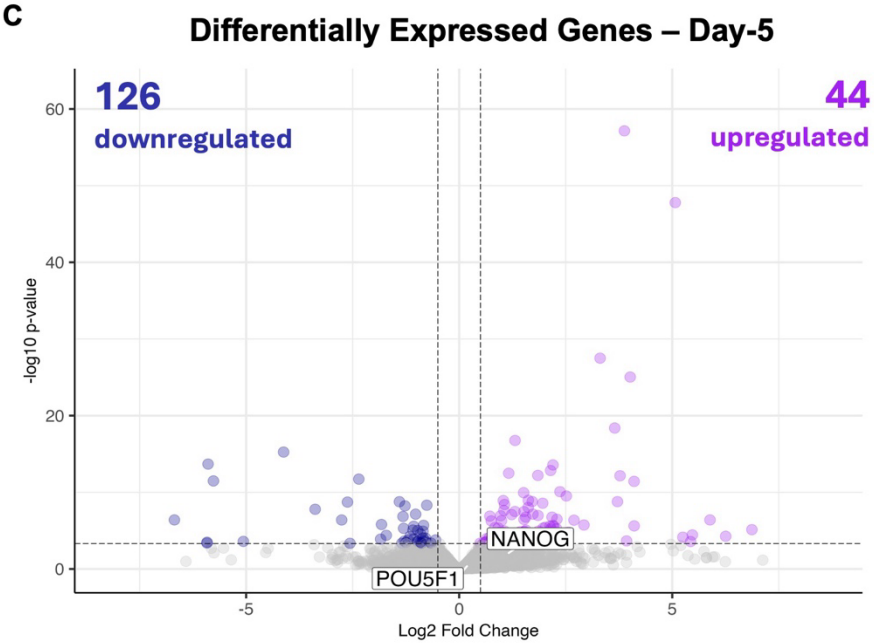
A


BRG1 KO-specific enhancers – Day-5
axonogenesis
regulation of neuron projection development
regulation of nervous system development
regulation of cellular component size
forebrain development
cell-cell signaling by wnt
cell growth
rhythmic process
cell junction assembly
small GTPase mediated signal transduction
0.0
0.5
1.0
1.5
2.0
Enrichment Ratio

B


BRG1 WT-specific enhancers – Day-5
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
Enrichment Ratio

C


Differentially Expressed Genes – Day-5
126
downregulated
44
upregulated
-log10 p-value
60
40
20
0
NANOG
POU5F1
-5
0
5
Log2 Fold Change

Supplementary Figure 7.

(A) GO term enrichment for genes near BRG1 ChIP-seq peaks gained in *SALL4*-het-KO cells. GO terms were determined using WebGestalt[1] over-representation pathway analysis, using affinity propagation as parameter for redundancy removal (FDR<0.05).

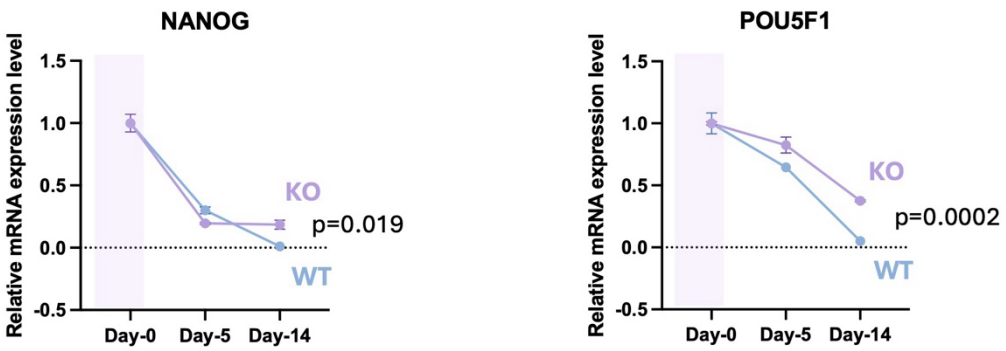
(B) GO term enrichment for genes near BRG1 ChIP-seq peaks lost in *SALL4*-het-KO cells. GO terms were determined using WebGestalt[1] over-representation pathway analysis, using affinity propagation as parameter for redundancy removal (FDR<0.05).

(C) Volcano plot of genes differentially expressed in *SALL4*-het-KO relative to *SALL4*-WT in neuroectoderm-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.

CONFIDENTIAL

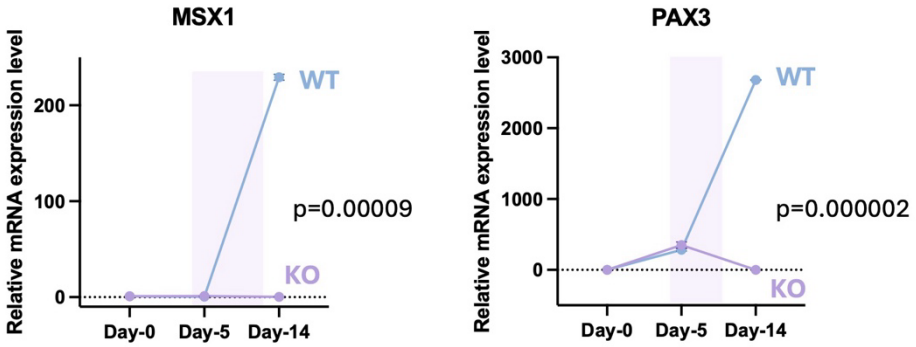
A

Pluripotency Markers



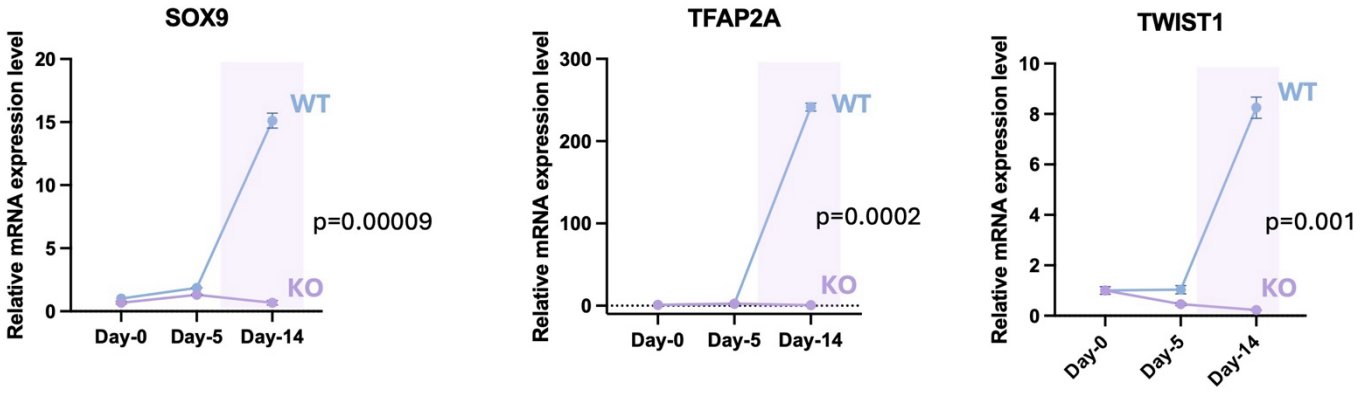
B

Neural Plate Border Markers

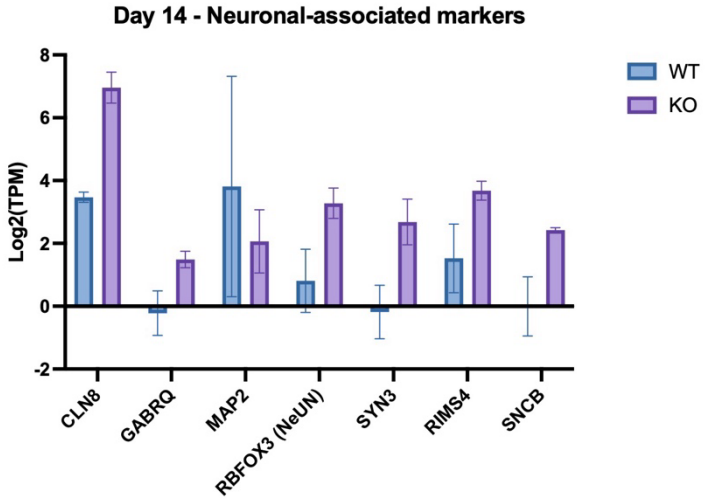


C

CNCC Specification Markers



D



Supplementary Figure 8.

Trajectory of stage-specific markers at Day-0 (n=2), Day-5 (n=2), and Day-14 (n=2). Displayed are the relative mRNA expression for:

(A) pluripotency markers (POU5F1 and NANOG),

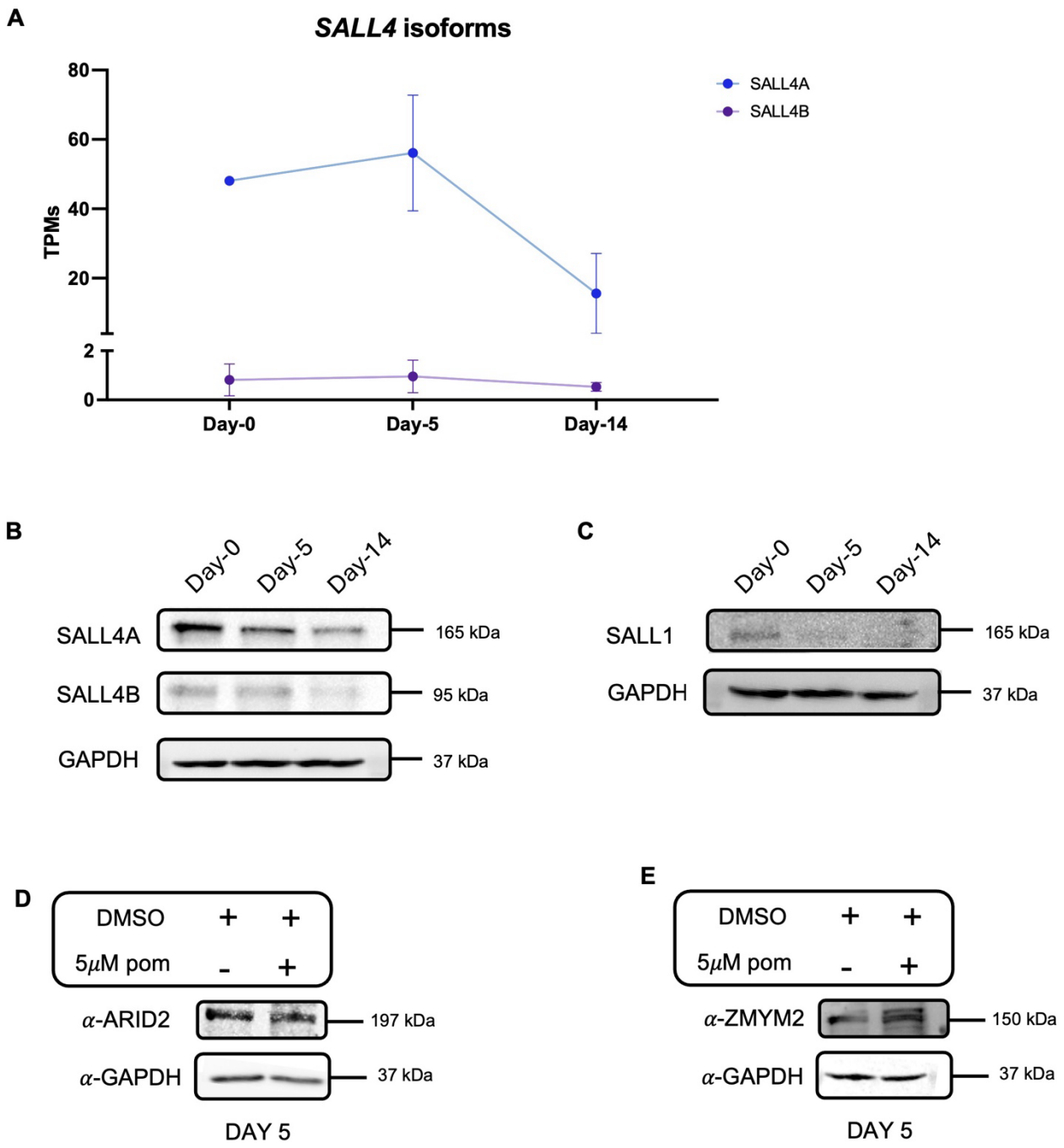
(B) neural plate border markers (MSX1, PAX3),

(C) CNCC specification markers (SOX9, TFAP2A).

Highlighted are the approximate stages in which each marker is conventionally expressed in *SALL4*-WT.

(D) Expression (Log2[TPMs]) of neuronal-associated markers at Day-14.

CONFIDENTIAL



Supplementary Figure 9.

(A) Trajectory of SALL4A and SALL4B RNA expression (TPMs) at significant iPSC-to-CNCC specification time-points: Day-0 (n=2); Day-5 (n=3); Day-14 (n=2).

(B) Time-course immunoblot for SALL4 in SALL4-WT at Day-0, Day-5, and Day-14. GAPDH was used as a loading control.

(C) Time-course immunoblot for SALL4 in SALL4-WT at Day-0, Day-5, and Day-14. GAPDH was used as a loading control.

(D) Immunoblot for ARID2 in SALL4-WT at Day-5 after 24-hour treatment with DMSO (control) or 5μM pomalidomide (5μM POM). GAPDH was used as a loading control.

(E) Immunoblot for ZMYM2 in SALL4-WT at Day-5 after 24-hour treatment with DMSO (control) or 5μM pomalidomide (5μM POM). GAPDH was used as a loading control.

CONFIDENTIAL